ÉCOLE DE TECHNOLOGIE SUPÉRIEURE
UNIVERSITÉ DU QUÉBEC

THESIS PRESENTED TO
ÉCOLE DE TECHNOLOGIE SUPÉRIEURE

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR
A MASTER'S DEGREE WITH THESIS IN ELECTRICAL ENGINEERING
M.A.Sc.

BY
Ibtihel AMARA

LOCAL QUALITY-BASED MATCHING OF FACES FOR WATCHLIST SCREENING APPLICATIONS

MONTREAL, JANUARY 11, 2016

BOARD OF EXAMINERS

THIS THESIS HAS BEEN EVALUATED

BY THE FOLLOWING BOARD OF EXAMINERS:

M. Eric Granger, thesis director
Department of Automatic Manufacturing Engineering at École de Technologie Supérieure

M. Abdenour Hadid, co-advisor
Center for Machine Vision Research at University of Oulu, Finland

M. Claude Thibeault, committee president
Department of Electrical Engineering at École de Technologie Supérieure

M. Luc Duong, board of examiners
Department of Software Engineering at École de Technologie Supérieure

THIS THESIS WAS PRESENTED AND DEFENDED

IN THE PRESENCE OF A BOARD OF EXAMINERS AND THE PUBLIC

ON DECEMBER 18, 2015

AT ÉCOLE DE TECHNOLOGIE SUPÉRIEURE

ACKNOWLEDGEMENTS

I am very much grateful to my research supervisor Professor Eric Granger for giving me the opportunity to work under his supervision. His guidance, support and constructive suggestions have contributed immensely to my progress on this project. His knowledge and dedication to research work have inspired me and have opened my mind to research in many different ways.

I would like also to thank my co-supervisor Professor Abdenour Hadid for his indispensable advice, information and support on different aspects of my project. I thank him for the time accorded to our Skype meetings despite the huge time difference between Montreal and Oulu, Finland. It has been an honor working directly with you during the summer of 2014 at École de technologie supérieure.

To all of my lab mates at the Laboratory for Imagery, Vision and Artificial Intelligence (LIVIA), I thank you all for your kindness, co-operation and for all the happy moments we shared. You have been a second family to me here in Montreal.

Special thanks to my friends Amel, Manel, Roghayeh, Faniya and Parisa for being always there for me whenever I needed a helping hand.

I am very grateful to have a loving and supportive family. I thank my parents and sister for their unconditional love and moral support throughout my entire life. They never stopped believing in me. I promise I will be the best I can be.

Last but not the least, to my beloved and caring husband, Khaled, thank you for standing up for me and supporting my dreams and ambitions. I am so lucky to have you in my life.

APPARIEMENT LOCAL DES MODÈLES DE VISAGES BASÉE SUR LA QUALITÉ DE L'IMAGE EN VIDÉO SURVEILLANCE

Ibtihel AMARA

SUMMARY

Les systèmes de vidéo surveillance occupent une place importante dans les organisations publiques et privées. En effet, leur utilisation se répend grâce à la démocratisation des appareils peu coûteux de vidéo surveillance. Une des applications importantes est la reconnaissance d'un individu appartenant à une liste noire (*watchlist screening*). Ce qui distingue cette application des autres systèmes de reconnaissance de visage (RV) en vidéo surveillance est le fait que les suspects sont abonnés au système de RV à partir d'une seule image statique.

La reconnaissance d'un individu appartenant à une liste noire utilise un nombre limité d'images de références (une seule image par personne dans notre situation) pour construire la galerie des modèles de visages. Ces derniers sont une série de représentation (formes, paramètres ou vecteurs caractéristiques) permettant de décrire un visage. Ce nombre limité d'informations rend le système de RV vulnérable et incapable de donner une décision correcte. Ce problème est appelé « seule échantillon par personne » (*single sample per personne*). Par ailleurs, on trouve aussi la présence des variations incontrôlables au niveau des captures de visages tels que les variations d'éclairage, un effet de flou et les changements de position de tête. En outre, parmi les difficultés qu'on trouve pour un système de RV, plus particulièrement pour les applications de RV dans une liste noire, est la différence au niveau de caméras utilisées : les images capturées pour les références sont souvent de haute qualité, tandis que celles capturées de la scène de surveillance sont souvent des images faibles en résolution et bruitées.

Il est certain qu'un éclairage uniforme est indispensable pour avoir de bonnes captures de visages. Néanmoins, dans un cas réel de vidéo surveillance, les visages capturés sont pauvres en illumination ce qui peut dégrader sévèrement la performance du système de RV. Pour cette raison, **la première partie** de ce mémoire consiste à explorer les différentes techniques de normalisation d'illumination qui seront appliquées au niveau du prétraitement du notre système de RV. Ensuite, une comparaison entre ces techniques de normalisation est effectuée pour pouvoir désigner la technique qui offre une meilleure invariance à l'illumination. La division en blocs des régions d'intérêt de visages pour l'appariement des modèles est adoptée dans ce travail car elle permet l'extraction des caractéristiques de manière discriminative. D'ailleurs, ces informations spatiales donnent plus de détails sur les différentes parties du visage. La division en bloc permet donc d'éviter les problèmes d'occlusions. Une étude approfondie est menée sur la manière d'appliquer ces techniques de normalisation sur l'image. Deux approches différentes sont comparées : l'approche globale dans laquelle on applique les techniques sur toute l'image et l'approche locale qui consiste à isoler les blocs, puis à appliquer sur chacun de ces blocs une technique de normalisation. Les résultats expérimentaux ont montré que l'approche Tan and Triggs (TT) et Multi-Scale Weberfaces (MSW) offrent une meilleure invariance d'illumination pour les systèmes de RV. En plus, ces deux techniques de normalisation appliquées localement ont aussi contribué à l'amélioration de la performance du système par rapport à l'approche globale.

Pour avoir un bon fonctionnement du système de RV, il faut que les données utilisées pour l'apprentissage et celles utilisées pour le test aient la même distribution. Dans notre application (RV dans une liste noire), les données pour l'apprentissage sont des images frontales, de haute qualité et haute résolution, alors que les données pour le test proviennent des vidéos de faible qualité, faible résolution et présentent des variations de position de tête. Pour surmonter ce décalage des domaines, on propose dans **la deuxième partie** de ce mémoire une nouvelle technique de pondération des régions locales tout en exploitant les concepts d'adaptation des domaines non supervisés (*unsupervised domain adaptation*) et les informations contextuelles avec les métriques de qualités d'images. La principale contribution est le calcul dynamique des pondérations qui sont spécifiques à une caméra (adaptation selon une vue de caméra). Cette étude contextuelle et adaptive selon les domaines offre une meilleure performance par rapport à la pondération statique et prédéfinie et par rapport aux systèmes sans pondération.

Ces expériences sont validées sur la base de données ChokePoint. Les performances du systèmes sont évaluées avec les mesures de performances, les courbes de Receiver operating characteristic (ROC) et les courbes de Precision-Recall.

Mots clés: reconnaissance de visages, classification locale, adaptation des domaines, information contextuelle, appariement local des modèles, normalisation d'illumination, pondération dynamique, qualité d'image, base de données chokePoint

LOCAL QUALITY-BASED MATCHING OF FACES FOR WATCHLIST SCREENING APPLICATIONS

Ibtihel AMARA

ABSTRACT

Video surveillance systems are often exploited by safety organizations for enhanced security and situational awareness. A key application in video surveillance is watchlist screening where target individuals are enrolled to a still-to-video Face Recognition (FR) system using single still images captured a priori under controlled conditions.

Watchlist Screening is a very challenging application. Indeed, the latter must provide accurate decisions and timely recognition using limited number of reference faces for the system's enrolment. This issue is often called the "Single Sample Per Person" (SSPP) problem. Added to that, uncontrolled factors such as variations in illumination pose and occlusion is unpreventable in real case video surveillance which causes the degradation of the FR system's performance. Another major problem in such applications is the camera interoperability. This means that there is a huge gap between the camera used for taking the still images and the camera used for taking the video surveillance footage in terms of quality and resolution. This issue hinders the classification process then decreases the system‘s performance.

Controlled and uniform lighting is indispensable for having good facial captures that contributes in the recognition performance of the system. However, in reality, facial captures are poor in illumination factor and are severely affecting the system's performance. This is why it is important to implement a FR system which is invariant to illumination changes. The **first part of this Thesis** consists in investigating different illumination normalization (IN) techniques that are applied at the pre-processing level of the still-to-video FR. Afterwards IN techniques are compared to each other in order to pinpoint the most suitable technique for illumination invariance. In addition, patch-based methods for template matching extracts facial features from different regions which offers more discriminative information and deals with occlusion issues. Thus, local matching is applied for the still-to-video FR system. For that, a profound examination is needed on the manner of applying these IN techniques. Two different approaches were conducted: the global approach which consists in performing IN on the image then performs local matching and the local approach which consists in primarily dividing the images into non overlapping patches then perform on individually on each patch each IN technique. The results obtained after executing these experiments have shown that the Tan and Triggs (TT) and Multi Scale Weberfaces are likely to offer better illumination invariance for the still-to-video FR system. In addition to that, these outperforming IN techniques applied locally on each patch have shown to improve the performance of the FR compared to the global approach.

The performance of a FR system is good when the training data and the operation data are from the same distribution. Unfortunately, in still-to-video FR systems this is not satisfied.

The training data are still, high quality, high resolution and frontal images. However, the testing data are video frames, low quality, low resolution and varying head pose images. Thus, the former and the latter do not have the same distribution. To address this domain shift, the **second part** of this Thesis consists in presenting a new technique of dynamic regional weighting exploiting unsupervised domain adaptation and contextual information based on quality. The main contribution consists in assigning dynamic weights that is specific to a camera domain.This study replaces the static and predefined manner of assigning weights. In order to assess the impact of applying local weights dynamically, results are compared to a baseline (no weights) and static weighting technique. This context based approach has proven to increase the system's performance compared to the static weighting that is dependent on the dataset and the baseline technique which consists of having no weights.

These experiments are conducted and validated using the ChokePoint Dataset. As for the performance of the still-to-video FR system, it is evaluated using performance measures, Receiver operating characteristic (ROC) curve and Precision-Recall (PR) curve analysis.

Keywords: still-to-video face recognition, watchlist screening, unsupervised domain adaptation, contextual information, local matching, illumination normalization, dynamic weighting, image quality assessment, chokePoint dataset

CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

LIST OF ALGORITHMS

LIST OF ABREVIATIONS

FR	Face Recognition
S2S	Still-to-Still Face Recognition
S2V	Still-to Video Face Recognition
V2V	Video-to-Video Face Recognition
SVM	Support Vector Machine
GMM	Gaussian Mixture Model
FD	Face Detection
FT	Face Tracking
FE	Feature Extraction
ROI	Region of Interest
SSPP	Single Sample per Person Problem
IQA	Image Quality Assessment
IN	Illumination Normalization
HOG	Histograms of Oriented Gradients
LBP	Local Binary Patterns
BLBP	BAyesian LBP
LPQ	Local Phase Quantization
BSIF	Binarized Statistical Image Features
SILTP	Scale Invariant Local Ternary Pattern

SVM	Support Vector Machines
SRC	Sparse Representation-base Classification
PSF	Point Spread Function
DCT	Discrete Cosine Transform
DOG	Difference of Gaussian
SIFT	Scale-Invariant Feature Transform
NIR	Near-Infrared
FAR	False Acceptance Rate
STFT	Short-Term Fourier Transform
MSE	Mean Squared Error
PSNR	Peak to Signal to Noise Ratio
SSIM	Structural Similarity Index
QM	Quality Metrics
ROC	Receiver Operating Characteristics
PR	Precion-Recall
AUC	Area Under the Curve
AUPR	Area Under the Precision-Recall curve
pAUC	Partial Area under the ROC curve
ASSR	Adaptive Single-Scale Retinex
LSSF	Large and Small-Scale Features

MSSQ	Multi-Scale Self Quotient
ID	Isotropic Diffusion
MAD	Modified Anisotropic Diffusion
TT	Tan and Triggs
MSW	Multi-Scale Weberfaces
ANLM	Adaptive Non Local Means
RM	Retina Modeling
WD	Wavelet Denoising
H	Homomorphic

LISTE OF SYMBOLS AND UNITS OF MEASUREMENTS

I	Image sample
R	Reflectance
L	Luminance
C_{RMS}	Contrast value
V_{IQA}	Image quality value
$V_{IQA_{norm}}$	Normalized image quality value
I_{still}	Still image sample
χ^2	Chi square
Nb_{ind}	Total number of target individuals enrolled in the gallery
N_{calib}	Number of individuals used for system calibration
N_{IN}	Total number of illumination normalization techniques
N_{feat}	Size of a feature vector
NT_{feat}	Total length of a facial model for one ROI
N_{TP}	Number of total local regions in a ROI
$N_{size} \times N_{size}$	Pixel size of an ROI
$\mathbf{m}$	Feature template of the still reference ROI
$\mathbf{f}$	Feature template of the probe ROI
$\mathbf{B}$	Set of blocks of an ROI
$\mathbf{B}_{IN}$	Set of blocks of an illumination normalized ROI

INTRODUCTION

Currently, the need to protect personal confidentiality and security has become important. Hence, biometric recognition technology has broken through as a key solution for access control, authentication, etc. It is now considered as a perfect substitute to passwords, pins, keys and tokens for identity recognition (Jain *et al.*, 2004) (Delac and Grgic, 2004).

There are several biometric traits to perform person recognition: fingerprints, hand geometry (Kumar *et al.*, 2003), voice (Rabiner, 1989) or even by combining all of the traits mentionned above (Ross and Jain, 2004). Cooperation from individuals is not always feasible. For that reason, faces are often used in surveillance applications for recognition because face captures can be inoperatively and passively acquired using distanced cameras compared to the other biometric means which need a direct interaction and full cooperation from the user.

In literature, there are three functions of FR: verification, identification and surveillance. Verification systems are also known as one-to-one matching (Jafri and Arabnia, 2009) in which a face region of an individual is verified and compared to the legitimate one. As for the identification process, called one-to-many matching, is performed by comparing the face region obtained a priori from all individuals and attempting to give the proper identity of the present individual.

Facial models are patterns often compact that should provide a robust representation of an individual's face to real world variations. These patterns are used at the classification level of a FR system. They can incorporate a set of features/templates used in template matching or a set of parameters for instance support vectors (weights,bias) in Support Vector Machines (SVM), weights or kernels in neural nets and (mean, variance) for Gaussian Mixture Models (GMM).

There are mainly two types of FR. The first type is the still-based recognition or the still-to-still FR where the gallery and the probe sets consist of facial models obtained from still images. But what really concerns us is the second type, video-based FR which is the most appropriate for video surveillance applications.

Video-based FR is often times adopted for security operations, more distinctly for surveillance purposes. Video surveillance is mainly used by governments and companies to ensure public safety. Owning video surveillance cameras is becoming more and more affordable and its demand has increased tremendously. Nowadays, these cameras are omnipresent and are able to capture any abnormal presence of a possible target individual or any other suspicious activities. Manual recognition can not provide fast real-time decisions due to the huge amount of data (videos) to be processed and more specifically agents behind these monitors are susceptible to get tired and miss certain information. For these simple reasons, the need of automatic video-based FR software for video surveillance is becoming very essential.

Video-based FR is divided into two categories: video-to-video in applications such as face re-identification where facial models are obtained from reference videos and still-to-video FR in applications such as watchlist screening where facial models are obtained from single still reference images.

Video-to-Video Face Recognition

A specificity of video-to-video FR is that the facial models in the gallery, during system's enrolment, extracted from reference video sequences, are matched against facial models obtained from the input video frames. Video-to-video is seen to be a very challenging task (Jafri and Arabnia, 2009) (Zhao *et al.*, 2003) (Vijayakumari, 2013). Applications such as person re-identification, search and retrieval are one of the most imminent utilization of video-to-video FR.

Undeniably, faces captured in video frames are often poor in quality and resolution which hinders the recognition process of the video-to-video FR system. Likewise, faces are obtained under uncontrolled conditions. This leads to facial variations due to environmental changes namely bad lighting conditions, varying head pose and occlusions or physiological changes like aging or facial expressions.

There are numerous existing methods to cope up with these issues. In (Barr *et al.*, 2012) existing techniques are surveyed and categorized into two groups: set-based techniques that neglect temporal information and sequence-based technique that exploits temporal information from the video sets. The first type proceeds in combining information over the collected samples of facial captures whether before or after matching. One of the proposed method is super-resolution. The goal is to construct high resolution images from low resolution samples in order to recover the high frequency content that was lost during acquisition. This technique has shown to improve video-to-video FR performance based on the works presented in (Al-Azzeh *et al.*, 2008) and (Gunturk *et al.*, 2003). Linear subspace, also called non linear manifolds, are employed to video-to-video FR systems in which the main objective is to measure distances and the common variations between the probe and the sets in the gallery. Many research work have adopted this method (Hadid and Pietikäinen, 2009b) (Takahashi *et al.*, 2009) and (Wang *et al.*, 2008). Each of these works concentrated on proposing different frameworks and distance metrics. Other techniques relies on the information already available in the video sets. Thus, the idea of performing frame selection techniques is implemented. Approaches such as quality based oriented frame selection (Berrani and Garcia, 2005) are proposed to detect and exclude frames that are poor quality in terms of illumination, pose variations and blur which can cause severe recognition errors. Weighting techniques are also used for frame selection (Stallkamp *et al.*, 2007) to reduce the influence of the probe frames that are different from those saved in the gallery. Also, clustering approaches (Hadid and Pietikainen, 2004) divide the set of images into a collection of groups with related appearance. Frame selection has proven to show improvement in video-to-video FR systems. As for the second type, sequence-based, the majority of the techniques are dependent on temporal information. One approach focuses on improving the tracker since spatio-temporal techniques rely on tracking. (Kim *et al.*, 2008) focuses on overcoming abrupt changes and possible occlusions by adding visual constraints such as facial pose and alignment to better acquire facial ROIs. Neural network classifiers are also used to estimate identity overtime through variations (Gorodnichy, 2005) (Barry and Granger, 2007) (Connolly *et al.*, 2012). Spatio-temporal features are also proposed to enhance video-to-video FR performance. In (Hadid and Pietikäinen, 2009a) an extended version of Local Binary Patterns (LBP) called Volume LBP proceeds in using pixels from three the neighbouring frames. Ensemble-based and multi classifiers have also proven to enhance video-to-video FR as shown in the works in (Pagano *et al.*, 2012) (Pagano *et al.*, 2014) (De-la Torre *et al.*, 2015).

Still-to-Video Face Recognition

The second category of video-based FR is the **still-to-video** FR which is the main focus in this Thesis. The gallery of these types of systems contains facial models that are extracted from one or many reference still images while the probe set incorporates facial models obtained from video frames.

A key application of **still-to-video** FR in video surveillance is **watchlist screening** where a limited amount of still images of a target individual (one single still image for watchlist screening application) is available for training and recognition process. Facial regions of interests (ROI) are isolated from the single reference images of the targets often captured a priori under controlled conditions and are enrolled to the system. During operations, ROIs corresponding to faces detected in the video surveillance cameras that furnishes low quality images are matched against the reference ROIs of each target individual in the watchlist.

Issues can be found in **still-to-video** FR systems. Definitely, nuisance factors such as illumination, motion blur, occlusion and different head positions have a negative impact on the face appearance. Other challenges are directly linked to the limited number of training samples. As a fact, there is a direct relationship between the number of samples and the recognition rate. Generally, the higher the number of samples for training the higher the recognition rate. In still-to-video FR systems, especially for watchlist screening applications, only a single representation of an individual is available. This problem is treated as the "*Single Sample Per Person*" (SSPP) problem. Another problem that can be considered is the presence of a gap between the distribution of data used for enrolment (source domain) (still, high quality and high resolution images) and the distribution of data used for testing (target domain) (frames, poor quality and low resolution). This issue is often called the domain shift. Hardware incompatibilities or in other words camera interoperability are also one of the major problems (acquisition perspective) in still-to-video FR systems. Moreover, other challenges related to processing time, lack of memory resources due to huge algorithms and data can also affect badly on the performance.

Only a handful of existing methods are available for problems related to still-to-video FR. In literature each issue is addressed differently. Here we are going to provide a global idea of existing methods tackling the issues mentioned above. More comprehensive details can be found in Section 1.2 Chapter 1. To manage the SSPP problem, multiple face representation is proposed. Multiple representations are obtained by using multiple feature extraction techniques on the reference sample (e.g. local binary pattern (LBP), local phase quantization (LPQ), etc.) or either by implementing a patch-based method for local matching. For further knowledge about existing patch configurations in literature refer to Chapter 2 of this Thesis. This approach is implemented in (Bashbaghi *et al.*, 2014). In addition, domain adaptation solutions are proposed to handle the gap between both source and target domains. In (Ho and Gopalan, 2014) facial variations such as illumination and blur are modelled into the source domain to match the target domain for one sample FR system. Enlarging the training set is introduced to surpass the SSPP problem in still-to-video FR systems. Approaches like generating synthetic faces from the single reference facial ROI of an individual using image warping (Kamgar-Parsi and Lawson, 2011) is introduced. 3D modeling has also contributed to the improvement of the still-to-video FR system (Park and Jain, 2007). Moreover, to compensate occlusion issues found in still-to-video FR systems, SRC-based methods are extended (Yang *et al.*, 2013). In (Wei and Wang, 2015) Robust Auxiliary Dictionary Learning by means of SRC was proposed. In this context, an external database is exploited to extract possible occlusion variants then the latter are applied to the single reference sample generating newer synthesized samples for matching. Additionally, the work presented in (Mokhayeri *et al.*, 2015) also exploits an external dictionary to extract different illumination variations. Eventually, IQA techniques has played an important role in improving still-to-video systems. In (Huang *et al.*, 2013), quality alignment module is implemented in selecting "good" quality and "best" aligned input probe ROIs compared to the still reference facial image enrolled in the system.

Although these mentionned techniques have promising ability to achieve high recognition performance, some challenges can be associated with these existing still-to-video FR systems. In fact, the majority of these systems have complex frameworks. For example in the case of multiple representation, this requires multiple face descriptors in a single framework. As for the case of enlarging the dataset through synthetic samples, there are insufficient representative samples. Therefore, it is difficult to predict all variations of faces. Finally, these proposed still-to-video FR systems in literature do not exploit camera domain information.

Objectives and Motivation

The main objective of this Thesis is to investigate and propose a robust still-to-video FR system. This system addresses varying capture conditions such as illumination variations, blur, head pose and resolution through enhancing feature representation by applying illumination normalization techniques and through enhancing classification task by using domain adaptation, local matching, contextual information and regional weighting.

Illumination normalization techniques can handle illumination variations. Thus, its application at a pre-processing level can enhance the feature representation of the captured ROIs. In addition, with still-to-video FR system especially watchlist screening the distribution of data between the training and testing process differ and present a domain shift problem. Therefore applying a domain adaptive approach may lessen the mismatch between domains. The performance of the still-to-video FR is dependent on the quality of the input images. Thereupon, it is important to have a system that contextually understands and exploits the image quality. Finally, local matching provides discriminative facial models that can handle the SSPP problem by providing multiple representation and it deals with partial occlusion often witnessed in most facial captures in video surveillance systems. Consequently, considering all these approach would be beneficial to build a robust still-to-video FR system.

Contributions

In order to realize the main objective of this Thesis, the work is divided into two separate yet complementary contributions:

- The first part of this Thesis is a discussion on different IN techniques applied to a still-to-video FR system and their manner of application (globally or locally);
- The second part of the work consists in implementing a new technique for regional weighting based on contextual information and domain adaptation solution to give importance on specific regions. Two experiments are held at this level. The first one reposes in implementing the approach without considering illumination issues. The second experiment consists in adding the outperforming IN technique concluded in the first part of the Thesis to attain illumination invariance of the still-to-video system. Table 0.1 provides a global idea of both contributions to be presented in this thesis.

Table 0.1 Brief descriptions of the contributions proposed in this Thesis

	Chapter 4: Contribution 1	**Chapter 5**: Contribution 2	
		Experiment 1	Experiment 2
Goal	Compare the effects of applying IN LBP-based still-to-video FR systems.	Propose a new quality-based weighting approach for local matching of still-to-video FR sytems.	
System module	Pre-processing level	Classification level	Pre-processing and Classification level

Document organization

The rest of this document is organized as the following structure. **Chapter one** provides a general overview of the still-to-video FR system detailing its each component, problems encountered with these types of systems, some existing still-to-video FR systems in literature. **Chapter two** shows a survey of existing techniques used for face recognition starting from the pre-processing, feature extraction to classification methods. **Chapter three** outlines the common experimental set up in both research contributions such as the description of the dataset and the measures for assessing a system's performance. **Chapter four** presents in detail the method implementing an illumination invariant still-to-video FR system and its experimental results. **Chapter five**, shows the second strategy in boosting still-to-video FR systems at the classification level. This chapter explains the protocol used in implementing the dynamic weighting module for local matching and discusses the obtained results. Finally, **conclusion and recommendations** are presented at the end of this Thesis.

CHAPTER 1

APPLICATION: STILL-TO-VIDEO FACE RECOGNITION SYSTEMS FOR WATCHLIST SCREENING

This chapter presents a global view on the state-of-the-art in still-to-video FR as needed for watchlist screening applications. Descriptions of each module of still-to-video FR system is explained. Comprehensive surveys on general techniques in FR for each module is given. Finally, existing systems of still-to-video FR in literature and their challenges are exposed.

1.1 General Framework

Target individuals are detected and recognized under uncontrolled camera capture conditions.

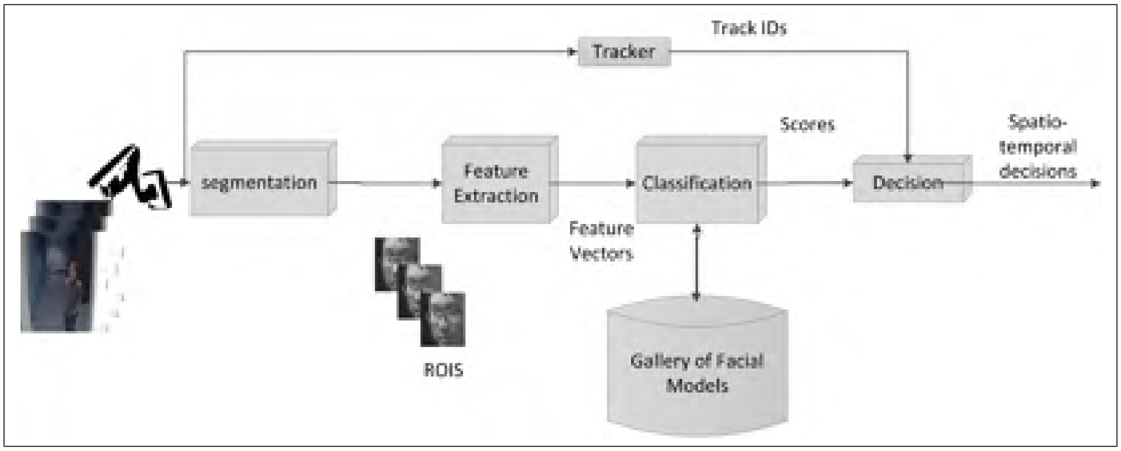


Figure 1.1 Generic system for Face Recognition

Figure 1.1 shows a generic still-to-video FR system. During enrolment to a watchlist, the segmentation process isolates ROIs from the reference still images of high quality that are previously captured under controlled conditions. Afterwards, facial features are then extracted and assembled into discriminate and compact ROI patterns in order to obtain unique facial models $\mathbf{t} = \{t_1, ...t_{N_{feat}}\}$. The latter is stored in a gallery as facial models for face matching.

Facial models are ROI patterns used for classification. These models differ from one classifier to the other and are stored in a gallery. In fact, if classification is performed using template matching the stored patterns are in a form of feature templates. Besides, the stored facial models can be in a form of a set of parameters. For example, if support vector machines (SVMs) are used, after system training, the stored outputs in the gallery are the values of weights and bias of the decision boundary which are often called support vectors. Another example is performing neural network for classification, the parameters to be stored are the weights. So, these facial models depend on the nature of classification used for recognition.

During operations, video streams are captured using video surveillance cameras that provides in most cases, low resolution and quality images. The segmentation mechanism isolates ROIs captured in successive video frames. A tracker is often triggered when a new ROI appears and detected differently from those already present in each frame. The tracker follows the movement or expression of distinct faces across consecutive frames using appearance, position and motion information. Features **a** are then extracted into ROI pattern for matching versus the face models already stored in the gallery. A positive prediction is ensured if the score $S(\mathbf{a},\mathbf{t})$ of the matching process surpasses an individual-specific threshold γ. Finally, at the spatio-temporal decision module the tracks and scores are combined for recognition. Details about functional module (segmentation, tracker, classification, and spatio-temporal fusion) are described below.

1.1.1 Segmentation

This module detects facial ROIs in each frame in a video sequence. In other words, each video frame undergoes a set of treatments such as face detection (FD) and scaling which gives an output of sets of facial ROIs from that given frame. FD withstand two main procedures: firstly, isolating the regions that are identified as faces then estimation of the position of the regions found and their sizes and delimitation of this region. Face detection has always been challenging in video surveillance due to many variations such as face pose and orientation, scale changes, different lighting conditions, presence of occlusions, skin color, etc.

A lot of approaches have been proposed for FD. First surveys (Yang *et al.*, 2002) began to regroup detection methods into primary categories: **knowledge-based** methods that uses human biased rules for face detection. For example in (Zhang and Lenders, 2000), face and eyes are detected by finding two dark round shaped and similar areas which are knowledge-based properties. **Feature invariant** methods are based on locating invariant features like texture, shape or color. In (Wang and Sung, 1999) skin-color approach is proposed to detect the facial region. **Template matching** methods which consist in exploiting both features extracted in the input image frame using extraction methods such as edge detectors or filters and the predefined facial features stored as templates by performing a correlation between them in order to determine the presence of a face in the image frame. **Appearance-based** methods are based on adopting machine learning approaches to classify regions whether it is a face or not a face. The common used technique for FD is the Viola-Jones method (Viola and Jones, 2001). The latter is based on applying Haar features across a given input frame which gives a certain number of features computed by the algorithm rapidly and less costly using the 'Integral Image' approach. Then it uses adaboost technique in order to narrow down the calculated features and remove any probable redundancy. Then finally, cascade training is performed to classify the best representative facial features and tell whether a face is detected or not.

Latest developments in face detection have appeared since the last proposal of the Viola-Jones technique which have led into newer categories for face detection. A detailed survey about advanced face detection techniques can be found in (Zhang and Zhang, 2010). **Feature-based** methods consist in extracting special features and exploit them for detection. In (Levi and Weiss, 2004) statistic-based features using edge orientation histograms are used for FD. Others used shape features such as edgelet (Wu and Nevatia, 2005) and shapelet features (Sabzmeydani and Mori, 2007). Another new category for FD methods is **boosting-based** methods. These methods are mostly inspired by the proposal of the Viola-Jones method (Viola and Jones, 2001). For a multi-view FD structural cascades were proposed for boosting such as parallel cascades (Wu *et al.*, 2004) and tree decisions (Fröba and Ernst, 2003). Other categories for face detection are also worth mentioning that are also based on machine learning are **SVM-based** methods (Wang and Ji, 2004) and **neural network-based** methods (Garcia and Delakis, 2004).

1.1.2 Face tracker

Face Tracking (FT) grants the system the ability to follow each motion and movement of an individual or set of individuals located in different positions in a video sequence. The FT module may be triggered by the detection of a new ROI in a new area of a frame. The process of detecting and tracking can be done in two manners either separately or jointly (Yilmaz *et al.*, 2006) . For the first mentioned, FD technique provides the regions of the possible faces then the tracker compares this region over the next frames. As for the second case, FD and the comparison of ROI in the upcoming frames are done simultaneously by updating the location information of the ROI based on the previous frames. FT allows to regroup facial ROIs of a same person in the scene over time. Indeed, a trajectory of faces is defined as a set of isolated and segmented ROIs that corresponds to a same track of an individual across consecutive frames and is linked to a $trackID$.

Many tracking techniques were introduced depending on the aimed application and the nature of the object to be tracked. There are mainly three categories for object tracking: **point tracking** where objects are detected in successive frames. These objects (faces in case of face tracking) are presented as points. The grouping of these points is based on the previous state (position and motion). These type of trackers are often triggered by an external detector module (face detection). A lot of works have concentrated in point tracking. In (Chan *et al.*, 1979) objects are tracked by estimating the motion using statistical correspondence given by Kalman filters. Multiple Hypothesis Tracking (MHT) (Cox and Hingorani, 1996) is also introduced as point tracking based on iteration and hypothesis and enumeration of all possible motion associations for tracking. **Kernel tracking** is related to its shape and appearance. Tracking is performed by calculating the motion (translation, rotation and affine) through the consecutive frames. Approaches like Mean-Shift (Cheng, 1995) and Camshift (Bradski, 1998) used for this category of tracking. As for the third type of tracking which is the **silhouette tracking**, it is assured by predicting the region of the object in each frame. In (Huttenlocher *et al.*, 1993) object tracking is performed using model-based techniques more specifically by decomposing a moving object into two components shape changes and motion. Contour tracking in (Chen *et al.*, 2001) is implemented using elliptical contour and joint probability data association filters to estimate the transitions.

1.1.3 Feature Extraction

The Feature Extraction (FE) takes in an input ROI and produces ROI patterns **a** that are usually compact providing a distinctive description of the facial ROI. This generation of feature values is often considered to be a loss of information but beneficial for classification process. Indeed, reducing the number of redundant bits representing certain facial characteristics such as overall pixel color of the human skin in that input ROI, allows the classification module to perform rapidly and more efficiently.

Feature representation of a target region are classified into **holistic FE** and **local FE** (Fasel and Luettin, 2003). **Holistic FE** has always been the frequently used method for FE especially for FR systems. As a matter of fact with this method large data describing a region can be reduced and transformed into a compressed information. An example of global FE technique is the Eigenface where the recognition process is based on projecting the region patch to each of the learned Eigen subspace. More details about how Eigenfaces for FR can be found in (Turk and Pentland, 1991b). Although holistic methods provide less extensive computational cost thanks to the small number of calculated feature values, it is still considered as an impoverished method for region description especially for uncontrolled condition scenarios.

Local FE, consists in representing local regions of the target input ROI by dividing the rearmost into local sub images for better and more efficient way to distinguish unique characteristics. Some common local FE techniques are Gabor Wavelets (Kyrki *et al.*, 2004), Discrete Cosine Transform (DCT) which also based on image decomposition into sub-images and the most used FE approach in FR systems is the Local Binary Pattern (LBP) (Ojala *et al.*, 2001) (Ahonen *et al.*, 2004). The latter is further explained in the upcoming subsection along with some of its variants.

1.1.4 Classification

The classification module allows the FR system to assign probe faces to face models **a** (or classes) of individuals enrolled to the gallery. Due to some limitations that FE techniques provide for individual description, the classifier can't ideally perform. That is why, in most cases of classification, the latter tends to calculate a probability or estimation of the possible classes one individual belongs to. The performance of a classifier depends on the number of information it can provide and the ability of the FE to give precise detailed description of an individual in the presence of random factors originating from uncontrolled open set sensors.

Many classification techniques have been introduced in order to provide better categorization of individuals. They have evolved from simple **template matching** calculating the similarity or dissimilarity between two templates. Distance measures (Goshtasby, 2012) belongs to this type of classification. In addition, classification can also be **statistical-based** where a probability model is calculated in order to estimate whether an input belongs to a certain class. As an example methods such as Bayesian rules (Moghaddam *et al.*, 2000) or density estimation using K-nearest neighbours (K-NN) (Denoeux, 1995) are used in pattern recognition. **Machine learning based** methods are mainly based on logical and binary operations and learning on a set of examples. An example of this classification category would be the implementation of decision trees (Maturana *et al.*, 2011) to divide classes. **Neural network based** classification has also emerged from the inspiration of the human abilities to understand, predict and distinguish. The complexity of this approach is a combination between machine learning and statistical methods. A survey on neural network techniques used for classification can be found in (Zhang, 2000). Finally, **modular** classification have recently been developed. It is based on combining multiple classification techniques (whether statistical or learning based). An example is the implementation of ensemble based classification for face recognition in (Pagano *et al.*, 2012).

1.1.5 Spatio-temporal fusion

This module combines the face trajectories from the tracker module and the scores from the classification module of the same ROI. It accumulates the scores across the trajectory and the latter are compared to a threshold. Spatio-temporal techniques have been used a lot in recent studies because of its advantage of increasing contextual information in videos by exploiting temporal information.

Here are some recent examples of spatio-temporal recognition. In (Ekenel *et al.*, 2010) a video-based FR system based spatio-temporal decision module was implemented. The latter combines the matching score using fusion technique the s̈um ruleöf all the sequences in order to perform recognition. The work presented in (De-la Torre *et al.*, 2015) consists in implementing adaptive ensemble based video-to-video FR system using spatio-temporal fusion. This is done by accumulating positive predictions from each ensemble of classifiers. If the value of this accumulation for each facial trajectory surpasses a detection threshold then a list of possible target individuals are given.

1.2 Still-to-Video Face Recognition Systems

Existing still-to-video FR systems can be categorized into different groups based on the adopted approach. Some systems are based on generating multiple representation in terms of facial models of an individual. For example in (Bashbaghi *et al.*, 2014), multiple representations are created using the combination of two different techniques: first using patch-based method for local matching and second using various feature extraction methods such as LBP, LPQ, HOG and HAAR. By this, each individual enrolled in the still-to-video FR system would have discriminative facial models. These representations are shown to improve the system's performance in terms of recognition because they are proven to be robust to occlusion thanks to local patch division and also robust to other nuisance factors such as illumination and blur thanks to the characteristics of the feature extraction method applied (i.e LPQ is known to be robust to blur).

Another proposed technique is the 3D modeling approach that exploits multiple sets of facial captures to estimate the geometric structure of the face in terms of head pose variations and illumination changes. Work in (Park and Jain, 2007) is based on extracting pose and illumination conditions from the probe images then then inject those informations to the face models in the gallery of the still-to-video FR system to match those from the probe.

In addition to that, enlarging training sets like generating synthetic faces for better training and classification of the still-to-video FR system is proposed. An example is the work of (Kamgar-Parsi and Lawson, 2011) in which 2D morphing techniques were used to build synthetic samples (positives and negatives) for training the system's classifier. The morphing technique resides in taking the still image of an individual then merge it with other facial images providing different facial morphologies. More specifically, in this work a limit of acceptable morphed images of a single individual was studied in order to obtain a boundary of positives and negatives for classification and training of the FR system.

Furthermore, enlarging the training set using synthetic face image generation is also used in (Wei and Wang, 2015). Unlike the previous technique, this second one concentrated on the variations of facial appearances due to changes in illumination and contrast. By using their proposed system, an extended version of morphing approach synthetic images of a single still ROI were generated under various illumination conditions which were taken from different cases of camera viewpoints. These techniques were proven to enhance the performance of still-to-video FR systems.

Some still-to-video FR techniques involves applying frame selection and weighting based on quality. In (Huang *et al.*, 2013), captures ROIs from video frames enter a quality alignment module which assesses the degree of similarity in terms of alignment with the still reference ROI using sparse representation and clustering. Then it provides weights to the frames depending on its importance and quality. Based on their results, this approach has also achieved better performance for still-to-video FR systems.

Techniques based on tracking systems for online face modeling has also been implemented to improve still-to-video FR systems as shown in (Dewan *et al.*, 2016).

Quality oriented frame selection approach has also gained popularity for still-to-video FR systems. In (Huang *et al.*, 2013), quality alignment module is implemented in selecting "good" quality and "best" aligned input probe ROIs compared to the still reference facial image enrolled in the system.

Finally, domain adaptation solutions are proposed. In (Ho and Gopalan, 2014) a model-driven approach using tensor geometry is used to model facial variations such illumination and blur in the source domain to match the target specificities in the target domain.

1.3 Challenges in Still-to-Video Face Recognition

There are many problems and challenges related to still-to-video FR systems. To start with, these types of systems especially for applications like watchlist screening, usually use one single still image per person for extracting facial models and enrolling it to the gallery. However, a single sample is not sufficient in order to obtain representative facial model of an individual (SSPP problem).

Besides that, in watchlist applications, facial models that are enrolled on a source domain (enrolment domain) originated from high resolution, high quality still images that are taken under controlled conditions. Per contra, the facial models obtained from the target domain (testing domain) are from frames that are poor in quality, low in resolution and containing uncontrolled variations (the domain gap or domain mismatch). In fact, in most still-to-video FR systems the properties of the testing data are somehow different from those ones used for enrolment which causes a decrease in the system's performance.

On the other hand, a concentration on issues that are related to the information retrieved from the sensors (facial captures) is important. In real life video surveillance, individuals are often captured under varying head positions, illumination changes that may result in occluded facial captures and also varying resolution and scaling properties. To add on that, a human face is completely uncontrollable. facial captures are vulnerable to abrupt behavioural and expression changes.

All of these uncontrolled variations degrades the performance of the still-to-video FR system.

CHAPTER 2

BACKGROUND TECHNIQUES FOR FACE RECOGNITION

In this chapter a detailed survey of some specific techniques is provided. More specifically, techniques like texture descriptors for feature extraction, local matching for classification, Illumination normalization (IN) for pre-processing and Image Quality Assessment for complementary modules are given importance.

2.1 Texture Descriptors

Finding the most representative facial descriptors is one of the most important processes in FR. There are many descriptors in literature: from geometric-based (Kanade, 1973) to appearance based (Turk and Pentland, 1991a). However, what broke through the most is those representations based on local extraction of information. In fact, local descriptors have played a huge role in FR because small details of a facial appearance can be captured without having the risk of information distortion. Methods in this group include Local Binary Patterns (LBP) and their derivatives. LBPs are efficient local texture descriptors and are numerously implemented in FR applications (Ahonen *et al.*, 2006) (Rodriguez and Marcel, 2006) (Heusch *et al.*, 2006) thanks to its ability to provide discriminative facial characteristics. In this section, local texture descriptors are detailed and explained.

2.1.1 Local Binary Patterns

The Local Binary Pattern is a texture analysis operator that was introduced primarily by (Ojala *et al.*, 2001). It is considered as a gray-scale invariant texture metric that uses general definition of texture in local neighbourhood of pixels. LBP is considered to provide discriminative information, computational simplicity and it is capable of tolerating monotonic gray-scale changes.

