Face Recognition's Grand Challenge: uncontrolled conditions under control

Bas Boom

Face Recognition's Grand Challenge: uncontrolled conditions under control

UNIVERSITY OF TWENTE.

Bas Boom

Signals and Systems Group

University of Twente

De promotiecommissie:

voorzitter en secretaris:	
Prof.dr.ir. A.J. Mouthaan	Universiteit Twente
promotor:	
Prof.dr.ir. C.H. Slump	Universiteit Twente
assistent promotors:	
dr.ir. R.N.J. Veldhuis	Universiteit Twente
dr.ir. L.J. Spreeuwers	Universiteit Twente
referent:	
dr. M. Brauckmann	L1 Identity Solutions
leden:	
Prof.dr.ir. A. Stein	Universiteit Twente
Prof.dr. M. Junger	Universiteit Twente
Prof. L. Akarun	Bogazic University Istanbul
Prof.dr. R.C.Veldkamp	Universiteit Utrecht

CTIT Dissertation Serie No. 10-185
Center for Telematics and Information Technology (CTIT)
P.O Box 217 - 7500AE Enschede - the Netherlands
ISSN: 1381-3617

Signals & Systems group,
EEMCS Faculty, University of Twente
P.O. Box 217, 7500 AE Enschede, the Netherlands

Printed by Gildeprint B.V., Enschede, The Netherlands
Typesetting in LaTeX2e
Images on the cover are from the FRGC database and www.hfs-info.com

ISSN 1381-3617, No. 10-185
ISBN 978-90-365-2987-7
DOI 10.3990/1.9789036529877

FACE RECOGNITION'S GRAND CHALLENGE: UNCONTROLLED CONDITIONS UNDER CONTROL

DISSERTATION

to obtain
the degree of doctor at the University of Twente,
on the authority of the rector magnificus,
prof.dr. H. Brinksma,
on account of the decision of the graduation committee,
to be publicly defended
on Friday the 3th of December 2010 at 16.45.

by

Bastiaan Johannes Boom
born on the 25th of April 1981
in Rotterdam, The Netherlands

Dit proefschrift is goedgekeurd door:

De promotor:	Prof.dr.ir. C.H. Slump
De assistent promotors:	dr.ir. R.N.J. Veldhuis
	dr.ir. L.J. Spreeuwers

CONTENTS

1 Introduction **1**
1.1 Camera Surveillance . . . 2
1.1.1 Social Opinion . . . 3
1.1.2 Legal Aspects . . . 3
1.1.3 Users . . . 4
1.1.4 Scenarios . . . 4
1.1.5 Characteristics of CCTV systems . . . 5
1.2 Biometrics . . . 6
1.2.1 Face as Biometric . . . 6
1.2.2 Example of Face Recognition . . . 7
1.2.3 Terminology . . . 8
1.2.4 Face Recognition System . . . 10
1.2.5 Requirements . . . 12
1.3 Purpose of our research . . . 13
1.4 Contributions . . . 14
1.5 Outline of Thesis . . . 14

2 Face Recognition System **17**
2.1 Introduction . . . 17
2.2 Face Detection . . . 18
2.2.1 Foreground and Background Detection . . . 19
2.2.2 Skin color Detection . . . 20
2.2.3 Face Detection based on Appearance . . . 20
2.2.4 Combining Face Detection Methods . . . 23
2.3 Face Registration . . . 24
2.3.1 The Viola and Jones Landmark Detector . . . 25
2.3.2 MLLL and BILBO . . . 25
2.3.3 Elastic Bunch Graphs . . . 26
2.3.4 Active Shape and Active Appearance Models . . . 27
2.4 Face Intensity Normalization . . . 28
2.4.1 Local Binary Patterns . . . 29
2.4.2 Local Reflectance Perception Model . . . 30
2.4.3 Illumination Correction using Lambertian reflectance model 30
2.5 Face Comparison . . . 31

2.5.1 Holistic Face Recognition Methods 31
2.5.1.1 Principle Component Analysis 32
2.5.1.2 Probablistic EigenFaces 32
2.5.1.3 Linear Discriminant Analysis 33
2.5.1.4 Likelihood Ratio for Face Recognition 34
2.5.1.5 Other Subspace methods 35
2.5.2 Face Recognition using Local Features 35
2.5.2.1 Elastic Bunch Graphs 35
2.5.2.2 Adaboost using Local Features 36

I Resolution 39

3 The effect of image resolution on the performance of a face recognition system 41
3.1 Introduction 41
3.2 Face Image Resolution 42
3.3 Face Recognition System 43
3.3.1 Face Detection 43
3.3.2 Face Registration and Normalization 43
3.3.2.1 MLLL 43
3.3.2.2 BILBO 44
3.3.2.3 Face Alignment 44
3.3.2.4 Face Normalization 44
3.3.3 Face Recognition 44
3.4 Experiments and Results 44
3.4.1 Experimental Setup 44
3.4.2 Experiments 45
3.4.2.1 Face Recognition 45
3.4.2.2 Face Registration 45
3.4.2.3 Face Registration and Recognition 46
3.4.2.4 Face Recognition using erroneous landmarks . . . 46
3.4.3 Results 46
3.4.3.1 Face Recognition (Experiment 1) 46
3.4.3.2 Face Registration (Experiment 2) 48
3.4.3.3 Face Recognition and Registration (Experiment 3) 50
3.4.3.4 Face Recognition by using erroneous landmarks (Experiment 4) 50
3.5 Conclusion 51

Conclusion Part I 53

II Registration **55**

4 Automatic face alignment by maximizing the similarity score **57**
4.1 Introduction . . . 57
4.2 Matching Score based Face Registration . . . 58
4.2.1 Face Registration . . . 58
4.2.2 Search for Maximum Alignment . . . 59
4.2.3 Face Recognition Algorithms . . . 60
4.3 Experimental Setup . . . 61
4.4 Experiments . . . 61
4.4.1 Comparison between recognition algorithms . . . 62
4.4.2 Lowering resolution . . . 63
4.4.3 Training using automatically obtained landmarks . . . 64
4.4.4 Improving maximization . . . 65
4.4.4.1 Using a different start simplex . . . 65
4.4.4.2 Adding noise to train our registration method . . . 66
4.5 Conclusion . . . 66

5 Subspace-based holistic registration for low resolution facial images **69**
5.1 Introduction . . . 69
5.2 Face Registration Method . . . 71
5.2.1 Subspace-based Holistic Registration . . . 71
5.2.2 Evaluation . . . 73
5.2.2.1 Evaluation to a user specific face model . . . 73
5.2.2.2 Using edge images to avoid local minima . . . 74
5.2.3 Alignment . . . 75
5.2.4 Search Methods . . . 75
5.2.4.1 Downhill Simplex search method . . . 75
5.2.4.2 Gradient based search method . . . 76
5.3 Experiments . . . 76
5.3.1 Experimental Setup . . . 77
5.3.1.1 Face Database . . . 77
5.3.1.2 Face Detection . . . 77
5.3.1.3 Low Resolution . . . 78
5.3.1.4 Face Recognition . . . 78
5.3.1.5 Landmark Methods for Comparison . . . 79
5.3.2 Experimental Settings . . . 80
5.4 Results . . . 81
5.4.1 Comparison with Earlier Work . . . 81
5.4.2 Subspace-based Holistic Registration versus Landmark based Face Registration . . . 82
5.4.3 User independent versus User specific . . . 86
5.4.4 Comparing Search Algorithms . . . 86
5.4.5 Lower resolutions . . . 88
5.5 Conclusion . . . 89

5.6 Appendix: Gradient based search method 89

Conclusion Part II **91**

III Illumination 93

6 Model-based reconstruction for illumination variation in face images **95**
6.1 Introduction 95
6.2 Method 96
6.2.1 Lambertian model 96
6.2.2 Overview of our correction method 97
6.2.3 Learning the Face Shape Model 97
6.2.4 Shadow and Reflection Term 98
6.2.5 Light Intensity 98
6.2.6 Estimation of the Face Shape 99
6.2.7 Evaluation of the Face Shape 99
6.2.8 Calculate final shape using kernel regression 100
6.2.9 Refinement 101
6.3 Experiments and Results 101
6.3.1 Face databases for Training 101
6.3.2 Determine albedo of the Shape 102
6.3.3 Face Recognition 102
6.3.4 Yale B database 103
6.3.5 FRGCv1 database 104
6.4 Conclusion 105

7 Model-based illumination correction for face images in uncontrolled scenarios **107**
7.1 Introduction 107
7.2 Illumination Correction Method 108
7.2.1 Phong Model 108
7.2.2 Search strategy for light conditions and face shape 109
7.2.3 Estimate the light intensities 109
7.2.4 Estimate the initial face shape 110
7.2.5 Estimate surface using geometrical constrains and a 3D surface model 110
7.2.6 Computing the albedo and its variations 111
7.2.7 Evaluation of the found parameters 111
7.3 Experiments and Results 111
7.3.1 3D Database to train the Illumination Correction Models . 112
7.3.2 Recognition Experiment on FRGCv1 database 112
7.4 Discussion 114
7.5 Conclusion 114

8 Combining illumination normalization methods **115**
8.1 Introduction . . . 115
8.2 Illumination normalization . . . 116
8.2.1 Local Binary Patterns . . . 116
8.2.2 Model-based Face Illumination Correction . . . 117
8.3 Fusion to improve recognition . . . 117
8.4 Experiments and Results . . . 118
8.4.1 The Yale B databases . . . 119
8.4.2 The FRGCv1 database . . . 120
8.5 Conclusions . . . 121

9 Virtual Illumination Grid for correction of uncontrolled illumination in facial images **123**
9.1 Introduction . . . 123
9.2 Method . . . 125
9.2.1 Reflectance model . . . 125
9.2.2 Face Shape and Albedo Models . . . 127
9.2.3 Illumination Correction Method . . . 128
9.2.4 Estimation of the illumination conditions . . . 129
9.2.5 Estimation of the crude face shape . . . 130
9.2.6 Estimation of the surface . . . 130
9.2.7 Estimation of the albedo . . . 131
9.2.8 Evaluation of the obtained illumination conditions, surface and albedo . . . 131
9.2.9 Refinement of the albedo . . . 132
9.3 Experiments . . . 133
9.3.1 Training VIG . . . 133
9.3.2 Experimental Setup . . . 134
9.3.3 Face Recognition Results on CMU-PIE database . . . 134
9.3.4 Face Recognition Results on FRGCv2 database . . . 136
9.3.5 Fusion . . . 137
9.4 Discussion . . . 139
9.4.1 Limitations . . . 139
9.4.2 Accuracy of the Depth Maps . . . 139
9.5 Conclusions . . . 141

Conclusion Part III **142**

10 Summary & Conclusions **143**
10.1 Summary . . . 143
10.2 Conclusions . . . 145
10.3 Recommendations . . . 147

References **149**

Samenvatting **159**

Dankwoord **161**

1

INTRODUCTION

The number of cameras increases rapidly in squares, shopping centers, railway stations and airport halls. There are hundreds of cameras in the city center of Amsterdam as is shown in Figure 1.1. This is still modest compared to the tens of thousands of cameras in London, where citizens are expected to be filmed by more than three hundred cameras of over thirty separate Closed Circuit Television (CCTV) systems in a single day [85]. These CCTV systems include both publicly owned systems (railway stations, squares, airports) and privately owned systems (shops, banks, hotels). The main purpose of all these cameras is to detect, prevent and monitor crime and anti-social behaviour. Other goals of camera surveillance can be detection of unauthorized access, improvement of service, fire safety, etc. Since the terrorist attack on 9/11, detection and prevention of terrorist activities especially at high profiled locations such as airports, railway stations, government buildings, etc, has become a new challenge in camera surveillance. In order to process all the recording from CCTV systems, smart solutions are necessary. It is unthinkable that human observers can watch all camera views and analyzing the surveillance footage afterward is a time consuming task. So the great challenge is the automatic selection of interesting recordings. For instance, focussing on well-known shoplifters instead of the shop owner behind the counter. In these cases, the identity of a person gives important information about the relevance of the scene. In order to establish the person's identity, camera surveillance can be combined with automatic face recognition. This allows us to search for possible well-known

offenders automatically. Combining face recognition with CCTV systems is difficult, because of the low resolution of recordings and the changing appearance of faces through different scenes. This research focusses on solving some of the fundamental technical problems, which arise when performing face recognition on video surveillance footage. In order to solve these problems, techniques from research on computer vision, image processing and pattern classification are used. These techniques are used to identify a person based on unique biological or behavioral characteristics (biometrics). In this case, the biometric is the face, other famous examples of biometrics are fingerprint and the iris. To recognize the face in surveillance footage, we investigate effects which resolution and illumination can have on existing face recognition systems. We also developed technical methods to improve the recognition rates under these conditions.

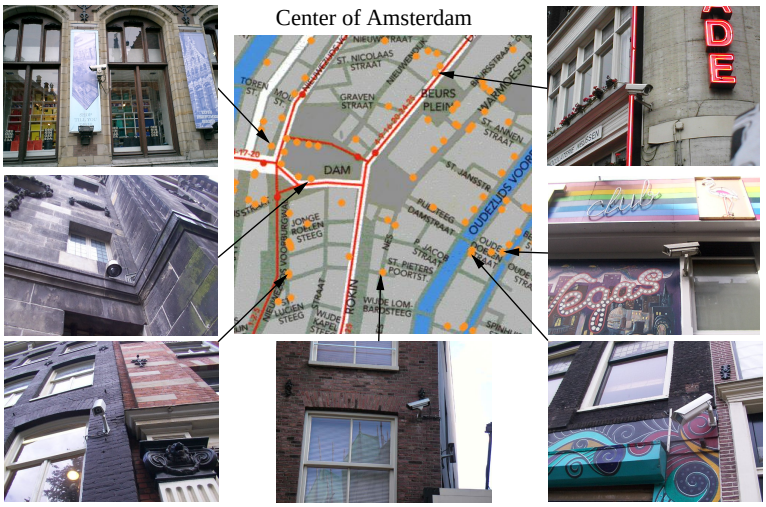


Figure 1.1: CCTV systems in the City Center of Amsterdam - The Orange Dot's are locations of camera which make recordings of the public streets of Amsterdam, from Spot the Cam (www.spotthecam.nl)

1.1 Camera Surveillance

Camera surveillance is conceptually more than a monitor connected to some cameras. Camera surveillance can be seen as a powerful technology to monitor and control social behaviour. This raises concerns on multiple levels, which are discussed in Section 1.1.1, where we also mention the public opinion. Camera surveillance is regulated by laws, which we summarize in Section 1.1.2. In Section 1.1.3 and 1.1.4, we describe the users of CCTV systems and categorize several different scenarios for camera surveillance. Finally, we determine the characteristics of CCTV systems and provide technical details that can be relevant for face recognition.

1.1.1 Social Opinion

The increased use of camera surveillance is partly caused by the public. "People feeling unsafe" is one of the reasons of the city of Amsterdam for installing CCTV systems [47]. In other Dutch local governments, citizens even requested camera surveillance to secure their neighborhoods [56]. In a large European investigation on the subject of camera surveillance, citizens were asked several questions concerning privacy. In this research [53], two thirds agree with the statement: "who has nothing to hide, has nothing to fear from CCTV". On the other hand, more than half of these citizens believed that recordings of CCTV systems can be misused and 40% believe that it invades their privacy. Concerning the most common goal of camera surveillance, namely the prevention of serious crime, 56% doubted that camera surveillance really works.
The acceptance of camera surveillance depends heavily on the location. Most people support camera surveillance in banks, railway stations, shopping malls, but people draw the line by camera surveillance in changing rooms and public toilets, and also outside the entrance of their homes. Another important related issue is the persons who have access to the footage. In most countries people agree that the police should be able to watch the recordings, but other access for instance by media or for commercial interests should be restricted.
The investigation in [53] shows that many people overestimate the technological state of camera surveillance. 36% of the people believed that most CCTV systems are able to make close up images of their faces and 29% believe that most CCTV systems have integrated automatic face recognition. Although these ideas are commonly used in television series like Crime Scene Investigation, Bones and NCIS, a survey of the used CCTV systems in European cities shows that most CCTV systems are small, isolated and not technologically advanced.

1.1.2 Legal Aspects

In Europe, article 8 (right to privacy) of the European Convention of Human Rights plays an important role for camera surveillance. Furthermore, the Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002, concerning the processing of personal data and the protection of privacy in the electronic communications sector applies to recordings of CCTV systems, making it illegal to disclose pictures captured from CCTV systems. This Directive, however, makes an exception for purposes of national security, public safety and criminal investigations. The precise implementation into national laws differs for every European country. Next to regulations, some countries, like the UK, have published guidelines [35] for CCTV systems, showing also clearly the obligations under the law. Even though these laws are in place, CCTV systems are often in violation of the national data protection acts. From [53], we know that one out of two CCTV systems is not notified by signage. In [53], it is also reported that there are often problems with the responsibility and ownership of the CCTV systems, especially in case of the smaller systems.

1.1.3 Users

We already mentioned some of the different uses of the camera surveillance. This also means that we have various user groups in camera surveillance. For example, a CCTV system in a shopping mall is used by a security officer to detect shoplifters. However, the police can afterward request the recordings as evidence of this theft. The users of other camera surveillance scenarios, like local government, banks, shops, etc, all differ a lot. For this reason, we choose to make a distinction in the manner in which the system is used. Our previous example shows that there are two kinds of users:

- **Active surveillance users** are for instance the security officer who looks for offences and guides other officers based on his observations.
- **Re-active surveillance users** can be the police asking for evidence. In this case, an action is taken in reaction on certain events, where the recordings are studied afterwards, for evidence or more information.

At this moment, most systems only support having a re-active surveillance user [53]. For this reason, our research focusses more on the re-active surveillance. Note that the requirements of the two users are different, allowing us in this case to ignore the real-time requirements that are necessary in case of active use of surveillance systems. However, problems in re-active surveillance are far from solved. Searching through CCTV footage is still manual labour, especially if suspicions or the suspect are not well defined.

1.1.4 Scenarios

Today's CCTV systems are used in various institutional settings, like in shops, banks, ATM's, railway/metro stations/airports, streets, metro's/buses/trains, highway patrol, building security, hospital, etc. We distinguish three global scenarios, which cover most of the CCTV systems, namely:

- **Overview Scenario:** In this case, a camera is installed in such manner that it observes as much of the surrounding as possible. A common example is cameras in public streets, for the detection of criminal behaviour. In this case, a clear picture of the body of the subject is more important than a picture of the face. The disadvantage of this scenario is that the facial resolution is usually very low. Also occlusions and extreme pose of the face occur in these recordings.
- **Entrance Scenario:** At entrances of government buildings, shops or stations, cameras are installed, which are more tuned to person identification. A nice property of most entrances is that they only allow a few people to enter at the same time. This also allows us to focus the camera, giving us a higher facial resolution. Surveillance cameras near entrances also record frontal face images, because the viewing direction is usually similar to the walking direction.

- **Compulsory Scenario:** In the compulsory scenario, the person has to look into a camera because it is necessary or polite. An obvious example is an ATM which contains a camera in the monitor. People have to look at the monitor to input their PIN code, which usually gives the security camera nice frontal facial images with a high resolution. Another example is a cash deck, where people have to pay for products and look in the direction of the cashier.

We decide to focus our research efforts more on the last two scenarios. In these scenarios remain enough challenges especially if we compare this with access control. In the case of access control, the person cooperates with the system by looking into the camera, giving the system a second attempt if the first fails. In the case of camera surveillance person rather avoid camera and have no benefit in being recorded. In many of the institutional settings, the overview scenario is combined together with an entrance scenario or a mandatory scenario. In these cases, the face recognition is usually performed with the higher resolution facial images obtained from the last two scenarios.

1.1.5 Characteristics of CCTV systems

It is important to determine the characteristics of camera surveillance in order to perform automatic face recognition on CCTV systems. This can provide insight into possible problems. Although face recognition is a promising technique for person identification, it is still far from perfect. On high resolution mugshot images, computers nowadays outperform humans, as is shown in the Face Recognition Grand Challenge [89]. However, in more uncontrolled situations, automatic face recognition has failed to achieve reliable results in multiple occasions (for example at the airports in Dallas, Fort Worth, Fresno and Palm Beach County, [132]). For this reason, we performed an assessment of the expected problems in face recognition for CCTV systems, from which we conclude that the following reasons might cause problems in face recognition:

- **Quality of the Recordings:** The research of [53] shows that most CCTV systems are far from advanced. CCTV systems often have a symbolic use rather than performing permanent and exhaustive surveillance. Although video recordings are taken, the number of frames is often low due to the camera's or limited storage.
- **Face Resolution:** A well-known issue of camera surveillance footages is the image resolution. Because the resolution of the recordings is often low, the regions containing the face contain a small number of pixels (32×32). This has consequences on the performance of the overall face recognition, making accurate recognition extremely difficult.
- **Illumination Conditions:** Although humans hardly have problems with changing illumination conditions, in computer vision, this is still largely an unsolved problem. Due to illumination, the appearance of a face changes dramatically making face comparison very difficult.

- **Poses:** The face is a 3D object, which is able to rotate in different directions. Although computers are really good in the classification of mugshots, like on a passport, comparing frontal face images with images of faces with different poses is extremely difficult, because of occlusions and registration problems.
- **Occlusions:** In the previous problem, we already mention the occlusions due to poses, but there are various other reasons for occlusions, like caps, sun glasses, scarfs, etc. This makes face recognition difficult because important features are sometimes missing.

1.2 Biometrics

The automatic identification of humans based on their appearance becomes increasingly popular. Secure entrances based on fingerprints, iris or face become accepted by the public and are sometimes even obliged (for instance to enter the USA). Furthermore, the use of biometrics increases, where in the past only police used fingerprints to trace criminals, nowadays fingerprints can also be used to access, for instance, a laptop. There are many kinds of biometrics, e.g. fingerprint, iris, face (2D or 3D), DNA, hand, speech, signatures, gait, ear, etc. The most popular biometrics are fingerprint, iris and face. One of the main reasons that fingerprint and iris recognition are popular is the accuracy of the authentication. This is mainly because the appearance of the biometric is very stable. But like in every biometric, the appearance slightly changes at every recording, so robust methods are necessary to deal with these changes.

1.2.1 Face as Biometric

In comparison with fingerprint and iris recognition, the appearance of the face is very unstable (see section 1.1.5), making it difficult to achieve a good accuracy in authentication. But unlike most other biometrics, a face can be captured without the cooperation of the person. For humans, faces are also the most common biometric to identify other humans. For this reason, several official documents, like the drivers license and passport, contain a facial image. In most western societies, covering your face is not accepted and usually associated with an immediate assumption of guilt [62]. For these reasons, faces are popular as a biometric, although the accuracy of identification is lower in comparison to fingerprint and iris recognition.
A big disadvantage of face recognition is that identity theft is also very easy. Making a photograph of a person without being noticed is not difficult, while for many other biometrics, anonymous retrieval of biometric data requires specialized equipment. This makes face recognition not the most suitable biometric for access applications. But in the case of camera surveillance, face recognition is one of the few biometrics that can be used. Furthermore, human observers can easily verify the automatic findings.

1.2.2 Example of Face Recognition

Based on research performed in [54], we introduce face recognition by means of an example. We show that face recognition is far from simple, also for a human observer. From [89], we already know that humans perform worse on uncontrolled frontal images, while in [54] research was performed on the capability to identify persons in CCTV recordings. In order to evaluate the human performance, a robbery was staged and recorded with both a CCTV camera and broadcast-quality footage. We show the two robbers in broadcast-quality footage from [54] in Figure 1.2. Notice that CCTV systems usually produce facial images with far lower

Figure 1.2: Footage of a Robbery (Robber 1 and Robber 2) - Face images of two robbers of a staged robbery, left Robber 1 and right Robber 2, (from [54] with permission)

resolutions. In order to select the criminal, a line-up is arranged, which is similar to a gallery in face recognition. Figure 1.3 shows the line-up with the question: "is one of the depicted faces the robber's?" In this example, Robber 1 is person 8 of the line-up. After viewing the broadcast-quality footage, 60% of the persons pick the correct face, 13% pick another face and 27% thought that the face was not present in the line-up. In Figure 1.4, the line-up of robber 2 is shown, where person 5 is robber 2. In this case, 83% recognized the person, 10% picked an incorrect person and 7% thought he was not in the line-up. By leaving the correct image out of the line-up in case of robber 1, 47% was correct and 53% selected another face. The experiment with the footage from a real CCTV system in [54] shows even worse results are achieved, where 21% selected the correct person in the case of robber 1 and 19% in the case of robber 2. We think that this experiment shows how difficult face recognition truly is, even for humans. This example also gives

an impression of the application to which our research contributes.

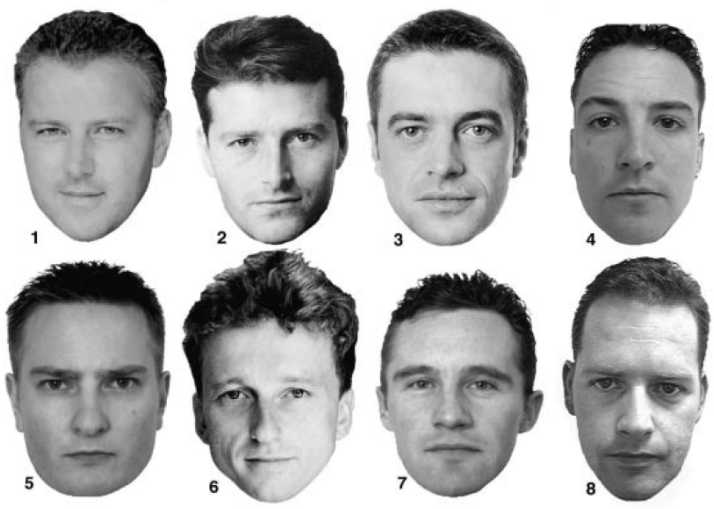


Figure 1.3: Gallery for Robber 1 - Does one of the faces belong to robber 1, if so which face? (from [54] with permission)

1.2.3 Terminology

Based on the previous example, we will now introduce some of the terminology used in biometrics.
In Figure 1.2, we show two images of the surveillance footage, which are called probe images. Because we evaluate our face recognition system usually on thousands of images, we denote this set of images as probe/query/test set. The images in Figures 1.3 and 1.4 are called gallery images, where the set of images is denoted by gallery/target/enrollment set. In order for a face recognition system to learn the appearance of a face, a training set is used to make a model of the face. In order to perform fair experiments, there should be no overlapping images between the training set and the target/test set.
The purpose of face identification is to determine the identity of the person in the probe images based on the gallery images. There is open-set identification, which is the general case, assuming that the person in the probe image can be in the gallery set, but it is also possible that the person is not present in the gallery set. In closed-set identification, the person is always present in the gallery set. Next to face identification, there is also face verification. In this case, there is an identity claim of the user and we only have to verify if this claim is correct. We compare the probe image with only one image from the gallery. If the person identity claim is correct, we call it a genuine user, while if the person claims to be someone else, he is an impostor.