The LBP operator creates labels on the image pixels by thresholding a 3 by 3 neighbourhood of each pixel with the value of the center pixel of that 3 by 3 window to provide a binary pattern as shown in Figure 2.1. Some works have dedicated in extending this concept by changing the number of neighbourhood used for threshold. Indeed, using a circular neighbourhood and a bilinear interpolation of values at the non-integer coordinates does not restrict the size of radius and the number of neighbouring pixels. The notation (P,R) is often used, where P is the number of sampling points on a radius R. The LBP code for a pixel is:

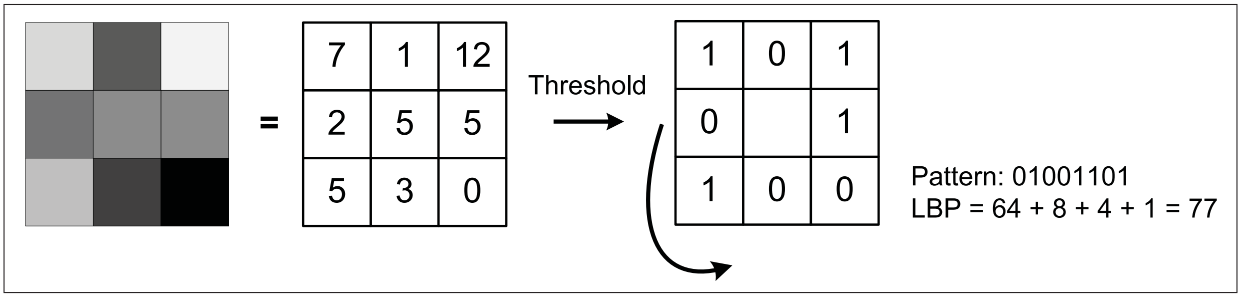


Figure 2.1 The basic LBP operator
Taken from Amara *et al.* (2014)

$$\mathrm{LBP}_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c)2^p, \tag{2.1}$$

where g_c corresponds to the gray value of the center pixel (x_c, y_c), g_p refers to gray values of P equally spaced pixels on a circle of radius R, and s defines a threshold function as follows:

$$s(x) = \begin{cases} 1, & \text{if } x \geq 0; \\ 0, & \text{otherwise.} \end{cases} \tag{2.2}$$

Since some binary patterns can occur frequently in a texture image than others, uniform pattern was introduced (Varma and Zisserman, 2005). A LBP pattern is considered uniform if the binary pattern comprises of two bitwise transitions from 0 to 1 and vice versa. This uses only the uniform patterns and labels the remaining with a single label. $LBP_{P,R}^{u2}$ stands for uniform pattern where (P,R) is the neighbourhood pixels and radius while $u2$ is for uniform pattern.

Another extension of the LBP would be the Bayesian LBP (BLBP) (He *et al.*, 2008). This was implemented to reduce the sensitivity of an image to illumination variations. This uses a framework called filtering, labeling and statistics for texture descriptors. This LBP uses prior and likelihood information and decreases the sensitivity to noise.

Scale Invariant Local Ternary Pattern (SILTP) is introduced to deal with the gray-scale intensity change in complex background (Liao *et al.*, 2010). It has been proven to be robust when there is a sudden change in illumination. Besides that, it has shown that it is robust when a shadow is occluding a region because even though a shadow is present it still conserves the texture information but tends to be slightly darker with a scale factor. However, this LBP version is not considered to be monotonic to gray-scale variations and has longer feature vectors than the original version of LBP.

2.1.2 Local Phase Quantization

Spatial blurring g is expressed as a convolution between the image intensity f and a point spread function (PSF) of the blur h.

$$g = f * h \tag{2.3}$$

In the Fourier domain, for a frequency u this is expresses as:

$$G(u) = F(u).H(u) \tag{2.4}$$

For this, only the phase relation is considered having:

$$\angle G(u) = \angle F(u) + \angle H(u) \tag{2.5}$$

When the PSF h is centrally symmetric, its Fourier transform would be only a real value having

$$\angle H(u) = \begin{cases} 0, & \text{if } H(u) \ \geq \ 0; \\ \pi, & \text{if } H(u) \ < \ 0; \end{cases} \tag{2.6}$$

The shape of H for a regular PSF is approximately close to a Gaussian or a sinc function which ensures low frequencies of $H(u)$ when positive. In this case for values of $H(u)$ positives

$$\angle G(u) = \angle F(u) \tag{2.7}$$

causing to be the image F to be invariant to blur.

LPQ uses a 2D discrete Short-Term Fourier Transform (STFT) that is computed locally. In analogy with the LBP method, LPQ is computed locally more specifically it is calculated at each pixel position within a local neighborhood N_x of an image $f(x)$, where x is a pixel position on the image having $x = (x_1, x_2)$. The local spectrum is computed using a discrete STFT defined by:

$$F(u,x) = \sum_y f(y) w_k(y-x) \exp - j2\pi u^T y \tag{2.8}$$

where $w(x)$ is a window function delimiting the neighborhood N_x. For regular LPQ, w_R is a $N_R by N_R$ rectangle given as $w_R(x) = 1$ if $|x_1|, |x_2| < \frac{N_r}{2}$ and 0 otherwise. The local Fourier is computed at four frequencies which results in a vector:

$$\mathbf{F}(x) = [F(u_1,x), F(u_2,x), F(u_3,x), F(u4,x)] \tag{2.9}$$

where $u_1 = [a,0]^T$, $u_2 = [0,a]^T$, $u_3 = [a,a]^T$and $u_4 = [a,-a]^T$ and a is a small scalar value that satisfies the positivity of $H(u)$. The binarization 2.10of of the phase information is obtained by thresholding the signs of the real and imaginary part of each component found in $\mathbf{F}(x)$.

$$b_j = \begin{cases} 1, & \text{if } g_j \geq 0; \\ 0, & \text{otherwise}; \end{cases} \tag{2.10}$$

where g_j is the j-th component of the vector $G(x) = [ReF(x), ImF(x)]$. This eight binary coefficients b_j are transformed into a integer value using an eight bit coding then are mapped into a histogram of 256 values.

In additional to the regular LPQ, a newer version has appeared giving the blur insensitive feature descriptor a rotational invariance characteristics (Ojansivu *et al.*, 2008). At the low frequency of the Fourier domain, this descriptor is obtained after performing two different processes which are the local characteristic orientation estimation and the quantification into a binary value. This last is then mapped into a histogram of 256 values. LPQ descriptor has been introduced to the FR field by Ahonen T. et al. in their work (Ahonen *et al.*, 2008)

2.1.3 Binarized Statistical Image Features

The Binarized Statistical Image Features (BSIF) descriptor is inspired by the previous texture descriptors already mentioned (LBP and LPQ) and their variants (Kannala and Rahtu, 2012). It performs locally by describing the neighborhood of a center pixel value with a binary value. In order to perform BSIF, a patch P of the size $l \times l$ of the image is needed for local description and especially the need of a linear filter $W'$of the same size $l \times l$. The response of a linear filter is obtained:

$$rlf_j = \sum u, vW_i'(x,y)P(x,y) = \mathbf{w'}_j\mathbf{p}_j \tag{2.11}$$

If n linear filters are used, they can all be stored into a matrix **W'** and all the responses can be computed simultaneously.

$$\mathbf{RLF} = \mathbf{W'p} \tag{2.12}$$

Finally, the binary code is obtained by performing thresholding:

$$b_j = \begin{cases} 1, & \text{if } rlf_j \ > \ 0; \\ 0, & \text{otherwise;} \end{cases} \tag{2.13}$$

b_j is the j-th component of b. This binary code is transformed into a integer value then represented into a histogram .

Table 2.1 Summary of the presented texture descriptors

Texture descriptors	Shared properties	Unshared properties
LBP	Performs locally within a neighborhood of a center pixel. Describes each computed characteristics with a binary code. Maps the integer value from the binary code into a histogram	Manipulation on pixel values. Thresholding is performed by comparing the value of the center pixel by its neighbors' pixel values which produces a binary digits
LPQ		Manipulation on the phase of pixels at low frequency values in a neighborhood. Calculation over 4 frequency points of the blurred image using STFT at the blur invariant range. Thresholding of phase by observing the real and imaginary component of the image resulting into a binary value.
BSIF		Manipulation on the pixel values along with a linear filter of the size of the neighborhood. Thresholding based on the response of the linear filter resulting into a binary value.

2.2 Local Template Matching Techniques

Holistic matching consists in exploiting the whole facial region as the input face recognition. This was abundantly used during the first introduction of the Eigenface approach (Zhang *et al.*, 1997) by using projection techniques such as Principal Component Analysis (PCA) (Wold *et al.*, 1987) subspace for FR process.

A newer technique that involves the use of local regions was introduced which is local matching. There are many methods to perform local matching. As a matter of fact, some proceeded in locating specific regions on the face. Then these regions are used for classification. This method is called the part-base or it can also be called as component-based local matching. This first appeared when the Eigenface technique was extended to a local representation by combining the eigenfaces with other modules such as eigeneyes and eigenmouths (Pentland *et al.*, 1994). Other works are based on scale-invariant feature transform (SIFT) operator for FR (Luo *et al.*, 2007). Local parts of the face are located from the images and are matched. This approach has shown to be quite performing for images that are pre-aligned and faster in terms of matching. However, this technique may present some disadvantages. Indeed, SIFT matching may provide high matching score to non- target individuals rather than real target due to few feature representations of each individual in the gallery for face screening applications.

As a whole, the part-based local matching presents certain disadvantages especially when the faces from the captured scenes are partially occluded and the important key facial features for FR are impossible to locate due to these occlusions.

For this, other research works have focused on local matching involving simple subdivision onto blocks or patches based on the facts that certain local features of the face do not vary with pose, illumination and expressions. As a fact, in (Ahonen *et al.*, 2004) the faces are divided into small regions then are matched locally using the weighted Chi-square method. In that same work, some sub-regions are given importance based on the location of the most distinctive facial features. That is why, weights are assigned on the local regions for local matching. This technique has shown to be more efficient and performing in terms of recognition in comparison with the holistic approach. The research presented in (Martínez, 2002) is based on region subdivision. Unlike the previous technique this is performed by first warping the facial image into a pre-defined face. Once the captured face is fitted and aligned to the standard face , the latter is divided into six local regions. Within these local regions, local matching is performed then combined.

Other local matching techniques do not involve locating certain facial components or dividing the face into sub-regions. Instead, local matching is ensured by combining two techniques: the Gabor filter and the LBP method to perform FR. In this case (Zhang *et al.*, 2005), the face image undergoes several process of Gabor filter process at five different scales and at eight different orientations. As for the LBP operator, it is applied on all 40 obtained Gabor filtered images which gave a pattern called: Local Gabor Binary Pattern Histogram Sequence (LGBPHS) for local matching.

Local matching have shown promising results in FR applications. A comparative study about local matching and how it can be best exploited can be found in (Zou *et al.*, 2007a) and a short survey on partial FR can be found in (Liao *et al.*, 2013).

Speaking of matching, it is considered as a simple approach for classification of pattern. Base on (Jain *et al.*, 2000), template matching falls into the category of similarity/dissimilarity approach. In FR, this allows individuals having the same features are regrouped into one class. The most popular metrics are the *Euclidean Distance* and the *Chi Square measure*.

2.2.1 Distance Measures

Euclidean Distance

This measure tries to assess the minimal distance between two templates. In a still-to-video system, let **m** be the template of the still ROI saved in the gallery and **f** be the template of the probe ROI having the size of N_{feat}. The equation of the Euclidean distance would be:

$$D(\mathbf{m},\mathbf{f}) = \sqrt{\mathbf{f}-\mathbf{m}} = \sqrt{\sum_{i=1}^{N_{feat}} (f_i - m_i)^2} \tag{2.14}$$

When local matching is applied, Euclidean distance metric is performed on each local regions providing a local value or score. These scores are then combined and averaged. Let N_{TP} be the number of total local regions in a ROI and j the corresponding number of the local region having $j = 1..N_{TP}$. The Euclidean distance for local matching would be:

$$D(\mathbf{m},\mathbf{f}) = \sqrt{\sum_{i,j}(f_i - m_i)^2} \tag{2.15}$$

Chi Square Measure

Chi Square (χ^2) metric is also considered to be a distance template matching that estimates the degree of dissimilarity in patterns.

$$\chi^2(\mathbf{f},\mathbf{m}) = \sum_{i}^{N_{feat}} \frac{(f_i - m_i)^2}{f_i + m_i)} \tag{2.16}$$

f and **m** are respectively the templates of the probe and the target ROI saved in the gallery.

If local matching is applied for a N_{TP} specific number of regions, the χ^2 is written as following:

$$\chi^2(\mathbf{f},\mathbf{m}) = \sum_{j}\sum_{i} \frac{(f_{ij} - m_{ij})^2}{f_{ij} + m_{ij})} \tag{2.17}$$

where j corresponds to the number of local regions.

Sometimes during local matching, some regions tend to be more important than the others. For that, weights are attributed and added to the previous equation 2.17 giving a more general expression of the χ^2:

$$\chi^2(\mathbf{f},\mathbf{m}) = \sum_j w_j \sum_i \frac{(f_{ij} - m_{ij})^2}{f_{ij} + m_{ij})} \tag{2.18}$$

where w_j is the attributed weight for region j.

2.3 Illumination Normalization

Changes in ambient illumination, and the resulting variations to facial appearance, are known to significantly deteriorate the performance of FR systems. Accordingly, several techniques have been proposed for illumination invariant FR (Sharma *et al.*, 2014). Zou et al. (Zou *et al.*, 2007b) presented a survey of techniques according to passive and active approaches. *Passive approaches* focus on the visible spectrum images where face appearance has been altered by illumination variations. They include illumination variation modeling, illumination invariant features, photometric normalisation, and 3D morphable model techniques. *Active approaches* employ active imaging techniques to obtain face images captured under consistent illumination condition, or images of illumination invariant modalities. Additional devices (optical filters, active illumination sources or specific sensors) are usually involved to actively obtain different modalities of face images that are insensitive to or independent of illumination change. Those modalities include 3D face information and face images in those spectra other than visible spectra, such as thermal infrared image and near-infrared hyper-spatial image.

There are three main types of techniques to produce illumination invariant facial images under the passive approaches: those applied at the pre-processing, feature extraction and classification levels (Struc and Pavesic, 2011). Pre-processing techniques seek to produce facial images that are free of illumination induced facial variations prior to feature extraction. They can be applied within any FR system, since they make no prior assumption that influences feature extraction or classification procedures. Feature extraction techniques seek to compensate for appearance variations in facial images using descriptors or representations that are stable under different illumination conditions. However, different empirical studies with LBP, Gabor wavelet-based features, and other descriptors have shown (Marcel *et al.*, 2006) that none of these can ensure illumination invariant FR given severe illumination changes. Classification-level techniques compensate for illumination changes according to the type of face model or classifier employed in the FR system. First, some assumptions regarding the effects of illumination on face models or classification procedure are made, and then based on these assumptions, counter measures are undertaken to obtain illumination invariant face models or illumination insensitive classification procedures. Managing the effects of illumination at the feature extraction level is debatable. Classification level techniques may impose difficult requirements on the design data. They may provide the most efficient approach to illumination invariant FR. However, a large training set must usually be acquired under a number of lighting conditions and is computationally expensive. In this section a brief overview on some of the passive illumination normalization techniques is explained.

Most of the existing passive illumination normalization techniques are related to the retinex theory (Land and McCann, 1971). The latter is based in comprehending the main process of the image formation and perception. Mathematically speaking, an image $I(x,y)$ can be presented as a product of the reflectance $R(x,y)$ and the luminance $L(x,y)$

$$I(x,y) = R(x,y).L(x,y) \tag{2.19}$$

$R(x,y)$ consists in the characteristic of the object in the image and it is based on the reflectivity. Meanwhile, $L(x,y)$ is based on the amount of illumination in the image. By this, $R(x,y)$ is the representation of the original image that is not dependent to the illumination (i.e invariant). The luminance is considered to vary slowly with the spatial position and can therefore be estimated as a smoother version of the image $I(x,y)$.

$$ln(R(x,y)) = ln(I(x,y)) - ln(L(x,y)) \tag{2.20}$$

$$R(x,y) = \frac{I(x,y)}{L(x,y)} \tag{2.21}$$

For example the Single Scale Retinex (SSR) algorithm calculates the luminance factor $L(x,y)$ by performing smoothing with a single Gaussian filter into the image $I(x,y)$. The reflectance $R(x,y)$ is calculated with the expression using algorithm in equation . However this technique seems to be insufficient. As a fact, images having large illumination discontinuities may have visible halo on the reflectance. Then a newer extension is the Adaptive Single Scale Retinex algorithm (ASSR) which is based on adapting the process of smoothing using iterative convolution that applies two discontinuity measures: local inhomogeneity and spatial discontinuity (Park *et al.*, 2008). Another extension of this work was the Multi-Scale Retinex (MSR) (Jobson *et al.*, 1997) where in this case multiple Gaussian filters are implemented in terms of widths and the different values of the reflectance for each case are then combined providing a global reflectance $R(x,y)$.

Self Quotient image (Wang *et al.*, 2004a) was proposed as another technique for combating illumination issues. It combines the image processing technique of edge preservation filtering with the theory of retinex. A newer version of this approach was introduced in the same work called Multi-Scale Image (SQI) where instead of using Gaussian filter the use of anisotropic filter was used for smoothing.

The Non-local mean (NL) consists in smoothing an image by computing at each pixel value a weighted average of surrounding pixels. The weight is a similarity function that calculates the tendency of similarity between the neighbouring pixels and the target one having constant variables. As for the newer version of this approach called Adaptive Non-local means (ANL), the smoothing parameter found in the weight function is dependent on local contrast instead of having preselected values In this case, more smoothing is performed when contrast is low at a region and vice versa. Details can be found in (Gross and Brajovic, 2003).

In the past few years recent normalization techniques have appeared such as the "Tan and Triggs" (TT) technique (Tan and Triggs, 2010). It merges multiple approaches: robust illumination normalization that contains series of stages in order to cope up with variations such as local shadowing and highlights. The stages can be resumed in applying Gamma correction as a start then filtering with Difference of Gaussian (DoG), finally performing masking in order to mask out irrelevant variables. Afterwards, local textured base feature extraction Local Ternary Pattern (LTP) which is based on the principle of LBP is employed. Finally, a feature fusion technique is applied. The TT technique shows to perform well when applied to FR especially using LBP feature representation.

2.4 Image Quality Assessment

Image quality is based on a visible distortion that can occur in an image. An example of quality would be blurriness, color contrast, Gaussian noise, etc. In order to assess quality in an analytical way, an accurate and meaningful as the human perception quantification of these distortions is needed. In literature, Image Quality Assessment (IQA) has been categorized into different groups. Perhaps the most popular distinction of these technique would be into two major groups: full-reference and no reference IQA.

2.4.1 Full-Reference

The full-reference IQA is in need of a reference original image to be able to compare the input image and estimate its quality based on the original one. These full-reference IQA can also be divided into sub-groups based on the manner of estimating the distortion in an image. It can be a mathematical based such as Squared Mean Error (MSE) (Tuchler *et al.*, 2002) which is a way to assess quality relatively to the reference image by calculating the error signal (difference) and then doing its average.

$$MSE(x,y) = \frac{1}{N}\sum_{i=1}^{N}(x_i - y_i)^2 \tag{2.22}$$

where x_i and y_i are respectively the input image and the reference image. The signal error is expressed in $e_i = x_i - y_i$. Another mathematical and full-reference IQA is the peak to Signal to Noise Ratio (PSNR) (Huynh-Thu and Ghanbari, 2008) which is inversely proportional to MSE.

$$PSNR = 10log_{10}\frac{L^2}{MSE} \tag{2.23}$$

Where L is the dynamic range of pixels. The advantage of PSNR is that it is useful for different dynamic range of an image.

Aside from these mathematical approach, other full-reference techniques have also been introduced such as the Structural similarity Index (SSIM) (Wang *et al.*, 2004b). This method is a measure of structural change of information between the original image and the distorted one.

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_x y + C_2)}{(\mu_x^2 + \mu_y^2 + C1)(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{2.24}$$

where $\mu_x, sigma_x$ and $\mu_y, sigma_y$ are respectively the mean intensity and standard deviation of the image x and the image y and $\sigma_x y$ is their cross validation. As for $C1$ and $C2$ they are constants having a low value to omit the problem having the denominator going to zero.

In literature full-reference IQA methods are considered to provide fidelity according to the original image instead of global quality.

2.4.2 No-Reference

No-reference IQA, also called "blind" IQA is a measure that is capable to return a quality estimation without having a reference image. Unlike the full-reference one, this type is more useful for applications where the reference images are not available. An example of these applications would be memory card management of a digital camera where this last mentioned must be able to assess which of the photos are of good quality for storage and which are not.There are a lot of techniques for no-reference IQA depending on the type of distortion. In this part we are going to present some facial related no-reference IQA which are : head pose, contrast and sharpness.

Head Pose Estimation

It is important to mention that IQA for head pose is divided into two types: IQA involving facial features and IQA dealing with the face as a whole. The first type consists in identifying specific features from the facial image then calculating certain measures. For example in a given facial image, the position of the eyes is located then the distance between both eyes is calculated to estimate the pose of the face. This category of pose quality does not always work very well and not reliable especially in conditions where the rotation of the head is very huge that it is hard to locate both eyes. For the second type, it deals with the face as a whole element. In (Abdel-Mottaleb and Mahoor, 2007) a QM for Head Pose is proposed. IQA involves facial features. To assess the facial image quality, three facial feature points were located: the center of the eyes and the mouth and use an algorithm for skin color discrimination for the left part S_L and the right part S_R.

$$pose = \frac{S_L - S_R}{\min(S_L, S_R)} \tag{2.25}$$

Another head pose estimation was proposed in (Nasrollahi and Moeslund, 2008) where the whole face image is exploited instead of locating specific features as the previous technique. In order to estimate the head pose series of tasks are needed to be accomplished. The first task is to calculate and locate the coordinates of the center of the mass of the image using the equations below.

$$xm = \frac{\sum_{i=1}^{N} \sum_{j=1}^{M} ib(i,j)}{A} \tag{2.26}$$

$$ym = \frac{\sum_{i=1}^{N} \sum_{j=1}^{M} jb(i,j)}{A} \tag{2.27}$$

(x_m, y_m) are the coordinates of the center of mass of the image. $b(i, j)$ is the binary version of the original image I. MxN is thr size of the image abd A is the area of the detected region.

Then, the next task is to locate the coordinates of the center of the face region detected.

$$x_c = \frac{x_2 - x_1}{2} \tag{2.28}$$

$$y_c = \frac{y_2 - y_1}{2} \tag{2.29}$$

x_1 and x_2 are respectively the most right and the most left pixel of the face region. y_1 and y_2 are respectively the lowest and the most top pixel of the detected face as seen in (Nasrollahi and Moeslund, 2008) 2.2. The knowledge of these values involves the usage of Gradient analysis in order to collect these positions from the facial image.

Once these coordinates (x_m, y_m) and (x_c, y_c) are calculated, it is now possible to assess the

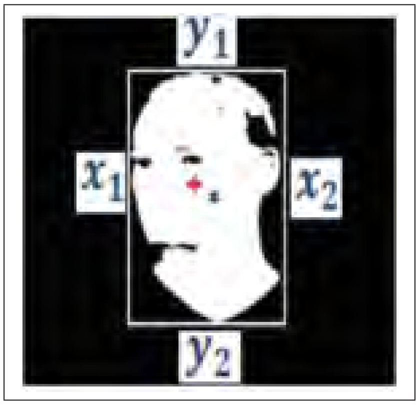


Figure 2.2 Head pose estimation with center of mass (+) and center of the region (*) Taken from Nasrollahi and Moeslund (2008)

position of the head as:

$$pose = \sqrt{(x_c - x_m)^2 + (y_c - y_m)^2} \tag{2.30}$$

The closer this value to zero the closer the face to be frontal.

Sharpness

In video surveillance, the individuals in the scene are all moving. So, it easy to have an affected image by motion blur where the facial image becomes blurred and useless for further treatment. It is then obvious to define a sharpness feature. In (Weber, 2006), the approach consists of applying a low-pass filter to the facial image first. Afterwards, calculate the sharpness of the face by calculating the average value of pixels of the original image $I(x,y)$ and the filtered image $lfI(x,y)$.

$$sharpness = |lfI(x,y) - I(x,y)| \tag{2.31}$$

The higher this value the better and the sharper the image is.

Sharpness is also defined in the QM standards for facial images ISO/IEC (Sang *et al.*, 2009). The QM is based on the evaluation of the frequency domain DCT. First thing to be done is to apply the IDCT operation on the original image $R(x,y)$. So the sharpness would be expressed as the equation 2.32

$$sharpness = \frac{1}{M \times N} \sqrt{\sum i = 1M \sum i = 1N (R(i,j) - I(i,j))^2} \tag{2.32}$$

Contrast

Contrast is an important measure to distinguish relative differences in terms of the intensity of an image. An expression of this image quality is given in (Abaza *et al.*, 2012).

$$C_{RMS} = \sqrt{\frac{\sum_{x=1}^{M} \sum_{y=1}^{N} [I(x,y) - \mu]^2}{MN}} \tag{2.33}$$

C_{RMS} is the contrast value. $I(x,y)$ is a test image of size $M \times N$. μ is the mean value of intensity of the image. I_{min} and I_{max} are respectively the minimum and maximum intensity values.

Illumination/Brightness

Images that are poor in illumination tend to be useless for further processing especially for recognition. This subsection provides a simpler way to assess brightness of an image without the use of any reference image.

This was highlighted and used by (Nasrollahi and Moeslund, 2008). They assumed that the region of the image is small and that the average value of the pixels values of that region is the illumination measure.

2.5 Domain Adaptation

In pattern recognition, systems are often confronted to a dilemma: the data distribution during training also called the source domain is different from the distribution of the data during testing. This issue is considered in literature as domain shift or domain adaptation problem. The latter is a specific case of transfer learning (TL) (Pan and Yang, 2010). Domain adaptation can be specifically found in uncontrolled settings of systems. It is firstly introduced in the fields of natural language processing (NLP) and data mining (Pan *et al.*, 2011). Then, it influenced other communities especially object recognition Gopalan *et al.* (2011) and face recognition (Ho and Gopalan, 2014) (Kan *et al.*, 2014).

When referencing to domain adaptation, two main categories can be discussed based on the nature of the target data. For a given labeled data from the source domain, if the data from the target domain are also labeled then this falls onto the group of semi-supervised domain adaptation. Whereas, if the data from the target domain are unlabeled then this category falls onto the group of unsupervised domain adaptation. This last category is more likely to occur for real-life visual recognition and especially face recognition in video surveillance applications.

In this section, we are going to provide existing domain adaptation solutions in general fields and also those related to face recognition.

Domain adaptation techniques were first implemented in semi-supervised scenarios. For example, a technique that is based on common features or common subspace was proposed in (Daume III and Marcu, 2006). The key idea is to extract common components from the source domain and the target domain to create a new subspace for classification. Another approach of domain adaptation is based on classifiers. In (Dai *et al.*, 2007) a semi-supervised variants is used using naive Bayes classifiers or even using auxilary classifiers like support vector machines (SVMs) in (Duan *et al.*, 2009). In object recognition, domain adaptation is performed using learning domain shifts through a metric learning (Saenko *et al.*, 2010).

However, recent systems are opting for unlabeled data for the source domain due to extensive labeling costs. Therefore, unsupervised domain adaptation solutions are adopted.

Some direct techniques for unsupervised domain adaptation is based on instance-weighting approach. The idea is to re-weight the source domain data samples whether to reduce the gap between both distribution or to encourage and promote certain components. Works presented in (Zadrozny, 2004) and (Sugiyama *et al.*, 2008) follow this technique. In addition, feature-based or sub-common space was also implemented for unsupervised domain adaptation. In (Gopalan *et al.*, 2014), domain shift for object recognition is ensured by creating an intermediate data representation that consist mainly in minimizing the geodesic patch between both domains. Besides that, for a face recognition system, a common subspace is also implemented in (Kan *et al.*, 2014) where this common space is created based on the knowledge shared between the source and the target domain and also based on the specific knowledge in the target domain. Moreover, in (Blitzer *et al.*, 2006) pivot features are used to extract the frequently appearing features in both domains.

Other than creating subspace for both domains, some unsupervised domain adaptation techniques focused on implementing auxiliary classifiers to cope up with the existing domain shift. In (Vázquez *et al.*, 2012), an unsupervised domain adaptation for pedestrian detection is proposed to adapt the data distribution obtained from the virtual world training set at the source domain and the data distribution at the target domain obtained from the real world set.

Dictionary-based domain adaptation has also gained a lot of attention these past few years thanks to the efficiency found in sparse representation for recognition. In (Shekhar *et al.*, 2013) a common dictionary is learnt representing both source and target domains. To overcome domain shift in terms of illumination variations and pose, a method is proposed in (Qiu and Chellappa, 2013). It consists in categorizing each domain shift into different dictionaries then proceed to learning these obtained dictionaries.

In this chapter, a detailed survey of techniques is provided. State-of-the-art local texture descriptors such as LBP, LPQ and BSIF and some methods for local matching are presented. Image quality metrics whether non-reference or full reference metrics are described. Finally, the importance of domain adaptation is highlighted and existing techniques are surveyed.

CHAPTER 3

EXPERIMENTAL METHODOLOGY

The experimental protocol of each contribution in this Thesis is presented in their corresponding chapters. In this chapter we are only going to provide the common methodology in both works.

3.1 ChokePoint Dataset

Both implemented works presented in this thesis are trained and tested with the Chokepoint dataset (Wong *et al.*, 2011). The latter is designed to mimic real-world surveillance conditions. This dataset consists of 25 subjects (19 males and 6 females). Each individual walks through several portals. The videos are recorded by an array of three cameras placed above each portal Figure 3.1. In consequence, the captured faces may present various distortions such as illumination conditions, head pose, blurriness, etc.

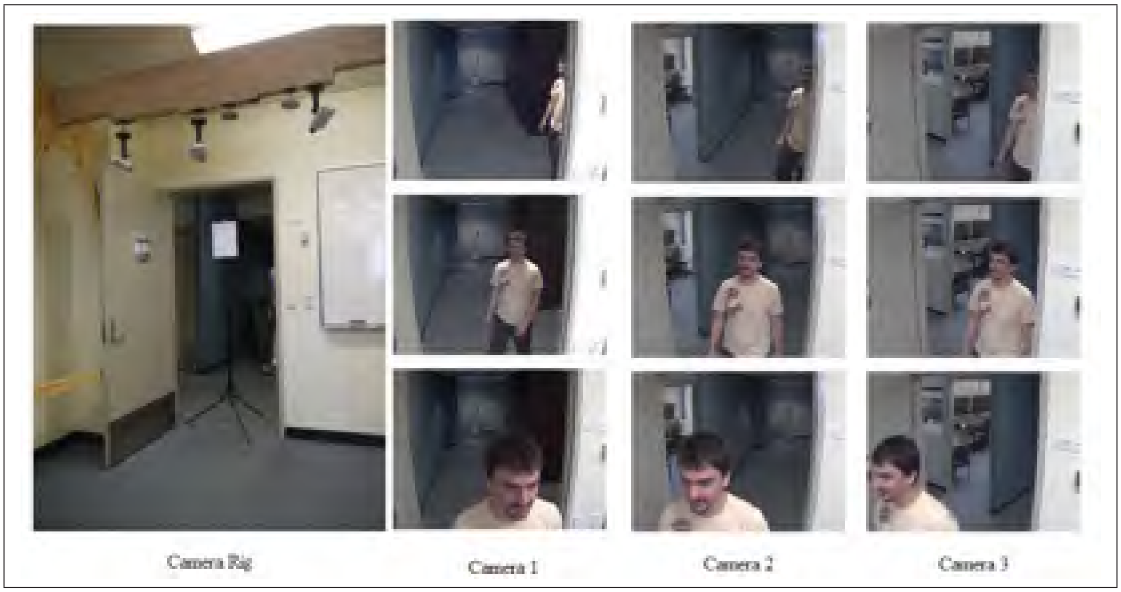


Figure 3.1 Recording setup used for the ChokePoint dataset
Taken from Wong *et al.* (2011)

The dataset contains 2 portals (P1 and P2), 2 states (E: entering and L: leaving), 4 sequences (S1, S2, S3 and S4) for each camera (C1, C2 and C3). Figure 3.2 shows the different recording conditions and their naming.

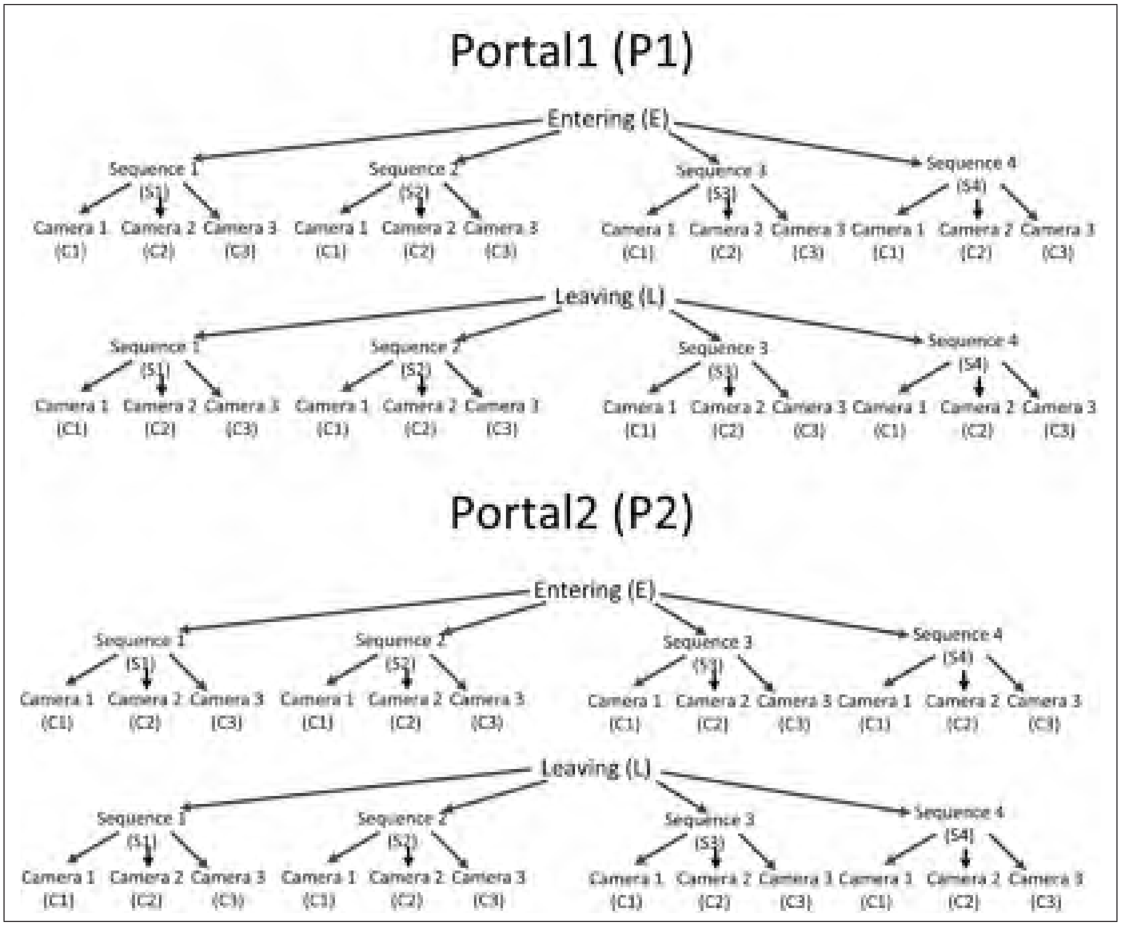


Figure 3.2 Namings of videos according to the recording conditions in the chokePoint dataset

For more challenging real-world surveillance issues, the dataset offers additional sequence (S5) which was recorded having a crowded scenario. This sequence was not exploited in this Thesis. Added to that, the Chokepoint dataset caters to its users high resolution single photos of the 25 subjects. These still photos are used as reference images for the watchlist system's training.

For both contributions presented in this Thesis, the ChokePoint dataset was used to prototype the proposed still-to-video FR systems. However, each work has its own manipulation of the dataset (e.g number of video sets taken). That is why, the dataset exploitation is presented in its corresponding chapter.

3.2 Difficulties found in ChokePoint Dataset

In this section, image quality assessment (IQA) metric explained in Chapter 2 Section 2.4 are applied on a sample video taken from the presented dataset. The goal of this part is to describe the challenging variations of some factors that can be found in the ChokePoint dataset.

Video P1E_S1_C1 is selected as a video sample for this section. The values presented below are normalized within all the values of a face trajectory in the video sequence. In other terms, if V_{IQA} is an IQA value of an image in a face trajectory then the normalized value $V_{IQA_{norm}}$ would be:

$$V_{IQA_{norm}} = \frac{V_{IQA}}{V_{IQA_{max}}} \tag{3.1}$$

An individual from the video scene walks through a hall way then passes through a portal on which the video surveillance camera is mounted.

Facial trajectory of individual # 1 presents variations in terms of headpose. As we can see in Figure 3.3, the first frames representing the individual are not likely to be frontal which causes the pose estimation quality metric to be low. However, as the individual gets half way, more specifically, closer to the camera's view point, the position of the head of the individual appears to be more frontal resulting in a high head pose estimation quality score. Then as the individual goes through the portal and almost under the camera the faces seems to be compressed. Therefore, head pose quality scores are low in the last few frames.

In the beginning of the facial trajectory presented in Figure 3.4 better lighting is available because in the video P1E_S1_C1 artificial lighting are placed at the head of the hall way.

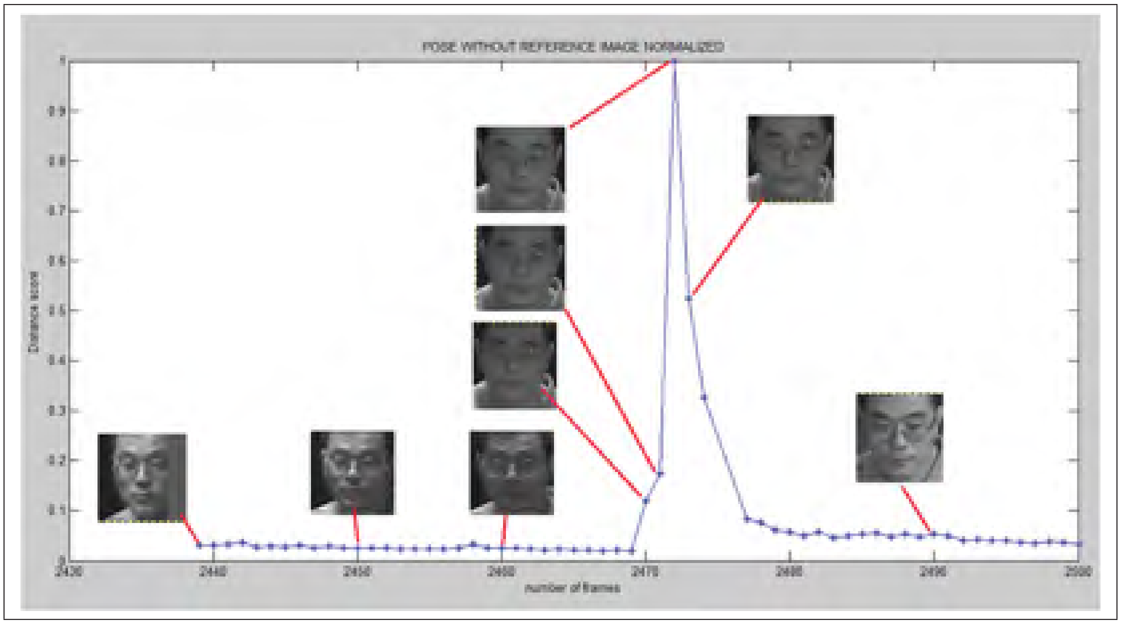


Figure 3.3 Pose variation using Equation 2.30 over a face trajectory of individual #1 in P1E_S1_C1 video of the chokePoint datatset

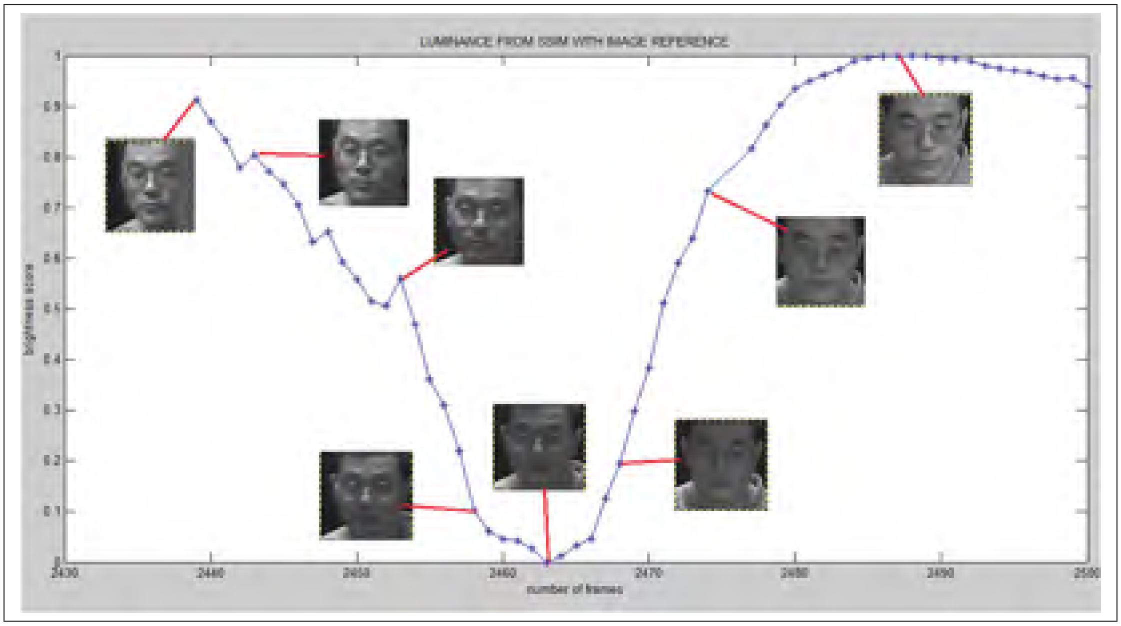


Figure 3.4 Illumination using the luminance component of the SSIM in Equation 2.24 variation over a face trajectory of individual #1 in P1E_S1_C1 video of the chokePoint datatset

But as the person walks through the midway of that hall way and gets closer to the camera's viewing point, lighting tends to be poor which causes a bad illumination quality score for the most frontal captured faces. Once the individual goes through the portal under the camera lighting gets better because of the presence of artificial lighting at the beginning of the door.