Figure 1.4: Gallery for Robber 2 - Does one of the faces belong to robber 2, if so which face? (from [54] with permission)

In face verification, there are two kinds of errors. For instance, we can claim that robber 2 is person 5 of the line-up. This is a genuine attempt, so the face recognition system can either accept or reject the claim, where rejection is the first error. We can also claim that robber 2 is person 3 of the line-up, this is not true so it is an impostor attempt, making acceptance of this claim the second error. In Table 1.1, we show the four different outcomes of the face recognition system. Most face recognition systems assign a similarity score to a certain probe image,

	Genuine Attempt	**Imposter Attempt**
Claim accepted	True Positive	False Positive
Claim rejected	False Negative	True Negative

Table 1.1: Confusion Matrix - The four different outcomes of a face recognition system, namely True Positive, False Positive, False Negative, True Negative

which indicates the confidence that the face is of the same person. For the final decision, a threshold is used to separate the genuine and impostor claims. Based on the similarity score, we make a graph showing the probability densities of both genuine and imposter attempts (see Figure 1.5). Because face recognition is a difficult problem as concluded in the previous section, the densities of genuine and imposter overlap. This means that there are usually some incorrect classifications, wherever the threshold is placed. The False Reject Rate (FRR) is the fraction of genuine attempts which is below the threshold and thus erroneously rejected. The False Accept Rate (FAR) is the fraction of imposter attempts which succeed.

Access to a vault of a bank, requires a low FAR, while a higher FRR is acceptable. The higher FRR results in more false alarms, alarming for instance a security guards. There are also scenarios where a very low FRR is necessary. In the case of Grip Pattern recognition on a police gun [106], a high FRR means that the police gun might refuse to fire creating a life threatening situation for the police officer. In Figure 1.5, we show the Detection Error Tradeoff (DET), which is very similar to a Reciever Operating Characteristic (ROC). Instead of using the FRR, the verification rate is also often used, which is $(1 - FRR)$. These curves give the relation between FRR and FAR for all thresholds. In order to compare face recognition methods, multiple ROCs are plotted, where the best curve is the one which is closest to the axis. To summarize the performance of the face recognition system by a single number, the Equal Error Rate (EER) is often used, which is the point where the FAR and FRR are equal. Another possibility is to measure the performance for a certain FAR or FRR, depending on the system requirements. The Face Recognition Grand Challenge [89] measures the performance in Verification Rate at 0.1% FAR, which basically means the number of genuine users that are correctly classified if 1 out of 1000 impostor attempts succeed.

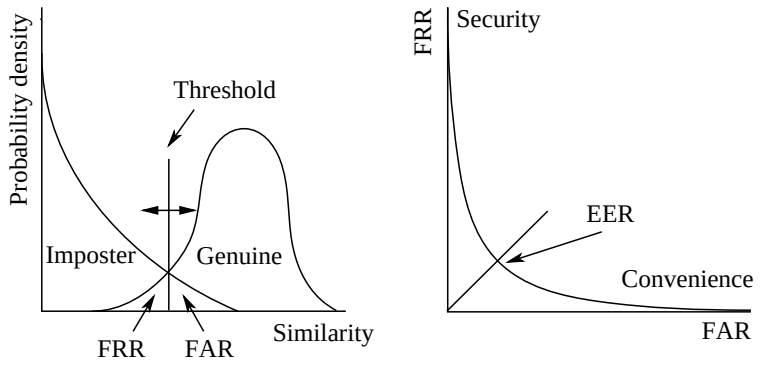


Figure 1.5: Explaination on the DET - Left: Probability densities of genuine and imposter scores, Right: The DET curve

1.2.4 Face Recognition System

An automatic face recognition system has to perform several tasks to successfully recognize a face. Although face recognition can be organized in several ways [63; 142], we defined the following four components: face detection/localization, face registration, face intensity normalization and face comparison/recognition. Face Detection determines the location of the faces if present in the image or video. The Face Registration aims to more accurately localize the face or landmarks in the face (eyes, nose, mouth, etc), allowing faces to be aligned to a common reference coordinate system. During the face registration, we perform geometrical transformation to the face image in order to make the comparison easier. Next

to the geometrical transformation, radiometrical transformation has to be applied normalizing the intensity values of the images for the camera setting and the illumination variations in images. The correction for these effects is performed by the Face Intensity Normalization component, making for instance faces under different illumination conditions more comparable. The Face Comparison (often also called Face Recognition) compares the face against the gallery of faces, which results usually in similarity scores. Based on the similarity scores, we can determine if the face is found in the gallery. A schematic representation of an automatic face recognition system is shown in Figure 1.6. Although we make a clear distinction

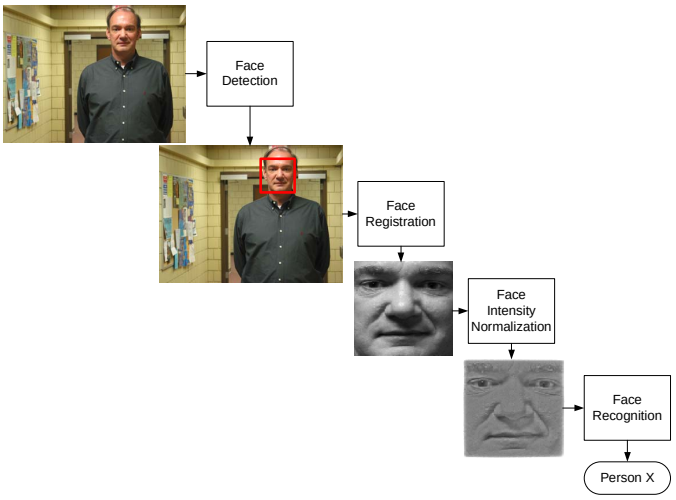


Figure 1.6: Schematic Representation of Face Recognition System - consisting of four components: Face Detection, Face Registration, Face Intensity Normalization and Face Comparison

between the four components, as can be observed from Figure 1.6, the different components can overlap each other. Our registration method discussed in chapters 4 and 5 is an example of a method, where different components overlap each other. This registration method finds the best geometrical transformation by optimizing similarity score of the face comparison. In this case, the face registration component uses the face comparison component, causing overlap between the components in the system. A component in a face recognition system also depends on the previous components. For instance, if the face detection fails, the other components in the face recognition system are not executed. Another example is that if face registration aligns a face incorrect, the face intensity normalization can fail because it expects to normalize the pixels of the eye, but instead normalizes the pixels belonging to the nose.

1.2.5 Requirements

Most face recognition systems are tested for the application of access control. An example is the access control at airports, where the identity of the person is verified with the photograph on his passport. The access control application has clearly defined requirements. For instance, a FRR of 1% at FAR of 0.1% has to be achieved. In this case, the FRR indicates the fraction of genuine persons who are falsely refused access causing inconvenience. The FAR in access control is the fraction of persons that enter with a false identity claim, which is a security risk. Camera surveillance differs from access control, because it is based on a blacklist. This blacklist contains the suspects who need to be recognized. In this case, the FRR indicates the fraction of suspects that are not caught, implying that the verification rate gives the probability that a suspect will be caught. Here, the FAR is the fraction of persons who are falsely recognized as suspects. In camera surveillance, the FAR quantifies the inconvenience, because in the case of a false accept, a security officer has to examine the identity of the person.
In Section 1.2.3, we have shown that face recognition in CCTV footage is a difficult task for humans. Although we admit that computers cannot perform perfect recognition, we believe that computers can support humans in narrowing the search through CCTV footage. For example, consider a security officer who has to monitor multiple entrances. In practice, the probability that he detects a suspect at an entrance is small. Now, we added a face recognition systems which has an FAR rate of 1% and a verification rate of 60%. Out of the hundred persons entering the building, one person gives a false alarm. In this case, the security officer can make a decision based on earlier recordings of both the possible suspect and CCTV footage. This changes the role of the human in the system making him the specialist. An advantage of this approach is that human observer can usually look beyond face identification, toward other behavioral characteristics which are difficult to determine for computers. For suspects, there is already a large risk (6 out of 10) that they will be recognized, which is probably better than the a single security officer looking at the monitors. It is difficult to define a FRR and FAR for camera surveillance, because it depends on the scenario. To illustrate this, we defined a couple of common CCTV systems where face recognition can be used:

- Small Shop: In a small shop, a camera system, that detects suspects, can help the owner of the shop to focus on known shoplifters. A FRR $\leq 50\%$ at FAR of 0.1% can already be sufficient. In this case, the person behind the counter can watch certain suspects more closely. This also deters suspects from entering the shop, because there is a large probability that they get caught.
- Shopping mall: In a shopping mall, the number of persons that enter increases in comparison with a single shop. On the other hand, there are usually security officers, who have the duty to detect criminal behavior. In this case, a face recognition system with a FRR $\leq 50\%$ at FAR of 0.1% can already be sufficient, where we reduce the FAR so that an officer takes a recognition of a suspect still serious.
- Police searching in CCTV recordings: The police usually wants to be sure

if a criminal is in recordings, instead of setting the FAR, which shops find important because they do not want to cause too much inconvenience. The police will set the FRR at around 95% to be sure that most suspects are found. They can increase the FRR by shifting the threshold if the results are not satisfying. The disadvantage in this case is that there is big increase in FAR, but this is always better than searching through all the recordings without face recognition software.

We have shown that different CCTV systems have different requirements. For this reason, the more global goal, which benefits all CCTV systems, is to achieve improvements in the ROC curves, focussing on faces recorded under uncontrolled conditions.

1.3 Purpose of our research

Automatic face recognition solves the difficult task of recognizing a person based on the appearance in an image. In order to perform face recognition, several additional components like registration and intensity normalization are necessary. We have investigated all components of the face recognition system for the application of camera surveillance. In order to improve a face recognition system for this application, we looked at specific problems which arise in camera surveillance. In Section 1.1.5, we already discussed the specific characteristics that cause problems for the face recognition (e.g. low resolution, illumination, pose and occlusions). This gives the following general research question:

- Can we improve face recognition for camera surveillance by optimizing the components mentioned in Section 1.2.4 for the application specific characteristics defined in Section 1.1.5?

In order to answer this question, insight in both the components of face recognition system and the effects that application specific characteristics have, are necessary. We choose to look at different characteristics separately, where we focus on the relative low resolutions of facial images and the varying illumination conditions in facial images. In this case, we ask the following specific questions:

- What is the effect of both low resolution and illumination on the different components (Face Detection, Face Registration, Face Intensity Normalization and Face Comparison) of the face recognition system?
- Which measures can be taken to improve the face recognition system for low resolution facial images and images captured under uncontrolled illumination conditions?
- How much improvement of the face recognition performance is obtained with the before mentioned measures?

1.4 Contributions

In this thesis, our aim is to improve face recognition in CCTV system. In order to achieve this, we examine the entire chain necessary to perform face recognition. Our contribution consists of extensive research on the performance of face recognition systems for camera surveillance applications. Our contribution can be divided in three parts:

- **Resolution:** One of the most well-known problems in CCTV recording is the low resolution of the facial images. Although multiple investigations are performed on the effect that resolution has on the face comparison component, no research was performed on the effect that resolution has on the other components in the system. In our investigation on the effects resolution has on the face recognition system, we look at the entire face recognition system. This also shows the effects on the face registration, that influence the final results even more than the face comparison.
- **Registration:** An important step in the face recognition system is the face registration. Face registration on high resolution images is often performed by landmark finding methods. These methods, however, become less accurate or fail if the face resolution decreases. We have developed an holistic registration methods for low resolution face images. The accuracy of this face registration method is better than the landmark based registration methods, while it also achieves a better accuracy on lower resolutions.
- **Illumination:** Faces recorded in uncontrolled conditions contain illumination variations, which often cause large variations in the appearance. These variations are often larger than the difference in appearance between persons. In the literature, different illumination correction methods are developed that partially solve these problems. These methods are usually tested on face images recorded in laboratory conditions. Our focus is on the uncontrolled conditions and on correcting the illumination in these images by modelling the illumination. For this reason, we investigate both local and global correction methods and combine their strengths. We have also developed our own illumination correction methods focusing on common problems that we discovered in faces illuminated under uncontrolled conditions, like ambient illumination and multiple light sources.