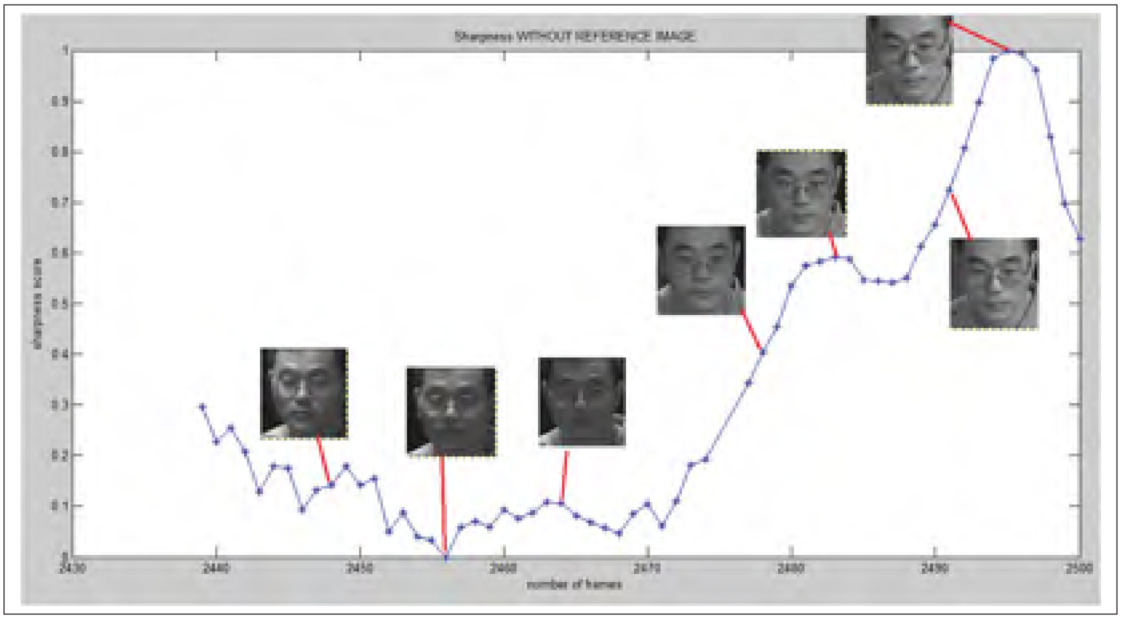


Figure 3.5 Sharpness variation using Equation 2.31 over a face trajectory of individual #1 in P1E_S1_C1 video of the chokePoint datatset

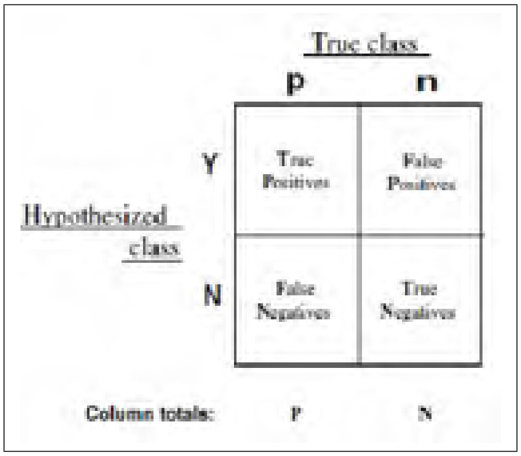
Figure 3.5 shows the variation of sharpness of the images in a trajectory. The faces at the first few frames show a lot of of blurriness which results to low sharpness quality score. As it gets nearer to the portal, the focus of the camera enhances. Thus, the facial images seems sharper which explains the increase of sharpness quality score over the trajectory.

The figures above prove the difficulties (variations in illumination, pose, etc.) found in a video surveillance scenario (especially in the ChokePoint dataset) which makes the recognition task more difficult especially when a single still reference image is available for matching.

3.3 Performance Measures

Performance measures are crucial in evaluating the quality of response and the performance of a classifier. There are many existing measure in literature depending on the type of classification performed: binary, multi-class, hierarchical, etc. Details about the classification tasks and some existing performance metrics adapted to each classification method can be found in (Sokolova and Lapalme, 2009).

A classification model consists in mapping predicted class instances to the true ones. This defines a matrix called confusion matrix 3.6 containing the number of correctly and incorrectly classified samples for each class.



		True class	
		p	n
Hypothesized class	Y	True Positives	False Positives
	N	False Negatives	True Negatives
Column totals:		P	N

Figure 3.6 Confusion matrix
Taken from Fawcett (2006)

Based on this matrix, there are four possible outputs:

- If the sample is predicted to be positive and it is classified as positive. It is then counted as true positives (TP);
- If the sample is predicted to be positive then after classification it turned out to be negative then it is considered as false positive (FP);

- If it is predicted as negative then turned out to be negative. It is counted as true negative (TN);
- Finally, if it is predicted as negative and turned out to be positive then it is counted as false negative (FN).

With this confusion matrix evaluation measures for classifiers can be put out. One of the most used measures would be the accuracy rate (Acc). This permits the evaluation of the effectiveness of a classifier based on the number of correct predictions.

$$Acc = \frac{TN + TP}{FN + FP + TN + TP} \tag{3.2}$$

Another common used measure would be the error rate (Err) having it to be the inverse of Acc.

$$Err = 1 - Acc \tag{3.3}$$

The true positive rate (TPR) or also called recall and the specificity (Spe) can also estimate the efficiency of a classifier in a two-class classification problem. TPR is the proportion of positive and correctly classified samples over the total positives.

$$TPR = R = \frac{TP}{P} = \frac{TP}{TP + FN} \tag{3.4}$$

The Spe is the proportion of negative samples that are correctly predicted as negatives

$$Spe = \frac{TN}{FP + TN} \tag{3.5}$$

The precision (P) is a measure which estimates the probability that a positive prediction is correct.

$$P = \frac{TP}{TP + FP} \tag{3.6}$$

All of the above mentioned measures contribute in assessing a classifier's performance. However, in this thesis, a concentration on the use of Receiver Operating Characteristics (ROC) as another evaluation performance measure for classification problem which relates the recall and the specificity. This evaluation approach has been given in details by Fawcett (Fawcett, 2006)

3.3.1 Receiver Operating Characteristic Curve Analysis

ROC curves have been adapted in signal theory detection then became more and more popular in analyzing and providing medical decisions. The first adaptation of ROC curves in machine learning and pattern recognition was with (Spackman, 1989). Since then, the usage of these curves in these fields has been more and more abundant. A ROC curve is a 2D representation that is obtained by plotting the FPR on the x axis and the TPR on the y axis. In order to assess the classifier's efficiency, a deeper knowledge about the ROC space and specific curves on that space is crucial. A point in a ROC space is represented as a pair of (FPR,TPR). The classifier represented as the point (0,0) shows that it does not commit FP errors and also never increases its TP. As for the point (0,1) represents the best case of classification task where the classifier does not commit FP errors but gain a lot of TP.

In general, a classifier is considered to be better when the TPR is higher than the FPR. Meaning that, a classifier represented as a point in a ROC space should be located in the high left side (north west) of the space. Classifiers can perform random guessing task. This is presented as the diagonal line having the equation of $FPR = TPR$. In this case the classifier can't discriminate between the classes. As a matter of fact, if the prediction of positive classes (TP) is 50% the down side is that the FP errors are also at 50%. Thus, in a ROC space, avoiding this random curve is very important. This random curve divides the space into two triangle areas. Any classifier below this diagonal or on the lower triangle region is said to be a non efficient. However, any ROC curves above the random curve (diagonal) are considered to be better.

It is certain that ROC curves may provide better comparison on the classifier's performance. However, it is more likely to quantify this comparison by using a single value representing the percentage of performance. This scalar is called Area Under the ROC Curve (AUC) introduced in (Bradley, 1997). The AUC is a portion of the area of the unit square so its value is between

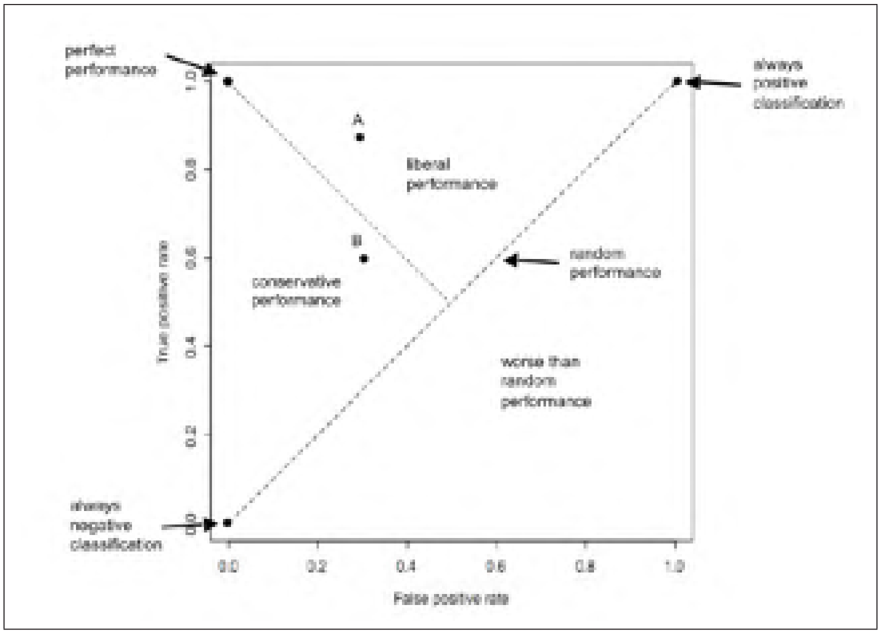


Figure 3.7 Regions of ROC curves
Taken from Hamel (2009)

0 and 1. A random classification task would have an AUC equals to 0.5 since it is dividing the unit square into two equivalent triangles. Knowing that a "good" classifier must have a ROC curve higher than the diagonal of the ROC space (random classification). Its AUC must also be higher than the random 0.5. In certain cases, ROC curves tend to perform better in a portion of the ROC space compared to other classifier's ROC graphs. In other words, a classifier having a low value of AUC can be assessed "good" compared to another classifier that has higher value of AUC in a restraint region. For this, it is preferable to limit the percentage of FPR then calculate the new area under the partial curve called partial Area under the curve (pAUC).

3.3.2 Precision-Recall Curve Analysis

Precision-Recall (PR) curves are used when faced to highly imbalanced classes. Indeed, it may expose differences between the performance of classifiers that are mostly not apparent to ROC curves. Sometimes classifiers present almost similar ROC curves at the ROC space. However, to perfectly assess the performance it would be clearer to see the difference once the Precision-Recall curve is projected. For ROC curves, when a huge variation of false positives FP may not lead to a huge variations in the fpr that is why it may seem to us that the ROC responses are almost similar. But since Precision-Recall compares the TP with FN and FP rather than TN. Consequently, this captures the effects of having larger values of negative samples on the system's performance. The goal in the ROC space is to achieve curves situated at the upper-left corner of the space. Per contra, the goal in the PR space is to achieve curves situated at the upper-right corner of the PR space.

Some direct relationship has been shown between the ROC curves and the Precision-recall curves (Davis and Goadrich, 2006).

- ROC curves and Precision-Recall curves share the same basis of confusion matrix;
- If a curve in a ROC space dominates then it should dominate in the Precision-Recall space and vice-versa.

In analogy to the AUC of ROC curves, for PR curves Area under the PR (AUPR) curve can also be calculated (Boyd *et al.*, 2013). The higher this value is the better the performance of the classifier.

Class imbalance problem, where the number of data in a positive class is less than the total number of data in a negative class, is very common in still-to-video FR system especially for watchlist screening applications. Therefore, for better assessment of these types of systems in terms of performance, it is preferable to give importance on the values of AUPR since both parameters precision and recall do not consider TN.

In this Thesis, the performance of the classifier is shown on the ROC space analysis and PR space analysis. A higher importance is accorded to AUPR values due to the presence of class imbalance in the implemented application.

In this chapter, a description of the ChokePoint dataset is given. The challenges found in this dataset are shown using image quality metrics. Finally, a detailed explanation of the performance measures implemented for the performance assessment of a system is provided.

CHAPTER 4

ON THE EFFECTS OF ILLUMINATION NORMALIZATION WITH LBP-BASED WATCHLIST SCREENING

Recent developments in image analysis and recognition have shown that LBP provides a simple yet powerful approach to represent faces for human computer interaction, recognition, security and surveillance (Pietikäinen *et al.*, 2011). As presented in the first section of Chapter 2 of this thesis, LBP, a gray-scale invariant texture operator has become a well-established feature extraction technique in FR (Ahonen *et al.*, 2006) thanks to its discriminative power, tolerance to monotonic gray-scale changes and computational efficiency. However, it is well known that LBP and other variants are sensitive to severe illumination changes. Indeed, variations in facial appearance caused by changes in ambient illumination conditions play an important role in the performance of any FR system applied to video surveillance. It has been shown that face images of different individuals appear to be more similar than images of the same individual under severe illumination variations (Štruc and Pavešić, 2011).

Several techniques have been proposed in literature for illumination invariant FR (Liu *et al.*, 2005). Zou et al. (Zou *et al.*, 2007b) presented a survey of techniques to manage variations in face appearance due to illumination changes using passive and active approaches. Passive approaches focus on the visible spectrum images, where face appearance has been altered by illumination variations, while active ones employ active imaging techniques to capture face images under consistent illumination conditions, or images of illumination invariant modalities.

Among passive techniques, some are specialized at either the pre-processing, the feature extraction or the classification level (Štruc and Pavešić, 2011). At the preprocessing level, normalization techniques seek to transform facial images such that facial variations due to illumination are removed. These approaches can be adapted for use with any FR algorithm. Techniques at the feature extraction technique seek to achieve illumination invariance by using features or representations that are stable under different illumination conditions. However, some empirical studies have shown that no descriptor can ensure illumination invariant FR in the presence of severe illumination changes. Finally, classification level techniques compensate for the illumination based on the type of face model or classifier employed for FR.

4.1 Goals and Contributions

The main motivation behind this work is to implement a still-to-video FR system that is not only robust to SSPP problem but also tough on illumination issues that degrades a FR system.

We are applying IN at the pre-processing for still-to-video FR systems especially for watchlist screening where only a single image is available for training the system. Added to that, state-of-the-art techniques were selected (refer to Table 4.2).

The main goal of this work is to pinpoint the most suitable passive IN techniques that boosts a still-to-video FR by applying N_{IN} selected illumination techniques (see Table 4.2) by assessing their performance on a FR system.

Some texts, figures and results are published in a conference paper (Amara *et al.*, 2014).

4.2 Framework and Algorithms

The framework of this study would be quite similar to the generic framework of a still-to-video FR presented in Figure 1.1. Whereas, the only difference would be an additional module for pre-processing ROI images with illumination invariance techniques. The block diagram in Figure 4.1 shows the modules implemented in order to perform the desired goal of performance comparison.

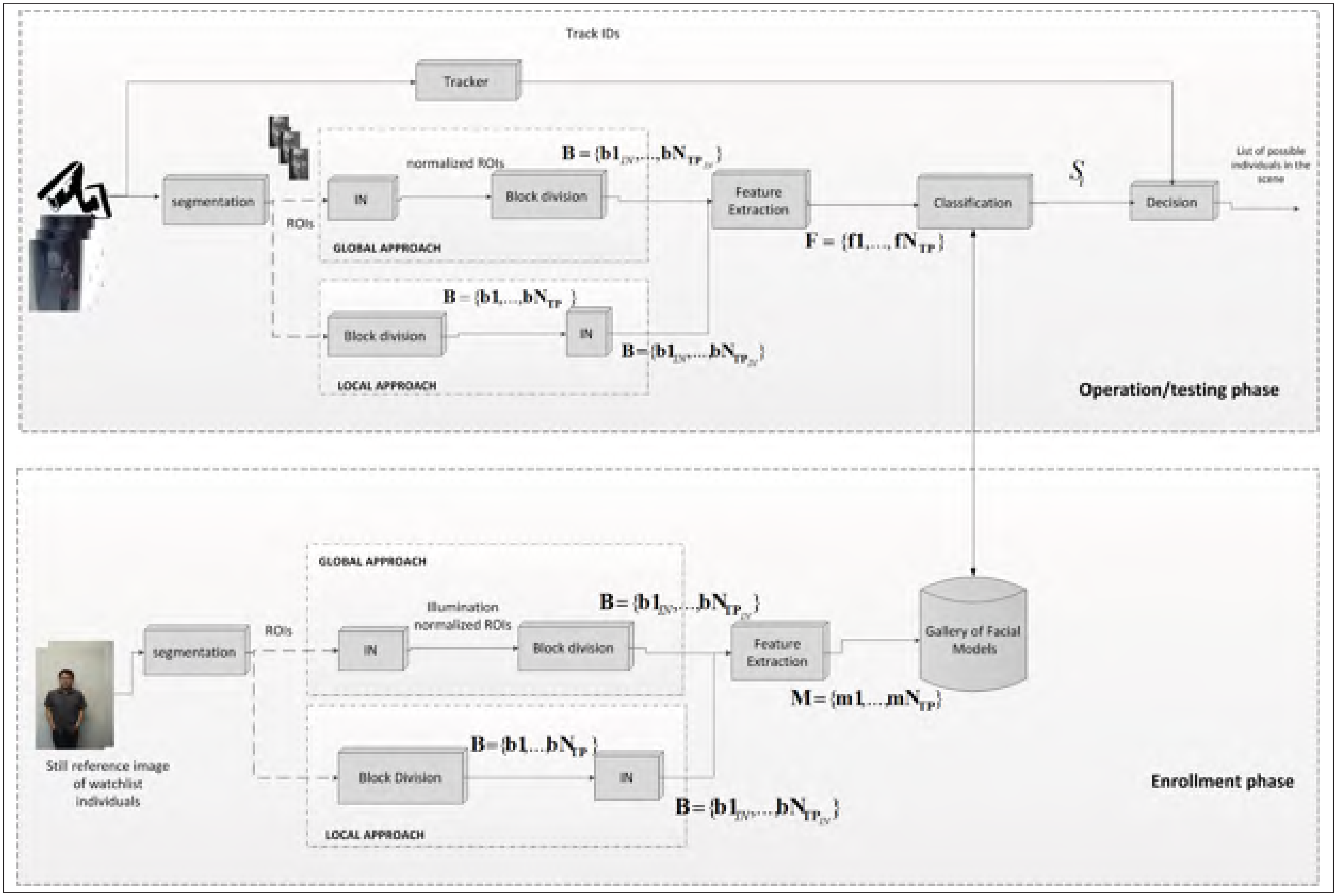

Track IDs
Tracker
segmentation
ROIs
IN
normalized ROIs
Block division
GLOBAL APPROACH
$\mathbf{B} = \{\mathbf{b1}_{IN},\ldots,\mathbf{bN}_{\mathbf{TP}_{IN}}\}$
Block division
$\mathbf{B} = \{\mathbf{b1},\ldots,\mathbf{bN}_{\mathbf{TP}}\}$
IN
$\mathbf{B} = \{\mathbf{b1}_{IN},\ldots,\mathbf{bN}_{\mathbf{TP}_{IN}}\}$
LOCAL APPROACH
Feature Extraction
$\mathbf{F} = \{\mathbf{f1},\ldots,\mathbf{fN}_{\mathbf{TP}}\}$
Classification
S
Decision
List of possible individuals in the scene
Operation/testing phase
Still reference image of watchlist individuals
segmentation
ROIs
GLOBAL APPROACH
IN
Illumination normalized ROIs
Block division
$\mathbf{B} = \{\mathbf{b1}_{IN},\ldots,\mathbf{bN}_{\mathbf{TP}_{IN}}\}$
Block Division
$\mathbf{B} = \{\mathbf{b1}\ldots\mathbf{bN}_{\mathbf{TP}}\}$
IN
$\mathbf{B} = \{\mathbf{b1}_{IN},\ldots,\mathbf{bN}_{\mathbf{TP}_{IN}}\}$
LOCAL APPROACH
Feature Extraction
$\mathbf{M} = \{\mathbf{m1},\ldots,\mathbf{mN}_{\mathbf{TP}}\}$
Gallery of Facial Models
Enrollment phase

Figure 4.1 Framework of the still-to-video FR system with illumination invariance

Algorithm 4.1 Illumination invariant still-to-video FR: Enrolment phase

```
Input: Still images $\{\mathbf{I\_still}_t : t = 1,...,Nb_{ind}\}$ from dataset.
Output: Gallery-face-models per IN technique $\{\mathbf{M}_t^{T_i} : t = 1,...,Nb_{ind}\}$ of target
        individuals
for each $\boldsymbol{I\_still}_t$ with $t = 1,...,Nb_{ind}$ do
    // Segmentation
    Apply Face detection algorithm to detect facial ROIs, $ROI_t^{'}$
    Resize $ROI_t^{'}$ into $N_{size} \times N_{size}$ pixels.
    Perform grey-scale transformation giving a resized and grey-scaled facial capture:
    $ROI_t$

    //pre-processing module
    for each technique $T_i$ with $i = 1,...,N_{IN} + 1$ do
        //Global approach Apply IN technique $T_i$ on $ROI_t$ producing $IN\_ROI_t^{T_i}$
        Divide into $N_{TP}$ uniform and non overlapping blocks producing set of blocks
        $\mathbf{b}_{IN_t}^{T_i} = \{\mathbf{b}1_{IN_t}^{T_i},...,\mathbf{b}N_{TP}{}_{IN_t}^{T_i}\}$

        // Local approach Divide $ROI_t$ into $N_T P$ uniform and non overlapping blocks
        producing set of blocks $\mathbf{b}_t = \{\mathbf{b}1_t,...,\mathbf{b}N_{TP}\}$. for each component of $\boldsymbol{b}_t$ do
            Apply IN technique $T_i$ producing normilized local patches
            $\mathbf{b}_{IN_t}^{T_i} = \{\mathbf{b}1_{IN_t}^{T_i},...,\mathbf{b}N_{TP}{}_{IN_t}^{T_i}\}$
        end
        // Feature Extraction and saving into gallery
        for each component of $\boldsymbol{b}_{IN_t}^{T_i}$ do
            Apply feature extraction. Giving a set of concatenated features
            $\mathbf{M}_t^{T_i} = \{\mathbf{m}1_t^{T_i},...,\mathbf{m}N_{TP}{}_t^{T_i}\}$
            Save $\mathbf{M}_t^{T_i}$ in the gallery.
        end
    end
end
```

As it can be seen, the additional module has been applied before performing FE technique in both phases (enrolment and operation phase). Prior to experiments, the single reference still image of the watchlist individuals (refer to dataset section in Chapter 3.1) are of high-quality mug-shots and are used to design the face model of each person for this still-to-video FR system.

Algorithm 4.2 Illumination invariant still-to-video FR: testing phase

```
Input: Input frames $\{\mathbf{I}_r : r = 1,...,R\}$ from dataset. Gallery-face-models per IN
        technique $\{\mathbf{M}_t^{T_i} : t = 1,...,Nb_{ind}\}$ of target individuals enlisted in the gallery
Output: Similarity scores per technique $S_t^{T_i}$
for each technique IN technique $T_i$ with $i = 1,...,N_{IN} + 1$ do
    for each $\mathbf{I}_r$ with $r = 1,...,R$ frames do
        // Segmentation module
        Apply face detection to detect $ROI_r$ corresponding to faces in a frame $r$

        // Tracker module
        Initialize tracker for detected $ROI_r$ giving it a TrackID

        //GLOBAL APPROACH
        // Pre-processing technique
        Apply IN/baseline technique $T_i$ onto $ROI_r$ producing normalized ROI, $ROInorm_r$
        // Block division
        Divide $ROInorm_r$ into $N_{TP}$ uniform and non overlapping blocks
        producing $\mathbf{b}_{IN_r}^{T_i} = \{\mathbf{b}1_{IN_r}^{T_i},..,\mathbf{b}N_{TP}{}_{IN_r}^{T_i}\}$ blocks
        //LOCAL APPROACH
        // Block division
        Divide $ROI_r$ into $N_{TP}$ uniform and non overlapping blocks
        producing $\mathbf{b}_r^{T_i} = \{\mathbf{b}1_r^{T_i},..,\mathbf{b}N_TP_r^{T_i}\}$ blocks // Local IN application for each
        component of $\boldsymbol{b}_r^{T_i}$ do
            Apply IN/baseline technique $T_i$ producing $\mathbf{b}_{IN_r}^{T_i} = \{\mathbf{b}1_{IN_r}^{T_i},..,\mathbf{b}N_{TP}{}_{IN_r}^{T_i}\}$ blocks
        end
        // Feature Extraction
        for each component of $b_{IN_r}^{T_i}$ do
            Apply feature extraction technique producing set of concatenated features
            $\mathbf{F}_r^{T_i} = \{\mathbf{f}1_r^{T_i},...,\mathbf{f}N_{TP}{}_r^{T_i}\}$
        end
        //Local matching
        for each face-model in the Gallery $\{\boldsymbol{M}_t^{T_i} : t = 1,...,5\}$ using the IN technique $T_i$
        do
            Compute similarity scores using Euclidean distance TM between $\mathbf{F}_r^{T_i}$ and
            $\mathbf{M}_t^{T_i}$ producing scores per individual and per IN technique $T_i$: $S_t^{T_i}$
        end
    end
end
```

The enrolment of each target individual t involves isolating the facial ROI from the reference still image by performing face detection algorithm and converting it into gray scale then cropping it to a common size of $N_{size} \times N_{size}$ pixels. These processes are done at the **segmentation** module of the system. At the **pre-processing** level, two possible choices are available:

- Global approach where a number of illumination normalization techniques N_{IN} are firstly applied (see Table 4.2) on the global image and producing different representation of normalized ROIs for each individual. Then **block division** is performed. Each representation of an ROI undergoes a division into $N_p \times N_p$ non-overlapping patches, having by then, a total of N_{TP} blocks ($b_{IN_i}, i = 1..N_{TP}$);
- Local approach where the obtained ROI from the segmentation module of the system undergoes firstly **block division** provding a set of blocks ($b_i, i = 1..N_{TP}$). Afterwards, IN is applied locally on each local patch b_i of the image giving a set of blocks $\mathbf{B}_{IN} = \{b_{IN_i}, .., b_{IN_{N_{TP}}}\}$

Feature extraction technique is applied on each block b_{IN_i}. Afterwards, all N_p feature vectors from one image ROI and one representation are concatenated. N_{feat} features are extracted from each patch then assembled into a ROI pattern for matching. So, overall a feature vector for one ROI contains $N_{Tfeat} = N_{feat} \times N_{TP}$ giving $\mathbf{M} = \{\mathbf{m}_1, ..., \mathbf{m}_{N_{TP}}\}$. The latter is saved into the **gallery of facial models**. By the end of this enrolment phase, the gallery would have $N_{IN} + 1$ different templates for each person in the watchlist. Let T_i be the set of techniques to be applied at the *pre-processing* level of this framework $T_i; i = 1..N_{IN} + 1$ represents the set of all N_{IN} IN techniques and 1 baseline technique (without application of IN). The last phase of the system would be the testing phase. Each captured facial ROI from video frames undergo the same processes **segmentation**, **pre-processing**, **block division** and **feature extraction** described at the enrolment phase.

For each IN technique, the corresponding feature vector $\mathbf{F} = \{\mathbf{f}_1, ..., \mathbf{f}_{N_{TP}}\}$ obtained in the testing phase would be compared using template matching with the corresponding feature vector of the same technique of the individuals saved in the Gallery. Finally, the scores produced with template matching S_i at the **classification** level is normalized to an interval of [0 1].

The performance of the still-to-video FR system in this work is evaluated at the transactional level using the performance measures presented in Section 3.3.

The algorithms of both enrolment and testing phase are provided.

4.3 Research Questions and Hypotheses

The main purpose of this study is to conduct a comparison between several IN techniques applied at the pre-processing level of a still-to-video FR system. It should be answering one main question: What are the effects of applying IN into images at the pre-processing level of a still-to-video FR? Since local matching is implemented, other research questions became important to answer:

- What are the effects of applying IN locally on each patch of an image?
- What are the effects of having higher number of patches on the still-to-video FR system?

By this, some hypotheses can be drawn out:

- Applying IN techniques on an image-level for both probe and gallery images removes the negative impact of illumination. So, an improvement of the system's performance may be observed;
- Local application of IN may provide higher results than applying IN globally on the image;
- Having higher number of patches for **block division** provides more spatial information, thus better feature representation which improves the system's performance.

Table 4.1 Exploited Videos from the ChokePoint dataset for the first study

Portal 1	
Entering	**Leaving**
P1E_S1_C1	P1L_S1_C1
P1E_S2_C2	P1L_S2_C2
P1E_S3_C3	P1L_S3_C3
P1E_S4_C1	P1L_S4_C1
Portal 2	
Entering	**Leaving**
P2E_S1_C3	P2L_S1_C1
P2E_S2_C2	P2L_S2_C2
P2E_S3_C1	P2L_S3_C3
P2E_S4_C2	P2L_S4_C2

4.4 Experimental Methodology

4.4.1 Dataset Exploitation

For this study $N_{ind} = 5$ watchlist individuals are randomly chosen out of 29 subjects from the ChokePoint dataset. Concerning the video sets for implementation, 16 out of 48 videos are taken into consideration. The videos from the ChokePoint dataset considered for this work is represented in Table 4.1 with their corresponding portal number, sequence type and camera number.

These video sets are chosen for a single main reason that they provide the most recorded frontal views based on (Wong *et al.*, 2011).

4.4.2 Experimental Protocol

During enrolment, the **segmentation** module consists in extracting faces from the still images of the watchlist individuals using the *Viola-Jones* method (Viola and Jones, 2001) for face detection. Then the detected faces are then converted into grayscale images and resized into a common size of 48×48 ($N_{size} = 48$). The **pre-processing** module comprises of 12 techniques where $N_{ind} = 11$ different IN techniques and 1 no IN application. Table 4.2 presents the specific

Table 4.2 Illumination Normalization Techniques

Family	**Specific Technique**
Retinex	Adaptive Single-Scale Retinex (ASSR)
Retinex	Large and Small-Scale Features (LSSF)
Self Quotient	Multi-Scale Self Quotient (MSSQ)
Diffusion	Isotropic Diffusion (ID)
Diffusion	Modified Anisotropic Diffusion (MAD)
Filter	Tan and Triggs (TT)
Gradient	Multi-Scale Weberfaces (MSW)
Mean Denoising	Adaptive Non Local Means (ANLM)
Retina	Retina Modeling (RM)
Wavelet	Wavelet Denoising (WD)
Frequency	Homomorphic (H)

techniques from literature that are considered in this study. These techniques are selected for the simplest reason that they are newer and more representative techniques from different families. These IN techniques are gathered in the INface toolbox (Štruc and Pavešić, 2011) and (Štruc and Pavešić, 2009) where all most state-of-the-art IN techniques are explained and implemented. So, once an ROI is segmented, the global approach consists in applying 12 pre-processing techniques providing $N_{ind} = 11$ normalized face images and 1 normal face image of the still ROI. Figure 4.2 shows the obtained results after performing IN on the facial ROI for individuals ID03 and ID04 of the ChokePoint dataset. Then, each representation undergoes **block division**. This process consists in extracting non-overlapping patches on the normalized facial ROI. As for the local approach, **block division** is applied on the segmented ROI then IN techniques are performed locally on each block giving a set of normalized patches. At this stage comes the two different experiments:

- First experiment consists in dividing the 48×48 pixel image ROI into a 3×3 $(N_p = 3)$ non overlapping patches, having a total of block $N_{TP} = 9$. A single patch would have the size of 16×16 pixels $(p_s = 16)$;

- The second experiment consists in dividing the same 48×48 pixel size ROI into 4×4 ($N_p = 4$) smaller non overlapping patches having the size of 12×12 pixels each patch $p_s = 12$ but having a higher number of total patches $N_{TP} = 16$.

Feature extraction is performed on each patch individually using *LBP* technique (see Section 2.1.1). This descriptor once applied provides feature vectors having the size of $N_{feat} = 59$ representing one local patch of the image. For one image ROI a total of $59 \times 9 = 531$ features are concatenated for experiment 1 of this study per contra $59 \times 16 = 944$ for experiment 2. These features are then saved into the **gallery**.

Table 4.3 Summary of techniques implemented on each module in the first study for both local and global approach

Methodology and Techniques implemented in Study 1		
Modules	**Experiment 1**	**Experiment2**
Segmentation	Face detection: Viola-Jones Image gray-scale transformation Resize image into 48×48 pixel	
Pre-processsing	1- No IN technique (baseline) 2- Adaptive Single Scale Retinex 3- Large and Small Scale Features 4- Multi-Scale Self-Quotient 5- Isotropic Diffusion 6- Modified Anisotropic Diffusion 7- Tan and Triggs 8- Multi-Scale Weberfaces 9- Adaptive Non-Local Means 10- Retina Modeling 11- Wavelet Denoising 12- Hommomorphic	
Block division	Division into $N_{TP} = 9$ non-overlapping patches	Division into $N_{TP} = 16$ non-overlapping patches
Feature extraction	LBP on each patch Number of features: $N_{feat} = 531$	LBP on each patch Number of features: $N_{feat} = 944$
Classification	Template matching: Euclidean Distance	

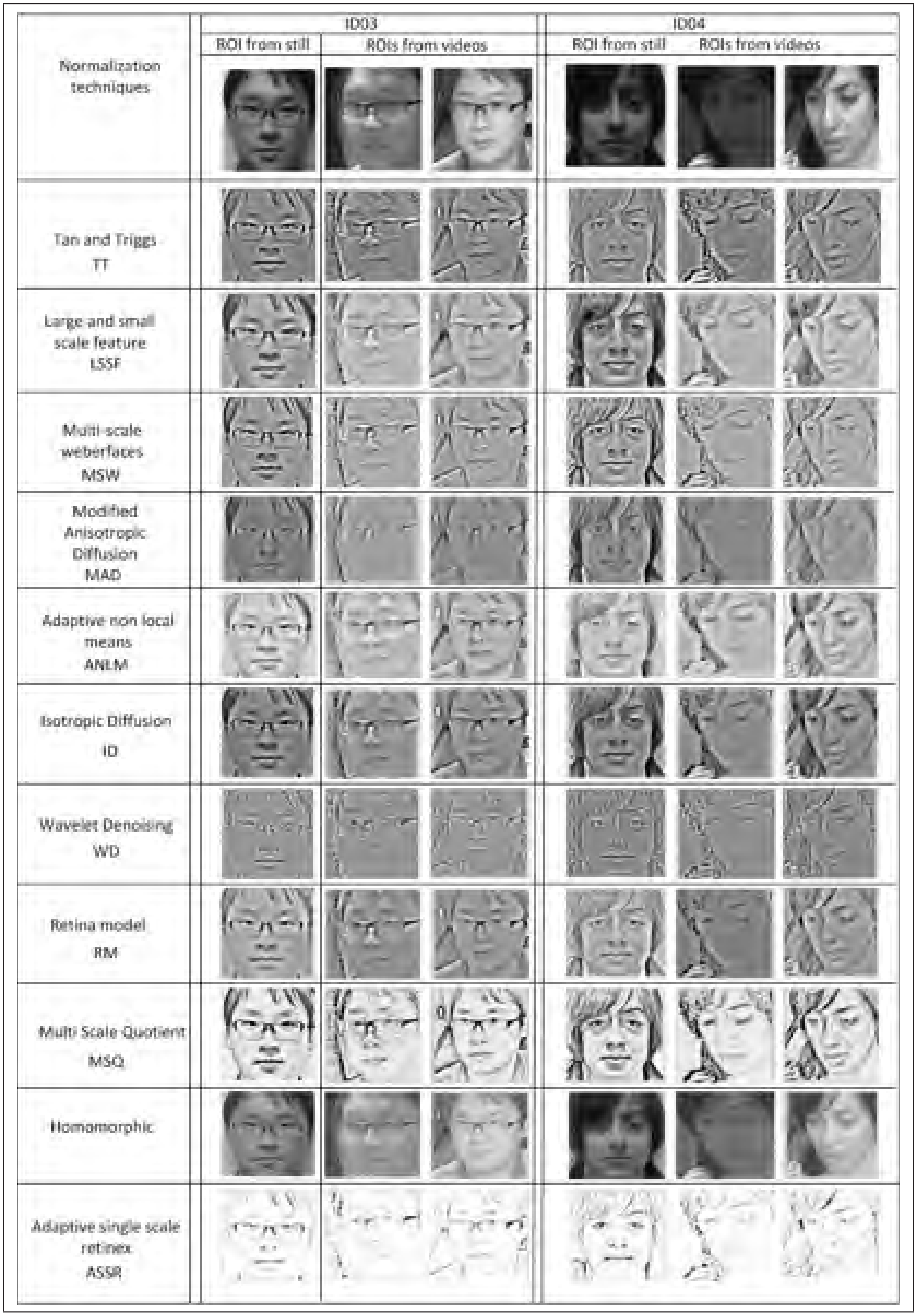


Figure 4.2 Examples of face images obtained after illumination normalization is applied to ROIs in stills and videos from individuals ID03 and ID05
Taken from Amara *et al.* (2014)

During operational phase, same processes: *Viola-Jones* for FD, 12 pre-processing IN techniques, **block division** (experiment 1 $N_{TP} = 9$ and experiment 2 $N_{TP} = 16$ blocks. **Feature extraction** using *LBP* is performed on each patch producing $N_{feat} = 531$ for experiment 1 and $N_{feat} = 944$ for experiment 2. As for the **classification** module, template matching is applied using *Euclidean Distance* measure giving scores for each matching for each technique. A summary table Table 4.3 shows the methodology and techniques implemented in this first work.

4.4.3 Validation of Results

5 independent replications are held to validate the results of experiments of this study. This is also done in order to take out the possibility of having randomly produced results. On each replication, 5 different and randomly chosen watchlist individuals from the ChokePoint dataset are chosen. Then, the implementation of the proposed system is done on the same video sets for each chosen watchlist. The final results shows an average over all 5 replications for each technique. The results are evaluated using the ROC space and the PR space (Section 3.3).

4.5 Results and Discussions

Experiment 1

4.5.1 Global Approach

Tables 4.4 and 4.5 show the transaction-level performance. More specifically, these tables present respectively the pAUC(20%) and the AUPR along with the standard deviation obtained using 11 IN techniques for each individual from the watchlist over one single video P1E_S1_-C1 of the dataset ChokePoint using a division of 9 blocks. Besides, these tables display the average over individuals in order to assess the performance of each IN technique regardless of the person of interest.

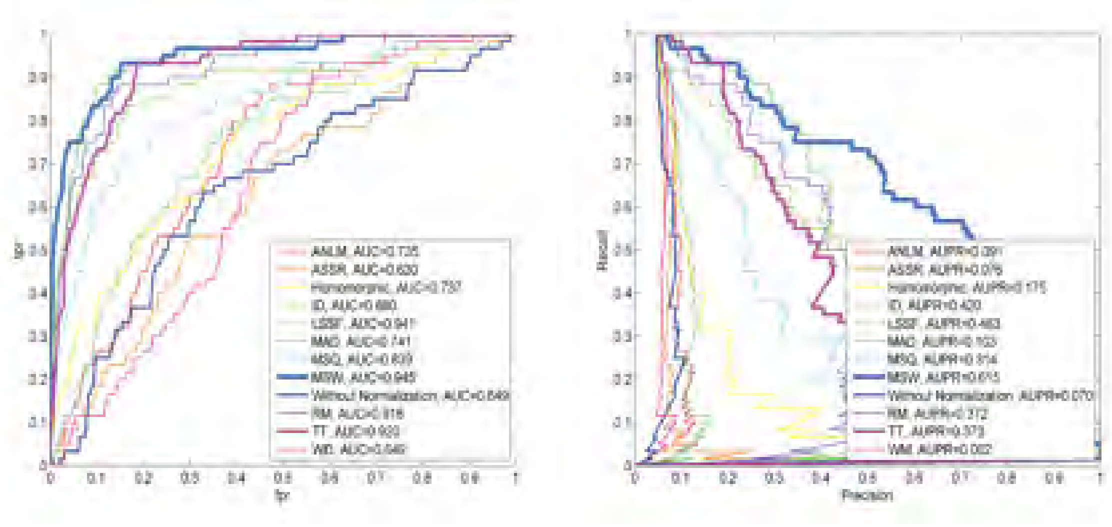
Table 4.4 pAUC(20%) performance (with standard error) for each watchlist individual in P1E_S1_C1 with illumination normaliztion techniques using $N_{TP} = 9$ blocks

pAUC(20%) performance						
Techniques	**ID3**	**ID4**	**ID7**	**ID9**	**ID12**	**AVG**
No normalization	0.18	0.80	0.23	0.79	0.91	0.58 ± 0.16
Adaptive Single Scale Retinex	0.16	0.63	0.20	0.41	0.82	0.45 ± 0.13
Large and Small Scale Features	0.76	0.50	0.17	0.91	0.80	0.63 ± 0.13
Multi Scale Self-Quotient	0.53	0.48	0.23	0.65	0.97	0.58 ± 0.12
Isotropic Diffusion	0.66	0.57	0.25	0.72	0.54	0.55 ± 0.08
Modified Anisotropic Diffusion	0.29	0.66	0.61	0.55	0.77	0.58 ± 0.08
Tan and Triggs	0.66	0.68	0.45	0.85	0.99	**0.73 ± 0.09**
Multi Scale Weberfaces	0.80	0.66	0.47	0.86	0.99	**0.76 ± 0.09**
Adaptive Non-Local Means	0.23	0.66	0.31	0.53	0.80	0.50 ± 0.11
Retina Modeling	0.70	0.47	0.39	0.80	0.98	0.67 ± 0.11
Wavelet Denoising	0.12	0.53	0.06	0.60	0.62	0.39 ± 0.12
Hommomorphic	0.34	0.86	0.57	0.69	0.95	0.68 ± 0.11

Table 4.5 AUPR performance (with standard error) for each watchlist individual in P1E_S1_C1 with illumination normalization techniques using $N_{TP} = 9$ blocks

AUPR performance						
Techniques	**ID3**	**ID4**	**ID7**	**ID9**	**ID12**	**AVG**
No normalization	0.07	0.41	0.09	0.44	0.81	0.36 ± 0.13
Adaptive Single Scale Retinex	0.07	0.44	0.06	0.10	0.67	0.27 ± 0.12
Large and Small Scale Features	0.46	0.18	0.07	0.67	0.61	0.40 ± 0.11
Multi Scale Self-Quotient	0.31	0.17	0.07	0.22	0.92	0.34 ± 0.15
Isotropic Diffusion	0.42	0.26	0.09	0.44	0.39	0.32 ± 0.07
Modified Anisotropic Diffusion	0.10	0.50	0.32	0.20	0.64	0.35 ± 0.22
Tan and Triggs	0.37	0.40	0.20	0.69	0.95	**0.52 ± 0.13**
Multi Scale Weberfaces	0.61	0.50	0.16	0.70	0.95	**0.58 ± 0.13**
Adaptive Non-Local Means	0.09	0.49	0.11	0.15	0.70	0.31 ± 0.12
Retina Modeling	0.37	0.14	0.15	0.39	0.92	0.39 ± 0.14
Wavelet Denoising	0.06	0.25	0.04	0.22	0.49	0.21 ± 0.08
Hommomorphic	0.17	0.58	0.37	0.28	0.87	0.45 ± 0.12

By looking through the averages of pAUC obtained in Table 4.4, the first thing that can be observed is that the results varies a lot from one technique to the other. Some results show that applying illumination normalization can increase the system's performance such as "Large and Small Scale Features" (LSSF) pAUC(20%)=0.63, "Hommomorphic" pAUC(20%)=0.68



a) ROC curves b) PR curves

Figure 4.3 ROC and PR curves for ID04 in video P1E_S1_C1

which are higher than the baseline technique (No normalization) pAUC(20%)=0.58 . The technique which outperforms would be "Multi Scale Weberfaces" (MSW). Indeed its pAUC has reached 0.76 which is an increase of performance of about 0.17 (17%) compared to the baseline technique. It is also impossible to deny the ability of the "Tan and Triggs" (TT) IN technique in increasing the performance of the still-to-video FR system by almost 0.15 (15%).

Looking at the AUPR values at Table 4.5, still, Multi Scale Weberfaces outperforms overall then Tan and Triggs ranking as the second best technique. If a deeper review is given for person ID3, the highest recorded pAUC(20%) comes from the Multi Scale Weberfaces technique. However, for ID4, aside from the Hommomorphic technique which is the highest one recorded in terms of performance, all of the values of pAUC(20%) are less than the baseline technique (pAUC(20%)=0.80) as it can be seen in the ROC space and PR space respectively in Figure 4.3 over a single video P1E_S1_C1 of individual ID04.

Looking at the Tables 4.6 and 4.7 which represent respectively the average pAUC and the average AUPR over 5 replications of the still-to-video FR system using 5 sets of randomly chosen target individuals. The values are presented by portals and by type (entering and leaving). An

Table 4.6 Average pAUC(20%) performance after 5 replications for global approach (with standard error) in all videos having the most frontal views of the ChokePoint dataset with different illumination normalization techniques using $N_{TP} = 9blocks$

Average pAUC (20%) performance					
Techniques	**Portal 1**		**Portal 2**		
	Entering	Leaving	Entering	Leaving	AVG
No normalization	0.49±0.060	0.66±0.046	0.21±0.042	0.43±0.048	0.45±0.049
Adaptive single Scale Retinex	0.38±0.052	0.48±0.046	0.21±0.040	0.38±0.046	0.36±0.046
Large and Small Scale Features	0.51±0.058	0.47±0.056	0.22±0.046	0.51±0.065	0.43±0.056
Multi-Scale Self-Quotient	0.51±0.063	0.50±0.065	0.18±0.027	0.50±0.065	0.42±0.055
Isotropic Diffusion	0.49±0.052	0.56±0.054	0.23±0.040	0.48±0.044	0.44±0.048
Modified Anisotropic Diffusion	0.49±0.050	0.52±0.052	0.21±0.042	0.49±0.050	0.43±0.049
Tan and Triggs	0.55±0.058	0.57±0.062	0.21±0.050	0.59±0.058	0.48±0.057
Multi Scale Weberfaces	0.53±0.062	0.58±0.066	0.23±0.052	0.61±0.058	0.49±0.060
Adaptive Non-Local Means	0.43±0.050	0.50±0.047	0.21±0.036	0.40±0.052	0.39±0.046
Retina Modeling	0.51±0.061	0.50±0.064	0.21±0.048	0.53±0.062	0.44±0.059
Wavelet Denoising	0.35±0.049	0.41±0.044	0.18±0.028	0.36±0.042	0.33±0.041
Hommomorphic	0.50±0.059	0.59±0.044	0.21±0.036	0.41±0.041	0.43±0.045

Table 4.7 Average AUPR performance after 5 replications for global approach (with standard error) in all videos having the most frontal views of the ChokePoint dataset with different illumination normalization techniques using $N_{TP} = 9blocks$

Average AUPR performance					
Techniques	**Portal 1**		**Portal 2**		AVG
	Entering	Leaving	Entering	Leaving	
No normalization	0.30±0.052	0.47±0.052	0.11±0.026	0.26±0.046	0.29±0.044
Adaptive single Scale Retinex	0.21±0.042	0.27±0.040	0.11±0.024	0.20±0.034	0.20±0.035
Large and Small Scale Features	0.33±0.054	0.30±0.050	0.12±0.032	0.32±0.056	0.27±0.048
Multi-Scale Self-Quotient	0.31±0.052	0.31±0.044	0.09±0.013	0.28±0.051	0.25±0.040
Isotropic Diffusion	0.30±0.050	0.36±0.050	0.11±0.024	0.28±0.048	0.26±0.043
Modified Anisotropic Diffusion	0.30±0.048	0.32±0.048	0.10±0.026	0.27±0.050	0.25±0.043
Tan and Triggs	0.37±0.060	0.40±0.062	0.13±0.040	0.41±0.070	0.33±0.058
Multi Scale Weberfaces	0.38±0.064	0.43±0.066	0.14±0.042	0.44±0.064	0.35±0.059
Adaptive Non-Local Means	0.24±0.048	0.30±0.042	0.10±0.022	0.22±0.048	0.22±0.040
Retina Modeling	0.34±0.060	0.34±0.060	0.13±0.042	0.37±0.066	0.30±0.057
Wavelet Denoising	0.17±0.034	0.20±0.034	0.08±0.010	0.16±0.024	0.15±0.026
Hommomorphic	0.29±0.044	0.40±0.048	0.11±0.022	0.23±0.036	0.26±0.038

overall average is also presented to provide a global performance regardless of the recording conditions.

It is obvious from the results that both techniques Multi Scale weberfaces and Tan and Triggs have a positive impact overall on the still-to-video FR system's performance.

Thoroughly at Table 4.6, some recording conditions like Portal 1 Leaving seems to decrease the system's performance when IN techniques is applied compared to the baseline method.

In addition, an ascertainment can also be made for the Wavelet Denoising technique. The latter shows very low performance compared to the baseline. In Table 4.6, the Wavelet Denoising has an average of pAUC(20%)=0.33 and AUPR= 0.15 after 5 replications which are very low values. If a glance is taken on the examples of face images after performing IN onto ROIs in Figure 4.2, the normalized image of Wavelet Denoising technique shows a very low quality representation. As a matter of fact, after applying the normalization, the images seem to have lost distinctive information and features which leads to generating poor LBP representation, thus, low recognition performance.

Experiment 2

Table 4.8 pAUC(20%) performance (with standard error) for each watchlist individual in P1E_S1_C1 with illumination normalization techniques using $N_{TP} = 16$ blocks

pAUC(20%) performance						
Techniques	**ID3**	**ID4**	**ID7**	**ID9**	**ID12**	**AVG**
No normalization	0.39	0.55	0.12	0.56	0.86	0.49 ± 0.12
Adaptive Single Scale Retinex	0.55	0.56	0.19	0.25	0.71	0.45 ± 0.10
Large and Small Scale Features	0.56	0.35	0.10	0.70	0.63	0.47 ± 0.11
Multi Scale Self-Quotient	0.59	0.52	0.13	0.39	0.87	0.50 ± 0.12
Isotropic Diffusion	0.40	0.26	0.17	0.67	0.70	0.44 ± 0.11
Modified Anisotropic Diffusion	0.33	0.36	0.23	0.58	0.70	0.44 ± 0.08
Tan and Triggs	0.55	0.58	0.40	0.86	0.91	**0.66 ± 0.10**
Multi Scale Weberfaces	0.58	0.56	0.33	0.86	0.91	**0.65 ± 0.11**
Adaptive Non-Local Means	0.37	0.52	0.29	0.56	0.82	0.51 ± 0.08
Retina Modeling	0.38	0.33	0.33	0.76	0.73	0.51 ± 0.10
Wavelet Denoising	0.42	0.50	0.15	0.56	0.67	0.46 ± 0.08
Hommomorphic	0.41	0.65	0.33	0.42	0.95	0.55 ± 0.11

Tables 4.8 and 4.9 respectively show pAUC(20%) and the AUPR performance of a still-to-video FR system under different IN techniques. However, for this experiment the number of total patches is $N_{TP} = 16$ patches of size 12×12 pixels each patch. Again, what can be di-

Table 4.9 AUPR performance (with standard error) for each watchlist individual in P1E_S1_C1 with illumination normalization techniques using $N_{TP} = 16$ blocks

AUPR performance						
Techniques	**ID3**	**ID4**	**ID7**	**ID9**	**ID12**	**AVG**
No normalization	0.18	0.42	0.06	0.19	0.75	0.32 ± 0.10
Adaptive Single Scale Retinex	0.42	0.31	0.06	0.06	0.36	0.24 ± 0.07
Large and Small Scale Features	0.25	0.09	0.06	0.39	0.30	0.22 ± 0.06
Multi Scale Self-Quotient	0.32	0.17	0.05	0.10	0.64	0.26 ± 0.11
Isotropic Diffusion	0.18	0.06	0.07	0.34	0.49	0.23 ± 0.08
Modified Anisotropic Diffusion	0.12	0.10	0.08	0.37	0.32	0.20 ± 0.06
Tan and Triggs	0.33	0.32	0.14	0.56	0.73	**0.42 ± 0.10**
Multi Scale Weberfaces	0.37	0.22	0.10	0.57	0.72	**0.39 ± 0.11**
Adaptive Non-Local Means	0.16	0.34	0.12	0.19	0.75	0.31 ± 0.12
Retina Modeling	0.17	0.09	0.11	0.45	0.35	0.23 ± 0.16
Wavelet Denoising	0.16	0.19	0.05	0.15	0.52	0.21 ± 0.08
Hommomorphic	0.15	0.47	0.11	0.12	0.89	0.35 ± 0.15

rectly concluded is that both Multi Scale Weberfaces and Tan and Triggs methods have outperformed the other techniques including the baseline for the video P1E_S1_C1 with an average pAUC(20%)= 0.66 and average AUPR=0.42 for Tan and Triggs technique which is an improvement of almost 16% (0.16) and an average pAUC(20%) =0.65 and average AUPR= 0.39 which is an improvement of 15% (0.15) .

In order to draw better conclusion on the effects of having smaller sizes of patches, it is preferable to refer to Table 4.10 and Table 4.11 which respectively represent the average pAUC(20%) and AUPR for each portals after 5 system replication. In most cases (portal 1 entering, portal 1 leaving, portal 2 entering and portal 2 leaving), both techniques Tan and Triggs and Multi Scale Weberfaces outplay the other techniques. Comparing the AUPR of both Tan and Triggs and Multi Scale Weberfaces, a conclusion can be made that the second mentionned technique (Multi Scale Weberfaces) performs better. In fact, the average AUPR value of Multi Scale Weberfaces (0.38) is higher than the average AUPR value of Tan and Triggs (0.26) .

Unlike the case in Table 4.6 and in Table 4.6 which divides the ROI into 9 total blocks, the results obtained from this second experiment have shown to be more stable and coherent in

Table 4.10 Average pAUC(20%) performance for global approach (with standard error) in all videos having the most frontal views of the ChokePoint dataset with different illumination normalization techniques using $N_{TP} = 16blocks$

Average pAUC(20%) performance					
Techniques	**Portal 1**		**Portal 2**		AVG
	Entering	Leaving	Entering	Leaving	
No normalization	0.52±0.048	0.65±0.042	0.21±0.036	0.47±0.058	0.46±0.046
Adaptive single Scale Retinex	0.45±0.048	0.56±0.050	0.20±0.040	0.51±0.036	0.43±0.043
Large and Small Scale Features	0.55±0.042	0.58±0.048	0.19±0.032	0.54±0.026	0.47±0.037
Multi-Scale Self-Quotient	0.46±0.033	0.49±0.069	0.16±0.029	0.40±0.056	0.38±0.047
Isotropic Diffusion	0.49±0.046	0.64±0.038	0.21±0.040	0.50±0.048	0.46±0.043
Modified Anisotropic Diffusion	0.52±0.040	0.55±0.040	0.19±0.034	0.46±0.030	0.43±0.036
Tan and Triggs	0.65±0.042	0.68±0.044	0.24±0.046	0.67±0.036	**0.56±0.042**
Multi Scale Weberfaces	0.65±0.048	0.69±0.046	0.24±0.042	0.67±0.034	**0.56±0.042**
Adaptive Non-Local Means	0.44±0.050	0.52±0.046	0.20±0.027	0.44±0.044	0.40±0.042
Retina Modeling	0.57±0.036	0.59±0.046	0.21±0.042	0.47±0.030	0.46±0.038
Wavelet Denoising	0.46±0.046	0.55±0.052	0.22±0.032	0.47±0.038	0.43±0.042
Hommomorphic	0.58±0.046	0.66±0.048	0.21±0.038	0.43±0.052	0.47±0.046

Table 4.11 Average AUPR performance for global approach (with standard error) in all videos having the most frontal views of the ChokePoint dataset with different illumination normalization techniques using $N_{TP} = 16blocks$

Average AUPR performance					
Techniques	**Portal 1**		**Portal 2**		AVG
	Entering	Leaving	Entering	Leaving	
No normalization	0.29±0.050	0.43±0.060	0.09±0.014	0.31±0.058	0.28±0.046
Adaptive single Scale Retinex	0.24±0.036	0.31±0.044	0.09±0.016	0.25±0.032	0.22±0.032
Large and Small Scale Features	0.30±0.042	0.36±0.044	0.08±0.010	0.28±0.034	0.26±0.033
Multi-Scale Self-Quotient	0.26±0.046	0.28±0.040	0.09±0.018	0.21±0.042	0.21±0.037
Isotropic Diffusion	0.26±0.042	0.42±0.038	0.10±0.018	0.31±0.048	0.27±0.037
Modified Anisotropic Diffusion	0.31±0.040	0.36±0.040	0.08±0.016	0.21±0.030	0.24±0.032
Tan and Triggs	0.04±0.052	0.47±0.048	0.11±0.022	0.42±0.050	0.26±0.043
Multi Scale Weberfaces	0.45±0.054	0.51±0.056	0.11±0.024	0.45±0.044	0.38±0.045
Adaptive Non-Local Means	0.31±0.042	0.31±0.042	0.08±0.012	0.24±0.044	0.24±0.035
Retina Modeling	0.34±0.042	0.38±0.042	0.09±0.018	0.30±0.038	0.28±0.035
Wavelet Denoising	0.24±0.038	0.34±0.048	0.10±0.012	0.30±0.034	0.25±0.033
Hommomorphic	0.36±0.050	0.47±0.060	0.09±0.014	0.23±0.050	0.29±0.044

terms of results for all video recordings. In addition to the stability and coherency, using 16 patches (meaning higher total number of patches but of smaller sizes) can boost the still-to-video FR system's performance. Indeed, the performance of our still-to-video FR system has increased from 4% using 9 blocks of 16×16 pixels each to 10% using 16 blocks of 12×12

pixels each for Multi Scale Weberfaces technique. Also, it has shown an increase from 3% using 9 blocks to 10% using 16 blocks with Tan and Triggs technique.

At this stage, using smaller sizes of patches provides more accuracy and better performance to the still-to-video FR using the outrunning techniques (Multi Scale Weberfaces and Tan and Triggs).

The hypothesis stated in Section 4.3 can be approved. Indeed, we have seen in both experiments (1 and 2) held in this study some techniques have shown to be beneficial to the improvement of the system in terms of performance such as Multi Scale Weberface and Tan and Triggs techniques. However, others seem to decrease the performances due to some quality issues of the captured facial ROIs or even due some loss of information after the application of illumination normalization process.

In this section we have seen the impact of applying IN techniques globally on a LBP-based still-to-video FR system. The techniques which has outrunned the others are the Multi Scale Weberfaces and Tan and Triggs. It has also been proven that for this still-to-video FR system, the smaller the patch for block division the better the performance rate.

4.5.2 Local Approach

In the global approach, IN techniques were processed prior to the division of ROI into blocks. However, for the local approach a division of the ROI image is primarily applied then IN techniques are applied locally on each obtained patch. Same experiments are done at this level concerning the variations of number of patches $N_{TP} = 9$ and $N_{TP} = 16$. In this section only 4 out of 11 IN techniques are shown. The choice of these techniques is based on the outcome of the global approach previously detailed. These techniques have been considered because their average pAUC and AUPR performances are higher than the baseline's pAUC and AUPR performance in both experiments ($N_{TP} = 9$ and $N_{TP} = 16$) . http://www.immigration-quebec.gouv.qc.ca/publications/fr/peq/A-0520-GF.pdf

Experiment 1

Table 4.12 shows the average pAUC(20%) performance after 5 replications. An obvious observation can be concluded that the techniques Tan and Triggs and Multi Scale Weberfaces are still outperforming overall having an improvement of 4% (0.4) compared to the baseline technique. If a comparison is done between local and global approach for the same number of blocks ($N_{TP} = 9$), the pAUC performance of Tan and Triggs technique has slightly increased by 1% (0.1) whereas the pAUC perforance of the Weberfaces did not vary much. It has practically remained stagnant at 49% (0.49) .

Table 4.12 Average pAUC(20%) performance for local approach (with standard error) in all videos having the most frontal views of the ChokePoint dataset with different illumination normalization techniques using $N_{TP} = 9blocks$

Average pAUC(20%) performance					
Techniques	**Portal 1**		**Portal 2**		AVG
	Entering	Leaving	Entering	Leaving	
No normalization	0.49±0.060	0.66±0.046	0.21±0.042	0.43±0.048	0.45±0.049
Large and Small Scale Features	0.48±0.074	0.47±0.069	0.18±0.028	0.47±0.064	0.40±0.059
Tan and Triggs	0.58±0.090	0.59±0.086	0.19±0.030	0.60±0.084	**0.49±0.073**
Multi Scale Weberfaces	0.54±0.080	0.60±0.083	0.22±0.033	0.60±0.081	**0.49±0.069**
Hommomorphic	0.48±0.059	0.61±0.077	0.17±0.030	0.38±0.047	0.41±0.053

Table 4.13 Average AUPR performance for local approach (with standard error) in all videos having the most frontal views of the ChokePoint dataset with different illumination normalization techniques using $N_{TP} = 9blocks$

Average AUPR performance					
Techniques	**Portal 1**		**Portal 2**		AVG
	Entering	Leaving	Entering	Leaving	
No normalization	0.30±0.052	0.47±0.052	0.11±0.026	0.26±0.046	0.29±0.044
Large and Small Scale Features	0.29±0.052	0.32±0.052	0.10±0.008	0.28±0.050	0.25±0.040
Tan and Triggs	0.40±0.056	0.44±0.064	0.13±0.019	0.32±0.063	0.32±0.051
Multi Scale Weberfaces	0.38±0.061	0.43±0.086	0.14±0.024	0.41±0.057	0.34±0.057
Hommomorphic	0.27±0.038	0.40±0.041	0.10±0.020	0.21±0.030	0.25±0.032

Since the results in terms of performance of the system are likely to be similar between both approaches (global and local) for a block division of $N_{TP} = 9$, it is then preferable to use the global approach for less computational cost and time processing.

Experiment 2

Table 4.14 provides the average pAUC(20%) performance for local approach after 5 replications of the system using a block division into $N_{TP} = 16$ blocks. All of the techniques shown have witnessed improvements compared to the baseline technique (No normalization). Tan and Triggs and Multi Scale Weberfaces are the outperforming techniques. With the local approach Tan and Triggs technique has increased the systems performance by almost 12% (0.12) compared to the baseline and 13% (0.13) using Multi Scale Weberfaces.

Table 4.14 Average pAUC(20%) performance for local approach (with standard error) in all videos having the most frontal views of the ChokePoint dataset with different illumination normalization techniques using $N_{TP} = 16blocks$

Average pAUC(20%) performance					
Techniques	**Portal 1**		**Portal 2**		AVG
	Entering	Leaving	Entering	Leaving	
No normalization	0.52±0.049	0.65±0.048	0.22±0.036	0.47±0.058	0.46±0.048
Large and Small Scale Features	0.45±0.059	0.53±0.050	0.41±0.053	0.49±0.056	0.47±0.055
Tan and Triggs	0.61±0.052	0.63±0.047	0.49±0.053	0.57±0.063	0.58±0.054
Multi Scale Weberfaces	0.60±0.047	0.67±0.041	0.51±0.040	0.60±0.047	0.59±0.044
Hommomorphic	0.50±0.058	0.64±0.054	0.40±0.060	0.41±0.062	0.49±0.059

Table 4.15 Average AUPR performance for local approach (with standard error) in all videos having the most frontal views of the ChokePoint dataset with different illumination normalization techniques using $N_{TP} = 16blocks$

Average AUPR performance					
Techniques	**Portal 1**		**Portal 2**		AVG
	Entering	Leaving	Entering	Leaving	
No normalization	0.29±0.050	0.43±0.060	0.09±0.014	0.31±0.058	0.28±0.046
Large and Small Scale Features	0.23±0.048	0.33±0.053	0.19±0.034	0.28±0.062	0.26±0.049
Tan and Triggs	0.43±0.059	0.45±0.050	0.25±0.043	0.38±0.066	0.38±0.055
Multi Scale Weberfaces	0.36±0.048	0.47±0.049	0.21±0.029	0.39±0.058	0.36±0.046
Hommomorphic	0.30±0.051	0.45±0.067	0.19±0.033	0.24±0.058	0.30±0.052

If a comparison is done between the global and local approach for $N_{TP} = 16$ blocks, Tan and Triggs, Multi Scale Weberfaces and Hommomorphic have shown improvement compared to the global approach. Indeed, the pAUC of Tan and Triggs has varied from 0.56 with global approach to 0.58 with local approach. The same thing can be sais with MSW technique. Its pAUC performance has fluctuated from 0.56 to 0.59 . With Hommomorphic technique the system's performance has increased from 0.47 to 0.49 . However the Large and Small Scale Features technique did not show any significant increase in terms of pAUC performance (0.47). Even the AUPR performance of this same IN technique for the local approach did not vary much in comparison to the AUPR value for the global approach. In this case, if LSSF is chosen as IN technique it would be better to use the global approach rather than local for less computational costs.

Summary of experiments

Table 4.16 Summary table of results

	Techniques	**pAUC(20%)**		**AUPR**	
		GLOBAL	**LOCAL**	**GLOBAL**	**LOCAL**
whole image	No normalization	0.20±0.037		0.09±0.015	
	Tan and Triggs	0.21±0.035		0.09±0.013	
	Multi Scale Weberfaces	0.20±0.029		0.08±0.010	
4 blocks	No normalization	0.33±0.052	0.33±0.052	0.18±0.029	0.18±0.029
	Tan and Triggs	0.37±0.062	0.38±0.058	0.23±0.044	0.24±0.044
	Multi Scale Weberfaces	0.38±0.057	0.38±0.063	0.23±0.040	0.22±0.038
9 blocks	No normalization	0.45±0.049	0.45±0.049	0.29±0.044	0.29±0.044
	Tan and Triggs	0.48±0.057	0.49±0.073	0.33±0.058	0.32±0.051
	Multi Scale Weberfaces	0.49±0.060	0.49±0.069	0.35±0.059	0.34±0.057
16 blocks	No normalization	0.46±0.046	0.46±0.048	0.28±0.046	0.28±0.046
	Tan and Triggs	0.56±0.042	0.58±0.054	0.26±0.043	0.38±0.055
	Multi Scale Weberfaces	0.56±0.042	0.59±0.044	0.38±0.045	0.36±0.046

To correctly assess the impact of having different block division, a final summary table is given in Table 4.16. The latter shows pAUC and AUPR performances for both global and local approach and for various block division N_{TP} using Tan and Triggs and Multi Scale Weberfaces techniques.

A first observation can be made about the number of patches and the performance of the system. The larger the number of patch used the better the performance of the system. This statement can be affirmed for both local and global approach. As a matter of fact with Tan and Triggs technique in the global approach the average pAUC performance has increased from 0.21 (for $N_{TP} = 1$ to 0.37 (for $N_{TP} = 4$), then from 0.37 to 0.48 (for $N_{TP} = 9$) and finally from 0.48 to 0.56 (for $N_{TP} = 16$). More specifically an improvement of 35% (0.35) .

With Multi Scale Weberfaces technique the pAUC performance boosted from 0.20 (using $N_{TP} = 1$) to 0.56 (using $N_{TP} = 16$). For this same IN technique in the global approach, the average aUPR performance has definitely increased from 0.09 (for $N_{TP} = 1$) to 0.24 (for $N_{TP} = 4$), then from 0.24 to 0.32 (for $N_{TP} = 9$) and finally from 0.32 to 0.38 (for $N_{TP} = 16$).

Same things can be said about the local approach. With Tan and Triggs the pAUC performance has an improvement of 37% from 0.21 using $N_{TP} = 1$ to 0.58 $N_{TP} = 16$. With Multi Scale Weberfaces, the pAUC performance has also improved by 39% from 0.20 (using $N_{TP} = 1$) to 0.59 (using $N_{TP} = 16$).

A second observation can be made about the choice of approach to use when applying IN. Based on the results, using the local approach is more beneficial in terms of improvement of the system's performance. This is particularly observed when using $N_{TP} = 16$ blocks. Indeed, there is at least 3% enhancements for each techniques between the global and the local approach.

To sum up, the series of experiments held in this chapter has led to some specific conclusions:

- Out of all the 11 chosen IN techniques, Tans and Triggs and Multi Scale Weberfaces provides the best results for illumination invariance of the FR system;
- The higher the number of block division N_{TP} (i.e: the smaller the size of patches used for block division) the better the performance of the system. Indeed, having smaller sizes of patches (or having larger number of patches) offers better information coverage once feature extraction is applied locally giving the FR system a performance boost;

- Systems having IN techniques applied locally on patches performs better than the system having a global approach where IN are applied on the global image.

Table 4.17 Average processing time of 1 ROI in seconds (segmentation, pre-processing, feature extration (LBP), matching (Euclidean distance)) on a i7 2.30 GHz processor

Patch configuration	**Techniques**	**Global**	**Local**
whole image (1 block)	No normalization	0.015 s	
	Tan and Triggs	0.018 s	
	Multi Scale Weberfaces	0.054 s	
4 blocks	No normalization	0.045 s	
	Tan and Triggs	0.046 s	0.054 s
	Multi Scale Weberfaces	0.082 s	0.101 s
9 blocks	No normalization	0.089 s	
	Tan and Triggs	0.093 s	0.115 s
	Multi Scale Weberfaces	0.130 s	0.169 s
16 blocks	No normalization	0.153 s	
	Tan and Triggs	0.157 s	0.191 s
	Multi Scale Weberfaces	0.198 s	0.299 s

Table 4.17 shows the average time to process a single ROI using different patch configurations on a i7 2.30 GHz processor. To calculate this average time, a total number of ROI (100 ROIs) is processed into the system and an overall processing time is estimated. The latter, is divided by the total number of ROIs.

It can be concluded that the increase of the number of blocks leads to an increase of processing time. Besides that, the processing time is multiplied by two from global to local approach. As for the IN techniques, Multi Scale Weberfaces has longer processing time than the Tan and Triggs.

There is a dilemma between processing time and recogntion performance. As we stated before, the local approach provides better pAUC and AUPR performance compared to the global approach. However, this increase in performance causes longer processing time.

CHAPTER 5

DYNAMIC QUALITY-BASED WEIGHTING OF TEMPLATE-BASED MATCHER FOR WATCHLIST SCREENING

Local matching for classification has achieved its popularity especially with video-based FR systems. There are many types of local matching in existing research works (See Chapter 2) but the most imminent one would be the non overlapping patch-based methods. Indeed, patch-based local matching (or block division) is considered to be efficient with still-to-video FR systems. Firstly, it helps in overcoming some facial variations such as occlusion due to illumination conditions or others. Besides that, it also helps in providing information on the spatial structure of the face. Finally, non-overlapping patch-based method is less complicated and less expensive in terms of computational costs compared to other techniques and it can be applied on very low quality input image.

Through local matching especially with patch-based technique, assigning weights on local regions became a necessity. As a fact, weighting local regions can enhance FR systems as proven in the work of (Ahonen *et al.*, 2006). As a matter of fact, by giving weights on local blocks of an image, we provide importance and priority to those weighted regions in an image. The importance is high when the assigned weight value is high and vice versa. The technique of assigning weights presented in (Ahonen *et al.*, 2006) is done statically with a prior knowledge of the dataset and the application (verification application). Facial ROIs are divided into uniform non-overlapping local patches. Higher weights are assigned to those regions where more distinctive features such as eyes, nose and mouth are located. Figure 5.1 shows their manner of assigning weights. Based on their results, the system has shown an improvement when compared to non-weighting local matching.

A limitation to their proposed technique is that it is biased on the exploited dataset which contains only frontal images. Therefore local regions containing special and unique features are likely to occupy the same locations from one facial image to the other (static location)

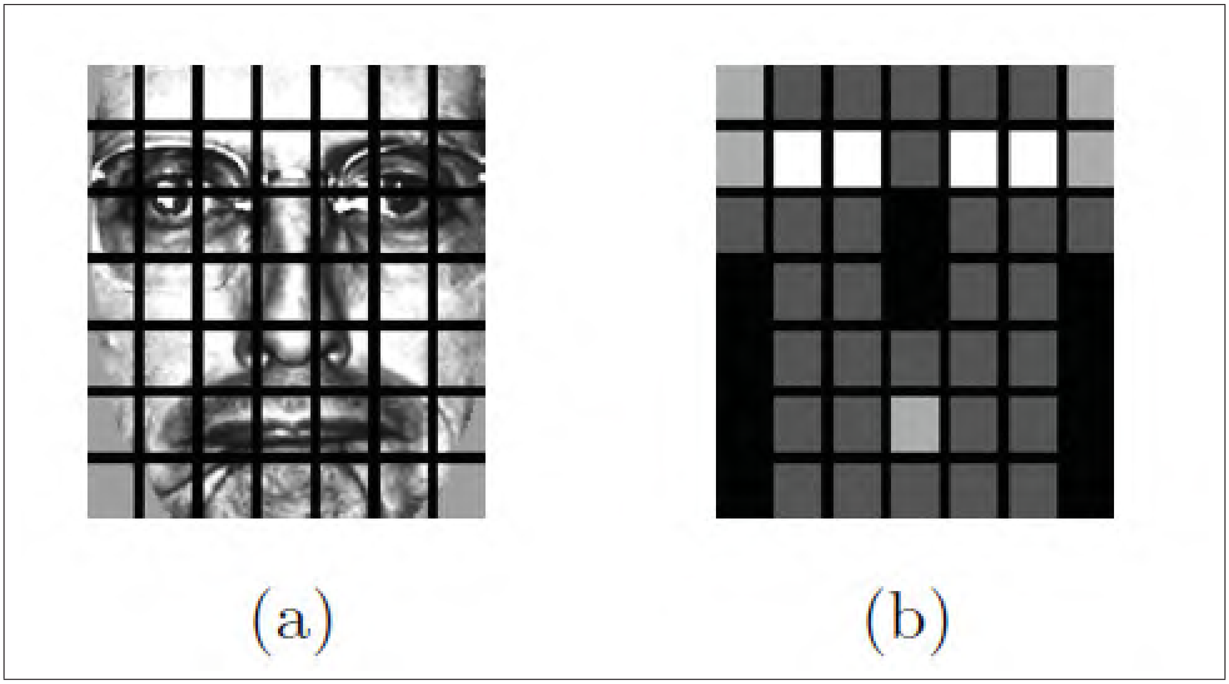


Figure 5.1 Static method of assigning local weights in a facial image using χ^2 dissimilarity measure, (a) Facial image divided into 7×7 patches, (b) Weights: black patches indicate a weight value of 0, dark gray 1, light gray 2 and white 4
Taken from Ahonen *et al.* (2006)

whereas in real video surveillance applications the area and locations of those features may vary from one image or may be occluded from one ROI to the other.

Another approach for local region weighting was proposed by Cheng and Chen (Cheng and Chen, 2014). Their concept was to apply local matching using Euclidean distance and performing their weighting scheme using multiple holistic-based classifiers. They exploited the fact that face regions have different significances when it comes to performing recognition. For that, they divided the facial regions into uniform non-overlapping patches. On each patch, a holistic algorithm is performed. The efficiency in terms of classification of that algorithm and on that region is taken as a weight value. That process was repeated with other several holistic algorithms. For a final weight value, each algorithm casts one vote using the strategy "one takes all". The weights are then obtained by regional majority voting.

Weighting regions was also embraced in (Yule and Chen, 2011). The difference between these two works would be that these weights were estimated dynamically using weighting equations derived from the Regret Minimization concept.

These two weighting approaches have proven to improve the recognition rate compared to no weighting method but the complexity in terms of implementation is complex.

On that account, one part of this Thesis suggests to put weights on local regions corresponding to contextual information that can be easily retrieved and may be fruitful for our still-to-video FR system.

Context was defined in (Zimmermann *et al.*, 2007) as a piece of information that one element can englobe but once acknowledged it may provide an estimation of a specific situation. Context aware techniques have recently emerged in the fields of information fusion (Snidaro *et al.*, 2015). Thus, contextualizing the weights of local patches based on the input data is needed.

When dealing with context-based methods, first thing to determine is the context source. Since quality information has proven to be very useful in most FR systems, Image Quality (IQ) is considered to be a perfect complementary module for recognition and a perfect context source. In previous research work, it has been used to select "good" and suitable facial ROIs at the preprocessing level before performing FE and classification because "poor" quality facial images can degrade the recognition performance of a FR system. In (Huang *et al.*, 2013), a "quality alignment" module is used for selecting good quality frames in their still-to-video FR system. In (Abboud *et al.*, 2009) they exploited no-reference image quality for their wavelet-based FR system and using Nearest Neighbour classification. Their work demonstrated a remarkable improvement in recognition rate. Similarly, in (Abboud and Jassim, 2012) they also used image quality for carefully selecting features/templates for better influence on the adaptive FR system's performance. The results have shown improvements compared to the non-adaptive system.

At this stage, exploiting the context source (IQ) in weighting the local patches would be a perfect blend to achieve a dynamic process for assessing importance of local regions of the available input data.

Another major axis that should be focused on when dealing with still-to-video FR systems is the fact that there is a gap between the facial captures obtained during training and those obtained for testing. This is called a domain shift between the source domain and the target domain. There are many approaches for domain adaptation (please refer to Section 2.5 in Chapter 2). In a still-to-video FR case, especially for video surveillance applications, the probe images obtained are unlabeled. Therefore, we are going to endorse the unsupervised domain adaptation.

Finally, the main goal of this work is to propose a more general and dynamic way of assigning weights that brings together both concepts contextual information through image quality and instance weighting motivated by domain adaptation for watchlist screening applications.

5.1 Framework and Algorithms

The framework of this contribution is given in Figure 5.2. As any FR system, this proposed system consists in having two main phases: the design phase and the operation (or testing) phase.

Through a quick glimpse into the proposed system's framework, the first observation would be directly related to the complementary module for classification. This module is responsible for assigning local weights on non overlapping patches for each input ROI. By looking deeply into this module, a new component, aside from IQA is used which is the *specialized window*. The latter is responsible for giving more generality for the weighting technique depending on the camera view point. Indeed, one of the main characteristics of the proposed system is to provide suitable local weights for classification for every camera view point that is dependent on the quality of the input ROIs.

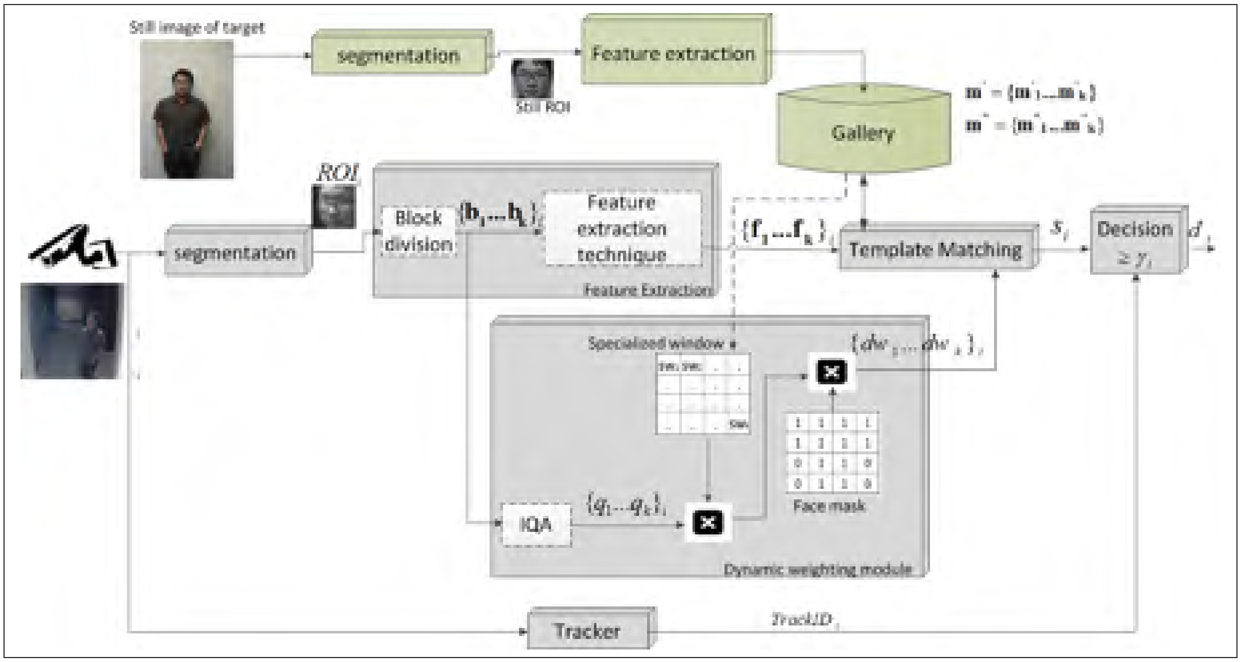


Figure 5.2 Global framework of the still-to-video FR system with dynamic weighting technique

To summarize, the proposed system assigns different weights according to the quality of the input ROI using IQA techniques and also according to a camera location and view point using the *specialized window*.

5.1.1 System construction

The design phase of the proposed system is slightly different from the traditional still-to-video FR systems. It is composed of two major parts:

Enrolment of the watchlist individuals

This resides in enrolling the target individuals into the system. Nb_{ind} individuals are randomly selected as watchlist. ROIs from the still images (mug-shots) are extracted using FD algorithm. A conversion into a gray scale and a resizing into a common size of $N_{size} \times N_{size}$ pixels is performed. The ROIs undergo *block division* into $N_p \times N_p = N_{TP}$ non overlapping patches and

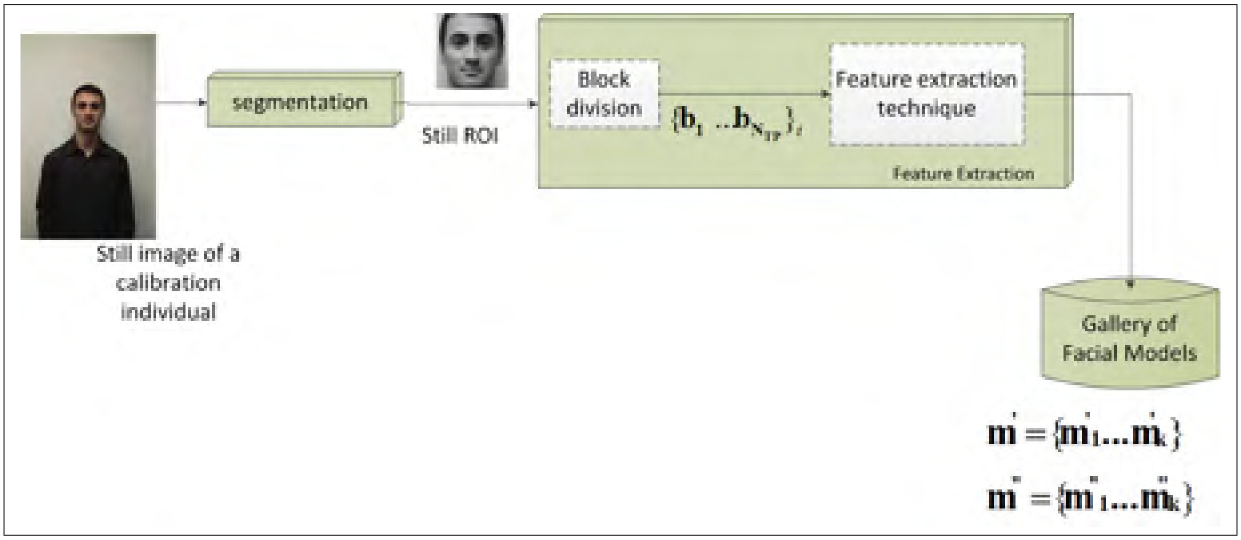

Still image of a calibration individual
segmentation
Still ROI
Block division
$\{\mathbf{b}_1 \ldots \mathbf{b}_{N_{TP}}\}_i$
Feature extraction technique
Feature Extraction
Gallery of Facial Models
$\mathbf{m}' = \{\mathbf{m}'_1 \ldots \mathbf{m}'_k\}$
$\mathbf{m}'' = \{\mathbf{m}''_1 \ldots \mathbf{m}''_k\}$

Figure 5.3 Enrolment process of the calibration individuals used during the design of the *specialized window*

each patch has the size of $p_s \times p_s$. Finally, each patch of an ROI is fed to the *feature extraction* module then are saved into a set of features $\mathbf{M} = \{\mathbf{m}_1, ..., \mathbf{m}_{N_{TP}}\}$ in the gallery.