1.5 Outline of Thesis

Next to the introduction, this thesis consist of a general introduction in automatic face recognition systems, three main parts and the conclusions. The three parts contain our contributions in resolution, registration and illumination. These parts begin with a general introduction which is followed by one or multiple chapters which contain the published or submitted papers and we finish these parts with a final section which contains the general conclusions of that part. In the introductions of the parts, we discuss the reason to investigate certain subjects in more

detail. Furthermore, these introductions describe the relationship between the underlying chapters. The conclusion of the parts discuss our contribution on the different subjects and place this in global context of face recognition in camera surveillance. This thesis contains the following chapters:
In Chapter 2, we give a general introduction in automatic face recognition systems. This introduction gives an overview of the four different components: face detection, face registration, face normalization and face comparison. We discuss several methods that we used throughout out the thesis for these components.
In Part I (Chapter 3), we investigate the effect of image resolution on the error rates of a face verification system. We do not restrict ourselves to the face comparison methods only, but we also consider the face registration. In our face recognition system, the face registration is done by finding landmarks in a face image and subsequent alignment based on these landmarks. To investigate the effect of image resolution we performed experiments where we varied the resolution. We investigate the effect of the resolution on the face comparison component, the registration component and the entire system. This research also confirms that accurate registration is of vital importance to the performance of the face recognition [21].
In Part II (Chapter 4), we propose a face registration method which searches for the optimal alignment by maximizing the score of a face recognition algorithm, because accurate face registration is of vital importance to the performance of a face recognition algorithm. We investigate the practical usability of our face registration method. Experiments show that our registration method achieves better results in face verification than the landmark based registration method. We even obtain face verification results which are similar to results obtained using landmark based registration with manually located eyes, nose and mouth as landmarks. The performance of the method is tested on the FRGCv1 database using images taken under both controlled and uncontrolled conditions [22; 26].
In Part II (Chapter 5), subspace-based holistic registration is introduced as an alternative to landmark based face registration, which has a poor performance on low resolution images, as obtained in camera surveillance applications. The proposed registration method finds the alignment by maximizing the similarity score between a probe and a gallery image. This allows us to perform a user independent as well as a user specific face registration. The similarity is calculated using the probability that the face image is correctly aligned in a face subspace, but additionally we take the probability into account that the face is misaligned based on the residual error in the dimensions perpendicular to the face subspace. We evaluated the registration methods by performing several face recognition experiments on the FRGCv2 database. Subspace-based holistic registration on low resolution images improved the recognition even in comparison with landmark based registration on high resolution images. The performance of subspace-based holistic registration is similar to that of the manual registration on the FRGCv2 database [24].
In Part III (Chapter 6), we propose a novel method to correct for an arbitrary single light source in the face images. The main purpose is to improve recognition results of face images taken under uncontrolled illumination conditions. We cor-

rect the illumination variation in the face images using a face shape model, which allows us to estimate the face shape in the face image. Using this face shape, we can reconstruct a face image under frontal illumination. These reconstructed images improve the results of face identification. We experimented both with face images acquired under different controlled illumination conditions in the laboratory and under uncontrolled illumination conditions [23].
In Part III (Chapter 7), we extend the previous method to correct for an ambient and a arbitrary single diffuse light source in the face images. Our focus is more on uncontrolled conditions. We use the Phong model which allows us to model ambient light in shadow areas. By estimating the face surface and illumination conditions, we are able to reconstruct a face image containing frontal illumination. The reconstructed face images give a large improvement in performance of face recognition under uncontrolled conditions [27].
In Part III (Chapter 8), we combine two categories of illumination normalization methods. The first category performs a local preprocessing, where they correct a pixel value based on a local neighborhood in the images. The second category performs a global preprocessing step, where the illumination conditions and the face shape of the entire image are estimated. We use two illumination normalization methods from both categories, namely Local Binary Patterns and the method discussed in Chapter 6. The preprocessed face images of both methods are individually classified with a face recognition algorithm which gives us two similarity scores for a face image. We combine the similarity scores using score-level fusion, decision-level fusion and hybrid fusion. In our previous work, we show that combining the similarity score of different methods using fusion can improve the performance of biometric systems. We achieve a significant performance improvement in comparison with the individual methods [28].
In Part III (Chapter 9), we improve our previous illumination correction methods to correct for multiple light sources. In order to correct for these illumination conditions, we propose a Virtual Illumination Grid (VIG) to reconstruct the uncontrolled illumination conditions. Furthermore, we use a coupled subspace model of both the facial surface and albedo to estimate the face shape. In order to obtain representation of the face under frontal illumination, we relight the estimate face shape with frontal illumination. We show that our relighted representation of the face gives better performance in face recognition. We have performed the challenging Experiment 4 of the FRGCv2 database, which compares uncontrolled probe images to controlled gallery images. By fusing our global illumination correction method with a local illumination correction method, significant improvements are achieved by using well-known face recognition methods [25].
In Chapter 10, we finish this thesis by summarizing our work. We also state and discuss our contributions and mention possible recommendations to extend this work.

2

FACE RECOGNITION SYSTEM

2.1 Introduction

An automatic face recognition system has to solve a difficult problem. A three-dimensional object with varying appearance due to illumination, pose, expressions, aging, and other variations has to be recognized from a two-dimensional image or a video recording. In video surveillance, this task becomes even more difficult, because of low resolution recordings and persons who deliberately hide their faces. Face recognition has received much attention during the past decades, not only for surveillance applications, but also in biometric authentication and human-computer interaction. Although many face recognition methods have been developed, many challenges remain especially when faces are recorded under uncontrolled conditions.
The goal of this chapter is to introduce the various components of a face recognition system. We discuss in more detail the tasks defined in section 1.2.4. Each component is an established research topic on which extensive literature is available. In this chapter, we give an overview of the most relevant implementations of these components as described in the literature. On some of the these methods, we will present a more detailed description.
In this chapter, we will discuss the different components of the face recognition system in separate sections. The Face Detection/Localization methods are intro-

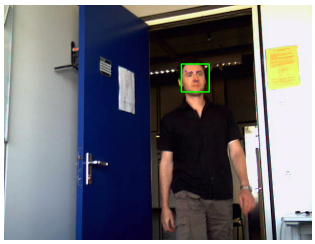
(a) Original Image (b) Background Subtraction

Figure 2.1: Background Subtraction of stationary scene - The left figure shows a stationary scene. Noise in the recording can create small foreground regions shown in the right image. Foreground, background and shadow areas are respectively denoted by white, black and grey)

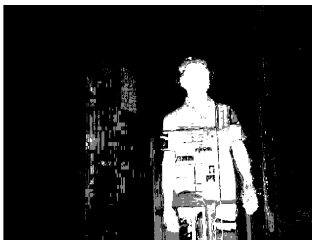
duced in Section 2.2. Face Registration is discussed in Section 2.3 and some Face Intensity Normalization methods are explained in Section 2.4. In Section 2.5, we finish this chapter with the Face Comparison/Recognition methods.

2.2 Face Detection

Face Detection or Localization detects whether there is a face in the image and locates it. It is the first step of the face recognition system, which needs to be reliable because it has a major influence on the remainder of the system. Face Detection remains a complicated problem because the appearance of a face is highly dynamic. For this reason, robust methods are needed to detect faces at different positions, scales, orientations, illuminations, ages and expressions in images or video recordings. Another desired property of a face detection method is that it should detect faces in real-time in order to deal with video streams.
Face detection can be performed using several clues in the video sequence or the image. If a person enters a room monitored by a surveillance camera, the image changes considerably. By remembering the background, which was an empty room, we can determine the foreground corresponding to the person in the image. We will briefly discuss the methods for foreground and background detection in Section 2.2.1. Another clue for face detection in color images can be skin color. This can vary due to illumination and racial differences. We will introduce skin color detection methods in Section 2.2.2. Face detection methods can also use the facial appearance in images, where these methods learn the difference between a face/non-face region. These methods classify each region in the image into regions containing a face and regions not containing a face. Section 2.2.3 will give an overview of several face detection methods based on appearance. In Section 2.2.4, we combine different methods and explain the possible advantages of combining face detection methods.



(a) Original Image

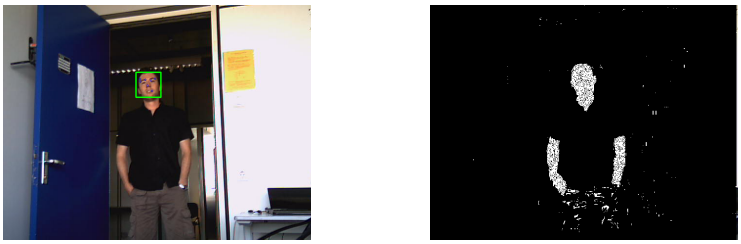


(b) Background Subtraction

Figure 2.2: Background Subtraction of a person entering the room - This scene shows the background subtraction results of a person entering the room. The region in the image where the person is located is marked as foreground (white), The shadow on the door is marked grey

2.2.1 Foreground and Background Detection

In video surveillance, a common setup is that a static camera observes the entrance. In this case, the scene is only interesting if someone enters, which will change the scene. Detecting these changes is essential in video surveillance. This reveals the location of the person in the image and it also allows us to ignore the part of video recordings where nothing happens. Methods used to detect intruding objects in a scene are known as "background subtraction methods". These background subtraction methods assume that the scene without intruding objects shows stationary behaviour where color and intensity change only slowly over time. This behaviour can be described by a statistical model, which models each pixel separately over time. In [134], a single Gaussian model is used to describe the background. Because pixel values often have more complex behaviour, the Gaussian Mixture Models (GMM) are also used for background subtraction [117; 145]. In Figures 2.1 and 2.2, examples of background subtraction with the GMM method described in [145] are shown. We observe that the stationary scene (Figure 2.1) contains almost no foreground, although some foreground pixels are visible due to noise in the video recordings. Once a person enters the room, this person is marked as foreground and can easily be detected in the image, as shown in Figure 2.2. Although background subtraction methods can locate a person entering the room, locating the face of the person requires more heuristics. Background subtraction methods can also fail when video recordings contain a constant motion; for example, a revolving door.
The use of background subtraction for face detection is usually in combination with other methods. Background subtraction, however, clearly shows the boundaries of the face, whereas face detection methods based on appearance usually only give a rough location.



(a) Original Image (b) Background Subtraction

Figure 2.3: Skin Color of a person - This scene shows the skin color detection results of [66] on a person, clearly detecting the regions in the image which contain skin color

2.2.2 Skin color Detection

Skin color can be an important clue for the presence of a face in an image. There are several methods in the literature that perform face detection based on skin color, for instance [1; 58; 116]. Although the skin color of people of different races varies, different studies show that the major difference is in intensity rather than in chrominance. Several color spaces have been used to label the human skin color including RGB (Red, Green, Blue)[66], HSV or HSI (Hue, Saturation, Intensity)[113], YIQ (Luma, Chorominance)[43] and YCrCb (Luma, Chroma Blue, Chroma Red)[58].
Many methods have been proposed to model skin color. Simple models define thresholds in a color space in order to determine if the image contains skin color [1; 58]. Other methods are based on Gaussian density functions [31; 135] or a Mixture of Gaussian densities [59; 66]. In [66], a large scale experiment is conducted with nearly one billion labeled skin color pixels. Using these images, a skin and a non-skin color model are constructed and the likelihood ratio is used for classification. Results of our implementation of this method are shown in Figure 2.3, where both head and arms are located.
Locating the skin color alone is usually not sufficient to locate the face. In Figure 2.3, other body parts are located as well with this technique and scenes can contain objects with similar colors. In [58], two other facial features (eyes,mouth) are also detected using their specific colors and the combination determines if a face is present. Others use a combination of shape analysis, skin color segmentation and motion information to locate the face, for instance in [49; 114].

2.2.3 Face Detection based on Appearance

Face detection based on appearance is basically a two class problem separating between faces and non-faces. The face is a complex 3D object which changes in appearance under different conditions. Pattern classification allows us to learn the differences in appearance between face and non-face regions, by using a training

set, which contains both examples of faces and non-faces. Furthermore, the speed of a face classification method is important, because almost every region (changing position, scales and sometimes rotation) of the images has to be classified.
In the literature, many pattern classification techniques have been used for face detection. In this section, we will distinguish between the linear and non-linear methods:

- **Linear:** In [123], Turk and Pentland describe a detection system based on principal component analysis (PCA). They model faces using the Eigenface representation, computing a linear subspace for faces. Moghaddam and Pentland [80] use both the face space and the orthogonal complement subspace, allowing them to calculate a distance in face space (DIFS) and distance from face space (DFFS). Combining the likelihood of both subspaces provides them with more accurate detection results. In [118], multiple face and non-face clusters are defined using multiple subspaces. An advantage of these linear methods is that they are relatively fast to compute in comparison with non-linear methods. However, they are sometimes not adequate to model the complex and highly variable face space, resulting in a lack of robustness against the highly variable non-face space.