Calibration: Design of the *specialized window* using domain adaptation

This calibration process should be performed off-line for every camera view point. This task consists in implementing a classification process using template matching between low quality, low resolution and unlabeled probe video frames in the target domain and the enrolled high quality resolution still images of the calibration individuals in the source domain.

The *specialized window* contributes iteratively to re-weight the patches or local regions based on prior knowledge performed in advance for each camera viewpoint. This process borrows the concept of instance-weighting domain adaptation technique to assign weights on local regions. The key idea is to overcome the existing domain shift by finding a camera-specific window based on the component features and quality properties of the data distribution in the target domain.

Figure 5.4 shows the steps to obtain a specialized window for a certain camera view point. The calibration individuals are first enrolled into the gallery as facial models using their still reference image and undergo the same traditional algorithm process for enrolment shown in Figure 5.3. Once these individuals are enrolled, the next step consists in the operational phase of this sub FR system. For one camera viewpoint, ROIs are extracted from the video sequence by performing the basic *segmentation* process (gray scale transformation and ROI resizing into $N_{size} \times N_{size}$ pixels). The segmented ROI goes through **block division** which divides the facial images into $N_p \times N_p$ having N_{TP} non overlapping blocks producing a set of blocks $\mathbf{b}_r = \{\mathbf{b}1, ..., \mathbf{b}_{N_{TP}}\}_r$.

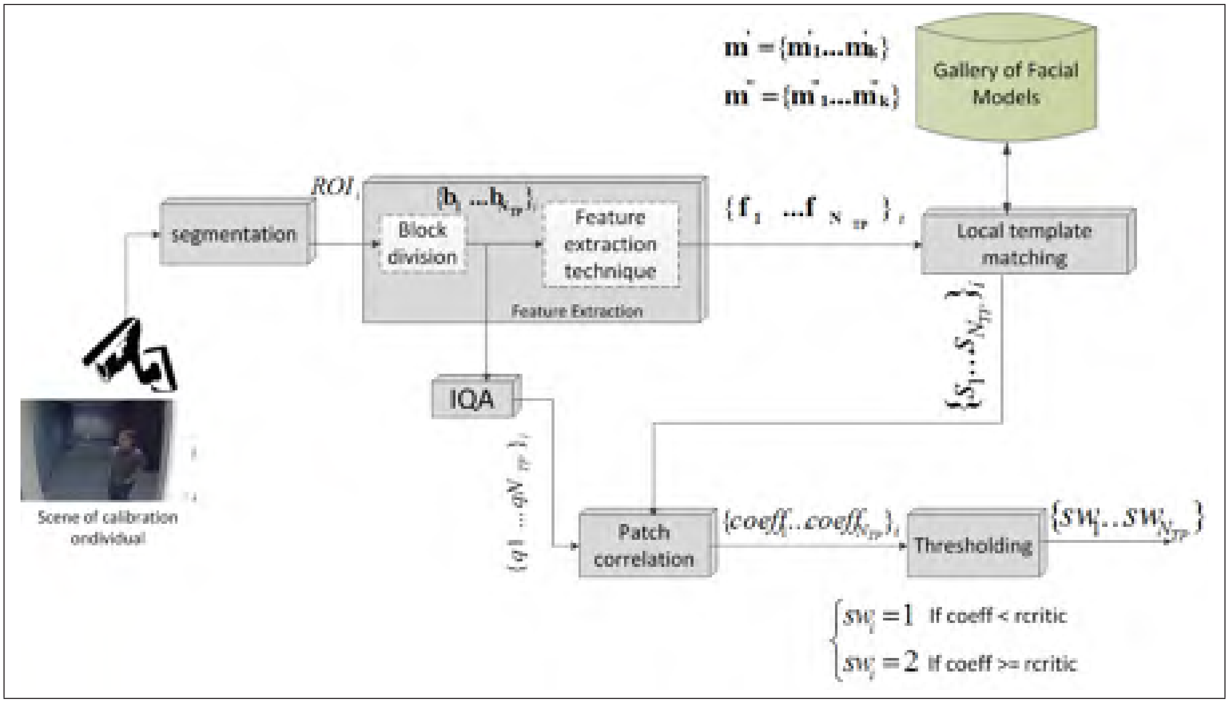


Figure 5.4 Design of the *specialized window*

At this stage a two-layer process should be executed in order to calculate the *specialized window*.

First layer: local matching and image quality assessment

For each calibration individual, a series of tasks should be done.

a. Local matching

This first task consists in performing direct matching. Each extracted ROI from the video frames goes through **segmentation** and **FE technique**. The obtained features are **locally matched** to the features from the still images of the calibration individuals saved in the gallery (one individual at a time) giving each ROI capture a set of local matching scores $\mathbf{S}_r^t = \{S_1, ..., S_{N_{TP}}\}_r^t$. For all frames from the video we would have a series of sets of local matching score.

b. Local IQA

This task consists in performing IQA locally on each patch or region of a captured ROI providing a set of local quality scores $\mathbf{q}_r^t = \{q_1, ..., q_{N_{TP}}\}_r^t$. For all frames from the video we would have a series of sets of local quality scores.

c. Patch correlation

At this stage, each individual used for calibration would have a series of sets of local matching scores and local sets of quality scores. These scores undergo **patch correlation**. This module considers one patch at a time for all frames in the video and assesses the relationship between the local matching scores and the local quality scores by providing correlation coefficient. So, for one calibration individual, a set of local correlation coefficient is obtained $\{coeff_1, ..., coeff_{N_{TP}}\}$.

A primary binary window (of values 1 and 2) is created by performing **thresholding**. Each local correlation coefficient is compared to a *critical correlation coefficient* [1] r_{critic} value. If the local correlation coefficient of one patch is higher than the critical coefficient r_{critic} then the value of the corresponding patch in the primary binary window is equals to 2 if not, then it is equals to 1 providing a final product: the *specialized window* for one calibration individual and for one camera viewpoint.

[1] refer to Appendix I

Second layer: camera specific *specialized window* using majority voting

A very important point that should be taken into consideration is that the design of the *specialized window* is done for each individual used for calibration per camera view point. So, the final camera specific *specialized window* that is integrated to the gallery of models (for a camera view point) is the combination of all 5 *specialized window* from all calibration people using the majority voting approach. The regions or patches containing more votes of value "2" are given that same value and correspondingly regions having more votes of value "1" are given that same value.

A basic flowchart is given in Figure 5.5 and the algorithm explaining step by step the process of designing the *specialized window* is provided in 5.1. A graphic example of this process on one recording condition is provided in Appendix II.

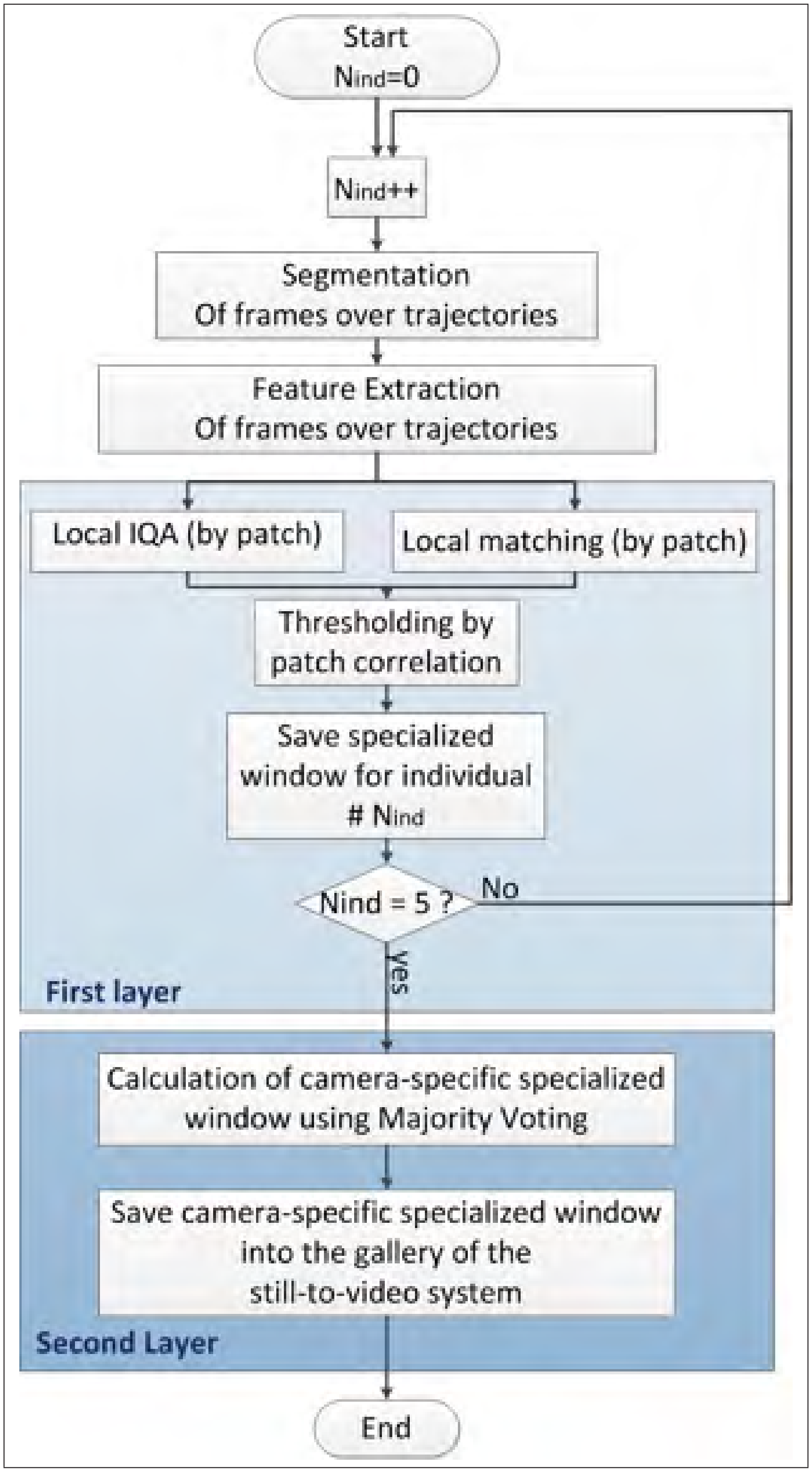


Figure 5.5 Flow chart for calculating the *specialized window*

Algorithm 5.1 Design of the specialized window per camera view point (design2) -part1

```
Input: Input frames from video sequence $I_r : r = 1,...,R$. Gallery-face-models $\{\mathbf{m}_t : t = 1,...,Nb_{calib}\}$ of
        calibration individuals enlisted in the gallery
Output: specialized window $\mathbf{sw} = \{sw_1,...,sw_{N_{TP}}\}$.
for each calibration individual $t$ with $t = 1..Nb_{calib}$ do
    for each $I_r$ input frame with $r = 1,..,R$ do
        // segmentation module: faced detection, resizing, gray scaling
        Apply segmentation to detect $ROI_r$ corresponding to faces in a frame $r$.
        //Block division
        Divide $ROI_r$ into $N_{TP}$ uniform and non-overlapping blocks producing $\mathbf{b}_r = \{\mathbf{b}1,...,\mathbf{b}_{N_{TP}}\}_r$
        blocks.
        //feature extraction
        for each component of $\boldsymbol{b}_r$ do
            Apply FE technique producing set of features $\mathbf{f}_r = \{\mathbf{f}1,...,\mathbf{f}_{N_{TP}}\}_r$.
        end
        //Local matching
        for each face model in the gallery $m_t : t = 1,...,N_{calib}$ do
            Compute dissimilarity scores $\mathbf{S}_r^t = \{S_1,...,S_{N_{TP}}\}_r^t$.
        end
        //Image Quality Assessment
        for each component of $\boldsymbol{b}_r$ do
            Apply IQA technique producing a set of quality scores $\mathbf{q}_r^t = \{\mathbf{q}1_r^t,...,\mathbf{q}N_{TP}{}_r^t\}$
        end
    end
    //patch correlation $stemp_p^t = []$ ; $qtemp_p^t = []$;
    //$N_s$: number of samples in a target to target trajectory
    for each sample $j$ with $j = 1,..,N_s$ do
        for each local patch $p$ with $p = \{1,...,N_{TP}\}$ do
            $stemp_p^t = stemp_p^t + S_{jp}^t$; $qtemp_p^t = qtemp_p^t + q_{jp}^t$ ;
        end
        //Calculate correlation using Pearson's rule
        $coeff_p = correlation\_pearson(stemp_p^t, qtemp_p^t)$;
        producing a set of coefficient for a trajectory
        $\mathbf{coeff}^t = \{coeff_1^t,...,coeff_{N_{TP}}^t\}$.
    end
    //Threshold
    Calculate Pearson's critical value $r_{critic}$
    for each patch $p$ with $p = 1,...,N_{TP}$ do
        if $coeff_p^t \geq r_{critic}$ then
            $swp_p^t = 2$
        end
        else
            $swp_p^t = 1$
        end
        producing a primary binary window $\mathbf{swp}^t = \{swp_1^t,...,swp_{N_{TP}}^t\}$
    end
end
```

Algorithm 5.1 Design of the specialized window per camera view point (design2) -part2

```
//Global combined specialized window
for $each\ patch\ p = \{1..N_{TP}\}$ do
    for $each\ individual\ t : 1, .., 5$ do
        if $sw_p^t = 2$ then
            $counter = counter + 1$
        end
        if $counter \geq 3$ then
            $sw_p = 2$
        end
        else
            $sw_p = 1$
        end
    end
    producing $\mathbf{sw} = \{sw_1, ..., sw_{N_{TP}}\}$
end
```

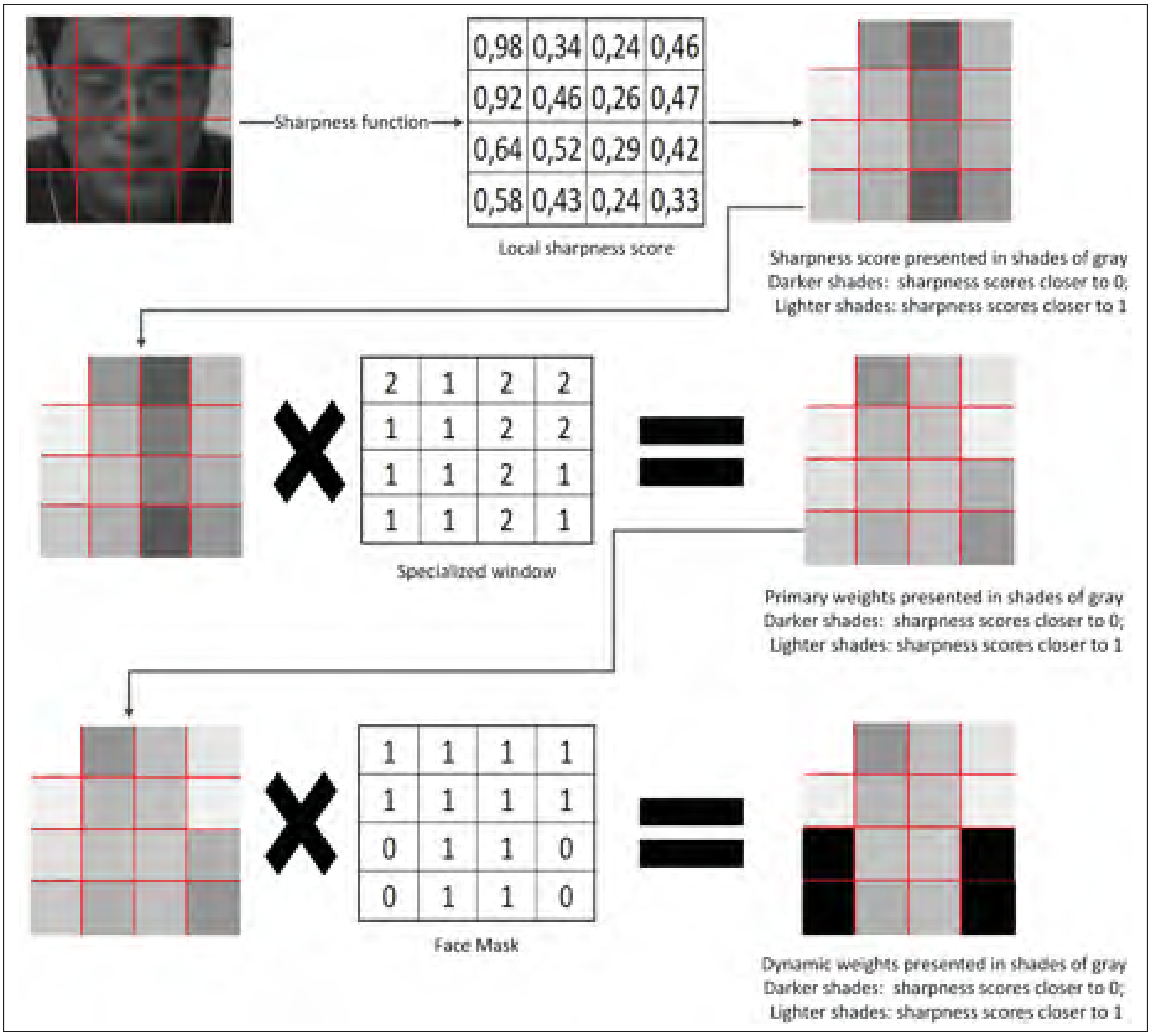
Through matching, correlation and majority voting a knowledge about the local regions is obtained. Indeed, for a single camera viewpoint, we were able to encourage and give importance to those regions/patches that are susceptible to contain high local matching scores and high local quality score by assigning values of "2" and to reduce the importance of certain regions by giving values of "1".

Through domain adaptation, we try to learn a better model for the source domain by:

- Getting knowledge from the target domain;
- Exploiting the particular information in the source domain.

5.1.2 Operation phase

The probe video frames goes through the same **segmentation** process described in the enrolment of the target individuals to obtain a resized and gray scaled facial ROI on each frame. By then, this ROI endures **block division** into N_{TP} non overlapping local patches. These patches are first taken through the **FE module** to produce a set of local features of an ROI. The most important part of this phase would be the weight calculation module in which a domain adaptive information (camera-specific specialized window) obtained from the target domain in the calibration process is exploited along with contextual information based on quality to provide dynamic regional weights.


Sharpness function
Local sharpness score
Sharpness score presented in shades of gray
Darker shades: sharpness scores closer to 0;
Lighter shades: sharpness scores closer to 1
Specialized window
Primary weights presented in shades of gray
Darker shades: sharpness scores closer to 0;
Lighter shades: sharpness scores closer to 1
Face Mask
Dynamic weights presented in shades of gray
Darker shades: sharpness scores closer to 0;
Lighter shades: sharpness scores closer to 1

0,98	0,34	0,24	0,46
0,92	0,46	0,26	0,47
0,64	0,52	0,29	0,42
0,58	0,43	0,24	0,33

2	1	2	2
1	1	2	2
1	1	2	1
1	1	2	1

1	1	1	1
1	1	1	1
0	1	1	0
0	1	1	0

Figure 5.6 Dynamic weight calculation for one single image of individual id1 in P2E_S1_C3

Weight Calculation using Contextual Information

For a camera viewpoint, Image Quality (IQ) is performed locally on each patch of the ROI providing a set of quality scores $\{q_1,...,q_k\}$. The latter is multiplied to the set of values of the camera-specific *specialized window* (retrieved from the gallery) of that camera viewpoint used during the operation $\{sw_1,...,sw_k\}$ producing a primary quality-based weights for each local area $\{q_1.sw_1,...,q_k.sw_k\}$. Due to the presence of unwanted background a **face mask** $\mathbf{f}_{mask} = \{fm_1,...,fm_k\}$ is applied to the primary attributed quality-based weights giving a final product of dynamic weights $\mathbf{dw} = \{dw_1,...,dw_k\} = \{q_1.sw_1.fm_1,...,q_k.sw_k,fm_k\}$. Dynamic weight calculation is illustrated for one single image for one specific camera recording condition in Figure 5.6. Finally, this window containing the set of dynamic weights is injected into the matching module for **classification** along with the corresponding set of features. The matcher produces a single value per ROI comparison between 0 and 1 (having 0: likely to match and 1 unlikely to match).

5.2 Contributions

The proposed system exploits:

- Domain adaptation at the calibration process of the system by defining a proper and specific knowledge (camera specific *specialized window*) to maximize regions that are susceptible to provide high local matching scores and high local quality scores that will be incorporated to the target domain;
- Contextual information with IQA at the operation phase giving a dynamic property to the still-to-video FR system. This also provides importance on local regions of the ROI based on the quality of the input data.

5.3 Research Hypotheses

Some hypotheses can be pulled out:

- Static weighting (i.e having weights pre-assigned) can not be generalized and it may not be suitable for video surveillance applications where a lot of variations are uncontrolled;
- Performing dynamic and contextual weighting that is dependant on the input image data provides better positive influence on the performance of still-to-video FR systems especially when quality information is exploited in assigning these weights.

These hypotheses are to be proved in the next sections.

5.4 Experimental Methodology

5.4.1 Dataset Exploitation

This second study considers $Nb_{calib} = 5$ randomly chosen individuals as calibration individuals out of 29 subjects from the ChokePoint dataset. Another $Nb_{ind} = 5$ individuals are chosen as person of interests for the watchlist out of the remaining 24 individuals. The remaining people are supposed to be unknown.

For this proposed system, all the video recordings from the ChokePoint dataset (Figure 3.2 are taken into consideration to estimate the versatility of the system at different camera viewpoints.

5.4.2 Experimental Protocol

At the design phase, more specifically, during the enrolment of the watchlist individuals. ROIs from the $Nb_{ind} = 5$ still images are extracted using Viola-Jones algorithm (Viola and Jones, 2001), resized into a common size of $48 \times 48, N_{size} = 48$. Based on the results obtained from the previous system (Chapter 4, larger number of patches for block division provides stability in results and improves the still-to-video FR system's performance. For this main reason and for less computational cost $N_{TP} = 16$ non overlapping patches is chosen. Feature extraction is executed on each patch using LBP and LPQ texture descriptors described in Section 2.1 giving a set of concatenated feature vectors for each method. In this study, a comparison of results using different texture descriptors is conducted to assess the effect of each one on the proposed quality based system. Finally, these features are saved into the gallery. Concerning the design of the specialized window per camera viewpoint, the sub FR system's enrolment using $Nb_{calib} = 5$ individuals follows the same protocol described previously for watchlist individuals: Viola-Jones for FD, resizing into 48×48 pixels, gray scale conversion, block division into $N_{TP} = 16$ non overlapping patches and feature extraction using LBP and LPQ.

Table 5.1 Summary of techniques implemented on each module for the proposed system: Dynamic weighting of local regions

Techniques implemented in Study 2		
Modules	**Experiment 1**	**Experiment 2**
Segmentation	Face Detection: Viola-Jones Image gray-scale transformation Image resizing into 48×48 pixels	
pre-processing	None	Illumination Normalization: Multi-Scale Weberfaces
Block division	$N_{TP} = 16$ non overlapping patches	
Feature Extraction	LBP and LPQ	
Classification: Sub FR for system design	Local Chi Square (Equation 2.17)	
Classification : Online operation phase	Weighted Chi Square (Equation 2.18)	
Image Quality Assessment	Sharpness (Equation 2.32)	
Patch correlation	Pearson's correlation rule	
Thresholding	Pearson's critical value	

As for the operational phase of this **sub FR system**, the ROIs extracted from the pre-recorded video sequence using Viola-Jones undergo the same series of processing (resizing, gray scale conversion, block division into 16 blocks). Local matching is then performed using the Chi Square dissimilarity measure found in Equation 2.17. Simultaneously, the 16 patches go through sharpness quality metric found in Equation 2.32 as IQA.

For patch correlation and for one individual of the calibration individuals, local scores and quality are evaluated using Pearson's correlation rule. In other terms, correlation is executed individually on each patch of the set of local Chi Square scores taken from target-to-target trajectory situation along with the corresponding local sharpness scores. To evaluate the conjunction of the local scores and the local sharpness on each patch, correlation coefficients is extracted and compared to a critical value r_{critic}. The latter depends on the number of target trajectory sample and is obtained using Pearson's critical value (Appendix I). If the local correlation coefficient is higher than the critical value the corresponding patch of the window would have the value of 2 else the value of 1. The final task of this window design is the application of the window mask for background factor elimination. A detailed example of the design of the specialized window for all calibration individuals is given in Appendix II. The global window specific to a camera viewpoint is obtained by combining the 5 specialized window using majority voting.

During the operational phase of the proposed system, **dynamic weighting** module takes the $N_{TP} = 16$ blocks of the segmented ROI from the video and apply sharpness metric for IQA providing a set of local sharpness scores. The latter is then multiplied with the global specialized window of the camera viewpoint giving a set of 16 dynamic weights. Classification is done on the extracted LBP and LPQ features using the weighted Chi Square method presented in Equation 2.18. Lastly, a global score is obtained.

An additional experiment is conducted with this proposed system. It is related to the results obtained in the previous study. IN technique is applied at the pre-processing level. The technique chosen Multi-Scale Weberfaces since it has outperformed the other IN techniques.

Details about the methods used in the whole system construction is given in Table 5.1.

5.5 Results and Discussion

Experiment 1

A comparison is held between three techniques: the baseline (without weights), static weights in analogy with the work presented in (Ahonen *et al.*, 2006) and the proposed technique (dynamic weighting with sharpness metric).

Table 5.2 pAUC(20%) and AUPR performance per individual in the watchlist in P1E_S1_C1

Average pAUC(20%) and AUPR											
Feature Extraction	**Techniques**	**id 3**		**id 4**		**id7**		**id9**		**id 12**	
		pAUC	**AUPR**	**pAUC**	**AUPR**	**pAUC**	**AUPR**	**pAUC**	**AUPR**	**pAUC**	**AUPR**
LBP	**No weights**	0.58	0.67	0.62	0.69	0.16	0.20	0.78	0.17	0.91	0.94
	Static weights	0.65	0.79	0.70	0.62	0.31	0.14	0.29	0.10	0.91	0.80
	Dynamic weights	0.69	0.45	0.75	0.55	0.46	0.10	0.48	0.07	0.98	0.88
LPQ	**No weights**	0.66	0.63	0.71	0.66	0.43	0.26	0.77	0.55	0.84	0.58
	Static weights	0.72	0.64	0.78	0.76	0.38	0.19	0.57	0.25	0.83	0.62
	Dynamic weights	0.76	0.73	0.83	0.78	0.45	0.21	0.64	0.28	0.98	0.78

Table 5.2 shows the performance values in terms of pAUC(20%) and AUPR for both feature extraction techniques LBP and LPQ. The proposed dynamic weighting technique gives a better recognition performance for the majority of the individuals in the watchlist. For example, for individual id 3 the dynamic weighting has achieved an important improvement of almost 10% using LBP and LPQ. The same observation can be made with id 4, id 7 and id 12 that the proposed method has enhanced the performance of the still-to-video FR system. Even so, individual id 9 has shown a decrease of performance with both LBP and LPQ.

To know more about the main reason of this decrease with individual id 9, a lower level analysis showing the trajectory scores and weight values over time is given in Figures 5.7 and 5.8.

Figure 5.7 provides a good case scenario with individual id 3 for frames from video sequence P1E_S1_C1 where our proposed system is performing well. A closer look onto the target-to-

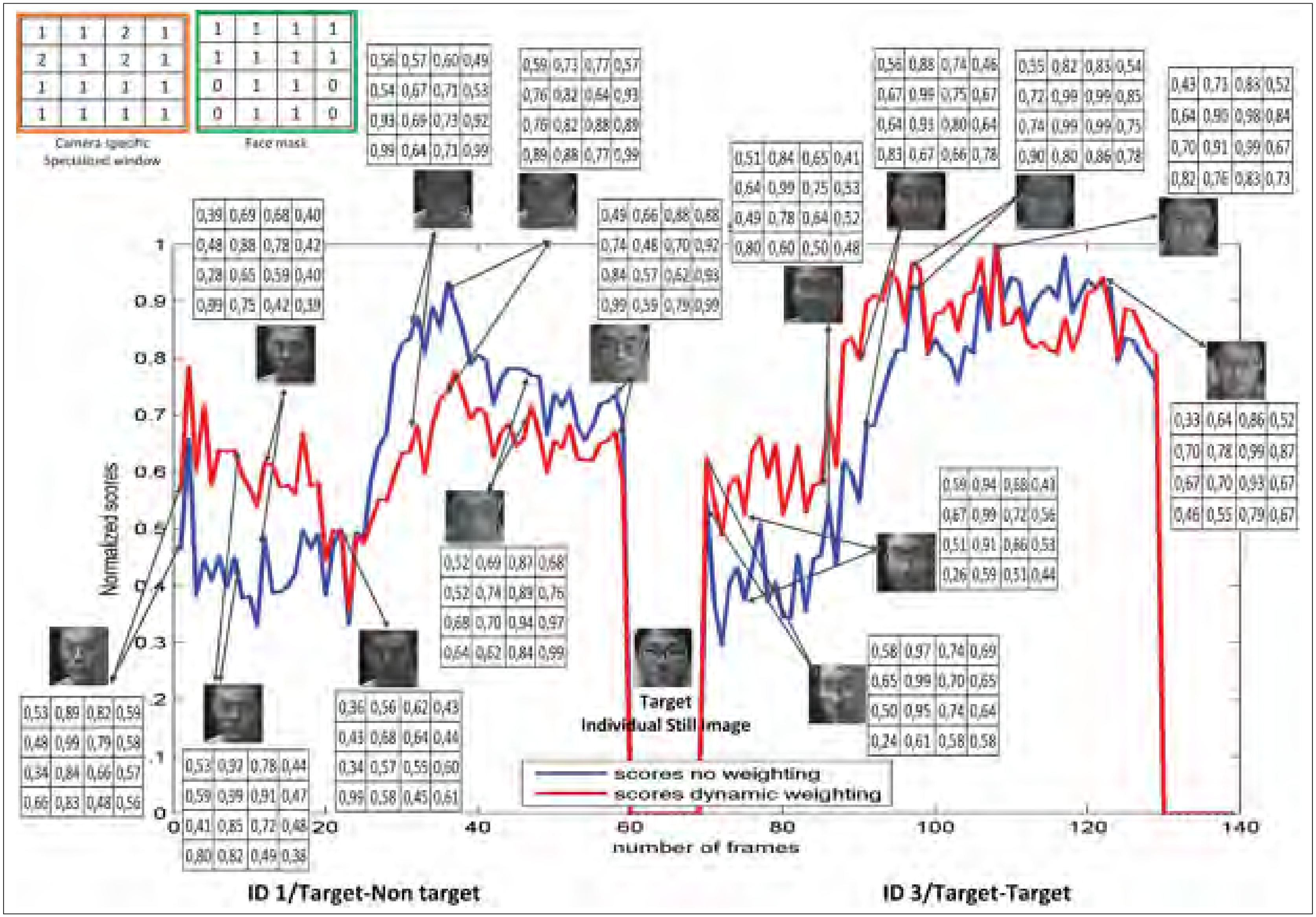
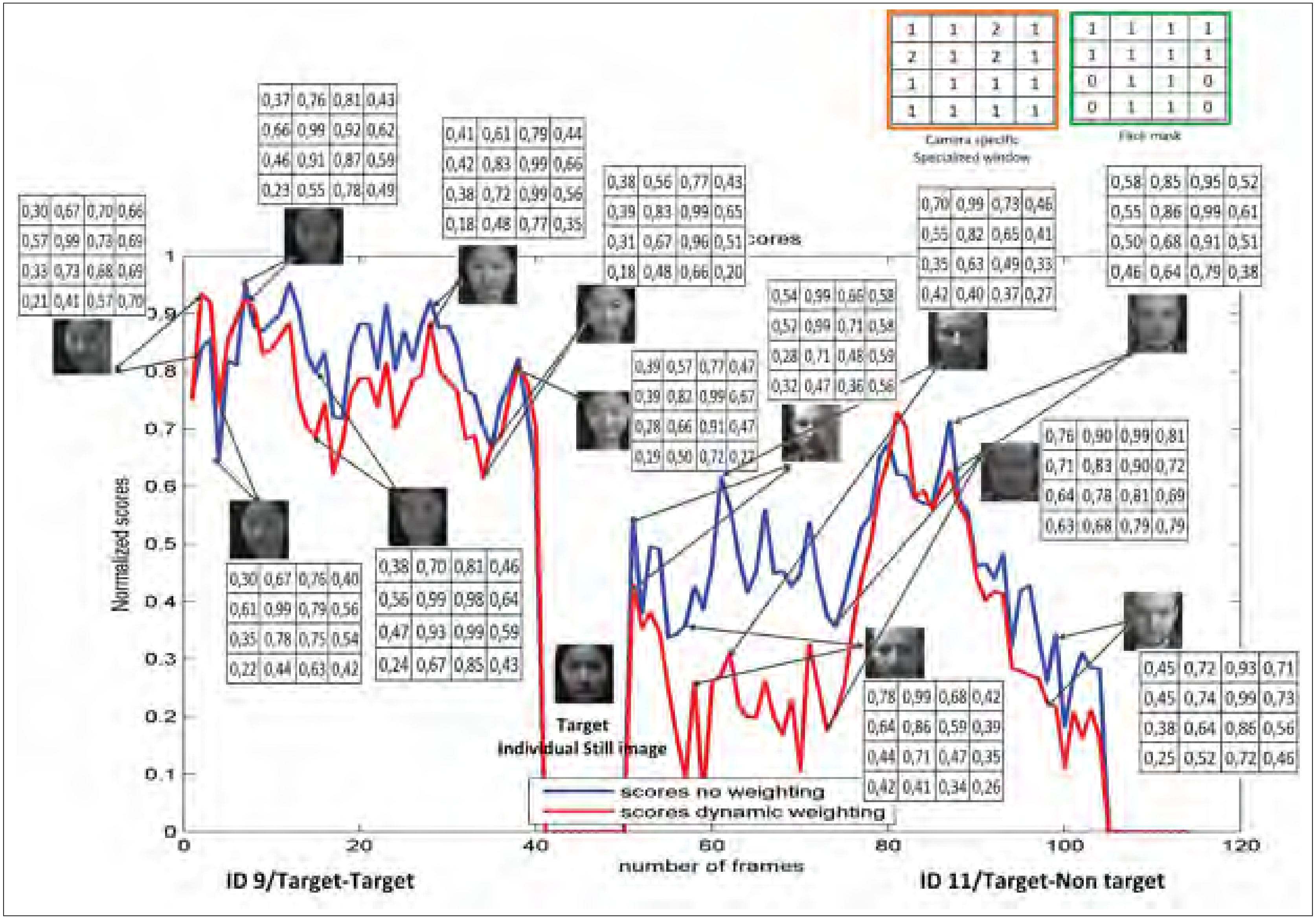


Figure 5.7 Trajectory of scores with sharpness values overtime; A good case



0,30	0,67	0,70	0,66
0,57	0,99	0,73	0,69
0,33	0,73	0,68	0,69
0,21	0,41	0,57	0,70

0,30	0,67	0,76	0,40
0,61	0,99	0,79	0,56
0,35	0,78	0,75	0,54
0,22	0,44	0,63	0,42

0,37	0,76	0,81	0,43
0,66	0,99	0,92	0,62
0,46	0,91	0,87	0,59
0,23	0,55	0,78	0,49

0,38	0,70	0,81	0,46
0,36	0,99	0,98	0,64
0,47	0,93	0,99	0,59
0,24	0,67	0,85	0,43

0,41	0,61	0,79	0,44
0,42	0,83	0,99	0,66
0,38	0,72	0,99	0,56
0,18	0,48	0,77	0,35

0,38	0,56	0,77	0,43
0,39	0,83	0,99	0,65
0,31	0,67	0,96	0,51
0,18	0,48	0,66	0,20

0,39	0,57	0,77	0,47
0,39	0,82	0,99	0,67
0,28	0,66	0,91	0,47
0,19	0,50	0,72	0,27

0,54	0,99	0,66	0,58
0,57	0,99	0,71	0,54
0,28	0,71	0,48	0,59
0,32	0,47	0,36	0,56

0,78	0,99	0,68	0,42
0,64	0,86	0,59	0,39
0,44	0,71	0,47	0,35
0,42	0,41	0,34	0,26

1	1	2	1
2	1	2	1
1	1	1	1
1	1	1	1

0,70	0,99	0,73	0,46
0,55	0,82	0,65	0,41
0,35	0,63	0,49	0,33
0,42	0,40	0,37	0,27

0,76	0,90	0,99	0,81
0,71	0,83	0,90	0,72
0,64	0,78	0,81	0,69
0,63	0,68	0,79	0,79

1	1	1	1
1	1	1	1
0	1	1	0
0	1	1	0

0,58	0,85	0,95	0,52
0,55	0,86	0,99	0,61
0,50	0,68	0,91	0,51
0,46	0,64	0,79	0,38

0,45	0,72	0,93	0,71
0,45	0,74	0,99	0,73
0,38	0,64	0,86	0,56
0,25	0,52	0,72	0,46

Figure 5.8 Trajectory of scores with sharpness values overtime; A bad case

target scores overtime, we could see that the trajectory scores from our technique are higher than the trajectory scores obtained from no weighting method. As we can also affirm, our dynamic weights of local regions through frames is giving proper importance to regions that are high in terms of quality sharpness which helps boosts the system matching process. Figure 5.8 is a bad case scenario individual id 9 where the system's performance seems to degrade. Indeed, the target-to-target scores from our proposed system overtime are slightly lower than the trajectory scores obtained from no weighting although the trajectory values of sharpness (weights) overtime are given correctly. An analysis of the dataset from individual id 9 from video sequence P1E_S1_C1 is necessary. As we can see from Figure 5.8 a majority of the captured frames from the probe video is poor in terms of illumination. Added to that, the still-image of the target individual id 9 presents distortions in terms of illumination creating a sort of occlusion. These reasons mentioned above may cause in obtaining less discriminative facial models, thus, degrading the matching process.

Tables 5.3 and 5.4 show the average pAUC(20%) performance after 5 system replications using different types of feature extraction techniques which are LBP and LPQ. The first observation that can be done is that the proposed system is outperforming overall whether using LBP (having an overall average of 0.52) or LPQ (having an overall average of 0.55).

Table 5.3 Average pAUC(20%) performance (with standard error) in all portals of the ChokePoint dataset after 5 replications using LBP

Average pAUC(20%)					
Techniques	**Portal 1**		**Portal 2**		
	Entering	Leaving	Entering	Leaving	AVG
No weights	0.47 ± 0.075	0.60 ± 0.091	0.38 ± 0.057	0.48 ± 0.066	0. 48 ± 0.072
Static weights	0.52 ± 0.085	0.60 ± 0.094	0.36 ± 0.056	0.50 ± 0.070	0.49± 0.076
Dynamic weights	**0.57 ± 0.088**	0.59 ± 0.091	**0.38 ± 0.059**	**0.53 ± 0.078**	**0.52 ± 0.079**

Some portal cases provides better improvements than the others. In **portal 1 entering** using LBP feature extraction technique a significant enhancement on the system's performance using the proposed method (dynamic weighting technique) is found by almost 10% compared to the

Table 5.4 Average AUPR performance (with standard error) in all portals of the ChokePoint dataset after 5 replications using LBP

Average AUPR					
Techniques	**Portal 1**		**Portal 2**		
	Entering	**Leaving**	**Entering**	**Leaving**	AVG
No weights	0.32 ± 0.045	0.45 ± 0.050	0.22 ± 0.034	0.31 ± 0.044	0.32 ± 0.043
Static weights	0.37 ± 0.045	0.47 ± 0.050	0.23 ± 0.036	0.34 ± 0.048	0.35 ± 0.045
Dynamic weights	**0.42± 0.051**	0.46 ± 0.055	**0.24 ± 0.038**	**0.38 ± 0.052**	**0.37 ± 0.049**

baseline (no weights) and by 5% compared to the static weighting. Besides that, improvements also can be found in **portal 2 leaving** where the proposed technique has improved by 5% compared to the baseline and 3% compared to the static weighting.

Table 5.5 Average pAUC(20%) performance (with standard error) in all portals of the ChokePoint dataset after 5 replications using LPQ

Average pAUC(20%)					
Techniques	**Portal 1**		**Portal 2**		
	Entering	Leaving	Entering	Leaving	AVG
No weights	0.54 ± 0.079	0.64 ± 0.095	0.42 ± 0.062	0.56 ± 0.078	0.54 ± 0.078
Static weights	0.55 ± 0.088	0.64 ± 0.098	0.41 ± 0.062	0.52 ± 0.073	0.53± 0.080
Dynamic weights	**0.59 ± 0.090**	0.60 ± 0.092	**0.44 ± 0.064**	**0.56 ± 0.081**	**0.55 ± 0.082**

Table 5.6 Average AUPR performance (with standard error) in all portals of the ChokePoint dataset after 5 replications using LPQ

Average AUPR					
Techniques	**Portal 1**		**Portal 2**		
	Entering	**Leaving**	**Entering**	**Leaving**	AVG
No weights	0.39 ± 0.044	0.50 ± 0.047	0.26 ± 0.038	0.38 ± 0.044	0.38 ± 0.043
Static weights	0.38 ± 0.047	0.49 ± 0.045	0.25 ± 0.034	0.34 ± 0.035	0.36 ± 0.040
Dynamic weights	**0.45± 0.049**	0.46 ± 0.047	**0.27 ± 0.040**	**0.39 ± 0.042**	0.39 ± 0.044

As for LPQ feature extraction technique improvements can be seen in **portal 1 entering** by 5% compared to the baseline and 4% compared to the static weighting. In addition, an increase is also found in **portal 2 entering** by 2% compared to the baseline technique and 3% compared to the static technique.

Globally, using LBP, the system has improved by 4% (from 0.42 to 0.46) as to the baseline and 3% (from 0.43 to 0.46) in comparison with the static weighting technique. Although the values have risen when using LPQ FE compared to those using LBP, the overall average improvement of the proposed system seems to be very minor. As a matter of fact, there is only 1% improvement between the proposed method and the baseline technique and 2% increase as to the static weighting technique.

At this stage, LBP feature extraction technique for the proposed system has more significant improvements in comparison with the system using LPQ. Indeed, to confirm this statement a better look on the AUPR performance in Tables 5.4 and 5.6 must be done. Using LBP, the average AUPR has increased by 5% compared to non weighting and 3% compared to the static weighting. However, using LPQ technique there is only 1% increase between the proposed technique and the static weighting. This finding helps to affirm one of the hypotheses stated in the previous section of this chapter that the dynamic local weighting using quality has better influence on the performance of the still-to-video FR system. Since LPQ is robust to blurriness, using sharpness as quality metric for weighting does not provide additional impact on the proposed system. Which explains the modest improvement of the proposed system.

In Table 5.3, more specifically in **portal 2 entering**, the average pAUC(20%) performance of the proposed system is almost similar to the baseline technique (no weights) which is at 0.38. To better assess the performance of the classification process, a better look into the AUPR values at the same portal case: **portal 2 entering** is needed. The AUPR value shown in Table 5.4 of the proposed system is higher (0.24) compared to both baseline (0.22) and static weighting (0.23). Therefore, the dynamic weighting has a better system performance.

Following the same reasoning mentioned above, in Table 5.5 with LPQ features in **portal 2 leaving**, a same observation can be made about the proposed dynamic weighting method having a better performance. In fact, although the average pAUC performance shows a similar value of 0.56 to the baseline technique, its AUPR value in Table 5.6 is slightly higher than the AUPR values of the baseline technique.

Table 5.7 Average pAUC(20%) and AUPR for portal 1 entering after 5 replications using LBP and LPQ

Feature Extraction	Techniques	**Camera 1 (C1)**		**Camera 2 (C2)**		**Camera 3 (C3)**	
		pAUC(20%)	AUPR	pAUC(20%)	AUPR	pAUC(20%)	AUPR
LBP	No weights	0.50 ± 0.076	0.33 ± 0.041	0.50 ± 0.081	0.37 ± 0.052	0.42 ± 0.067	0.27 ± 0.043
	Static weights	0.54 ± 0.084	0.39 ± 0.045	0.55 ± 0.091	0.41 ± 0.052	0.46 ± 0.079	0.31 ± 0.036
	Dynamic weights	**0.60 ± 0.090**	**0.46 ± 0.026**	**0.59 ±0.092**	**0.45 ± 0.056**	**0.51 ± 0.080**	**0.36 ± 0.043**
LPQ	No weights	0.56 ± 0.080	0.40 ± 0.040	0.56 ± 0.083	0.40 ± 0.051	0.50 ± 0.073	0.36 ± 0.040
	Static weights	0.53 ± 0.085	0.37 ± 0.047	0.57 ± 0.090	0.41 ± 0.054	0.53 ± 0.087	0.37 ± 0.040
	Dynamic weights	**0.61 ± 0.091**	**0.45 ± 0.052**	**0.62 ± 0.094**	**0.48 ± 0.054**	**0.54 ± 0.083**	**0.40 ± 0.039**

Let us present the results per camera domain. Table 5.7 shows the average performance after 5 replications using LBP and LPQ for **portal 1 entering**. For both feature extraction techniques, the proposed system has shown an improvement overall. In fact, using LBP, the pAUC(20%) improvement in camera 1 is 10%, 9% in both camera 1 and camera 2. As for using LPQ, an enhancement of 5% is seen in camera 1, 6% in camera 2 and 3% in camera 3.

If we look into the static weighting presented in the same Table 5.7, this method is not stable in terms of results. For example in **portal 1 entering camera 1** using LPQ, the average pAUC(20%) decreased (0.53) compared to the baseline having an average value of pAUC(20%) equals to 0.56. If we also look into the Table 5.5 in **portal 2 entering** and **portal 2 leaving** the value pAUC(20%) of the static weighting approach has decreased in comparison to the baseline (no weights) technique.

The static weighting has proven to be not stable in terms of results. Indeed, in some cases the pAUC performance of this approach is lower than the pAUC of the baseline. For example using LBP feature extraction in **portal 1 leaving** and **portal 2 entering** the static weighting technique is less performing. If we also look at the average pAUC (20%) after 5 replications using the LPQ, the static weighting is less performing.

Going back to both Tables 5.3 and 5.5, the performance of the proposed technique in **portal 1 leaving** has decreased of 1% using LBP features and 4% using LPQ features compared to the baseline method (no weighting).

To interpret this decrease, series of investigation is performed:

- Check how does the proposed system behaves using datasets having less pose variations. For this, the videos in 4.1 found in in Chapter 4 are used;
- Examine the illumination variation present in the global dataset per portals. For this, the no-reference illumination quality assessment in Section 2.4 is exploited;
- Examine the quality in terms of sharpness of the global dataset per portals;
- Finally, examine the performance per camera domain in **portal 1 leaving** (i.e. camera 1, camera 2, camera 3).

First, we view the effects of head pose variations in the dataset.

Table 5.8 Average pAUC(20%) and AUPR in portals having the most frontal views of the ChokePoint dataset using LBP

Feature Extraction	**Techniques**	**Portal 1**			
		Entering		**Leaving**	
		pAUC(20%)	AUPR	pAUC(20%)	AUPR
LBP	No weights	0.54 ± 0.060	0.40 ± 0.058	0.67 ± 0.101	0.52 ± 0.055
	Static weights	0.56 ± 0.056	0.44 ± 0.062	0.68 ± 0.106	0.55 ± 0.052
	Dynamic weights	0.67 ± 0.042	0.53 ± 0.056	0.69 ± 0.104	0.57 ± 0.058
		Portal 2			
		Entering		**Leaving**	
		pAUC(20%)	AUPR	pAUC(20%)	AUPR
	No weights	0.38 ± 0.052	0.22 ± 0.039	0.52 ± 0.073	0.38 ± 0.055
	Static weights	0.37 ± 0.037	0.22 ± 0.035	0.52 ± 0.076	0.39 ± 0.054
	Dynamic weights	0.43 ± 0.063	0.28 ± 0.039	0.54 ± 0.080	0.40 ± 0.056

Table 5.9 Average pAUC(20%) and AUPR in portals having the most frontal views of the ChokePoint dataset using LPQ

Feature Extraction	**Techniques**	**Portal 1**			
		Entering		**Leaving**	
		pAUC(20%)	AUPR	pAUC(20%)	AUPR
LPQ	No weights	0.63 ± 0.067	0.51 ± 0.066	0.72 ± 0.104	0.59 ± 0.050
	Static weights	0.63 ± 0.054	0.48 ± 0.055	0.70 ± 0.107	0.56 ± 0.046
	Dynamic weights	0.70 ± 0.040	0.62 ± 0.051	0.71 ± 0.104	58 ± 0.050
		Portal 2			
		Entering		**Leaving**	
		pAUC(20%)	AUPR	pAUC(20%)	AUPR
	No weights	0.45 ± 0.060	0.30 ± 0.049	0.58 ± 0.083	0.42 ± 0.048
	Static weights	0.44 ± 0.058	0.29 ± 0.045	0.53 ± 0.076	0.37 ± 0.039
	Dynamic weights	0.51 ± 0.071	0.33 ± 0.041	0.60 ± 0.088	0.45 ± 0.048

Tables 5.8 and 5.9 show the average pAUC(20%) and AUPR used on the videos from the ChokePoint dataset containing a majority of frontal facial views after 5 random replications using LBP and LPQ FE techniques. As we can see, the proposed dynamic weighting technique outperforms overall. So, head pose variations have a negative impact on the system's performance.

In addition to that, a concentration is given to the illumination problem. A closer inspection is held to assess the quality of the dataset in terms of illumination. Illumination quality measure is applied on all of the captured samples for all individuals. An average value is obtained for each portal. The Figure 5.9 shows the average illumination (brightness) score over all the samples for each portal. We can see that the minimum illumination score is found in **portal 1 leaving**. Thus, the captured samples in **portal 1 leaving** have poor illumination quality which have caused a negative impact on the proposed system's performance.

Since the dynamic weighting technique is related to sharpness a better look into the average values of the sharpness score in all sessions and all cameras for each portal is needed.

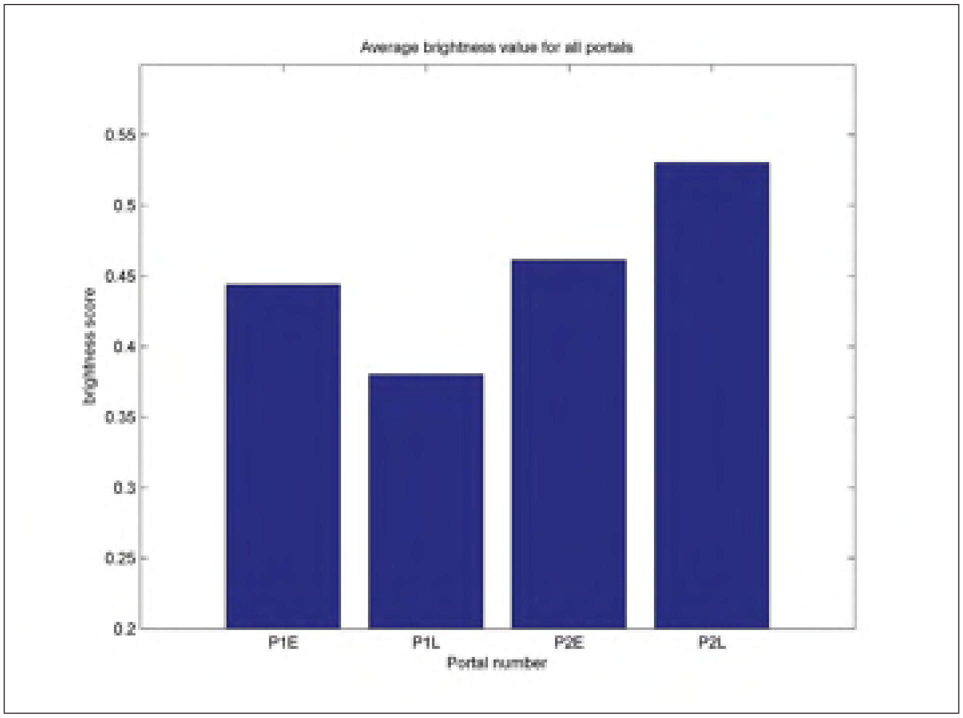


Figure 5.9 Average illumination score per portal

Figure 5.10 shows this variation of values within the samples in the ChokePoint dataset for each portal case. **Portal 1 leaving** has the least average value in terms of sharpness score compared to the other portals from the dataset.

At this stage, we can resume that severe decrease of quality information in terms of illumination/brightness and sharpness has a negative influence on the proposed system's performance.

Further investigation for camera domain is done in order to assess the performance of the FR system using the dynamic weighting technique. Table 5.10 shows the average pAUC(20%) and AUPR (after 5 replications) per camera of the **portal 1 leaving** case for both LBP and LPQ features. The system's performance using the dynamic weighting technique has shown a decrease of performance at camera 3 (C3) for both LBP and LPQ features.

The system's performance is low at camera 3 (C3) for both LBP and LPQ features. For that a comparison between these three cameras (camera 1, camera 2 and camera 3) of **portal 1 leav-**

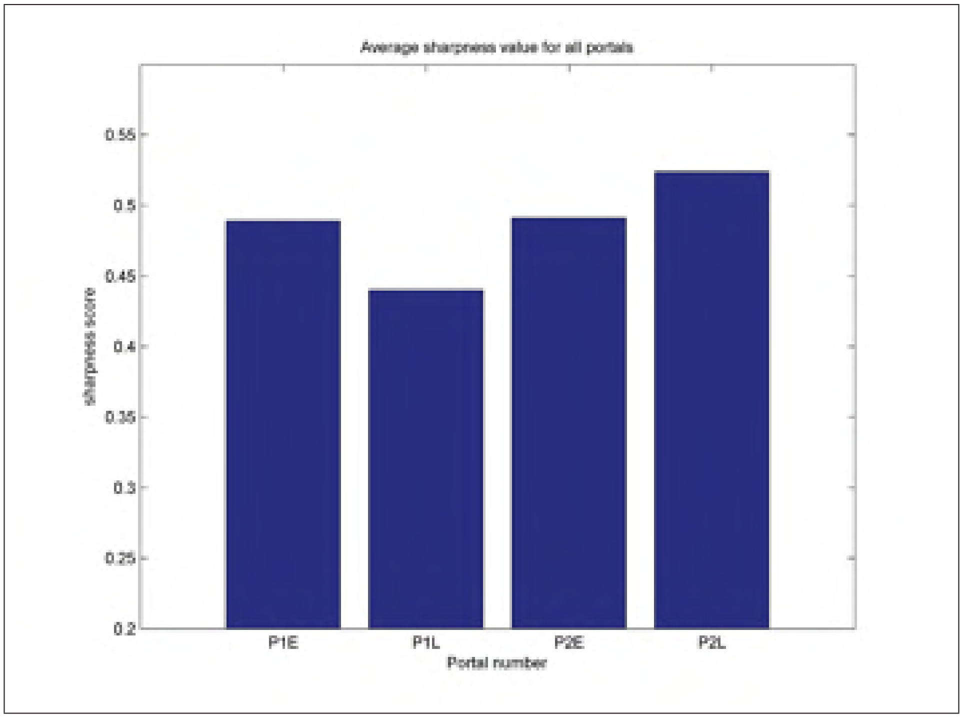


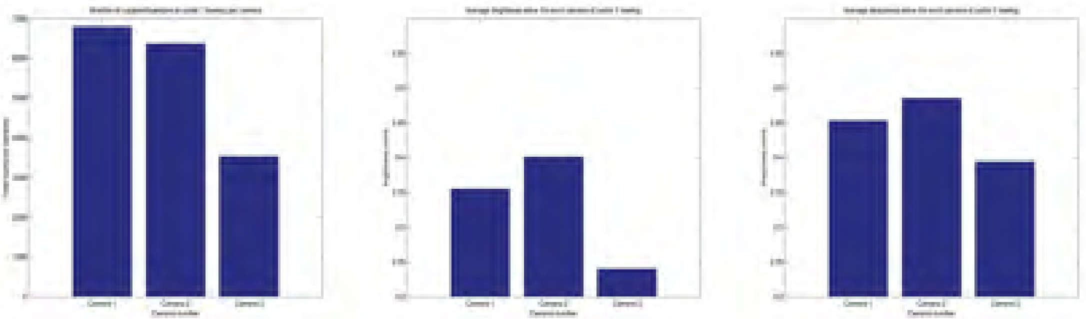
Figure 5.10 Average sharpness score per portal

Table 5.10 Average pAUC(20%) and AUPR for portal 1 leaving after 5 replications using LBP and LPQ

Feature Extraction	Techniques	**Camera 1 (C1)**		**Camera 2 (C2)**		**Camera 3 (C3)**	
		pAUC(20%)	AUPR	pAUC(20%)	AUPR	pAUC(20%)	AUPR
LBP	No weights	0.55 ± 0.087	0.40 ± 0.042	0.61 ± 0.099	0.45 ± 0.047	0.63 ± 0.087	0.49 ± 0.061
	Static weights	0.56 ± 0.091	0.42 ± 0.043	0.63 ± 0.100	0.50 ± 0.052	0.61 ± 0.090	0.48 ± 0.056
	Dynamic weights	**0.58 ± 0.092**	**0.45 ± 0.051**	**0.63 ±0.100**	**0.51 ± 0.058**	0.56 ± 0.081	0.41 ± 0.055
LPQ	No weights	0.61 ± 0.092	0.48 ± 0.042	0.64 ± 0.102	0.50 ± 0.043	0.67 ± 0.090	0.52 ± 0.054
	Static weights	0.61 ± 0.095	0.46 ± 0.041	0.65 ± 0.101	0.50 ± 0.042	0.66 ± 0.097	0.50 ± 0.049
	Dynamic weights	**0.62 ± 0.094**	**0.50 ± 0.046**	**0.66 ± 0.101**	**0.53 ± 0.049**	0.52 ± 0.081	0.35 ± 0.046

ing is held. This comparison is based on the number of captured samples, average illumination score and average sharpness score.

In Figure 5.11, camera 3 has the least number of samples, average illumination score and sharpness score which explains this degradation of performance of the system globally and specifically with the proposed dynamic weighting technique.



a) Total number of pcaptured facial samples per camera in portal 1 leaving

b) Average illumination score per camera in portal 1 leaving

c) Average sharpness score per camera in portal 1 leaving

Figure 5.11 Quality and quantity information per camera in portal 1 leaving in the Chokepoint dataset

Table 5.11 Average pAUC(20%) and AUPR of all sessions in Camera 3 (C3) Portal 1 leaving

Feature Extraction	**Techniques**	**Session 1 (S1)**		**Session 2 (S2)**		**Session 3 (S3)**		**Session 4 (S4**	
		pAUC(20%)	AUPR	pAUC(20%)	AUPR	pAUC(20%)	AUPR	pAUC(20%)	AUPR
LBP	No weights	0.13 ± 0.016	0.09 ± 0.014	0.16 ± 0.022	0.11 ± 0.015	0.17 ± 0.026	0.15 ± 0.016	0.14 ± 0.023	0.11 ± 0.015
	Static weights	0.13 ± 0.017	0.09 ± 0.013	0.15 ± 0.023	0.12 ± 0.016	0.18 ± 0.028	0.15 ± 0.013	0.14 ± 0.022	0.11 ± 0.015
	Dynamic weights	0.11 ± 0.014	0.07 ± 0.011	0.13 ± 0.021	0.09 ± 0.013	**0.19 ± 0.026**	**0.16 ± 0.015**	0.13 ± 0.019	0.10 ± 0.018
LPQ	No weights	0.13 ± 0.019	0.07 ± 0.009	0.15 ± 0.021	0.11 ± 0.014	0.17 ± 0.027	0.15 ± 0.017	0.15 ± 0.022	0.14 ± 0.013
	Static weights	0.14 ± 0.023	0.09 ± 0.009	0.15 ± 0.021	0.11 ± 0.014	0.17 ± 0.029	0.15 ± 0.013	0.15 ± 0.024	0.12 ± 0.012
	Dynamic weights	0.11 ± 0.016	0.06 ± 0.007	0.11 ± 0.016	0.06 ± 0.10	**0.18 ± 0.026**	**0.16 ± 0.014**	0.11 ± 0.020	0.08 ± 0.013

Thoroughly in Table 5.11 provides the average pAUC(20%) and AUPR for each session of camera 3 **portal 1 leaving**. Based on the previous observations previously about camera 3 in **portal 1 leaving** having poor quality data still in session 3 of this same portal (P1E_S3_C3) the proposed technique offers a slightly higher performance values compared to the baseline and the static weighting technique. This high score was compensated by the less variations in head pose in this recording (P1E_S3_C3). In fact (P1E_S3_C3) is considered as one of the recordings having a majority of frontal facial captures.

In summary, having a limited amount of captured samples from the testing phase and a severe low quality information in illumination and especially sharpness may cause the proposed system not to properly perform.

Experiment 2

This experiment consists in applying IN at the pre-processing level on all techniques (baseline: no weights,static weighting and dynamic local weighting). The IN technique used is the Multi-Scale Weberfaces and it is applied locally since based on the results in Chapter 4 local approach and Multi Scale Weberfaces outperforms over all the other IN techniques and over the global approach. In this experiment different feature extraction technique LBP and LPQ are implemented.

Table 5.12 Average pAUC(20%) performance of the illumination invariant still-to-video FR system (with standard error) in all portals of the ChokePoint dataset after 5 replications using LBP

Average pAUC(20%)					
Techniques	**Portal 1**		**Portal 2**		
	Entering	Leaving	Entering	Leaving	AVG
No weights	0.41 ± 0.053	0.61 ± 0.050	0.39 ± 0.054	0.46 ± 0.055	0.47 ± 0.053
Static weights	0.46 ± 0.050	0.64 ± 0.049	0.37 ± 0.048	0.48 ± 0.051	0.49± 0.049
Dynamic weights	**0.62 ± 0.044**	**0.72 ± 0.045**	**0.49 ± 0.054**	**0.61 ± 0.053**	**0.61 ± 0.049**

Table 5.13 Average AUPR performance of the illumination invariant still-to-video FR system (with standard error) in all portals of the ChokePoint dataset after 5 replications using LBP

Average AUPR					
Techniques	**Portal 1**		**Portal 2**		
	Entering	Leaving	Entering	Leaving	AVG
No weights	0.29 ± 0.045	0.47 ± 0.057	0.25 ± 0.042	0.33 ± 0.053	0.34 ± 0.049
Static weights	0.33 ± 0.050	0.51 ± 0.055	0.24 ± 0.044	0.32 ± 0.054	0.35± 0.051
Dynamic weights	**0.49 ± 0.054**	**0.58 ± 0.051**	**0.33 ± 0.048**	**0.46 ± 0.060**	**0.47 ± 0.053**

Table 5.14 Average pAUC(20%) performance of the illumination invariant still-to-video FR system (with standard error) in all portals of the ChokePoint dataset after 5 replications using LPQ

Average pAUC(20%)					
Techniques	**Portal 1**		**Portal 2**		
	Entering	Leaving	Entering	Leaving	AVG
No weights	0.58 ± 0.045	0.70 ± 0.032	0.44 ± 0.069	0.55 ± 0.064	0.57 ± 0.052
Static weights	0.60 ± 0.049	0.69 ± 0.035	0.42 ± 0.057	0.51 ± 0.052	0.55± 0.048
Dynamic weights	**0.63 ± 0.047**	**0.71 ± 0.034**	**0.51 ± 0.050**	**0.62 ± 0.046**	**0.62 ± 0.044**

Table 5.15 Average AUPR performance of the illumination invariant still-to-video FR system (with standard error) in all portals of the ChokePoint dataset after 5 replications using LPQ

Average AUPR					
Techniques	**Portal 1**		**Portal 2**		
	Entering	Leaving	Entering	Leaving	AVG
No weights	0.44 ± 0.046	0.56 ± 0.040	0.30 ± 0.056	0.41 ± 0.061	0.43 ± 0.051
Static weights	0.45 ± 0.051	0.55 ± 0.042	0.28 ± 0.051	0.46 ± 0.053	0.46± 0.049
Dynamic weights	**0.50 ± 0.050**	**0.57 ± 0.041**	**0.32 ± 0.050**	**0.46 ± 0.053**	**0.46 ± 0.049**

Tables 5.12 and 5.14 show the average pAUC(20%) performance after 5 system replications. The illumination invariant system with application of the Multi Scale Weberfaces locally on each patch has boosted the performances of all three techniques. This was confirmed in Chapter 4. The proposed technique dynamic local weighting surpasses the baseline and the static weighting.

With feature extraction technique LBP there is a significant improvement by 14% compared to the baseline (fom 0.41 to 0.61) and by 12% as to the static weighting (from 0.49 to 0.61). Nevertheless using LPQ with the dynamic local weights combined with the illumination invariant pre-processing task the system's performance for recognition has barely increased. Indeed,there is a slight improvement of 5% compared to the static weighting. In this context a similar conclusion can be drawn out about which feature extraction technique to use in order to have better enhancements: the LBP feature extraction offers higher boost in comparison with the baseline and static weighting technique. This also can be confirmed by looking on to the

AUPR tables (Refer to Tables 5.13 and 5.15). With LBP feature extraction the average AUPR of the proposed technique has jumped from 0.34 (baseline) to 0.47 (dynamic). As for the LPQ, the average AUPR has jumped from 0.43 (baseline) to 0.46 (dynamic). At this stage some statements can be concluded:

- The proposed dynamic weighting that exploits quality information of the local facial regions on the technique improved the still-to-video FR system;
- LBP feature extraction shows applied on the dynamic weighting technique shows better improvement and increase in the recognition performance.

Table 5.16 Average processing time of 1 ROI in seconds (segmentation, pre-processing, feature extration (LBP and LPQ), matching (weighted Chi-square)) on a i7 2.30 GHz processor

	Without pre-processing		With pre-processing (Multi Scale Webefaces)	
Techniques	**LBP**	**LPQ**	**LBP**	**LPQ**
No weights	0.161 s	0.021 s	0.252 s	0.109 s
Static weights	0.162 s	0.022 s	0.274 s	0.112 s
Dynamic weights	0.212 s	0.074 s	0.300 s	0.188 s

Table 5.16 shows the average time to process a single ROI of size 48 × 48 on a i7 2.30 GHz processor. To calculate this average time, a total number of ROI (100 ROIs) is processed into the system and an overall processing time is estimated. The latter, is divided by the total number of ROIs.

As we can see, the proposed dynamic weighting technique has a slight longer processing time of almost 0.05s per ROI using both LBP and LPQ. This increase on time is predictable due to the dynamicity of the proposed technique when calculating the weights with image quality metric (sharpness metric). However, this additional time lapse of 0.05s is still considered acceptable compared to the amount of process done on each patch of a ROI. When a pre-processing module is added to the system, the processing time increased by almost 0.1 s per ROI.

If we compare both FE techniques (LBP and LPQ) the process is executed faster with LPQ. Therefore, if we want to implement the proposed system onto a real time, it is preferable to use LPQ. Indeed, with LPQ not only does the recognition performance is high but also the processing time is shorter compared to LBP.

CONCLUSION

In this Thesis, a still-to-video FR was implemented mainly for watchlist screening applications. Two main studies were conducted. Each study concentrates on a specific level of the system.

In **Chapter 4**, the concentration was based at the pre-processing level where the goal is to implement an illumination invariant still-to-video system. A comparison of different illumination normalization techniques is performed to determine the suitable IN technique that provides better illumination invariance to our watchlist application. A series of experiments were held at this stage of the study: experiments on the approach of applying IN techniques and experiments on varying the total number of patches (4, 9, and 16 patches).

The global approach consists in applying the IN technique on the whole image before isolating the patches for patch-based local matching. This experiment have shown that IN technique have an impact on the performance in terms of recognition of the system. Moire specifically, Tan and Triggs and Multi-scale weberfaces techniques have recorded the highest pAUC(20%) rate. These values have increased by 36 % from using one block image to 16 blocks (from 0.20 using the 1 block and 0.56 using 16 blocks). Within the same number of blocks (6 blocks), there was a boost of performance by almost 10% for both Tan and Triggs and Multi-scale weberfaces techniques. Compared to the baseline (No normalization), the AUPR values with Multi Scale Weberfaces has increased from 0.09 (using 1 block) to 0.38 (using 16 blocks).

The local approach, consists in firstly isolating the patches. Then, on each individual patch IN techniques are applied. Two same techniques have outperformed which are the Tan and Triggs and the Multi-scale weberfaces. There was a significant increase of performance when the number of patches increases. Indeed, the pAUC performance has elevated using 1 block (whole image) from 0.21 with Tan and Triggs and 0.59 with Multi-scale weberfaces to respectively 0.58 and 0.59 using 16 blocks for local matching. As for the AUPR performance a significant increase is witnessed: from 0.09 (using 1 block) to 0.38 with Tan and Triggs and from 0.09 (using 1 block) to 0.36 with Multi Scale Weberfaces.

These results confirms the hypotheses stated for this study which are:

- IN techniques in our case Tan and Triggs and Multi-scale weberfaces compensate the negative impact of poor illumination conditions and help improve the system's performance;
- For both approaches, global and local IN application, results have shown that the performance is related to the number of blocks used fro the patch-based matching. The higher the number of total patches, higher the performance is achieved. Of course, by having more patches this provides more distinctive feature extraction which benefits the matching process;
- Finally, local application of IN has proven to be better than the global approach because IN applied locally can provide better illumination compensation on the facial regions.

In **Chapter 5** of this Thesis, the concentration is based at the classification level of the still-to-video FR system and is a continuation of the previous part. The goal was to propose a new dynamic weighting for local matching that is based on domain adaptation and image quality. Three methods of weighting were compared to each other in order to assess the impact of the proposed technique onto the still-to-video FR system. These techniques are the baseline technique (no weighting), the static weighting which emphasizes certain regions of the face based on previous knowledge (inspired by the configuration proposed in (Ahonen *et al.*, 2006)) and finally the proposed dynamic weighting using sharpness quality metric.

The results have shown an improvement of performance by 4% compared to the baseline method and 3% compared to the static method using LBP feature extraction overall (average performance through all portals from the dataset). As for the LPQ, there is an improvement of 1% compared to the baseline and 2% compared to the static weighting. This slight improvement compared to the baseline technique can be linked to the fact that LPQ is already robust to blur giving the system a better feature representation. The performance of the baseline technique using LPQ has higher pAUC value than the the baseline technique using LBP. Same thing can be said to the proposed dynamic technique. As a matter of fact, the performance have increased from 0.52 using LBP to 0.55 using LPQ.

In addition, for a second experiment, the previous pre-processing (illumination invariance) method was injected to this dynamic weighting for local matching system. The results have shown better results. Surely, using the LBP descriptor, the proposed system recorded an increase of 14% compared to the baseline method and an increase of 12% compared to the static weighting method. Per contra, with LPQ descriptors there was an increase of 5% compared to the baseline and 7% compared to the static weighting.

All of these results confirms the hypotheses made for this second study. These affirmations are:

- IN applied at the pre-processing level does alleviate the illumination issue and contributes positively in the improvement of the system's performance;
- Contextual information using quality is correlated to the system's performance and its exploitation is very useful for the improvement of the still-to-video FR system. In fact, sharpness quality used to emphasize certain regions of the face by assigning weights and using domain adaptation solution to adapt the source domain of the system has proven to enhance the system's performance;
- Assigning weights contextually based on the provided input gives stable results. In fact, the proposed system shows stability (outperforms overall) whereas the static weighting method fluctuates in terms of performance values. (i.e results are sometimes lower that the baseline method). The reason of this fluctuation in the static weighting is that the weights are knowledge based and are related to the dataset used to configure it (works better with frontal images). However, the proposed technique shows more versatility in assigning these weights.

Recommendations and future work

A list of recommendations can be proposed:

- An investigation should be done on the size limit of a patch configuration and the appropriate size of an ROI where the local descriptor can be applied and where the system performance is still improving by assessing different ROI sizes and patch configuration during local matching;

- An exploration of other image quality metrics with the proposed dynamic quality-based regional weighting system is advised in order to see the impact of each quality metric on the system's performance;

- Other matching methods or learning-based classifiers can be implemented to assess the system's performance. In this same idea of classification, the implementation of ensemble of classifiers, in which each ensemble uses different classifiers having themselves different dynamic quality weighting method (illumination based, contrast based, sharpness based, etc.) and all of the scores provided are combined using a score fusion module, is advisable;

- Since *super-resolution image reconstruction* [2] methods have become prominent in the field of face recognition, an interesting avenue for improving the proposed dynamic quality-based regional weighting system is to use these super-resolution techniques. Indeed, applying super-resolution on low quality and low resolution probe images and implementing the proposed dynamic weighting on the obtained images may improve the system's performance because the domain gap in terms of resolution is lessened: from matching between high resolution images and low resolution images (without super-resolution) to matching between high resolution images and the obtained high resolution images (using super-resolution);

- Based on the results in Chapter 5, the proposed method (dynamic quality based weighting) outperforms globally (over all the videos from the dataset). In some cases where the cap-

[2] This technique is the process of combining low resolution and successive images taken from video frames that often have complementary information using computational techniques.

tured faces are mostly frontal and very sharp the static weighting technique can be beneficial. So, an update of the system can be proposed by adding a module at the pre-processing level of the still-to-video FR that estimates the head position of the facial capture (frontal or non frontal). Then, based on this primary result a dynamic selection of weighting techniques (between the static and dynamic quality based technique) can be implemented. In other words, if the face appears to be more frontal then the static weighting method is used else the proposed dynamic quality based system is used;

- Another recommendation would be to assess the performance of the system using other datasets (other than ChokePoint) that provides a good projection of a real watchlist application for example the COX-S2V dataset[3] and the QUIS-CAMPI dataset[4];
- Finally, a high suggestion would be to implement this proposed system onto real-time FR system using state-of-the-art programming technologies and maybe with graphics processing units (GPUs).

[3] http://vipl.ict.ac.cn/resources/datasets/cox-face-dataset/COX-S2V

[4] http://quiscampi.di.ubi.pt/

APPENDIX I

CORRELATION WITH SIGNIFICANT LEVEL

After determining the correlation coefficient $coeff$ between both variables sharpness score and matching score for each patch of our study, the next thing is to see the likelihood of our correlation coefficient value to occur by chance (NULL variation). We must determine if the correlation does really exist in our data.

To do this, we must set a value of probability of the data to occur in random variations. This probability is often called the alpha level or the proportion in ONE Tail which is the level of being wrong when we state that there is a relationship (correlation) between two measured variables. The commonly used alpha level is $\alpha = 0.05$.

Having α already set, we need to assess whether the $coeff$ value that we found after performing correlation on our sample is significant or not, we need to use the *critical value table for Pearson's correlation coefficient* given in Figure I-1.

In order to use the table below, we need two pieces of information which are:

a. The value of the correlation coefficient for our study: $coeff$

b. The number of samples we have: $N_{samples}$ (In our case, this represents the number of ROI samples we have in our study.)

Reading and determining the critical value

To find the critical value, series of steps should be performed.

a. Determine the degrees of freedom (df) for a correlation study. The degree of freedom is equals to 2 less than the number: $df = N_{samples} - 2$

b. Use the critical value table and find the intersection of alpha level $\alpha = 0.05$ (columns) and the df degrees of freedom (rows). The value found at the intersection is the minimum correlation coefficient: r_{critic}

c. If $|coeff|$ is above r_{critic}, we reject the NULL hypothesis or we can say that there is no significant relationship between the matching score and the sharpness quality score. If $|coeff|$ is less than r_{critic} we can't reject the NULL hypothesis or we can say that there is no significant relationship between the variables.

   In general, when using the absolute value of $coeff$ we are ignoring the sign of the correlation. However, in this Thesis, we are looking for strong, significant and positive relationship (correlation) in our study. For that we considered the patches having $coeff >= r_{critic}$

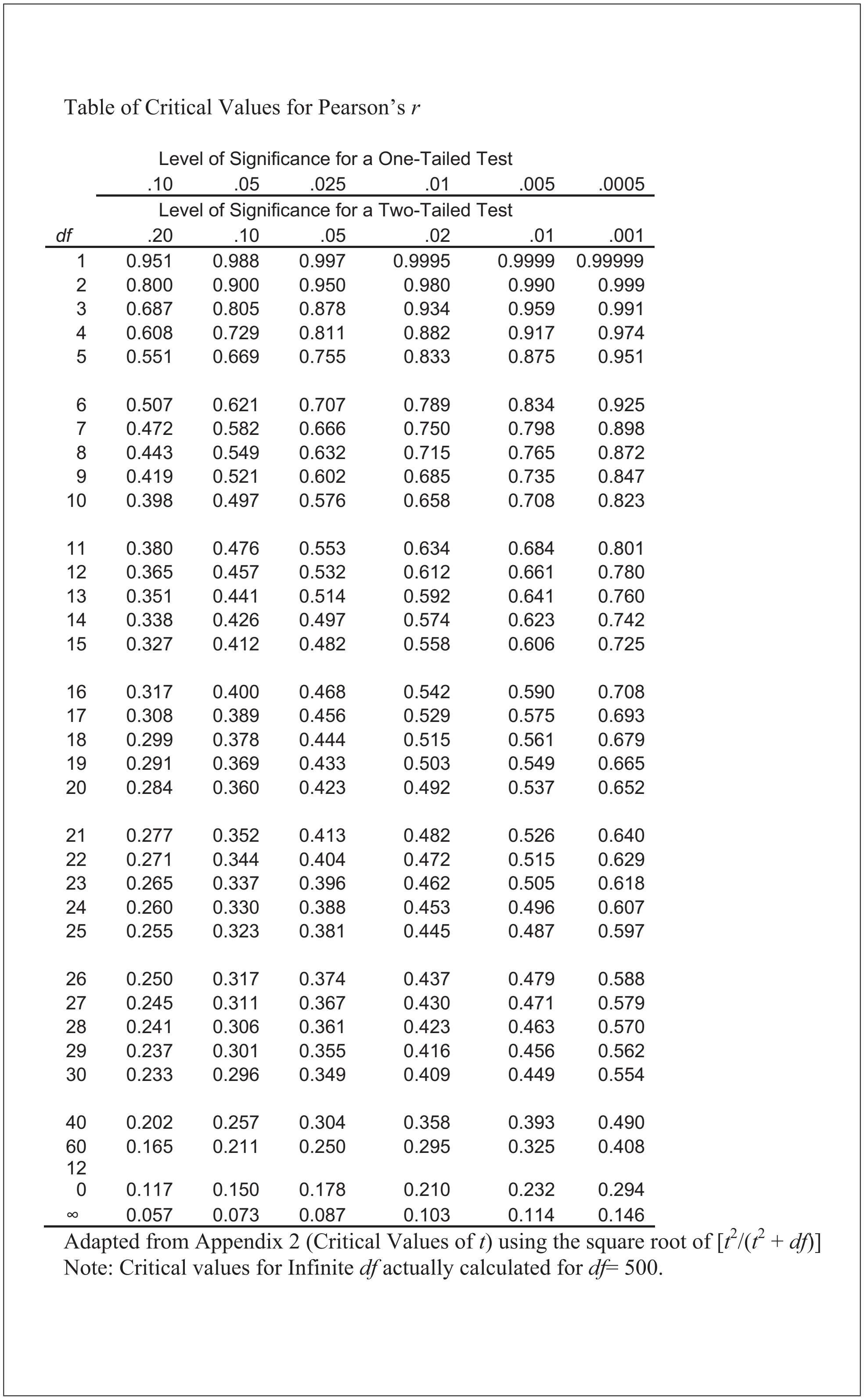


Table of Critical Values for Pearson's *r*

	Level of Significance for a One-Tailed Test					
	.10	.05	.025	.01	.005	.0005
	Level of Significance for a Two-Tailed Test					
df	.20	.10	.05	.02	.01	.001
1	0.951	0.988	0.997	0.9995	0.9999	0.99999
2	0.800	0.900	0.950	0.980	0.990	0.999
3	0.687	0.805	0.878	0.934	0.959	0.991
4	0.608	0.729	0.811	0.882	0.917	0.974
5	0.551	0.669	0.755	0.833	0.875	0.951
6	0.507	0.621	0.707	0.789	0.834	0.925
7	0.472	0.582	0.666	0.750	0.798	0.898
8	0.443	0.549	0.632	0.715	0.765	0.872
9	0.419	0.521	0.602	0.685	0.735	0.847
10	0.398	0.497	0.576	0.658	0.708	0.823
11	0.380	0.476	0.553	0.634	0.684	0.801
12	0.365	0.457	0.532	0.612	0.661	0.780
13	0.351	0.441	0.514	0.592	0.641	0.760
14	0.338	0.426	0.497	0.574	0.623	0.742
15	0.327	0.412	0.482	0.558	0.606	0.725
16	0.317	0.400	0.468	0.542	0.590	0.708
17	0.308	0.389	0.456	0.529	0.575	0.693
18	0.299	0.378	0.444	0.515	0.561	0.679
19	0.291	0.369	0.433	0.503	0.549	0.665
20	0.284	0.360	0.423	0.492	0.537	0.652
21	0.277	0.352	0.413	0.482	0.526	0.640
22	0.271	0.344	0.404	0.472	0.515	0.629
23	0.265	0.337	0.396	0.462	0.505	0.618
24	0.260	0.330	0.388	0.453	0.496	0.607
25	0.255	0.323	0.381	0.445	0.487	0.597
26	0.250	0.317	0.374	0.437	0.479	0.588
27	0.245	0.311	0.367	0.430	0.471	0.579
28	0.241	0.306	0.361	0.423	0.463	0.570
29	0.237	0.301	0.355	0.416	0.456	0.562
30	0.233	0.296	0.349	0.409	0.449	0.554
40	0.202	0.257	0.304	0.358	0.393	0.490
60	0.165	0.211	0.250	0.295	0.325	0.408
12 0	0.117	0.150	0.178	0.210	0.232	0.294
∞	0.057	0.073	0.087	0.103	0.114	0.146

Adapted from Appendix 2 (Critical Values of *t*) using the square root of [$t^2/(t^2 + df)$]
Note: Critical values for Infinite *df* actually calculated for *df*= 500.