- **Non-linear:** In [97], a retinally connected neural network is used to classify between faces and non-faces. A *bootstrapping* method is adopted, because the non-face space is much larger and more complex than the face space. This makes it difficult to collect a small representative set of non-faces to learn this space. Instead of learning all possible non-face patterns, the idea of bootstrapping is to perform the classification in several stages, where the first stages handle the "easy" patterns and the later stages classify the more difficult patterns. This can be achieved by introducing non-face samples, which are misclassified in previous iterations, into the training of the classifier for later stages. These misclassified samples are more difficult, needing emphasis from these classifiers. Other face detection methods based on neural networks are the probabilistic decision-based neural network (PDBNN) [75] and sparse network of winnows (SNoW) [96]. The Support Vector Machine (SVM) [124] is another non-linear classifier that is often used for face detection [73; 87]. The goal of SVM is to find the maximum separation (margin) between two classes by a hyperplane. This is achieved by finding a hyperplane where we can maximize the margin between the points nearest to the border, see Figure 2.4. Using the kernel trick [4], we can also fit the maximum-margin hyperplane to non-linear feature spaces. In [104], multi-resolution information is obtained using the wavelet transformation. This information gives us features to learn a statistical distribution with products of histograms. Using Adaboost, histograms are selected that minimize the classification error on the training set. One of the best-known methods is the framework for face detection proposed by Viola and Jones [127]. This method can be divided into three important components, namely the Haar-like features, the Adaboost learning method and the cascade classification structure. This method is especially developed for rapid face detection,

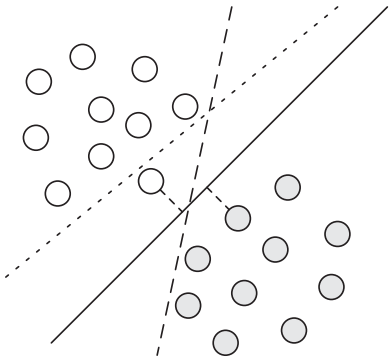


Figure 2.4: Separating two classes - The dotted line is not able to separate the two classes, the dashed line separate the classes with a small margin, while the solid line separate the classes with the maximum margin

where the Haar-like features can be computed quickly on multiple scales using an Integral Image. Adaboost selects the weak classifiers, which are Haar-like features together with a threshold, and combines the weak classifiers into a strong classifier. The cascade structure allows us to pay less attention to the easy background patterns, and spend more time in computation of difficult patterns, as in Figure 2.5. The first strong classifiers determined by Adaboost are simple and reject non-face patterns using only a few features, the last strong classifiers are have to separate difficult pattern which requires much more features. This method is able to process images very quickly, but it still can make a good distinction between the complex patterns of faces and non-faces.

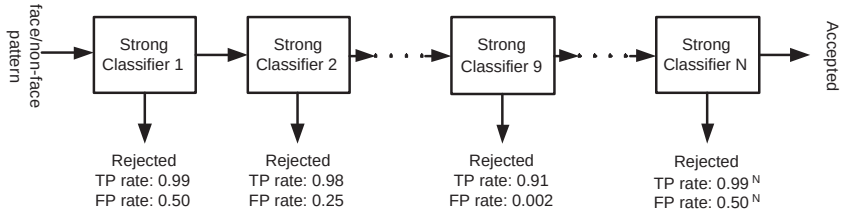


Figure 2.5: Cascade of Strong Classifiers - The strong classifiers reject the easy non-face patterns leaving the difficult patterns for the last stages, the total true positive (TP) and total false positive (FP) rate after each stage are shown if the individual strong classifiers have a TP rate of 99 % and a FP rate of 50%

2.2.4 Combining Face Detection Methods

In the previous sections, we discussed several methods to locate the face in video recordings and images. In this section, we combine some of the previously mentioned methods where our focus lies on the domain of video surveillance. In video streams of a surveillance camera, we are interested in changes which occur in the scene. For this reason, we apply the background subtraction method described in Section 2.2.1. Most of these methods are not computationally complex and always detect the region which contains the person reliably. If we use a camera which records color images, the pixels which are labelled by the background subtraction methods can be classified by a skin color method (see Section 2.2.2), narrowing the interesting regions even more.
Because these methods do not exclude detection of objects other than faces, we

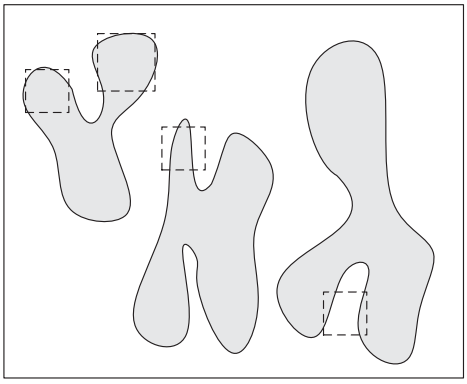


Figure 2.6: Selecting rectangular regions from blobs - This figure shows possible labelled regions (grey blobs) found with background subtraction and/or skin color detection, the task is to select rectangular regions at different scales to be used as input for appearance based face detection

finally use a face detection method based on appearance, searching only in the regions left by the previous methods. All appearance based face detection methods are computationally complex, even the framework of Viola and Jones. Reducing the search regions that can contain faces helps to reduce the number of computations. It however is not a straightforward task as can be observed in Figure 2.6, where both background subtraction and skin color detection give blob like regions, while the appearance based method uses a rectangular region. In this case, we define a mask containing a label 1 in case the pixel belongs to the foreground and contains skin color, while the other pixels get a label 0. By using an Integral Image, we can quickly determine the percentage of labelled pixels in a region. All regions containing more than 80 % labelled pixels are processed by the appearance-based method.
Combining the face detection methods reduces the computational complexity, because we only focus on areas which are worthwhile to investigate. However, combining these methods might introduce some false negatives, especially the skin

color detection is sometimes incorrect. The advantage is that it also reduces the number of false positives in comparison with only using appearance-based face detection.

2.3 Face Registration

For face recognition, it is necessary that the faces are aligned by transforming them to a common coordinate system. While face detection only finds a rough position of the face in an image, face registration refines the positioning and performs other transformations like scaling and rotation to make the comparison between facial images possible. It has been shown that accurate registration improves the performance of face recognition [95],[17]. Because public face databases usually contain manually labelled landmarks, which are used for registration in academic research, "optimistic" results are obtained compared to a fully automatic approach. There are several ways to register images, where the most common methods are based on locations of certain landmarks in the face. In this section, we will discuss some landmark-based registration methods. A simple method to find landmarks is based on the framework of Viola and Jones, see Section 2.3.1. More advanced methods, like MLLL and BILBO also use the relation between landmarks to correct for outliers (Section 2.3.2). Other registration methods perform an iterative search to correctly register the face. Examples of such registration methods are the Elastic Bunch Graphs (Section 2.3.3) and Active Appearance Models (Section 2.3.4). In Chapters 4 and 5, we introduce our own holistic face registration methods, which are developed for video surveillance applications. A comparison between a number of landmark based registration methods and the holistic face registration method can also be found in Chapters 4 and 5.

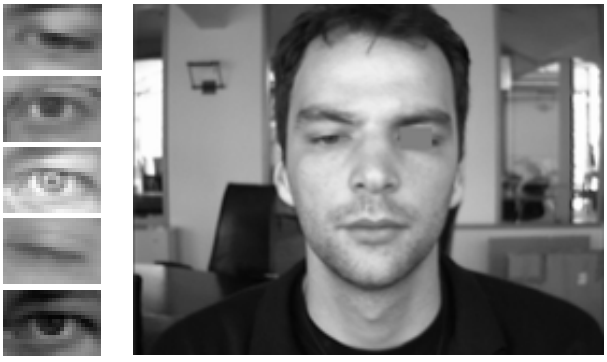


Figure 2.7: Positive and Negative Training Samples - (a) Positive examples of the Left Eye, with a region size of 30×20, (b) Negative examples are random regions of 30×20 selected from the face image where we mask out the left eye with a grey window

2.3.1 The Viola and Jones Landmark Detector

A popular method for finding landmarks in facial images is the framework of Viola and Jones [127]. In section 2.2.3, we already introduced this method for face detection, but this method can also be used to find facial landmarks. In order to train this method, we take the exact landmark region as positive examples (Figure 2.7). The negative examples are obtained from remaining parts of the face images (see Figure 2.7). The advantages of the Viola and Jones method are the computation time and the robustness. Disadvantages are that landmarks are sometimes not detected at all, that landmarks are detected in multiple regions or that landmarks are detected at incorrect locations (outliers). To overcome these problems, heuristics and constraints can be used to remove outliers and select the correct landmarks. If many landmarks are used, missing landmarks are no problem, because the alignment usually can be calculated from a subset of all the landmarks.

2.3.2 MLLL and BILBO

Subspace methods to locate facial landmarks are described in [15; 18; 46; 80]. In this section, we will discuss the Most Likely Landmark Locator (MLLL) [15; 18] in more detail. MLLL searches for landmarks by calculating for each location $\mathbf{p} = (x, y)^T$ a likelihood ratio that the landmark is present. The likelihood ratio is defined as follows

$$L_{\mathbf{p}} = \frac{P(\mathbf{x_p}|L)}{P(\mathbf{x_p}|\bar{L})} \tag{2.1}$$

where $\mathbf{x_p}$ are the vectorized gray-level values around the location $\mathbf{p}$. The likelihood ratio is the quotient of the probability density function $P(\mathbf{x_p}|L)$ of feature vector $\mathbf{x_p}$, given that the location contains a landmark, and the probability density function $P(\mathbf{x_p}|\bar{L})$ given that the same feature vector contains no landmark. Assuming that both probability density functions are normal, we can compute the log likelihood ratio as follows[15]

$$\begin{aligned} S_{\mathbf{p}} &= -(\mathbf{y_p} - \bar{\mathbf{y}}_L)^T \Sigma_L^{-1} (\mathbf{y_p} - \bar{\mathbf{y}}_L) \\ &\quad +(\mathbf{y_p} - \bar{\mathbf{y}}_{\bar{L}})^T \Sigma_{\bar{L}}^{-1} (\mathbf{y_p} - \bar{\mathbf{y}}_{\bar{L}}) \end{aligned} \tag{2.2}$$

In this equation, $\mathbf{y_p}$ is a dimensionality reduced feature vector of $\mathbf{x_p}$, using subsequently PCA 2.5.1.1 and LDA 2.5.1.3. $\bar{\mathbf{y}}_L$, Σ^L and $\bar{\mathbf{y}}_{\bar{L}}$, $\Sigma^{\bar{L}}$ are the reduced landmark mean and covariance matrices and the reduced non-landmark mean and covariance matrices. We obtain the means and covariance matrices from training based on examples of landmarks and non-landmarks, see Figure 2.7. By determining the log likelihood ratio score for each location, MLLL finds landmarks on the locations where the score is maximal.
Because landmark locations are sometimes incorrect, a method is developed to

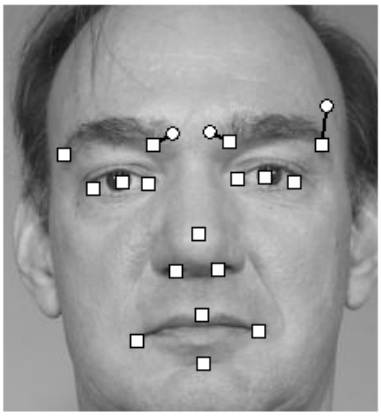


Figure 2.8: Example of landmark finding - Circles are incorrect landmarks found by MLLL and the squares depict the landmark after correction using BILBO

detect and correct these landmarks. To detect the incorrect landmarks, we use the relation between landmarks. A collection of relative landmark coordinates (x_i, y_j) form a shape $\mathbf{s} = (x_1, ..., x_n, y_1, ..., y_n)$ and we assume that correct shapes can be modelled by a subspace, while incorrect shapes are outside this subspace. Using PCA, we are able to learn a subspace of shapes from a training set of correct face shapes, giving us a basis P_s. Once a new shape $\mathbf{s}$ is determined by finding the landmarks, we can projected the shape to the subspace and back: $\mathbf{s}' = P_s P_s^T \mathbf{s}$ (BILBO), which results in the modified shape $\mathbf{s}'$. In the modified shape, the locations of the incorrectly found landmarks have changed significantly, while the other landmark locations change only slightly. The landmark locations, which differ significantly, are determined by thresholding. These landmark locations are corrected to the landmarks locations given by the modified shape s'. This procedure is repeated for a few iterations. The results of MLLL (dots) and some corrections by BILBO (squares) are shown in Figure 2.8.

2.3.3 Elastic Bunch Graphs

Elastic Bunch Graphs [133] are intended for both registration and recognition of faces. This method fits an Elastic Bunch Graph to the facial image. At each landmark location, a bunch of Gabor Jets is defined, which consists of 40 different Gabor features. Gabor features can be calculated using a convolution with a Gabor filter, with different orientations and frequencies. As an example, two Gabor filters with different frequencies and orientations are shown in Figure 2.9. For each landmark, a bunch of Gabor Jets is defined, which represents the different appearances of that landmark. For the eyes for instance, the different Gabor Jets may include open, closed, male, female eyes and glasses. The best fitting Gabor Jet is then selected to refine the search for the best location. These Gabor Jets are placed in a graph which limits the search space and also constrains the landmarks

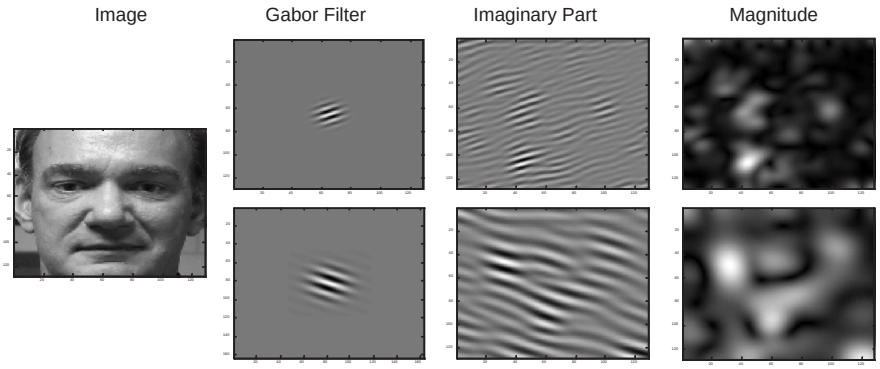


Figure 2.9: Gabor filters - The first column contains the original image, second column contains two examples of Gabor filters, third column shows the imaginary part after convolution of the filter with the face image, fourth column contains the magnitude after convolution of the filter with the face image

locations. Another advantage of the graph is that landmarks can also be placed at locations which are not clearly defined landmarks. By connecting different landmarks, like eyes, nose and mouth, intermediate points can be defined in the cheeks and on the forehead. Elastic bunch graphs also use the contour of the face, defining landmarks on the contour at the same horizontal and vertical axis as well-known landmarks like eyes, nose and mouth. The goal of the Elastic Bunch Graph is to find a graph which fits to the facial image. In this case, they search for the location which best matches the Gabor Jets but at the same time constrain the search by using the Elastic Bunch Graph which limits certain impossible landmark locations. A coarse to fine search is performed, because of the complexity of the search space and to reduce computation time.

2.3.4 Active Shape and Active Appearance Models

In order to register a face, models can be used to describe the different appearances of faces. Multiple landmarks of the face together form a shape $\mathbf{s}$, as is already discussed in Section 2.3.2. In the Active Shape Models (ASM) [37], a subspace model of the variations of the shape is computed using PCA, which gives us the following equation

$$\mathbf{s} = \bar{\mathbf{s}} - P_s \mathbf{b}_s \tag{2.3}$$

where P_s contains the eigenvectors that correspond to the largest eigenvalues and $\bar{\mathbf{s}}$ represents the average shape. The vector $\mathbf{b}_s$ defines the variations of the shape. For a given shape $\mathbf{s}$, these parameters are $\mathbf{b}_s = P_s^T(\mathbf{s} - \bar{\mathbf{s}})$. This shape model is obtained from a training set of facial images together with labelled landmark positions (where ASM and AAM usually use around 65 landmarks). Given the shape $\mathbf{s}$, we can also define a transformation $T_{\boldsymbol{\theta}}$ to the pixel values, where we

transform the normalized shape into the shape in the image. This real shape is defined as $\mathbf{S} = T_{\boldsymbol{\theta}}(\mathbf{s})$, where $\boldsymbol{\theta}$ consists of the parameters for translation, rotation and scaling. In order to register an Active Shape Model, the following iterative approach can be used:

- Examine a region of the image around each landmark $\mathbf{S}_i$ to find the best nearby match for the point $\mathbf{S}'_i$
- Update the model parameters $\mathbf{b}_s$ according to the newly found positions
- Repeat until convergence

Although this approach seems simple, there are many different approaches to match the landmarks, which we will not discuss in detail, because they are usually domain dependent. To give an idea, landmarks can be located using gray values, edges or by determining the profile of the face at multiple resolutions. For certain landmarks, we can also build a statistical model of the gray values, which is very similar to Active Appearance Models.
The Active Appearance Models (AAM) [38] can be seen as an extension of the Active Shape Model. In Active Appearance Models, we also model the texture using PCA, so we get the following equations:

$$\mathbf{s} = \bar{\mathbf{s}} - Q_s \mathbf{c} \tag{2.4}$$

$$\mathbf{x} = \bar{\mathbf{x}} - Q_x \mathbf{c} \tag{2.5}$$

In this case, the appearance model is modelled by the parameters $\mathbf{c}$, which controls both the shape and texture (appearance). Besides the appearance model parameters, other parameters are used in [38] to deal with translation, rotation and scale transformations of the shape ($\boldsymbol{\theta}$) and intensity scaling and offset ($\mathbf{u}$). Using the shape $\mathbf{s}$ obtained from the model parameters together with the parameters $\boldsymbol{\theta}$ and $\mathbf{u}$, we can create a shape free image $\mathbf{x}_s$ from the original image. The current model texture is given by $\mathbf{x}_m = \bar{\mathbf{x}} - Q_x \mathbf{c}$, where the difference between model and image measurements is $\mathbf{r}(\mathbf{q}) = \mathbf{x}_s - \mathbf{x}_m$. The goal is to minimize this difference $\mathbf{r}(\mathbf{q})$ with respect to the all parameters $\mathbf{q} = \{\mathbf{c}^T, \boldsymbol{\theta}^T, \mathbf{u}^T\}$. There are several approaches to find these parameters, for example [11; 38; 57]. Active Appearance Models can perform a very fine registration, but sometimes fail to converge, especially if the initial estimate is not precise enough.

2.4 Face Intensity Normalization

Face Intensity Normalization changes the intensities of the pixels in such a way that the faces become better comparable. Camera settings and variations in illumination make it very difficult to compare images using face recognition methods. In order to compare the faces, they have to become invariant to these effects. This can be achieved by finding an invariant representation. Another approach is to model the effects in order to correct the facial images to a standard representation. The face intensity normalization can also be learned by face recognition methods, making them to a certain extent invariant to these variations. We have observed

that especially in the case of more uncontrolled acquisition conditions large improvements in recognition results can be achieved by performing normalization.
In this section, we divide the face intensity normalization methods into two categories: The first category contains methods that normalize the face image based on a local region around a pixel position. Two examples of methods are Histogram Equalization [105] and (Simplified) Local Binary Patterns [55],[121]. More advanced methods that are especially developed to correct the illumination are [50; 119; 128], which usually make simple assumptions about the behavior of illumination. The second category contains methods that estimate a physical model of reflectance in faces based on the entire image. This category includes for instance the Quotient Image [107], Spherical harmonics [14], 3D morphable models [20].
In our research, we developed our own illumination corrections methods, which are discussed in Chapters 6 to 9. In order to compare these methods, we have used methods from both categories. In section 2.4.1 and 2.4.2, we will discuss Local Binary Patterns and a method developed by Gross et al [50], which are two face intensity normalization methods, that perform correction based on local image information. In section 2.4.3, a short explanation of the illumination correction method of Sim et al [110] is given, which is a method in the second category that uses the Lambertian reflectance model.

2.4.1 Local Binary Patterns

Local Binary Patterns (LBP) are introduced in [86] and it has been shown that they form a set of robust features for face recognition [3]. The standard LBP, shown in Figure 2.10, assigns to each of the pixels of a 3×3-neighborhood the value 0, if the pixel value is smaller than the center pixel value and 1 otherwise. This gives a 8-bit string, which can be used to represent the texture at that point. It is also possible to select larger neighborhoods, which allows us to capture larger scale structures.

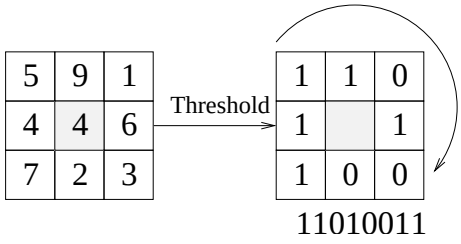


Figure 2.10: Local Binary Pattern - Example to determine the LBP value from a pixel neighborhood

These features can also be used as a preprocessing step in face recognition to obtain a representation independent of illumination variations [55]. An extension to LBP, called (Simplified) Local Binary Patterns is developed in [121] which can be used to become invariant to illumination conditions.

2.4.2 Local Reflectance Perception Model

There are several local methods to obtain illumination invariance [50; 119; 128], which are weakly based on reflectance properties. The method developed by Gross et al [50] is motivated by two assumptions about human vision: First, the human vision is mostly sensitive to reflectance and insensitive to illumination variations. Second, the human vision responds to local changes in contrast. This method uses the following reflectance model:

$$I(\mathbf{p}) = L(\mathbf{p})R(\mathbf{p}) \tag{2.6}$$

The image I at position $\mathbf{p}$ consists of the reflectance R and the illuminance L. In deep shadow areas, human perception makes small changes in these area result in big changes in perceived sensation. However, of illuminated areas, small changes in these areas result in smaller changes in the perceived sensation. This gives an inverse relation of the image intensity I with a local neighborhood Φ and the perceived sensation. The inverse relation can be formulated in the following equation: $1/I_\Phi$. The perceived sensation will be in our case the reflectance and intensity of local neighborhood gives the properties of the illuminance L. By replacing $L(\mathbf{p})$ with a smooth version of I_Φ, which is calculated using an anisotropic function, we can obtain $R(\mathbf{p})$ from Equation 2.6. The reflectance $R(\mathbf{p})$ gives us an illuminance free representation of the face, where the output of this face intensity normalization method is used in Figure 1.6.

2.4.3 Illumination Correction using Lambertian reflectance model

The Lambertian reflectance model is used in, for instance, the following papers [48; 107; 110; 128; 143; 144], to correct for illumination in facial images. Our own illumination correction methods also use the Lambertian reflectance model. The Lambertian reflectance model describes the image intensity $b \in \mathcal{R}$ at a certain position $\mathbf{p} = \{x, y\}$ as follows:

$$b(\mathbf{p}) = \rho(\mathbf{p})\mathbf{n}^T(\mathbf{p})\mathbf{s}i \tag{2.7}$$

where the surface normals $\mathbf{n} \in \mathcal{R}^3$ and the albedo $\rho \in \mathcal{R}$ together define the face shape $\mathbf{h}(\mathbf{p}) = \rho(\mathbf{p})\mathbf{n}(\mathbf{p})^T$. The direction of the light is a normalized vector given by $\mathbf{s} \in \mathcal{R}^3$, and the intensity of the light is given by $i \in \mathcal{R}$. Together they form the light conditions $\mathbf{v} = \mathbf{s}i$. Notice that shadows and specular reflections are not modelled in Equation 2.7.
In order to correct for illumination variations in facial images, the variables in Equation 2.7 have to be estimated. The method of Sim et al [110] uses a bootstrap database of faces to learn certain properties of the illumination. In order to model shadows and specular reflections, an error term $e(\mathbf{p}, \mathbf{v})$ is introduced, which depends on the light conditions $\mathbf{v}$. The method of Sim et al contains the following steps to correct for illumination in the face:

- Given the image, estimate the illumination conditions $\mathbf{v}$ using kernel regression. The kernel regression is trained on a bootstrap database of faces with labelled illumination conditions.
- Given the illumination conditions $\mathbf{v}$, compute the error term $e(\mathbf{p}, \mathbf{v})$ containing possible shadows and specular reflections.
- Estimate the face shape $\mathbf{h}(\mathbf{p})$ at each pixel $\mathbf{p}$ using a MAP estimator of the face shape.
- Synthesize a new image under a different illumination condition using the estimated face shape.

Illumination correction comes down to synthesizing the standard (e.g. frontal) illumination conditions in a face image.

2.5 Face Comparison

Face Comparison (often also called Face Recognition) is basically a pattern classification problem. In face comparison, there are two scenarios, namely face verification and face identification. In face verification, a person claims an identity and the system has to decide whether this claim is correct or not. This makes the verification problem a two-class classification problem. Face identification is more difficult, given a face image we need to determine the person's identity, where we have to solve a multi-class classification problem. In order to determine the identity of a person, a recorded face image (probe images) has to be compared to face images in a database (gallery images). In the case of verification, the face comparison assigns a score, which represents the probability that both face images contain the same person. For face identification, we can easily extend this scheme, assigning all persons in the gallery this score and selecting the best score.
In the literature, many methods have been proposed to perform a face comparison. We categorize these methods into two main groups, namely Holistic Face Recognition Methods and Face Recognition using Local Features. In the following sections, we discussed some of the most well-known face recognition methods together with the methods we used during our experiments. In Section 2.5.1, we elaborate on the holistic face recognition methods and the relationship between these methods. Most of the face recognition methods are holistic methods, while a smaller part of the face recognition methods use local features. We discussed two well-known face recognition methods which use local features in Section 2.5.2.

2.5.1 Holistic Face Recognition Methods

The idea of holistic face recognition is to model the global changes in the appearance of faces. By defining a face space, which is a subset of the image space, we can define for each face a position in this face space based on its appearance. The challenge is to define a face space, which separates the face of different persons, while ignoring other effects like expressions, illumination and pose changes.