Figure-A I-1 Table of critical values for Pearson's correlation, Taken from Radford

APPENDIX II

ILLUSTRATION AND ALGORITHM OF THE SPECIALIZED WINDOW

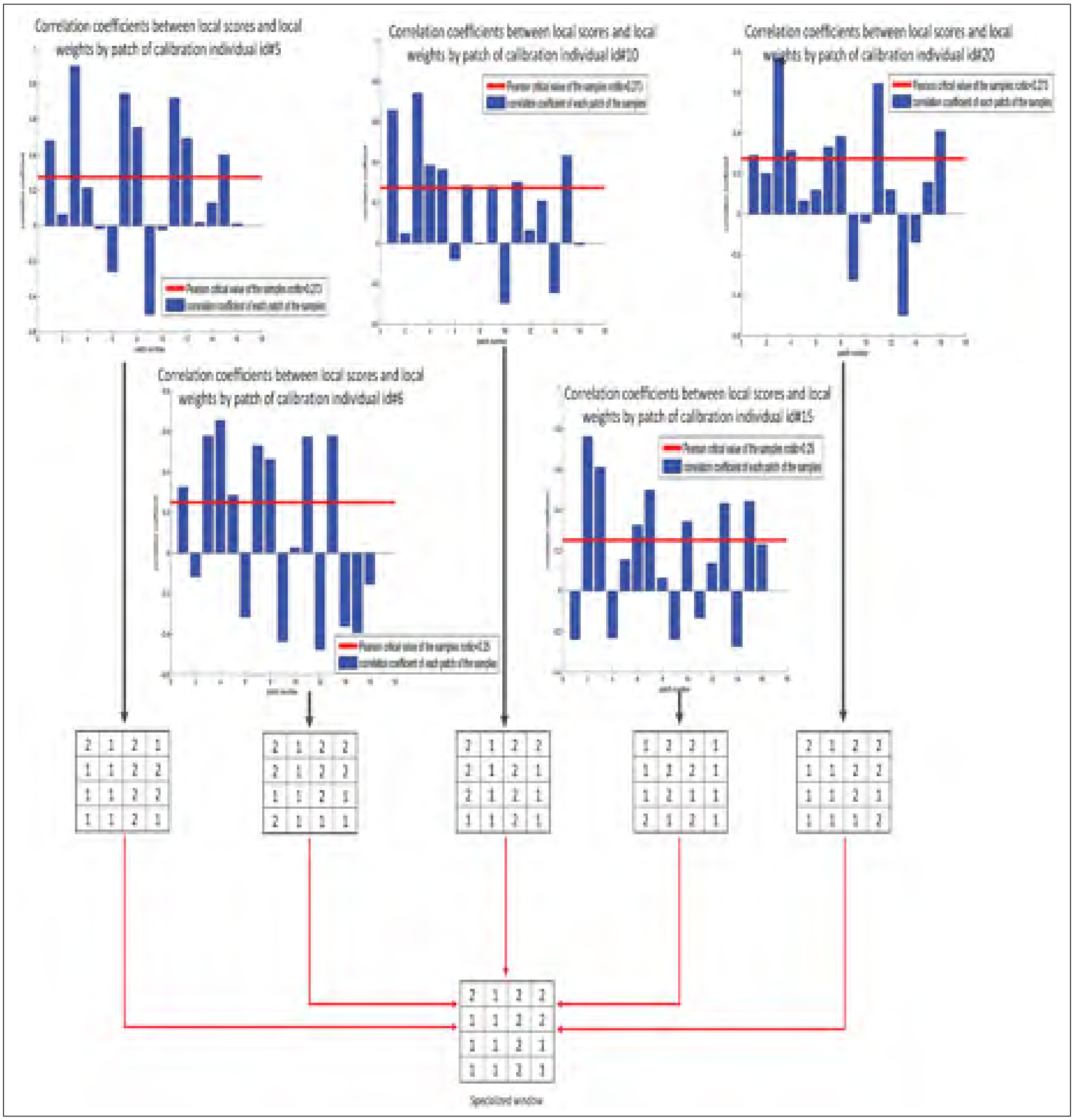

Correlation coefficients between local scores and local weights by patch of calibration individual id#5
Correlation coefficients between local scores and local weights by patch of calibration individual id#10
Correlation coefficients between local scores and local weights by patch of calibration individual id#20
Correlation coefficients between local scores and local weights by patch of calibration individual id#6
Correlation coefficients between local scores and local weights by patch of calibration individual id#15
2 1 2 1
1 1 2 2
1 1 2 2
1 1 2 1
2 1 2 2
2 1 2 2
1 1 2 1
2 1 1 1
2 1 2 2
2 1 2 1
2 1 2 1
1 1 2 1
1 2 2 1
1 2 2 1
1 2 1 1
2 1 2 1
2 1 2 2
1 1 2 2
1 1 2 1
1 1 1 2
2 1 2 2
1 1 2 2
1 1 2 1
1 1 2 1
Specialized window

Figure-A II-1 Calculation of the specialized window for recording P2E_S1_C3

APPENDIX III

ON THE EFFECTS OF ILLUMINATION NORMALIZATION WITH LBP-BASED WATCHLIST SCREENING

Ibtihel Amara[1], Eric Granger[1], and Abdenour Hadid[2]

[1] Department of Electrical Engineering, École de Technologie Supérieure,
1100 Notre-Dame West, Montreal, Quebec, (H3C 1K3) Canada

[2] Department of Computer Science and Engineering, Center of Machine Vision,
P.O. Box 4500, FIN-90014 University of Oulu Finland

This paper has been accepted for publication in
"ECCV 2014, Second International Workshop on Computer Vision with Local Binary Patterns Variants" in July 2014.

Abstract

Still-to-video face recognition (FR) is an important function in several video surveillance applications like watchlist screening, where faces captured over a network of video cameras are matched against reference stills belonging to target individuals. Screening of faces against a watchlist is a challenging problem due to variations in capturing conditions (e.g., pose and illumination), to camera inter-operability, and to the limited number of reference stills. To constrain time complexity, a holistic FR approach based on Local Binary Pattern (LBP) descriptors is often considered to represent facial captures and reference stills. Despite their efficiency, LBP descriptors are known as being sensitive to illumination changes. In this paper, the performance of still-to-video FR is compared when different passive illumination normalization techniques are applied prior to LBP feature extraction. This study focuses on representative retinex, self-quotient, diffusion, filtering, means de-noising, retina, wavelet and frequency-based techniques that are suitable for fast and accurate face screening. Simulation results obtained with videos from the Chokepoint dataset indicates that, although Multi-Scale Weberfaces and Tan and Triggs techniques tend to outperform others, the benefits of these techniques varies considerably according to the individual and illumination conditions. Results suggest that a combination of these techniques should be selected dynamically based on changing capture conditions.

Keywords: illumination normalization, local binary patterns, face screening, still-to-video face recognition, video surveillance

Introduction

In watchlist screening applications, systems for still-to-video FR are increasingly employed to automatically detect the presence of target individuals of interest for enhanced public security. Accurate and timely responses are required to recognize faces captured under semi-controlled or uncontrolled conditions, as found at various security checkpoint entries, inspection lanes, portals, etc. Under these conditions, face captures incorporate variations due to ambient illumination, pose, expressions, occlusion, scale, resolution and blur (Barr *et al.*, 2012) (De-la Torre *et al.*, 2014), and the performance of FR systems tend to deteriorate. Despite these challenges, it is generally possible to exploit spatiotemporal information extracted from video streams to improve system robustness and accuracy (Matta and Dugelay, 2009) (Dewan *et al.*, December 16-17, 2013).

Recent developments in image analysis and recognition have shown that the Local Binary Patterns (LBP) (Ojala *et al.*, 2002) provide a simple yet powerful approach to represent faces for human computer interaction, biometric recognition, surveillance and security, etc. (Ahonen *et al.*, 2006; Pietikäinen *et al.*, 2011). LBP is a gray-scale invariant texture operator which labels each pixel of an image by thresholding its neighborhood pixels with the intensity value of the center pixel. The resulting LBP labels can be regarded as local primitives such as curved edges, spots, flat areas, etc. The histogram of these labels over facial image can be then used as a face descriptor. Given its discriminative power, tolerance to monotonic grey-scale changes, and computational efficiently, LBP has become a well-established technique in FR[1], and has inspired many recent extensions and new research on related methods.

[1] See LBP bibliography at http://www.cse.oulu.fi/MVG/LBP_Bibliography

Variations in facial appearance caused by changes in ambient illumination conditions play an important role in the performance of any FR system applied to video surveillance. It has been shown that face images of different individuals appear more similar than images of the same individual under severe illumination variations (Struc and Pavesic, 2011). However, it is well known that LBP and other variants are sensitive to severe illumination changes.

Several techniques have been proposed in literature for illumination invariant FR (Sharma *et al.*, 2014). Zou et al. (Zou *et al.*, 2007b) presented a survey of techniques to manage variations in face appearance due to illumination changes according to passive and active approaches. Passive approaches focus on the visible spectrum images, where face appearance has been altered by illumination variations, while active ones employ active imaging techniques to capture face images under consistent illumination conditions, or images of illumination invariant modalities.

Among passive techniques, some are specialized at either the pre-processing, the feature extraction, or the classification level (Struc and Pavesic, 2011). At the pre-processing level, normalization techniques seek to transform facial images such that facial variations induced by illumination are removed. These approaches can be adapted for use with any FR algorithm. Techniques at the feature extraction level seek to achieve illumination invariance by using features or representations that are stable under different illumination conditions. However, some empirical studies have shown that no descriptor can ensure illumination invariant FR in the presence of severe illumination changes. Finally, classification level techniques compensate for the illumination based on the type of face model or classifier employed for FR. Assumptions regarding the effects of illumination on the face model or classifier are employed in counter measures to obtain illumination invariance.

In this study, the performance of several illumination normalization techniques is compared for representation of face captures in still-to-video FR systems using LBP descriptors, as seen in many watchlist screening applications. This empirical study focuses on passive techniques applied at the pre-processing level, and compares representative retinex, self-quotient, diffusion, filtering, means de-noising, retina, wavelet and frequency-based techniques in term of accuracy and computational complexity. The benefits of these approaches are assessed using faces captured in the Chokepoint video data set, with individuals walking through an array of cameras located above different portals.

Face Screening in Video Surveillance

Watchlist screening is an important application for decision support in video surveillance systems. It involves still-to-video FR according to the following steps (Chellappa *et al.*, 2010). During enrollment to a watchlist, the segmentation process isolates the regions of interest (ROIs) from reference still images (mugshots) that were previously captured under controlled conditions. Features are extracted and assembled into a discriminant and compact ROI patterns to design facial models[2]. These features are often image-based (e.g., LBP descriptors) or pattern recognition-based (e.g., PCA).

During operations, a video stream is captured using some video surveillance camera, and segmentation isolates the ROIs corresponding to faces captured in successive frames. A tracker is often initialized when an emergent ROI is detected far from other faces, and a track is defined to follow the movement or expression of distinct faces across consecutive frames using appearance, position and motion information. Features are extracted into ROI pattern for matching against the facial models of individuals enrolled to the watchlist. A positive prediction is produced if a matching score surpasses an individual-specific threshold. Finally, the decision function combines the tracks and classification predictions in order to recognize the most likely individuals in the scene.

Systems for still-to-video FR are typically modeled in terms of independent detection problems, each one implemented using a template matcher or classifier. These individual-specific detectors are designed with references face samples from target and non-target individuals (from a cohort or the background model). The advantages of modular architectures with

[2] A *facial model* of an individual is defined as a set of one or more reference ROI patterns (used for a template matching system), or parameters estimated from reference ROI patterns (for a classification system).

individual-specific detectors include the ease with which face models may be added, updated and removed from the systems, and the possibility of specializing pre-processing, feature extraction, matching and decision thresholds to each specific individual (Ekenel *et al.*, 2010; Pagano *et al.*, 2012).

The performance of state-of-the-art FR systems applied to video surveillance is limited by the difficulty in capturing and recognizing facial regions from video streams under semi-controlled and uncontrolled capture conditions (e.g., at inspection lanes, portals and checkpoint entries, in cluttered free-flow scenes at airports or casinos). In particular, performance is severely affected by the variations in ambient illumination, pose, expression, occlusion, scale, resolution, blur and ageing. Still-to-video FR is particularly challenging because very few reference samples are typically available for system design, and because of camera inter-operability – ROIs captured with still cameras (during enrollment) have different properties than those captured with video cameras (during operations). In pattern recognition literature, the situation where only one reference sample is available for system design are often referred to as a “single sample per person" (SSPP) or “one sample training" problem. Techniques specialized for SSPP in FR include multiple face representations, synthetic face generation, and enlarging the training set using auxiliary set (Kan *et al.*, 2013). It is worth noting that the still-to-video FR systems from literature assume that the single face reference is consistent and representative of the individuals in operational conditions.

Few specialized techniques have been proposed for still-to-video FR (Shaokang *et al.*, 2011). A framework based on local facial features has been proposed to match stills against video frames with different features (e.g., manifold to manifold distance, affine hull method, and multi-region histogram)(Shaokang *et al.*, 2011). These features are extracted from a set of stills utilizing spatial and temporal video information. More recently, partial and local linear discriminant analyses have been proposed using a high quality still and a set of low resolution video sequences of each individual (Huang *et al.*, 2013). Finally, a specialized feed-forward neural network is trained for each individual of interest in a watch-list to identify the decision regions of individual faces in the feature space, where morphology is employed to synthetically generate variations of a reference still (Kamgar-Parsi and Lawson, 2011).

LBP-based Face Recognition

The LBP texture analysis operator, introduced by Ojala et al. (Ojala *et al.*, 2002), is defined as a gray-scale invariant texture measure, derived from a general definition of texture in a local neighborhood. It is a powerful means of texture description and among its properties in real-world applications are its discriminative power, computational simplicity and tolerance against monotonic gray-scale changes.

The original LBP operator forms labels for the image pixels by thresholding the 3×3 neighborhood of each pixel with the center value and considering the result as a binary number. Fig.III-1 shows an example of an LBP calculation. The histogram of these $2^8 = 256$ different labels can then be used as a texture descriptor.

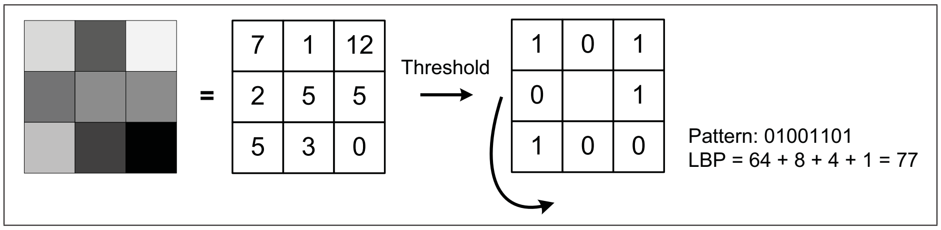


Figure-A III-1 The basic LBP operator

The operator has been extended to use neighborhoods of different sizes. Using a circular neighborhood and bilinearly interpolating values at non-integer pixel coordinates allow any radius and number of pixels in the neighborhood. The notation (P,R) is generally used for pixel neighborhoods to refer to P sampling points on a circle of radius R. The calculation of the LBP codes can be easily done in a single scan through the image. The value of the LBP code of a pixel (x_c, y_c) is given by:

$$\mathrm{LBP}_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c)2^p, \tag{A III-1}$$

where g_c corresponds to the gray value of the center pixel (x_c, y_c), g_p refers to gray values of P equally spaced pixels on a circle of radius R, and s defines a thresholding function as follows:

$$s(x) = \begin{cases} 1, & \text{if } x \geq 0; \\ 0, & \text{otherwise.} \end{cases} \tag{A III-2}$$

Another extension to the original operator is the definition of the so called *uniform patterns*. This extension was inspired by the fact that some binary patterns occur more commonly in texture images than others. A local binary pattern is called uniform if the binary pattern contains at most two bitwise transitions from 0 to 1 or vice versa when the bit pattern is traversed circularly. In the computation of the LBP labels, uniform patterns are used so that there is a separate label for each uniform pattern and all the non-uniform patterns are labeled with a single label. This yields to the following notation for the LBP operator: $\mathrm{LBP}_{P,R}^{u2}$. The subscript represents using the operator in a (P,R) neighborhood. Superscript $u2$ stands for using only uniform patterns and labeling all remaining patterns with a single label.

Each LBP label (or code) can be regarded as a micro-texton. Local primitives which are codified by these labels include different types of curved edges, spots, flat areas etc. The occurrences of the LBP codes in the image are collected into a histogram. The classification is then performed by computing histogram similarities. For an efficient representation, facial images are first divided into several local regions from which LBP histograms are extracted and concatenated into an enhanced feature histogram.

It is known that LBP is sensitive to severe illumination changes. As a consequence, several attempts have been made to overcome this sensitivity. For instance, Tan and Triggs (Tan and Triggs, 2007) developed a very effective preprocessing chain for compensating illumination variations in face images. It is composed of gamma correction, difference of Gaussian (DoG) filtering, masking (optional) and equalization of variation. This approach has been very successful in LBP-based face recognition under varying illumination conditions. When using it for the original LBP, the last step (i.e. equalization of variations) can be omitted due to LBP's invariance to monotonic gray scale changes.

Aiming at reducing the sensitivity of the image descriptor to illumination changes, a Bayesian LBP (BLBP) was developed by He et al.(He *et al.*, 2008). This operator is formulated in a Filtering, Labeling and Statistic (FLS) framework for texture descriptors. In the framework, the local labeling procedure, which is a part of many popular descriptors such as LBP and SIFT, can be modeled as a probability and optimization process. This enables the use of more reliable prior and likelihood information, and reduces the sensitivity to noise. The BLBP operator pursues a label image, when given the filtered vector image, by maximizing the joint probability of two images.

Liao et al. (Liao *et al.*, 2010) noticed that adding a small offset value (T) for comparison in LBP-like methods is not invariant under scaling of intensity values. The intensity scale invariant property of a local comparison operator is very important for example in background modeling, because illumination variations, either global or local, often cause sudden changes of gray scale intensities of neighboring pixels simultaneously, which would approximately be a scale transform with a constant factor. Therefore, a Scale Invariant Local Ternary Pattern (SILTP) operator was developed for dealing with the gray scale intensity changes in complex background. Assuming linear camera response, The SILTP feature is invariant if the illumination is suddenly changed from darker to brighter or vice versa. Besides, SILTP is robust when a soft shadow covers a background region, because the soft cast shadow reserves the background texture information but tends to be darker than the local background region with a scale factor. A downside of the methods mentioned above using one or two thresholds is that the methods are not strictly invariant to local monotonic gray level changes as the original LBP. The feature vector lengths of these operators are also longer.

In order to deal with strong illumination variations, Li et al. developed a very successful system combining near-infrared (NIR) imaging with local binary pattern features and AdaBoost learning (Li *et al.*, 2007). The invariance of LBP with respect to monotonic gray level changes makes the NIR images illumination invariant. The method achieved a verification rate of 90% at FAR=0.001 and 95% at FAR=0.01 on a database with 870 subjects.

Illumination Normalization

Table-A III-1 Illumination Normalization Techniques

Family	**Specific Technique**
Retinex	Adaptive Single-Scale Retinex (ASSR)
Retinex	Large and Small-Scale Features (LSSF)
Self Quotient	Multi-Scale Self Quotient (MSSQ)
Diffusion	Isotropic Diffusion (ID)
Diffusion	Modified Anisotropic Diffusion (MAD)
Filter	Tan and Triggs (TT)
Gradient	Multi-Scale Weberfaces (MSW)
Mean Denoising	Adaptive Non Local Means (ANLM)
Retina	Retina Modeling (RM)
Wavelet	Wavelet Denoising (WD)
Frequency	Homomorphic (H)

Changes in ambient illumination, and the resulting variations to facial appearance, are known to significantly deteriorate the performance of FR systems. Accordingly, several techniques have been proposed for illumination invariant FR (Sharma *et al.*, 2014). Zou et al. (Zou *et al.*, 2007b) presented a survey of techniques according to passive and active approaches. *Passive approaches* focus on the visible spectrum images where face appearance has been altered by illumination variations. They include illumination variation modelling, illumination invariant features, photometric normalisation, and 3D morphable model techniques. *Active approaches* employ active imaging techniques to obtain face images captured under consistent illumination condition, or images of illumination invariant modalities. Additional devices (optical filters, active illumination sources or specific sensors) are usually involved to actively obtain different modalities of face images that are insensitive to or independent of illumination change. Those modalities include 3D face information and face images in those spectra other than visible spectra, such as thermal infrared image and near-infrared hyperspatial image.

Passive approaches fall under three main types of techniques to produce illumination invariant facial images – those applied at the pre-processing, feature extraction and classification levels (Struc and Pavesic, 2011). Pre-processing techniques seek to produce facial images that are free of illumination induced facial variations prior to feature extraction. They can be applied within any FR system, since they make no prior assumption that influence feature extraction or classification procedures. Feature extraction techniques seek to compensate for appearance variations in facial images using descriptors or representations that are stable under different illumination conditions. However, different empirical studies with LBP, Gabor wavelet-based features, and other descriptors have shown (Marcel *et al.*, 2007) that none of these can ensure illumination invariant FR given severe illumination changes. Classification-level techniques compensate for illumination changes according to the type of face model or classifier employed in the FR system. First, some assumptions regarding the effects of illumination on face models or classification procedure are made, and then based on these assumptions, counter measures are undertaken to obtain illumination invariant face models or illumination insensitive classification procedures. Managing the effects of illumination at the feature extraction level is debatable, while classification level techniques may impose difficult requirements on design data. While they may provide the more efficient approach to illumination invariant FR, large training set must usually be acquired under a number of lighting conditions and are, furthermore, also computationally expensive.

This empirical study focuses on passive techniques for illumination normalization at the pre-processing level. Table III-1 presents the specific techniques from literature considered in this study. A more detailed description of these techniques may be found in (Struc and Pavesic, 2011). They are the newer and more popular techniques that are representative of different families of techniques, e.g., retinex, diffusion, wavelet, frequency-based techniques. Techniques that compensate for the illumination changes during the pre-processing level may be computationally simple, and effective at achieving illumination invariant FR. A common challenges among all theses techniques is that performance depends heavily on their implementation, and on the suitable selection of their parameters. Parameters must be set empirically. In this study, results were produced using default setting from the authors of respective techniques.

Experimental Analysis

Dataset and Experimental Protocol

Chokepoint is a video dataset (Wong *et al.*, 2011) that is employed to benchmark in video surveillance applications. An array of three cameras is placed above a series of portals to record persons walking through them one by one or simultaneously depending on the scenarios, where captured videos contain changes in illumination, pose, scale, blur and occlusion. To analyze the performance of system for still-to-video FR, all 48 video sequences from the center camera (in both entering and leaving cases) from the Chokepoint dataset have been considered. Each video sequence views 25 subjects walking through a portal.

The protocol of this work consists of performing a series of steps. To start with, 5 persons are randomly selected as target individuals in the watchlist, where just one reference still image (high-quality neutral mug-shot) is available to design each face model (template). The remaining people are assumed to be unknown (non-target individuals, and reflect the universal background model). Then, the enrollment phase of the randomly selected individuals from the watchlist is accomplished. This phase consists of performing a pre-processing task on each reference still image by capturing ROIs using Viola-Jones algorithms, converting the captured ROI into grayscale and rescaling it to a common size of 48x48 due to a limit processing time. For each still ROI of an individual in the watchlist, 11 treatments of chosen illumination normalization techniques, as mentioned in section 4 of this paper, were applied using the INFace toolbox[3]. At this level, 12 representations of one ROI are created in which 11 represent the

[3] http://luks.fe.uni-lj.si/sl/osebje/vitomir/face_tools/INFace/

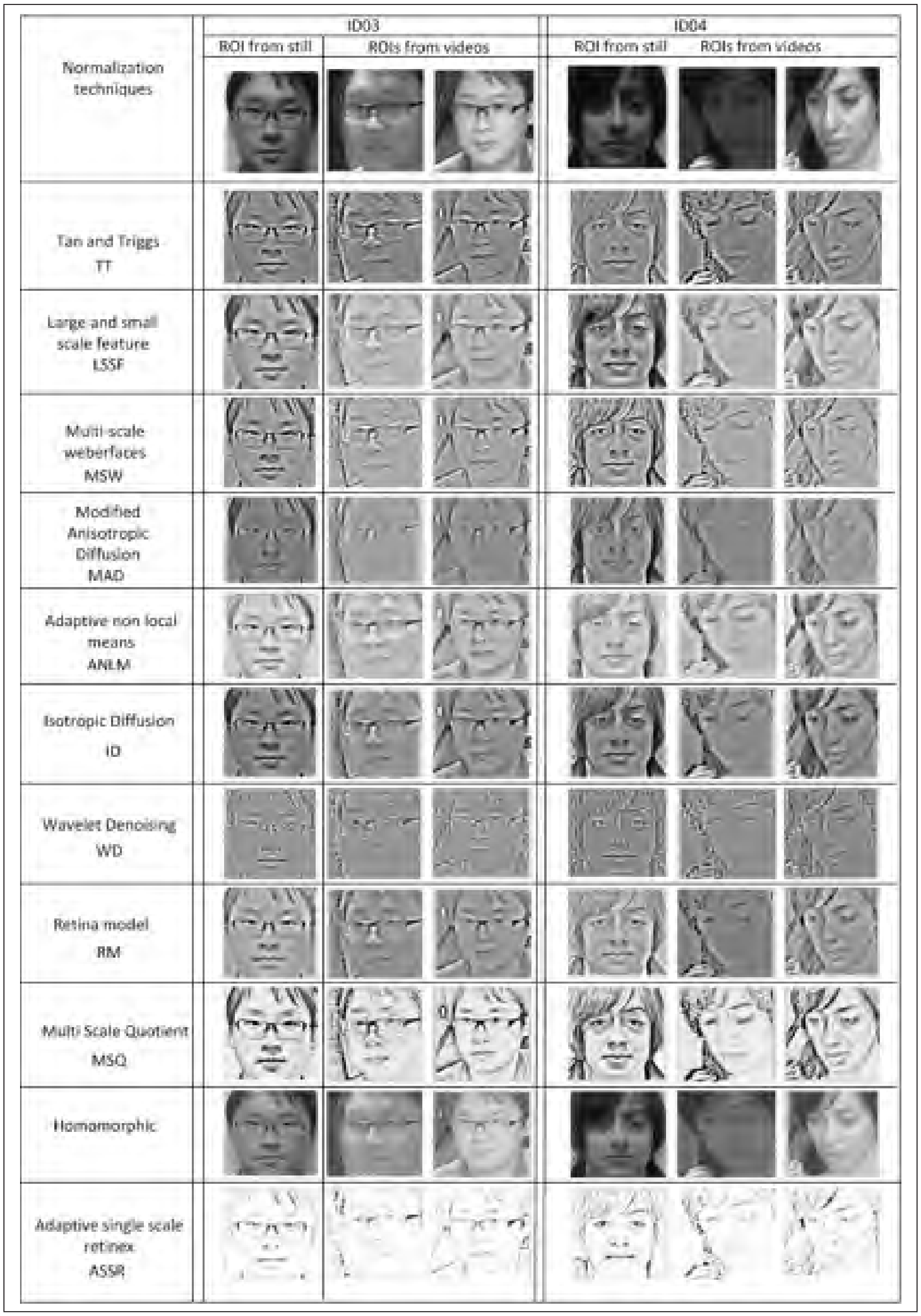


Figure-A III-2 Examples of face images obtained after illumination normalization is applied to ROIs in stills and videos from individuals ID03 and ID05

normalized ROI in terms of illumination and 1 represents the original ROI (without application of illumination normalization techniques). These representations can be found in Figure III-2.

A division into 3x3 non-overlapping patches (9 blocks bi, i=1..9) is performed on each 12 representations of each ROI. LBP feature extraction technique is applied on each block bi and all 9 feature vectors from one image ROI are then concatenated. In general,with patch-based methods, facial ROIs are divided into several overlapping or non-overlapping regions called patches, and then features are extracted locally from each patch for recognition purposes. Some specialized decision fusion techniques have been also introduced in (Topcu and Erdogan, 2010) for patch-based FR. A patch-based approach has been proposed to extract LBP texture features from each patch, and combined with the weighted majority vote for decision fusion (Nikan and Ahmadi, 2012). In this paper, 9 non-overlapping patches of size is 16x16 pixels are extracted from each ROI. Performance is provided for each individual of interest in the watch-list for all video sequences. With this method, 59 features are extracted from each patch using LBP, and assembled into a ROI pattern for matching. So, overall a feature vector for one ROI contains 531 features. The latter are then saved into a gallery of templates. By the end of this enrollment phase, the gallery of templates would have 12 different templates for each person in the watchlist. The last phase of the work protocol would be the testing phase in which video frames undergo the same pre-processing part described in the enrollment phase as well as the patches division, illumination normalization application and feature extraction using LBP on each captured ROI. For each technique, the corresponding feature vector in the testing phase would be compared using template matching with the corresponding feature vector of the same technique of the 5 individuals saved in the Gallery of templates. Finally, the scores produced with template matching is normalized to an interval of [0 1].

To assess the transaction-level performance of a watch-list system, *receiver operating characteristic (ROC)* space is considered. A ROC curve displays the the proportion of target ROIs that correctly detected as individual of interest over the total number of target ROIs in the sequence, the true positive rate (tpr), as a function of the proportion of non-target (imposter) ROI detected as individual of interest over the total number of non-target ROIs, the false positive rate (fpr). The area under ROC curve (AUC) is a global scalar performance measure that can be interpreted as the probability of classification over the range of tpr and fpr. In order to estimate the performance of the system based on target ROIs, the precision-recall (PROC) space is also considered. To measure the performance in the imbalanced data situation, recall is the tpr and precision is computed as follows $pr = TP/(TP+FP)$. The AUPR car illustrate system performance on targets given the imbalances proportions between target and non-targets (majority) in the skewed.

For trajectory-level analysis, a tracking module is employed to regroup ROIs captured over frames. Hence, captured ROIs of persons in the scene are grouped and processed individually.

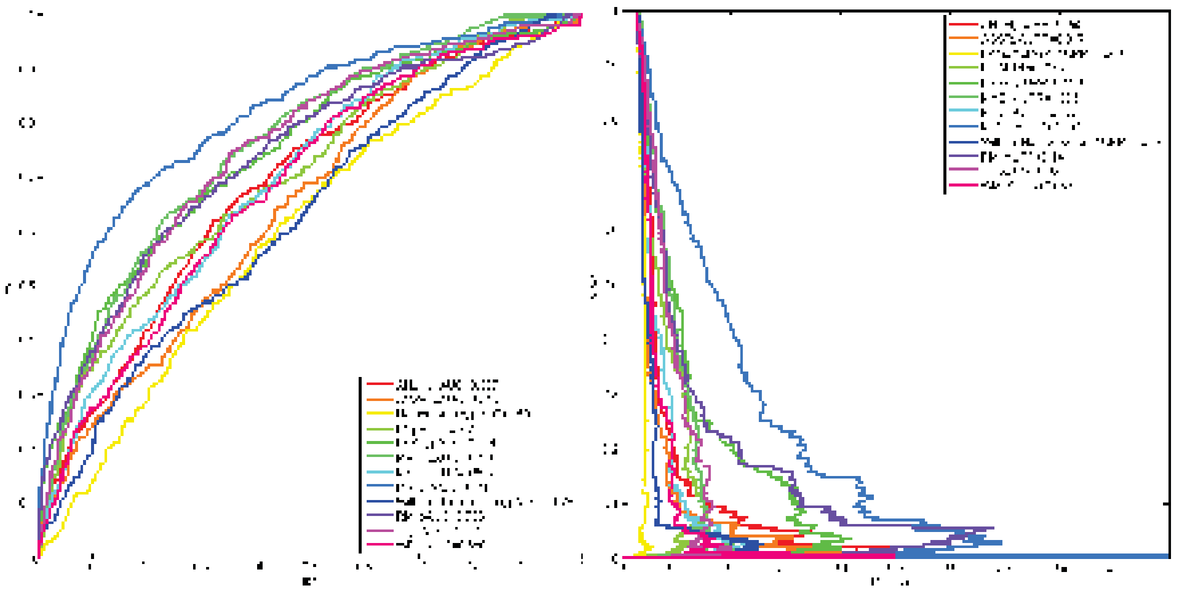
Results and discussion

Table-A III-2 Average pAUC(5%) performance with illumination normalization techniques for each individual from the watchlist

Illumination Normalisation	**ID # of Watchlist Individuals**					
	ID03	**ID04**	**ID07**	**ID09**	**ID12**	**Average**
Entering Videos						
No Normalization	0.66±0.04	0.96±0.01	0.72±0.02	0.84±0.03	0.91±0.02	0.82±0.02
Adaptive Single Scale Retinex	0.65±0.04	0.90±0.01	0.54±0.02	0.76±0.05	0.90±0.01	0.75±0.02
Large and Small Scale Features	0.72±0.06	0.89±0.03	0.69±0.02	0.89±0.03	0.92±0.03	0.82±0.03
Multi Scale Self-Quotient	0.69±0.04	0.88±0.03	0.67±0.05	0.87±0.02	0.93±0.02	0.81±0.03
Isotropic Diffusion	0.69±0.06	0.86±0.03	0.70±0.01	0.90±0.01	0.97±0.01	0.82±0.02
Modified Anisotropic Diffusion	0.74±0.05	0.85±0.03	0.74±0.02	0.80±0.03	0.94±0.02	0.81±0.03
Tan & Triggs	0.74±0.03	0.86±0.03	0.71±0.02	0.88±0.04	0.92±0.03	0.82±0.03
Multi Scale Weberfaces	0.82±0.02	0.83±0.03	0.73±0.03	0.88±0.05	0.91±0.03	0.83±0.03
Adaptive Non-Local Means	0.71±0.02	0.89±0.02	0.66±0.03	0.69±0.04	0.84±0.03	0.76±0.02
Retina Modeling	0.73±0.05	0.85±0.03	0.69±0.02	0.90±0.03	0.91±0.05	0.82±0.03
Wavelet Denoising	0.66±0.03	0.89±0.02	0.54±0.03	0.83±0.02	0.87±0.01	0.76±0.02
Homomorphic	0.62±0.04	0.94±0.01	0.73±0.02	0.81±0.05	0.91±0.01	0.80±0.02
Leaving Videos						
No Normalization	0.67±0.08	0.91±0.03	0.79±0.02	0.91±0.02	0.94±0.02	0.84±0.03
Adaptive Single Scale Retinex	0.73±0.03	0.89±0.02	0.66±0.01	0.89±0.02	0.92±0.01	0.82±0.01
Large and Small Scale Features	0.78±0.03	0.94±0.02	0.54±0.02	0.94±0.02	0.96±0.01	0.83±0.02
Multi Scale Self - Quotient	0.74±0.03	0.82±0.07	0.74±0.02	0.92±0.01	0.93±0.02	0.83±0.03
Isotropic Diffusion	0.82±0.03	0.83±0.05	0.75±0.02	0.91±0.02	0.95±0.01	0.85±0.02
Modified Anisotropic Diffusion	0.78±0.02	0.89±0.02	0.64±0.02	0.93±0.01	0.94±0.01	0.83±0.01
Tan & Triggs	0.80±0.03	0.93±0.01	0.61±0.03	0.97±0.01	0.96±0.01	0.85±0.01
Multi-Scale Weberfaces	0.85±0.03	0.92±0.01	0.73±0.02	0.95±0.01	0.95±0.01	0.88±0.01
Adaptive Non-Local Means	0.74±0.04	0.94±0.02	0.71±0.02	0.86±0.01	0.95±0.01	0.84±0.02
Retina Modeling	0.77±0.03	0.93±0.01	0.55±0.03	0.96±0.01	0.96±0.01	0.83±0.01
Wavelet Denoising	0.71±0.02	0.91±0.02	0.66±0.02	0.87±0.01	0.92±0.01	0.81±0.01
Homomorphic	0.65±0.03	0.90±0.02	0.78±0.01	0.87±0.02	0.91±0.01	0.82±0.01

Table-A III-3 Average AUPR performance with illumination normalization techniques for each individual from the watchlist

Illumination Normalisation	**ID # of Watchlist Individuals**					
	ID03	**ID04**	**ID07**	**ID09**	**ID12**	**Average**
Entering Videos						
Without Normalization	0.06±0.01	0.64±0.07	0.16±0.03	0.30±0.08	0.60±0.08	0.35±0.05
Adaptive Single Scale Retinex	0.09±0.03	0.28±0.03	0.06±0.08	0.17±0.04	0.46±0.07	0.21±0.05
Large and Small Scale features	0.18±0.06	0.40±0.04	0.14±0.03	0.51±0.08	0.63±0.01	0.37±0.04
Multi Scale Self-Quotient	0.12±0.05	0.45±0.07	0.15±0.02	0.31±0.04	0.57±0.09	0.32±0.05
Isotropic Diffusion	0.13±0.04	0.31±0.05	0.11±0.01	0.35±0.05	0.74±0.05	0.33±0.04
Modified Anisotropic Diffusion	0.13±0.04	0.34±0.07	0.17±0.02	0.28±0.06	0.68±0.09	0.32±0.05
Tan & Triggs	0.16±0.06	0.37±0.05	0.16±0.02	0.59±0.10	0.64±0.10	0.38±0.06
Multi-Scale Weberfaces	0.25±0.07	0.37±0.05	0.19±0.03	0.58±0.10	0.57±0.11	0.39±0.07
Adaptive Non-Local Means	0.11±0.04	0.51±0.06	0.14±0.02	0.07±0.01	0.53±0.07	0.27±0.04
Retina Modeling	0.22±0.07	0.32±0.05	0.16±0.02	0.63±0.09	0.66±0.10	0.40±0.06
Wavelet Denoising	0.08±0.02	0.37±0.05	0.06±0.01	0.14±0.02	0.32±0.06	0.19±0.03
Homomorphic	0.05±0.01	0.65±0.06	0.18±0.04	0.18±0.05	0.47±0.08	0.30±0.04
Leaving Videos						
Without Normalization	0.19±0.06	0.43±0.07	0.23±0.03	0.57±0.08	0.66±0.07	0.42±0.06
Adaptive Single Scale Retinex	0.14±0.02	0.26±0.06	0.11±0.01	0.41±0.04	0.58±0.03	0.30±0.03
Large and Small Scale features	0.22±0.03	0.49±0.07	0.07±0.01	0.67±0.06	0.71±0.06	0.43±0.04
Multi Scale Self-Quotient	0.11±0.01	0.31±0.09	0.19±0.04	0.59±0.05	0.66±0.04	0.40±0.04
Isotropic Diffusion	0.26±0.05	0.27±0.07	0.21±0.03	0.58±0.06	0.65±0.07	0.40±0.05
Modified Anisotropic Diffusion	0.16±0.03	0.29±0.04	0.08±0.01	0.60±0.07	0.61±0.08	0.35±0.04
Tan & Triggs	0.29±0.05	0.35±0.05	0.10±0.01	0.81±0.04	0.78±0.05	0.47±0.04
Multi-Scale Weberfaces	0.39±0.05	0.34±0.05	0.18±0.03	0.79±0.04	0.78±0.05	0.50±0.04
Adaptive Non-Local Means	0.22±0.05	0.58±0.09	0.17±0.04	0.37±0.04	0.69±0.03	0.41±0.05
Retina Modeling	0.23±0.03	0.38±0.05	0.09±0.02	0.75±0.06	0.68±0.06	0.43±0.04
Wavelet Denoising	0.09±0.01	0.37±0.08	0.09±0.01	0.30±0.03	0.46±0.06	0.26±0.03
Homomorphic	0.10±0.02	0.41±0.08	0.18±0.02	0.44±0.07	0.54±0.08	0.33±0.05



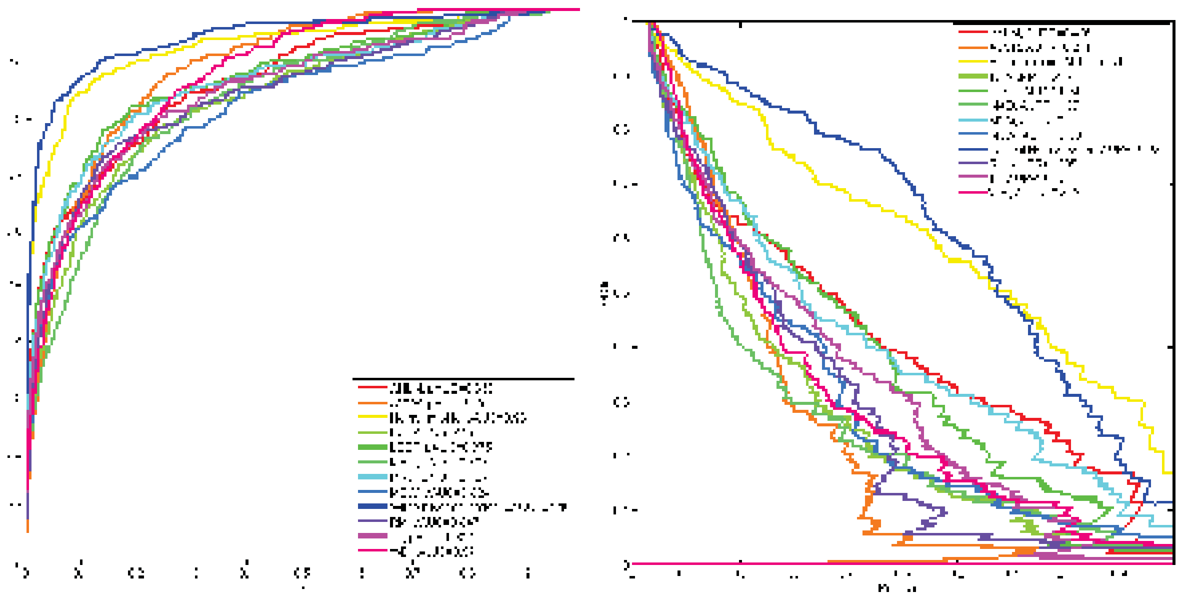
(a) ROC of ID03

(b) PR of ID03

Figure-A III-3 ROC curves *left* Inverted PR *right* of all twelve chosen illumination normalization techniques for person #3

Results in Tables III-2 and III-3 present the transaction-level performance. More specifically, these tables show the average pAUC (5%) and the AUPR performance (along with the standard deviation) obtained using 11 illumination normalization techniques for each individual from the watchlist over all entering and leaving videos of the dataset Chokepoint. Besides, these tables display the average over individuals in order to assess the performance of each illumination normalization technique regardless of the person of interest. By looking through these values, the first thing that can be observed is that the results vary significantly according to the individual and the capture conditions (sequence and portals). Indeed, looking at the results of person ID04 (Tables III-2 and III-3), applying illumination normalization decreases its performance compared to the results of "without normalization". Whereas, for person ID03, the pAUC(5%) and AUPR are much higher when normalization techniques are applied.