2.5.1.1 Principle Component Analysis

Principal Component Analysis (PCA) is a technique to reduce dimensionality of complex data. In [69], this technique is proposed for face analysis and representation, where Turk and Pentland [123] are the first to apply PCA in face recognition. PCA finds a linear combination of principle components, which have the highest variance and are orthogonal to the previous components. This allows us to reduce the high dimensional data (in case of face recognition images containing around 10000 pixel values) into a reduced feature vector (containing around 300 values). With these reduced feature vectors, we are then able to represent the face data. This representation of the face is also less sensitive to noise or blurring effects in the image. In order to find the principle component, a training set of N face images is needed, a vectorized representation of the two dimensional images is given by $\mathbf{x}_1, ..., \mathbf{x}_N$, where $\bar{\mathbf{x}}$ is the mean face vector and Σ is the covariance matrix. We find the eigenvectors Φ and a diagonal matrix with the eigenvalues $\lambda_1 \geq ... \geq \lambda_N$ of the covariance matrix Σ. After obtaining the eigenvectors, we are able to calculate the reduced feature vector as follows $\mathbf{y} = \Phi^T(\mathbf{x} - \bar{\mathbf{x}})$, which gives us the variations of the subspace. In [123], the distance between two reduced feature vectors is determined using the Euclidean distance. This measures if two faces have a similar variation from the mean face. There are also different distance measures than the Euclidean distance. Other commonly used distance measures in face recognition are the Mahalanobis distance or Mahalanobis-Cosine distance.

2.5.1.2 Probablistic EigenFaces

In the previous section, we discussed PCA for face recognition, measuring the similarity between two images I_1 and I_2 in a subspace. This work is extended by Moghaddam and Pentland [80]. We use this evaluation measure in chapter 5. They define a probablistic similarity measure based on the probability density function $P(\Delta|\Omega)$ where $\Delta = I_1 - I_2$ and Ω denotes the object class of face images for the same person. By using PCA, we divide into the face space in two subspaces, namely the principal subspace $F = \{\Phi_i\}_{i=1}^{M}$, which reduces the feature vector from N to M dimensions, and the orthogonal complement $\bar{F} = \{\Phi_i\}_{i=M+1}^{N}$ spanned by the remaining columns of Φ. In Figure 2.11, a graphical representation is given. We can decompose the probability density function into two orthogonal components and assume that they are both Gaussian densities:

$$\begin{aligned} \hat{P}(\Delta|\Omega) &= P_F(\Delta|\Omega)\hat{P}_{\bar{F}}(\Delta|\Omega) \\ &= \left[\frac{\exp\left(-\frac{1}{2}\sum_{i=1}^{M}\frac{y_i^2}{\lambda_i}\right)}{(2\pi)^{M/2}\prod_{i=1}^{M}\lambda_i^{1/2}}\right]\left[\frac{\exp\left(-\frac{\epsilon^2(\Delta)}{2\rho}\right)}{(2\pi\rho)^{(N-M)/2}}\right] \end{aligned} \tag{2.8}$$

where $P_F(\Delta|\Omega)$ is the true marginal density in F and $\hat{P}_{\bar{F}}(\Delta|\Omega)$ is an estimate marginal density in $\bar{F}$. Using PCA, we obtain the reduced feature vector $\mathbf{y} = \Phi^T\tilde{\mathbf{x}}$ and ϵ is the PCA reconstruction error given by:

$$\epsilon(\mathbf{x}) = ||\mathbf{x} - \sum_{i=1}^{M}(\boldsymbol{\phi}_i^T\mathbf{x})\boldsymbol{\phi}_i|| \tag{2.9}$$

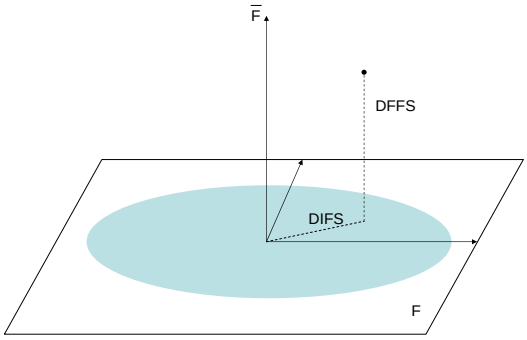


Figure 2.11: Schematic representation of the Subspaces - Showing the principle subspace F and its orthogonal complement $\bar{F}$

In Equation 2.8, ρ is the average of the $N - M$ smallest eigenvalues. In practise, we use log of the two density functions which gives us two distance measures. The first distance measure is the Distance In Feature Space (DIFS), which is the same as the Mahalanobis distance. The second distance measure gives us the Distance From Feature Space (DFFS), both distances are also depicted in Figure 2.11.

2.5.1.3 Linear Discriminant Analysis

Variations in face images are usually not only due to different identities, but they are also caused by for instance expression, illumination and poses of the face. The PCA method retains these variations, which means that measuring the similarity in this subspace does not have to be optimal for identity variations. In [16], Belhumeur et al. have proposed to solve face recognition using Linear Discriminant Analysis (LDA), also called Fisherfaces or Fisher's Linear Discriminant (FLD). LDA is supervised dimensionality reduction method, in contrast to PCA which is unsupervised. Using LDA, we try to minimize the distance between faces of the same person (within-class scatter) and maximize the distance between faces of different person's (between-class scatter). The within-class scatter is defined as

$$S_w = \sum_{c=1}^{C} \sum_{\mathbf{x} \in \omega_c} (\mathbf{x} - \bar{\mathbf{x}}_c)(\mathbf{x} - \bar{\mathbf{x}}_c)^T \tag{2.10}$$

and the between-class scatter is defined as

$$S_b = \sum_{c=1}^{C} N_c (\bar{\mathbf{x}}_c - \bar{\mathbf{x}})(\bar{\mathbf{x}}_c - \bar{\mathbf{x}})^T \tag{2.11}$$

where $\bar{\mathbf{x}}_c$ is the mean of class ω_c, $\bar{\mathbf{x}}$ is the total mean, N_c is the number of samples of class ω_c, and C is the number of classes. We now have to solve the following eigenvalue problem to find the projection matrix:

$$\Phi_{LDA} = \arg\max_{\Phi} \frac{|\Phi^T S_b \Phi|}{|\Phi^T S_w \Phi|} \tag{2.12}$$

In practise however, we first perform a PCA dimension reduction before applying LDA because S_w is usually singular. Furthermore, because LDA finds a reduction to separated the classes, it is shown in [126] that the dimensionality of Φ_{LDA} is at most N_c.

2.5.1.4 Likelihood Ratio for Face Recognition

In order to classify faces, we can use the likelihood ratio, which is an optimal statistic in the Neyman-Pearson sense [82]. The likelihood ratio is defined as

$$L(x) = \frac{p(\mathbf{x}|\omega)}{p(\mathbf{x}|\bar{\omega})} \tag{2.13}$$

where ω denotes a certain class and $\bar{\omega}$ denotes all other possible classes. By assuming an infinite number of classes in the sets, excluding a single class ω does not change the distribution of $\mathbf{x}$. This allows us to assume the following: $p(\mathbf{x}|\bar{\omega}) = p(\mathbf{x})$. In order to use the likelihood ratio, we need the probability density function of $p(\mathbf{x}|\omega)$ and $p(\mathbf{x}|\bar{\omega})$. In face recognition, a Gaussian distribution is often assumed to model faces. The multivariate Gaussian distribution is expressed by

$$p(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^d |\Sigma|}} \exp\left(-\frac{(\mathbf{x}-\bar{\mathbf{x}})^T \Sigma^{-1} (\mathbf{x}-\bar{\mathbf{x}})}{2} \right) \tag{2.14}$$

where d is the dimensionality of the feature vector. The log likelihood of Equation 2.13 will then result in:

$$\begin{aligned}
\ln L(x) &= \ln p(\mathbf{x}|\omega) - \ln p(\mathbf{x}|\bar{\omega}) \\
&= -\frac{1}{2}\big(\ln|\Sigma_w| + (\mathbf{x}-\bar{\mathbf{x}}_c)^T \Sigma_w^{-1} (\mathbf{x}-\bar{\mathbf{x}}_c) \big) + \\
&\quad \frac{1}{2}\big(\ln|\Sigma_t| + (\mathbf{x}-\bar{\mathbf{x}})^T \Sigma_t^{-1} (\mathbf{x}-\bar{\mathbf{x}}) \big) \\
&= \frac{1}{2}\big((\mathbf{x}-\bar{\mathbf{x}}_c)^T \Sigma_w^{-1} (\mathbf{x}-\bar{\mathbf{x}}_c) \ + (\mathbf{x}-\bar{\mathbf{x}})^T \Sigma_t^{-1} (\mathbf{x}-\bar{\mathbf{x}}) \big) + \\
&\quad \frac{1}{2}(\ln|\Sigma_w| + \ln|\Sigma_t|)
\end{aligned} \tag{2.15}$$

where $\bar{\mathbf{x}}_c$ is the mean of the user c, $\bar{\mathbf{x}}$ is the overall mean, Σ_w is the within covariance matrix and Σ_t is the total covariance matrix. The likelihood ratio is normally applied after dimensionality reduction, because the covariance matrices are often singular and to reduce computational costs. In this case, we first perform a reduction using PCA, this also allows us to whiten the total covariance matrix,

which becomes an identity matrix in reduced face space, see [126]. Then, the LDA transformation is performed, where we used the total covariance matrix instead of using the between covariance matrix to maximize Equation 2.12. After dimension reduction using subsequently PCA and LDA, we perform the log likelihood ratio (Equation 2.15) on the reduced feature vectors.

2.5.1.5 Other Subspace methods

In the previous sections, we have introduced some subspace methods which we also use in our research to perform face recognition. In the literature, there are however more subspace methods for face recognition. This section will shortly discuss some other well-known subspace methods, namely Independent Component Analysis (ICA), Kernal-PCA and Kernel-LDA.
Independent Component Analysis [36] is very similar to PCA, but where PCA minimizes only the second-order dependencies, ICA also minimizes higher-order dependencies, finding components that are non-Gaussian. Although this dimension reduction method finds a linear projection, it has a different outcome compared to PCA. In [13], ICA is first used for face recogntion, where ICA originates from solving the blind source separation problem, decomposing the input signal $\mathbf{x}$ into a linear combination of independent source signals.
Kernel Principal Component Analysis (KPCA) [103] is a nonlinear generalization of PCA, allowing us to model higher order correlations between the input vectors. This is achieved by using the same kernel trick [4] as is used for Support Vector Machines, projecting the feature to a higher dimensional nonlinear space. In the linear space, the kernel function is the inner product between two vectors, replacing this function by another kernel function allowing us to make a non-linear projection. Two kernel functions which are often used to replace the linear kernel are the polynomial and Gaussian kernel functions. A similar strategy can be applied on the LDA scheme, which results in KLDA.

2.5.2 Face Recognition using Local Features

Another way of performing face recognition is to separate faces based on local features. In many cases, individuals have distinct noses, mouths or eyes allowing us to match only a part of the face. Other features like moles, scars or freckles can also be used to compare different persons with each other. The holistic registration methods, however, are usually not able to model these details. For this reason, we use local features descriptors like the Haar features, Gabor wavelets or Local Binary Patterns (LBP).

2.5.2.1 Elastic Bunch Graphs

To perform face recognition based on local features, a good registration is necessary to ensure that similar features are compared. In Section 2.3.3, we already discussed the Elastic Bunch Graphs [133] for registration, locating different facial features.

The Gabor Jets and the Bunch Graph can also be used in face recognition. Gabor Jets are powerful features which record the details in a face image on different scales. Besides using the local information, global information can be obtained by comparing the behaviour of the obtained Bunch Graphs. In [133], a comparison is performed on the similarities of the Gabor Jets, by a simple summation on the outcome of all visible Gabor Jets. A schematic representation of holistic face recognition versus Elastic Bunch Graphs is given in Figure 2.12.

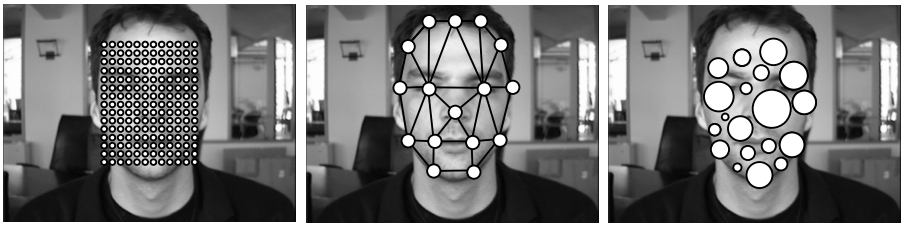


Figure 2.12: Schematic representation of the different feature selections and their importance - (a) depicts Holistic face recognition, (b) shows the feature selection of Elastic Bunch Graphs, (c) are differently weighted feature selected by Adaboost

2.5.2.2 Adaboost using Local Features

Another local feature approach is based on the Adaboost framework already mentioned in Section 2.2.3. This section contains a part of [29], where we have investigated this framework for face recognition. Several other papers [29; 51; 67; 136; 137; 138] also applied the boosting framework on the problem of face recognition. In these papers, various kinds of simple features like Haar-like features, Gabor wavelets or Local Binary Patterns (LBP) are used. Adaboost is the machine learning algorithm, which is used to combine features into a strong classifier. In the boosting framework, we train a face similarity function which determines if two faces belong to the same person or to different persons. The face similarity function is given here:

$$F(I_1, I_2) = \sum_{t=1}^{T} f_t(I_1, I_2) \tag{2.18}$$

In this equation, I_1 and I_2 are the face images which are compared to see if they belong to the same person. The function f_t represents a weak classifier used by Adaboost. The final classifier is a weighted sum over all the selected weak classifiers. A weak classifier f_j is given below, where α and β are given by Adaboost.