A simple glimpse on the transaction level scores for individual ID03 (Figure III-5), it can be observed that the scores after normalization are higher than the normal (without normalization) for the cases of target to target and non-target to target. As a matter of fact, the normalized



(a) ROC of ID04

(b) PR of ID04

Figure-A III-4 ROC curves *left* Inverted PR *right* of all twelve chosen illumination normalization techniques for person #4

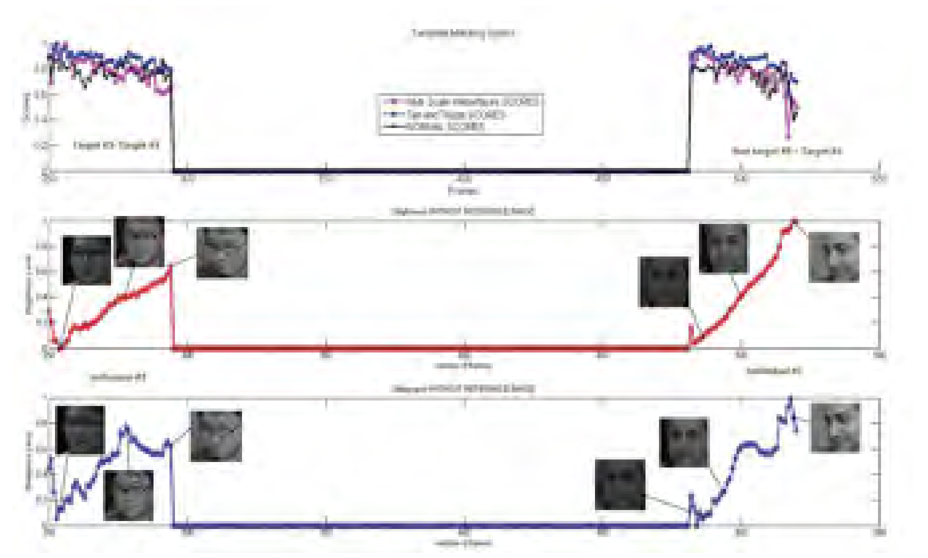


Figure-A III-5 Template Matching Scores, Brightness Level and Sharpness Level of person #3

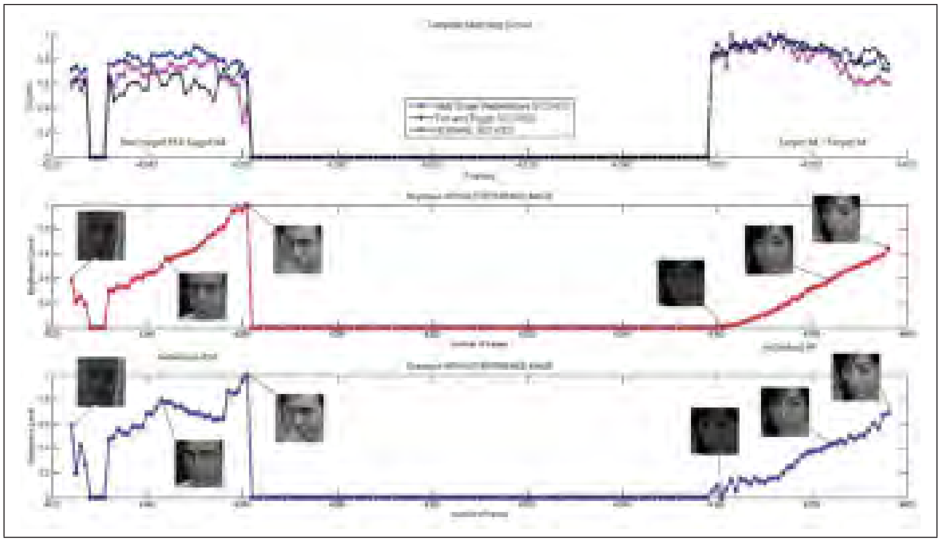


Figure-A III-6 Template Matching Scores, Brightness Level and Sharpness Level of person #4

scores are boosted. On the other hand, the normal (without normalization) transaction-level score of individual ID04 are already high (Figure III-6). Once illumination normalization is applied, most of the scores are then increased. Since the values of the scores are normalized between the 0 and 1, the boosted value scores would be very close to each other. Thus, it is hard to discriminate between the targets and the non-targets in a video sequence which causes the degradation of the ROC curves of ID04 (Figure III-4a). In this figure,“without normalization is outperforming”. Nonetheless, the precision values, when normalization technique is applied, are quite higher. For example, if a value of 0.1 (10%) false positive rate (fpr) on the ROC curve for cases “without normalization”, “MSW” and “TT”, the values of precisions on the PR curves are respectively 20%, 18% and 19% which are slightly higher compared to the 10% fpr. In addition, based on the average results by techniques, it can be concluded that both techniques Multi-Scale Weberfaces (MSW) and Tan and Triggs (TT) tend to outperform the rest of the techniques used in this work for both entering and leaving cases of the dataset.

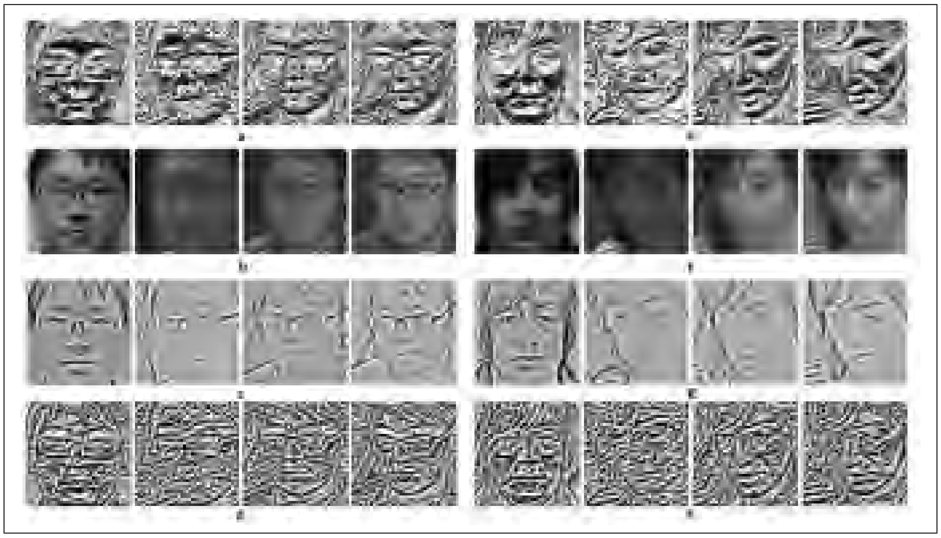


Figure-A III-7 LBP Projection of both individuals # 3 and #4 applied on the illumination normalized image using Multi Scale Weberfaces. a. original images of person #3, b. images of person #3 after applying Multi Scale Weberfaces normalization, c. lbp projection of the normalized images of person #3, d. original images of person #4, e.images of person #4 after applying Multi Scale Weberfaces normalization, f. lbp projection of the normalized images of person #4

Conclusion

The popular LBP-based approach to face analysis is known to be sensitive to severe illumination changes. Based on this observation, we investigated in this work the effect of different illumination normalization techniques in the case of LBP-Based Watchlist Screening. Watch-list screening is an important application for decision support in video surveillance systems. Extensive experimental analysis on videos from the benchmark Chokepoint dataset indicated the benefit of different illumination normalization techniques varies considerably according to the individual and illumination conditions. This suggests that a combination of these techniques should be selected dynamically based on changing capture conditions. Overall, Multi-Scale Weberfaces and Tan and Triggs techniques tend to provide interesting results outperforming those of some other techniques.

REFERENCES

Ahonen, T., A. Hadid, and M. Pietikäinen. 2006. "Face Description with Local Binary Patterns: Application to Face Recognition". *TPAMI*, vol. 28, n° 12, p. 2037-2041.

Barr, J. R., K. W. Bowyer, P. J. Flynn, and S. Biswas. 2012. "Face recognition from video: A review". *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 26, n° 05.

Chellappa, R., P. Sinha, and P. J. Phillips. 2010. "Face recognition by computers and humans". *Computer*, vol. 43, n° 2, p. 46-55.

De-la Torre, M., E. Granger, R. Sabourin, and D. O. Gorodnichy. 2014. "Partially-Supervised Learning from Facial Trajectories for Face Recognition in Video Surveillance". p. In Press, DOI: 10.1016/j.inffus.2014.05.006.

Dewan, M., E. Granger, F. Roli, R. Sabourin, and G.-L. Marcialis. December 16-17, 2013. "A Comparison of Adaptive Appearance Methods for Tracking Faces in Video Surveillance". In *The 5th International Conference on Imaging for Crime Detection and Prevention*.

Ekenel, H. K., J. Stallkamp, and R. Stiefelhagen. 2010. "A video-based door monitoring system using local appearance-based face models". *CVIU*, vol. 114, n° 5, p. 596-608.

He, C., T. Ahonen, and M. Pietikäinen. 2008. "A Bayesian local binary pattern texture descriptor". In *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*. p. 1–4. IEEE.

Huang, Z., S. Shan, H. Zhang, S. Lao, A. Kuerban, and X. Chen. 2013. Benchmarking still-to-video face recognition via partial and local linear discriminant analysis on cox-s2v dataset. *ACCV*, p. 589-600. Springer.

Kamgar-Parsi, B. and W. Lawson. 2011. "Toward development of a face recognition system for watchlist surveillance". *PAMI*, vol. 33, n° 10, p. 1925-1937.

Kan, M., S. Shan, Y. Su, D. Xu, and X. Chen. 2013. "Adaptive discriminant learning for face recognition". *Pattern Recognition*, vol. 46, n° 9, p. 2497–2509.

Li, S. Z., S. R. Chu, S. Liao, and L. Zhang. 2007. "Illumination invariant face recognition using near-infrared images". *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 29, n° 4, p. 627–639.

Liao, S., G. Zhao, V. Kellokumpu, M. Pietikäinen, and S. Z. Li. 2010. "Modeling pixel process with scale invariant local patterns for background subtraction in complex scenes". In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. p. 1301–1306. IEEE.

Marcel, S., Y. Rodriguez, and G. Heusch. 2007. "On the Recent Use of Local Binary Patterns for Face Authentication". *International Journal of Image and Video Processing, Special Issue on Facial Image Processing*, p. 469–481.

Matta, F. and J.-L. Dugelay. 2009. "Person recognition using facial video information: A state of the art". *Journal of Visual Languages & Computing*, vol. 20, n° 3, p. 180–187.

Nikan, S. and M. Ahmadi. 2012. "Human face recognition under occlusion using lbp and entropy weighted voting". In *Pattern Recognition (ICPR), 2012 21st International Conference on*. p. 1699–1702. IEEE.

Ojala, T., M. Pietikäinen, and T. Mäenpää. 2002. "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns". *TPAMI*, vol. 24, n° 7, p. 971-987.

Pagano, C., E. Granger, R. Sabourin, and D. O. Gorodnichy. 2012. "Detector ensembles for face recognition in video surveillance". In *IJCNN*. p. 1-8. IEEE.

Pietikäinen, M., A. Hadid, G. Zhao, and T. Ahonen, 2011. *Computer Vision Using Local Binary Patterns*.

Shaokang, C., M. Sandra, H. Mehrtash T, S. Conrad, B. Abbas, L. Brian C, et al. 2011. "Face recognition from still images to video sequences: A local-feature-based framework". *EURASIP journal on image and video processing*, vol. 2011.

Sharma, A., V. D. Kaushik, and P. Gupta. 2014. Illumination invariant face recognition. *Intelligent Computing Theory*, p. 308–319. Springer.

Struc, V. and N. Pavesic. 2011. Performance evaluation of photometric normalization techniques for illumination invariant face recognition. Zhang, Y., editor, *Advances in Face Image Analysis: Techniques and Technologies*. IGI Global, Hershey, USA.

Tan, X. and B. Triggs. 2007. Enhanced local texture feature sets for face recognition under difficult lighting conditions. *Analysis and Modeling of Faces and Gestures*, p. 168–182. Springer.

Topcu, B. and H. Erdogan. 2010. "Decision fusion for patch-based face recognition". In *ICPR*. p. 1348-1351. IEEE.

Wong, Y., S. Chen, S. Mau, C. Sanderson, and B. C. Lovell. 2011. "Patch-based probabilistic image quality assessment for face selection and improved video-based face recognition". In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2011 IEEE Computer Society Conference on*. p. 74–81. IEEE.

Zou, X., J. Kittler, and K. Messer. 2007. "Illumination invariant face recognition: A survey". In *Biometrics: Theory, Applications, and Systems, 2007. BTAS 2007. First IEEE International Conference on*. p. 1–8. IEEE.

BIBLIOGRAPHY

Abaza, A., M. A. Harrison, and T. Bourlai. 2012. "Quality metrics for practical face recognition". In *Pattern Recognition (ICPR), 2012 21st International Conference on*. p. 3103–3107. IEEE.

Abboud, A. J. and S. A. Jassim. 2012. "Biometric templates selection and update using quality measures". In *SPIE Defense, Security, and Sensing*. p. 840609–840609. International Society for Optics and Photonics.

Abboud, A. J., H. Sellahewa, and S. A. Jassim. 2009. "Quality based approach for adaptive face recognition". In *SPIE Defense, Security, and Sensing*. p. 73510N–73510N. International Society for Optics and Photonics.

Abdel-Mottaleb, M. and M. H. Mahoor. 2007. "Application notes-algorithms for assessing the quality of facial images". *Computational Intelligence Magazine, IEEE*, vol. 2, n° 2, p. 10–17.

Ahonen, T., A. Hadid, and M. Pietikäinen. 2004. Face recognition with local binary patterns. *Computer vision-eccv 2004*, p. 469–481. Springer.

Ahonen, T., A. Hadid, and M. Pietikainen. 2006. "Face description with local binary patterns: Application to face recognition". *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 28, n° 12, p. 2037–2041.

Ahonen, T., E. Rahtu, V. Ojansivu, and J. Heikkila. 2008. "Recognition of blurred faces using local phase quantization". In *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*. p. 1–4. IEEE.

Al-Azzeh, M., A. Eleyan, and H. Demirel. 2008. "PCA-based face recognition from video using super-resolution". In *Computer and Information Sciences, 2008. ISCIS'08. 23rd International Symposium on*. p. 1–4. IEEE.

Amara, I., E. Granger, and A. Hadid. 2014. "On the Effects of Illumination Normalization with LBP-Based Watchlist Screening". In *Computer Vision-ECCV 2014 Workshops*. p. 173–188. Springer.

Barr, J. R., K. W. Bowyer, P. J. Flynn, and S. Biswas. 2012. "Face recognition from video: A review". *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 26, n° 05, p. 1266002.

Barry, M. and E. Granger. 2007. "Face recognition in video using a what-and-where fusion neural network". In *Neural Networks, 2007. IJCNN 2007. International Joint Conference on*. p. 2256–2261. IEEE.

Bashbaghi, S., E. Granger, R. Sabourin, and G.-A. Bilodeau. 2014. "Watch-List Screening Using Ensembles Based on Multiple Face Representations". In *Pattern Recognition (ICPR), 2014 22nd International Conference on*. p. 4489–4494. IEEE.

Berrani, S.-A. and C. Garcia. 2005. "Enhancing face recognition from video sequences using robust statistics". In *Advanced Video and Signal Based Surveillance, 2005. AVSS 2005. IEEE Conference on*. p. 324–329. IEEE.

Beymer, D. and T. Poggio. 1995. "Face recognition from one example view". In *Computer Vision, 1995. Proceedings., Fifth International Conference on*. p. 500–507. IEEE.

Blitzer, J., R. McDonald, and F. Pereira. 2006. "Domain adaptation with structural correspondence learning". In *Proceedings of the 2006 conference on empirical methods in natural language processing*. p. 120–128. Association for Computational Linguistics.

Boyd, K., K. H. Eng, and C. D. Page. 2013. Area under the precision-recall curve: Point estimates and confidence intervals. *Machine Learning and Knowledge Discovery in Databases*, p. 451–466. Springer.

Bradley, A. P. 1997. "The use of the area under the ROC curve in the evaluation of machine learning algorithms". *Pattern recognition*, vol. 30, n° 7, p. 1145–1159.

Bradski, G. R. 1998. "Real time face and object tracking as a component of a perceptual user interface". In *Applications of Computer Vision, 1998. WACV'98. Proceedings., Fourth IEEE Workshop on*. p. 214–219. IEEE.

Chan, Y., A. G. Hu, and J. Plant. 1979. "A Kalman filter based tracking scheme with input estimation". *Aerospace and Electronic Systems, IEEE Transactions on*, , p. 237–244.

Chen, Y., Y. Rui, and T. S. Huang. 2001. "JPDAF based HMM for real-time contour tracking". In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*. p. I–543. IEEE.

Cheng, J. and L. Chen, 2014. *A Weighted Regional Voting Based Ensemble of Multiple Classifiers for Face Recognition*.

Cheng, Y. 1995. "Mean shift, mode seeking, and clustering". *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 17, n° 8, p. 790–799.

Connolly, J.-F., E. Granger, and R. Sabourin. 2012. "Evolution of heterogeneous ensembles through dynamic particle swarm optimization for video-based face recognition". *Pattern Recognition*, vol. 45, n° 7, p. 2460–2477.

Cox, L. J. and S. L. Hingorani. 1996. "An efficient implementation of Reid's multiple hypothesis tracking algorithm and its evaluation for the purpose of visual tracking". *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 18, n° 2, p. 138–150.

Dai, W., G.-R. Xue, Q. Yang, and Y. Yu. 2007. "Transferring naive bayes classifiers for text classification". In *Proceedings of the national conference on artificial intelligence*. p. 540. Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999.

Dalal, N. and B. Triggs. 2005. "Histograms of oriented gradients for human detection". In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*. p. 886–893. IEEE.

Daume III, H. and D. Marcu. 2006. "Domain adaptation for statistical classifiers". *Journal of Artificial Intelligence Research*, p. 101–126.

Davis, J. and M. Goadrich. 2006. "The relationship between Precision-Recall and ROC curves". In *Proceedings of the 23rd international conference on Machine learning*. p. 233–240. ACM.

De-la Torre, M., E. Granger, P. V. Radtke, R. Sabourin, and D. O. Gorodnichy. 2015. "Partially-supervised learning from facial trajectories for face recognition in video surveillance". *Information Fusion*, vol. 24, p. 31–53.

Delac, K. and M. Grgic. 2004. "A survey of biometric recognition methods". In *Electronics in Marine, 2004. Proceedings Elmar 2004. 46th International Symposium*. p. 184–193. IEEE.

Deng, W., J. Hu, and J. Guo. 2012. "Extended SRC: Undersampled face recognition via intra-class variant dictionary". *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 34, n° 9, p. 1864–1870.

Denoeux, T. 1995. "A k-nearest neighbor classification rule based on Dempster-Shafer theory". *Systems, Man and Cybernetics, IEEE Transactions on*, vol. 25, n° 5, p. 804–813.

Dewan, M. A. A., E. Granger, G.-L. Marcialis, R. Sabourin, and F. Roli. 2016. "Adaptive appearance model tracking for still-to-video face recognition". *Pattern Recognition*, vol. 49, p. 129–151.

Duan, L., I. W. Tsang, D. Xu, and T.-S. Chua. 2009. "Domain adaptation from multiple sources via auxiliary classifiers". In *Proceedings of the 26th Annual International Conference on Machine Learning*. p. 289–296. ACM.

Ekenel, H. K., J. Stallkamp, and R. Stiefelhagen. 2010. "A video-based door monitoring system using local appearance-based face models". *Computer Vision and Image Understanding*, vol. 114, n° 5, p. 596–608.

Fasel, B. and J. Luettin. 2003. "Automatic facial expression analysis: a survey". *Pattern recognition*, vol. 36, n° 1, p. 259–275.

Fawcett, T. 2006. "An introduction to ROC analysis". *Pattern recognition letters*, vol. 27, n° 8, p. 861–874.

Fröba, B. and A. Ernst. 2003. "Fast frontal-view face detection using a multi-path decision tree". In *Audio-and Video-Based Biometric Person Authentication*. p. 921–928. Springer.

Garcia, C. and M. Delakis. 2004. "Convolutional face finder: A neural architecture for fast and robust face detection". *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 26, n° 11, p. 1408–1423.

Gopalan, R., R. Li, and R. Chellappa. 2014. "Unsupervised adaptation across domain shifts by generating intermediate data representations". *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 36, n° 11, p. 2288–2302.

Gopalan, R., R. Li, and R. Chellappa. 2011. "Domain adaptation for object recognition: An unsupervised approach". In *Computer Vision (ICCV), 2011 IEEE International Conference on*. p. 999–1006. IEEE.

Gorodnichy, D. 2005. "Video-based framework for face recognition in video".

Goshtasby, A. A. 2012. Similarity and dissimilarity measures. *Image Registration*, p. 7–66. Springer.

Gross, R. and V. Brajovic. 2003. "An image preprocessing algorithm for illumination invariant face recognition". In *Audio-and Video-Based Biometric Person Authentication*. p. 10–18. Springer.

Gunturk, B. K., A. U. Batur, Y. Altunbasak, M. H. Hayes, and R. M. Mersereau. 2003. "Eigenface-domain super-resolution for face recognition". *Image Processing, IEEE Transactions on*, vol. 12, n° 5, p. 597–606.

Guo, G., S. Z. Li, and K. L. Chan. 2000. "Face recognition by support vector machines". In *Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on*. p. 196–201. IEEE.

Guo, G., S. Z. Li, and K. L. Chan. 2001. "Support vector machines for face recognition". *Image and Vision computing*, vol. 19, n° 9, p. 631–638.

Hadid, A. and M. Pietikainen. 2004. "Selecting models from videos for appearance-based face recognition". In *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*. p. 304–308. IEEE.

Hadid, A. and M. Pietikäinen. 2009a. "Combining appearance and motion for face and gender recognition from videos". *Pattern Recognition*, vol. 42, n° 11, p. 2818–2827.

Hadid, A. and M. Pietikäinen. 2009b. Manifold learning for video-to-video face recognition. *Biometric ID Management and Multimodal Communication*, p. 9–16. Springer.

Hamel, L. 2009. "Model Assessment with ROC Curves.".

He, C., T. Ahonen, and M. Pietikainen. Dec 2008. "A Bayesian Local Binary Pattern texture descriptor". In *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*. p. 1-4.

Heusch, G., Y. Rodriguez, and S. Marcel. 2006. "Local binary patterns as an image preprocessing for face authentication". In *Automatic Face and Gesture Recognition, 2006. FGR 2006. 7th International Conference on*. p. 6–pp. IEEE.

Ho, H. T. and R. Gopalan. 2014. "Model-driven domain adaptation on product manifolds for unconstrained face recognition". *International Journal of Computer Vision*, vol. 109, n° 1-2, p. 110–125.

Ho, T. K., J. J. Hull, and S. N. Srihari. 1994. "Decision combination in multiple classifier systems". *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 16, n° 1, p. 66–75.

Huang, K. S. and M. M. Trivedi. 2002. "Streaming face recognition using multicamera video arrays". In *Pattern Recognition, 2002. Proceedings. 16th International Conference on*. p. 213–216. IEEE.

Huang, Z., X. Zhao, S. Shan, R. Wang, and X. Chen. 2013. "Coupling alignments with recognition for still-to-video face recognition". In *Computer Vision (ICCV), 2013 IEEE International Conference on*. p. 3296–3303. IEEE.

Huttenlocher, D. P., J. J. Noh, and W. J. Rucklidge. 1993. "Tracking non-rigid objects in complex scenes". In *Computer Vision, 1993. Proceedings., Fourth International Conference on*. p. 93–101. IEEE.

Huynh-Thu, Q. and M. Ghanbari. 2008. "Scope of validity of PSNR in image/video quality assessment". *Electronics letters*, vol. 44, n° 13, p. 800–801.

Jafri, R. and H. R. Arabnia. 2009. "A Survey of Face Recognition Techniques.". *JIPS*, vol. 5, n° 2, p. 41–68.

Jain, A. K., R. P. Duin, and J. Mao. 2000. "Statistical pattern recognition: A review". *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 22, n° 1, p. 4–37.

Jain, A. K., A. Ross, and S. Prabhakar. 2004. "An introduction to biometric recognition". *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 14, n° 1, p. 4–20.

Jobson, D. J., Z.-U. Rahman, and G. A. Woodell. 1997. "A multiscale retinex for bridging the gap between color images and the human observation of scenes". *Image Processing, IEEE Transactions on*, vol. 6, n° 7, p. 965–976.

Kamgar-Parsi, B. and W. Lawson. 2011. "Toward development of a face recognition system for watchlist surveillance". *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 33, n° 10, p. 1925–1937.

Kan, M., J. Wu, S. Shan, and X. Chen. 2014. "Domain adaptation for face recognition: Targetize source domain bridged by common subspace". *International Journal of Computer Vision*, vol. 109, n° 1-2, p. 94–109.

Kanade, T. 1973. "Picture processing system by computer complex and recognition of human faces". *Doctoral dissertation, Kyoto University*, vol. 3952, p. 83–97.

Kannala, J. and E. Rahtu. 2012. "Bsif: Binarized statistical image features". In *Pattern Recognition (ICPR), 2012 21st International Conference on*. p. 1363–1366. IEEE.

Kim, M., S. Kumar, V. Pavlovic, and H. Rowley. 2008. "Face tracking and recognition with visual constraints in real-world videos". In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*. p. 1–8. IEEE.

Kittler, J., M. Hatef, R. P. Duin, and J. Matas. 1998. "On combining classifiers". *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 20, n° 3, p. 226–239.

Kumar, A., D. C. Wong, H. C. Shen, and A. K. Jain. 2003. "Personal verification using palmprint and hand geometry biometric". In *Audio-and Video-Based Biometric Person Authentication*. p. 668–678. Springer.

Kyrki, V., J.-K. Kamarainen, and H. Kälviäinen. 2004. "Simple Gabor feature space for invariant object recognition". *Pattern recognition letters*, vol. 25, n° 3, p. 311–318.

Lam, K.-M. and H. Yan. 1998. "An analytic-to-holistic approach for face recognition based on a single frontal view". *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 20, n° 7, p. 673–686.

Land, E. H. and J. McCann. 1971. "Lightness and retinex theory". *JOSA*, vol. 61, n° 1, p. 1–11.

Lee, K.-C., J. Ho, M.-H. Yang, and D. Kriegman. 2003. "Video-based face recognition using probabilistic appearance manifolds". In *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*. p. I–313. IEEE.

Lemieux, A. and M. Parizeau. 2003. "Flexible multi-classifier architecture for face recognition systems". In *The 16th International Conference on Vision Interface*. Citeseer.

Levi, K. and Y. Weiss. 2004. "Learning object detection from a small number of examples: the importance of good features". In *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*. p. II–53. IEEE.

Li, Q. 2012. "Literature Survey: Domain Adaptation Algorithms for Natural Language Processing".

Li, S. Z., R. Chu, S. Liao, and L. Zhang. 2007. "Illumination Invariant Face Recognition Using Near-Infrared Images". *IEEE T. PAMI*, vol. 29, n° 4, p. 627–639.

Liao, S., G. Zhao, V. Kellokumpu, M. Pietikainen, and S. Li. June 2010. "Modeling pixel process with scale invariant local patterns for background subtraction in complex scenes". In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. p. 1301-1306.

Liao, S., A. K. Jain, and S. Z. Li. 2013. "Partial face recognition: Alignment-free approach". *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 35, n° 5, p. 1193–1205.

Liu, D.-H., K.-M. Lam, and L.-S. Shen. 2005. "Illumination invariant face recognition". *Pattern Recognition*, vol. 38, n° 10, p. 1705–1716.

Lu, X., Y. Wang, and A. K. Jain. 2003. "Combining classifiers for face recognition". In *Multimedia and Expo, 2003. ICME'03. Proceedings. 2003 International Conference on*. p. III–13. IEEE.

Luo, J., Y. Ma, E. Takikawa, S. Lao, M. Kawade, and B.-L. Lu. 2007. "Person-specific SIFT features for face recognition". In *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on*. p. II–593. IEEE.

Marcel, S., Y. Rodriguez, and G. Heusch. 2006. *On the recent use of local binary patterns for face authentication*. Technical report.

Martínez, A. M. 2002. "Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class". *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 24, n° 6, p. 748–763.

Matta, F. and J.-L. Dugelay. 2009. "Person recognition using facial video information: A state of the art". *Journal of Visual Languages & Computing*, vol. 20, n° 3, p. 180–187.

Maturana, D., D. Mery, and A. Soto. 2011. Face recognition with decision tree-based local binary patterns. *Computer Vision–ACCV 2010*, p. 618–629. Springer.

Moghaddam, B., T. Jebara, and A. Pentland. 2000. "Bayesian face recognition". *Pattern Recognition*, vol. 33, n° 11, p. 1771–1782.

Mokhayeri, F., E. Granger, and G.-A. Bilodeau. 2015. "Synthetic face generation under various operational conditions in video surveillance". In *International Conference on Image Processing, Quebec, Canada*.

Nasrollahi, K. and T. B. Moeslund. 2008. Face quality assessment system in video sequences. *Biometrics and Identity Management*, p. 10–18. Springer.

Nguyen, H. V. and L. Bai. 2011. Cosine similarity metric learning for face verification. *Computer Vision–ACCV 2010*, p. 709–720. Springer.

Ojala, T., K. Valkealahti, E. Oja, and M. Pietikäinen. 2001. "Texture discrimination with multidimensional distributions of signed gray-level differences". *Pattern Recognition*, vol. 34, n° 3, p. 727–739.

Ojansivu, V. and J. Heikkilä. 2008. Blur insensitive texture classification using local phase quantization. *Image and signal processing*, p. 236–243. Springer.

Ojansivu, V., E. Rahtu, and J. Heikkilä. 2008. "Rotation invariant local phase quantization for blur insensitive texture analysis". In *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*. p. 1–4. IEEE.

Pagano, C., E. Granger, R. Sabourin, G. Marcialis, and F. Roli. 2014. "Adaptive ensembles for face recognition in changing video surveillance environments". *Information Sciences*, vol. 286, p. 75–101.

Pagano, C., E. Granger, R. Sabourin, and D. O. Gorodnichy. 2012. "Detector ensembles for face recognition in video surveillance". In *Neural Networks (IJCNN), The 2012 International Joint Conference on*. p. 1–8. IEEE.

Pan, S. J. and Q. Yang. 2010. "A survey on transfer learning". *Knowledge and Data Engineering, IEEE Transactions on*, vol. 22, n° 10, p. 1345–1359.

Pan, S. J., I. W. Tsang, J. T. Kwok, and Q. Yang. 2011. "Domain adaptation via transfer component analysis". *Neural Networks, IEEE Transactions on*, vol. 22, n° 2, p. 199–210.

Park, U. and A. K. Jain. 2007. 3d model-based face recognition in video. *Advances in Biometrics*, p. 1085–1094. Springer.

Park, Y. K., S. L. Park, and J. K. Kim. 2008. "Retinex method based on adaptive smoothing for illumination invariant face recognition". *Signal Processing*, vol. 88, n° 8, p. 1929–1945.

Parveen, P. and B. Thuraisingham. 2006. "Face recognition using multiple classifiers". In *Tools with Artificial Intelligence, 2006. ICTAI'06. 18th IEEE International Conference on*. p. 179–186. IEEE.

Patel, V. M., R. Gopalan, R. Li, and R. Chellappa. 2015. "Visual Domain Adaptation: A survey of recent advances". *Signal Processing Magazine, IEEE*, vol. 32, n° 3, p. 53–69.

Pentland, A., B. Moghaddam, and T. Starner. 1994. "View-based and modular eigenspaces for face recognition". In *Computer Vision and Pattern Recognition, 1994. Proceedings CVPR'94., 1994 IEEE Computer Society Conference on*. p. 84–91. IEEE.

Pietikäinen, M., A. Hadid, G. Zhao, and T. Ahonen, 2011. *Computer vision using local binary patterns*, volume 40.

Qiu, Q. and R. Chellappa. 2013. "Compositional dictionaries for domain adaptive face recognition". *arXiv preprint arXiv:1308.0271.*

Rabiner, L. 1989. "A tutorial on hidden Markov models and selected applications in speech recognition". *Proceedings of the IEEE*, vol. 77, n° 2, p. 257–286.

Radford, U. "Table of Critical Values for Pearson's r". <http://www.radford.edu/~jaspelme/statsbook/Chapter%20files/Table_of_Critical_Values_for_r.pdf>. Accessed: 2015-04-15.

Rodriguez, Y. and S. Marcel. 2006. Face authentication using adapted local binary pattern histograms. *Computer Vision–ECCV 2006*, p. 321–332. Springer.

Ross, A. and A. K. Jain. 2004. "Multimodal biometrics: An overview". In *Signal Processing Conference, 2004 12th European*. p. 1221–1224. IEEE.

Sabzmeydani, P. and G. Mori. 2007. "Detecting pedestrians by learning shapelet features". In *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*. p. 1–8. IEEE.

Saenko, K., B. Kulis, M. Fritz, and T. Darrell. 2010. Adapting visual category models to new domains. *Computer Vision–ECCV 2010*, p. 213–226. Springer.

Sang, J., Z. Lei, and S. Z. Li. 2009. Face image quality evaluation for iso/iec standards 19794-5 and 29794-5. *Advances in Biometrics*, p. 229–238. Springer.

Santamaría, M. V. and R. P. Palacios. 2005. "Comparison of illumination normalization methods for face recognition".

Sharma, A., V. D. Kaushik, and P. Gupta. 2014. Illumination invariant face recognition. *Intelligent Computing Theory*, p. 308–319. Springer.

Shekhar, S., V. M. Patel, H. Nguyen, and R. Chellappa. 2013. "Generalized domain-adaptive dictionaries". In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*. p. 361–368. IEEE.

Snidaro, L., L. Vati, J. Garcia, E. D. Marti, A.-L. Jousselme, K. Bryan, D. D. Bloisi, and D. Nardi. 2015. "A framework for dynamic context exploitation". In *Information Fusion (Fusion), 2015 18th International Conference on*. p. 1160–1167. IEEE.

Sokolova, M. and G. Lapalme. 2009. "A systematic analysis of performance measures for classification tasks". *Information Processing & Management*, vol. 45, n° 4, p. 427–437.

Spackman, K. A. 1989. "Signal detection theory: Valuable tools for evaluating inductive learning". In *Proceedings of the sixth international workshop on Machine learning*. p. 160–163. Morgan Kaufmann Publishers Inc.

Stallkamp, J., H. K. Ekenel, and R. Stiefelhagen. 2007. "Video-based face recognition on real-world data". In *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*. p. 1–8. ieee.

Struc, V. and N. Pavesic. 2011. "Performance evaluation of photometric normalization techniques for illumination invariant face recognition". *Advances in face image analysis: techniques and technologies. IGI Global.*

Suen, C. Y. and L. Lam. 2000. Multiple classifier combination methodologies for different output levels. *Multiple Classifier Systems*, p. 52–66. Springer.

Sugiyama, M., S. Nakajima, H. Kashima, P. V. Buenau, and M. Kawanabe. 2008. "Direct importance estimation with model selection and its application to covariate shift adaptation". In *Advances in neural information processing systems*. p. 1433–1440.

Takahashi, T., I. Ichiro, H. Murase, et al. 2009. "Incremental unsupervised-learning of appearance manifold with view-dependent covariance matrix for face recognition from video sequences". *IEICE transactions on information and systems*, vol. 92, n° 4, p. 642–652.

Tan, X. and B. Triggs. 2007. Enhanced local texture feature sets for face recognition under difficult lighting conditions. Zhou, S., Wenyi Zhao, Xiaoou Tang, and Shaogang Gong, editors, *Analysis and Modeling of Faces and Gestures*, volume 4778 of *Lecture Notes in Computer Science*, p. 168-182. Springer Berlin Heidelberg. ISBN 978-3-540-75689-7. doi: 10.1007/978-3-540-75690-3_13.

Tan, X. and B. Triggs. 2010. "Enhanced local texture feature sets for face recognition under difficult lighting conditions". *Image Processing, IEEE Transactions on*, vol. 19, n° 6, p. 1635–1650.

Tan, X., S. Chen, Z.-H. Zhou, and F. Zhang. 2006. "Face recognition from a single image per person: A survey". *Pattern recognition*, vol. 39, n° 9, p. 1725–1745.

Torres, L. and J. Vilá. 2002. "Automatic face recognition for video indexing applications". *Pattern recognition*, vol. 35, n° 3, p. 615–625.

Tuchler, M., A. C. Singer, and R. Koetter. 2002. "Minimum mean squared error equalization using a priori information". *Signal Processing, IEEE Transactions on*, vol. 50, n° 3, p. 673–683.

Turk, M. and A. Pentland. 1991a. "Eigenfaces for recognition". *Journal of cognitive neuroscience*, vol. 3, n° 1, p. 71–86.

Turk, M. A. and A. P. Pentland. 1991b. "Face recognition using eigenfaces". In *Computer Vision and Pattern Recognition, 1991. Proceedings CVPR'91., IEEE Computer Society Conference on*. p. 586–591. IEEE.

Varma, M. and A. Zisserman. 2005. "A statistical approach to texture classification from single images". *International Journal of Computer Vision*, vol. 62, n° 1-2, p. 61–81.

Vázquez, D., A. M. López, and D. Ponsa. 2012. "Unsupervised domain adaptation of virtual and real worlds for pedestrian detection". In *Pattern Recognition (ICPR), 2012 21st International Conference on*. p. 3492–3495. IEEE.

Vijayakumari, V. 2013. "Face Recognition Techniques: A Survey". *World Journal of Computer Application and Technology*, vol. 1, n° 2, p. 41–50.

Viola, P. and M. Jones. 2001. "Rapid object detection using a boosted cascade of simple features". In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*. p. I–511. IEEE.

Štruc, V. and N. Pavešić. 2009. "Gabor-based kernel-partial-least-squares discrimination features for face recognition". *Informatica (Vilnius)*, vol. 20, n° 1, p. 115-138.

Štruc, V. and N. Pavešić, 2011. *Photometric normalization techniques for illumination invariance*, p. 279-300. IGI-Global.

Wang, H., S. Z. Li, Y. Wang, and J. Zhang. 2004a. "Self quotient image for face recognition". In *Image Processing, 2004. ICIP'04. 2004 International Conference on*. p. 1397–1400. IEEE.

Wang, J.-G. and E. Sung. 1999. "Frontal-view face detection and facial feature extraction using color and morphological operations". *Pattern recognition letters*, vol. 20, n° 10, p. 1053–1068.

Wang, P. and Q. Ji. 2004. "Multi-View Face Detection under Complex Scene based on Combined SVMs.". In *ICPR (4)*. p. 179–182.

Wang, R., S. Shan, X. Chen, and W. Gao. 2008. "Manifold-manifold distance with application to face recognition based on image set". In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*. p. 1–8. IEEE.

Wang, Z., A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. 2004b. "Image quality assessment: from error visibility to structural similarity". *Image Processing, IEEE Transactions on*, vol. 13, n° 4, p. 600–612.

Weber, F. 2006. "Some quality measures for face images and their relationship to recognition performance". In *Biometric Quality Workshop. National Institute of Standards and Technology, Maryland, USA*.

Wei, C.-P. and Y.-C. F. Wang. 2015. "Undersampled Face Recognition via Robust Auxiliary Dictionary Learning". *Image Processing, IEEE Transactions on*, vol. 24, n° 6, p. 1722–1734.

Wold, S., K. Esbensen, and P. Geladi. 1987. "Principal component analysis". *Chemometrics and intelligent laboratory systems*, vol. 2, n° 1, p. 37–52.

Wong, Y., S. Chen, S. Mau, C. Sanderson, and B. C. Lovell. 2011. "Patch-based probabilistic image quality assessment for face selection and improved video-based face recognition". In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2011 IEEE Computer Society Conference on*. p. 74–81. IEEE.

Wright, J., A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma. 2009. "Robust face recognition via sparse representation". *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 31, n° 2, p. 210–227.

Wu, B. and R. Nevatia. 2005. "Detection of multiple, partially occluded humans in a single image by bayesian combination of edgelet part detectors". In *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*. p. 90–97. IEEE.

Wu, B., H. Ai, C. Huang, and S. Lao. 2004. "Fast rotation invariant multi-view face detection based on real adaboost". In *Automatic Face and Gesture Recognition, 2004. Proceedings. Sixth IEEE International Conference on*. p. 79–84. IEEE.

Xu, Y., A. Roy-Chowdhury, and K. Patel. 2007. "Integrating illumination, motion, and shape models for robust face recognition in video". *EURASIP Journal on Advances in Signal Processing*, vol. 2008, n° 1, p. 469698.

Yang, M., L. Van Gool, and L. Zhang. 2013. "Sparse variation dictionary learning for face recognition with a single training sample per person". In *Computer Vision (ICCV), 2013 IEEE International Conference on*. p. 689–696. IEEE.

Yang, M.-H., D. Kriegman, and N. Ahuja. 2002. "Detecting faces in images: A survey". *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 24, n° 1, p. 34–58.

Yilmaz, A., O. Javed, and M. Shah. 2006. "Object tracking: A survey". *Acm computing surveys (CSUR)*, vol. 38, n° 4, p. 13.

Yule, D. and L. Chen. 2011. "Face recognition through regional weight estimation". In *Image Processing (ICIP), 2011 18th IEEE International Conference on*. p. 1761–1764. IEEE.

Zadrozny, B. 2004. "Learning and evaluating classifiers under sample selection bias". In *Proceedings of the twenty-first international conference on Machine learning*. p. 114. ACM.

Zhang, C. and Z. Zhang. 2010. *A survey of recent advances in face detection*. Technical report.

Zhang, G. P. 2000. "Neural networks for classification: a survey". *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, vol. 30, n° 4, p. 451–462.

Zhang, J., Y. Yan, and M. Lades. 1997. "Face recognition: eigenface, elastic matching, and neural nets". *Proceedings of the IEEE*, vol. 85, n° 9, p. 1423–1435.

Zhang, L. and P. Lenders. 2000. "Knowledge-based eye detection for human face recognition". In *Knowledge-Based Intelligent Engineering Systems and Allied Technologies, 2000. Proceedings. Fourth International Conference on*. p. 117–120. IEEE.

Zhang, W., S. Shan, W. Gao, X. Chen, and H. Zhang. 2005. "Local gabor binary pattern histogram sequence (lgbphs): A novel non-statistical model for face representation and recognition". In *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*. p. 786–791. IEEE.

Zhao, W., R. Chellappa, P. J. Phillips, and A. Rosenfeld. 2003. "Face recognition: A literature survey". *Acm Computing Surveys (CSUR)*, vol. 35, n° 4, p. 399–458.

Zimmermann, A., A. Lorenz, and R. Oppermann. 2007. An operational definition of context. *Modeling and using context*, p. 558–571. Springer.

Zou, J., Q. Ji, and G. Nagy. 2007a. "A comparative study of local matching approach for face recognition". *Image Processing, IEEE Transactions on*, vol. 16, n° 10, p. 2617–2628.

Zou, X., J. Kittler, and K. Messer. 2007b. "Illumination invariant face recognition: A survey". In *Biometrics: Theory, Applications, and Systems, 2007. BTAS 2007. First IEEE International Conference on*. p. 1–8. IEEE.