- Given examples $(x_1, y_1), ..., (x_N, y_N)$ where $N = A + B$. A is the number of examples with $y_i = +1$ and B is the number of examples with $y_i = -1$
- $w_{1,i} = \frac{1}{2A}$ for those examples with $y_i = +1$ and $w_{1,i} = \frac{1}{2B}$ for those examples with $y_i = -1$
- For $t = 1, ..., T$

  - Normalize the weights, $w_{t,i} = \frac{w_{t,i}}{\sum_{k=1}^{N} w_{t,i}}$

  - For each ϕ_j find the ϕ_j which minimizes the error, ϵ_j, in Equation 2.22

  - Choose α and β according to Equations 2.20 and 2.21

  - Update the weights:

$$w_{t+1,i} = w_{t,i} e^{-f_t(x_i) y_i} \tag{2.16}$$

- The final strong classifier is:

$$F(x) = \sum_{t=1}^{T} f_t(x) \tag{2.17}$$

Figure 2.13: Adaboost algorithm

$$f_j(I_1, I_2) = \begin{cases} \alpha & \text{if } |\phi_j(I_1, I_2)| < t_j \\ \beta & \text{otherwise} \end{cases} \tag{2.19}$$

In this formula, ϕ_j is the feature output for the given image pair and t_j is the feature threshold. This means that for every feature ϕ_j, we first determine the optimal threshold t_j, by minimizing the weighted training error, Equation 2.22. α and β are given in Equations 2.20 and 2.21.

$$\alpha = \frac{1}{2} \log\left(\frac{\sum_{i:y_i=+1 \wedge \phi_j(x_i)<t_j} w_i}{\sum_{i:y_i=-1 \wedge \phi_j(x_i)<t_j} w_i}\right) \tag{2.20}$$

$$\beta = \frac{1}{2} \log\left(\frac{\sum_{i:y_i=+1 \wedge \phi_j(x_i)\geq t_j} w_i}{\sum_{i:y_i=-1 \wedge \phi_j(x_i)\geq t_j} w_i}\right) \tag{2.21}$$

To each example $x_i = (I_1^i, I_2^i)$, a label $y_i \in \{+1, -1\}$ and a weight w_i are assigned.

$$\epsilon_j = \sum_{i:y_i=+1 \wedge \phi_j(x_i) \geq t_j} w_i + \sum_{i:y_i=-1 \wedge \phi_j(x_i) < t_j} w_i \tag{2.22}$$

The goal of Adaboost is to select the weak classifiers which minimize the classification error ϵ_j. The training samples that are incorrectly classified by the selected weak classifiers get more weight. This makes it more important for the next weak classifier to classify these samples correctly. A weighted combination of selected weak classifiers results into a strong classifier, as shown in Figure 2.13 A schematic representation of the different feature selection strategies is given in Figure 2.12. The holistic face recognition algorithms usually select their features from an evenly weighted grid. Elastic Bunch Graph finds Gabor Jets on positions defined by the user, while Adaboost selects the features automatically and gives larger weight to the more important features for classification.

Part I

RESOLUTION

One of the most important characteristics of camera surveillance is the relatively low resolution of faces in the recordings. For this reason, the following specific research question is asked: What is the effect of low resolution on different components (Face Detection, Face Registration, Face Intensity Normalization and Face Comparison) of the face recognition system? The low resolution of facial images can have an impact on the performance of the various components. For this reason, we decided to investigate the effects of resolution on different components. Some research is already presented in other publications, but this research is limited to the face comparison component. Here, we have looked beyond the face comparison component and have also taken into account the other components. This part also gives an answer to the research question: What is the effect of illumination on the different components of the face recognition system? This question is answered through the use of two datasets. The first dataset contains face images recorded under controlled illumination conditions (High Quality) and the second dataset, face images recorded under uncontrolled illumination conditions (Low Quality). This allows us to investigate the impact of the illumination on the face recognition system from these experiments. This part contains one chapter, which is published in [21].
As already discussed in Chapter 2, our face recognition system is divided into four components. In order to perform this research, we decided to use the face recognition system developed in our group. For this system, we discussed the effects on the different tasks:

- **Face Detection Component:** This is performed using the framework of Viola-Jones [127]. For this framework and most other face detection methods, there is a certain theoretical minimum resolution of around 20×20 pixels, which is similar to the input of the training images. Performing face detection on lower resolution is not possible.

- **Face Registration Component:** The face registration is performed by finding landmarks using MLLL and BILBO [15; 18]. This registration method is able to find landmarks on relatively low resolutions. An important issue however is how accurately these landmarks are found. The accuracy of registration has effect on the recognition results.
- **Face Intensity Normalization Component:** Although more advanced face normalization methods require a minimum resolution to function properly, in our early face recognition system, we used a simple method to normalize the energy in the image. This method is not sensitive to face resolution differences.
- **Face Comparison Component:** The face comparison is performed using the log likelihood classifier, described in [126]. In the literature, some publications claimed that reasonable results could be reached using much lower resolutions. We have investigated these claims also in combination with the other components of a face recognition system.

The following chapter is published in [21]. We have focussed especially on the face registration and recognition components, because these components of the face recognition system appear most sensitive to low resolutions of facial images.

3

THE EFFECT OF IMAGE RESOLUTION ON THE PERFORMANCE OF A FACE RECOGNITION SYSTEM

3.1 Introduction

In video surveillance applications it is often difficult to obtain good quality recordings of the faces of the observed individuals. One important factor determining the quality of the recordings is the resolution of the images, which is often much lower than the resolutions typically used in face recognition. Therefore, we decided to investigate the lowest resolution at which a face recognition system still can achieve acceptable performance. In a face recognition system several processing steps are needed to recognize a person's face. Here, we investigate the sensitivity of the recognition part and the registration part to the image resolution.
Currently available face recognition systems usually require face images with more than 50 pixels between the eyes. In the literature several papers can be found which use much lower resolutions. Zhao et al[141] use a combination of PCA and LDA for face recognition on a resolution of 24×21 pixels and claim that their approach will even give good results on 19×17 pixels. Kukharev et al[70] report that the images should be larger than 28×23 pixels using PCA and LDA. Wang et al [129] investigate the effects of resolution on face recognition and conclude that results improve until a resolution of 64×48 pixels and remain constant for higher resolutions using both PCA and a combination of PCA and LDA. In [42], Czyz et

al mention that results only slightly decrease using face images of 16×16 pixels. Ekenel et al [45] investigate the frequency subbands that perform best. By first low pass filtering and subsequently downsampling, their results of PCA, ICA1 and ICA2 hardly get worse on face images with resolutions down to 16×16 pixels.
Other papers investigate the enhancement of the resolution of images using face hallucinating or super-resolution [9],[10]. This paper might give an indication to which resolution the image must be enhanced for reliable face recognition. In [9] it is suggested that the resolution required to find landmarks should probably be higher than the resolution used for performing face recognition. In [9] it is also noticed that not only face recognition depends on resolution but also the registration, which in many cases relies on finding facial landmarks.
The papers found in the literature only mention face recognition as a function of image resolution, while we also investigate the face registration as a function of image resolution. In comparison to other papers we use a larger dataset containing 3699 face image. While in [129] very limited research is done in the choice of the number of PCA/LDA components, we do more elaborate research into this subject.
This chapter is organized as follows. In section 2 we explain how we generate face images with different resolutions. In section 3 the methods we use for face registration and recognition are described. In section 4 the experiments and results are given on different resolutions. Finally, in section 5 conclusions are presented.

Figure 3.1: Different resolutions face images - The face images above are for the registration and below are for recognition

3.2 Face Image Resolution

In this chapter, we investigate the face registration and recognition as a function of the image resolution. Instead of taking images from faces which have different distances to the camera, we have tried to simulate the effect of lowering the resolution. By using simple downsampling or taking the mean pixel value, the image will still contain high frequency components. To simulate the effect of the camera

lens we used a Gaussian low pass filter with $\sigma = 0.375$, followed by downsampling the image with a decimation factor of 2 in both dimensions . This approach is similar to the Gaussian pyramid used in [129] and [9]. In our research we create the following resolutions: 128×128, 96×96, 64×64, 48×48, 32×32, 24×24, 16×16, 12×12 and 8×8. Since cutting out the region of interest (ROI) and finding the landmarks in the lowest resolution would lead to serious quantisation errors, we decided to upscale all images to the same resolution. After downscaling, we scaled the image up to 256×256 for face registration and 128×128 for face recognition using a bilinear interpolation. The results can be seen in Figure 3.1.

3.3 Face Recognition System

In this section the different steps of our face recognition algorithm are discussed. The following steps are performed in our system: face detection, registration, feature extraction and classification, the latter two are taken together as 'recognition'.

3.3.1 Face Detection

For face detection we used the OpenCV implementation [61] of the face detection algorithm first proposed by Viola and Jones [127]. The region found by this algorithm is then used in our face registration algorithm. We did not investigate the sensitivity to resolution of the face detection algorithm in this chapter.

3.3.2 Face Registration and Normalization

The methods for locating (MLLL:Most Likely Landmark Locator) and correcting (BILBO) the landmarks are published in [15; 18] and summarized below. Based on the corrected landmarks the image is aligned, the ROI is determined and the face image is normalized.

3.3.2.1 MLLL

This algorithm searches for landmarks, which are typical facial features, easily distinguishable by a human observer, in the region given by the face detection algorithm. In our case we search for 17 landmarks, see Figure 2.8. The algorithm searches in positions around the mean location of a landmark in the detection region. The landmark is found at the location where the likelihood ratio $L_{u,v}$ is maximum. The likilihood ratio is given by:

$$L_{u,v} = \frac{p(x_{u,v}|L)}{p(x_{u,v}|\overline{L})}, \tag{3.1}$$

Here the vector $x_{u,v}$ contains the gray-levels of a subimage at the location (u, v). The parameters of the probability densities $p(x_{u,v}|L)$ and $p(x_{u,v}|\overline{L})$ are respectively learnt from examples of manually labelled landmarks and non-landmarks.

3.3.2.2 BILBO

BILBO is used to correct the outliers produced by the MLLL algorithm. The landmarks which have been found by MLLL are put into a vector s, which we call the shape. The shape s is projected onto a subspace of trained correct shapes, resulting in a modified shape s'. The landmark positions which significantly changed in s' will be corrected. The complete method is described in [18]. We use the same parameters as can be found in [18] for our experiments.

3.3.2.3 Face Alignment

The face images are aligned using a rigid transformation based on the landmarks

3.3.2.4 Face Normalization

We remove background and hair by taking an ROI as can be seen at the bottom row of Figure 3.1 and then we normalize the energy of the image inside the ROI.

3.3.3 Face Recognition

For face recognition we do feature reduction by subsequently performing PCA [123] and LDA [16]. We use the algorithm proposed in [126] which uses the log-likelihood ratio to classify face images. For each class i the similarity score S is calculated by:

$$\begin{aligned} S_{y,i} &= -(y-\mu_{W,i})^T\Sigma_W^{-1}(y-\mu_{W,i}) \\ &\quad +y^T\Sigma_T^{-1}y - \log|\Sigma_W| + \log|\Sigma_T| \end{aligned} \tag{3.2}$$

Here y is a vector which is a representation of the face image after feature reduction, Σ_T is the total covariance matrix, Σ_W is the within class covariance matrix and $\mu_{W,i}$ is the class average.

3.4 Experiments and Results

3.4.1 Experimental Setup

In our experiments we use three datasets, namely the BioID [60], the high-quality FRCG and the low-quality FRGC [83]. The BioID dataset consists of 1521 images of frontal faces of 23 persons, where we use 17 landmarks which are manually labelled in this database. From FRGC version 1, we used 3699 images taken under controlled conditions which is the high-quality FRGC dataset and 1803 images taken under uncontrolled conditions which is the low-quality FRGC dataset, the FRGC version 1 consists of 271 individuals. In our experiments we use the regions found by the face detection algorithm. In the framework of this research we are

not interested in the performance of the face detection algorithm. Therefore, respectively 73 and 84 samples are removed from the high- and low-quality dataset in which the face is not correctly detected.
For training the landmark finder we use the BioID database [60]. To train MLLL, the positive examples are cut out of every image in the dataset. For every positive example 10 negative examples are taken around that facial landmark. The BILBO algorithm is trained on the shapes of all 1521 faces in the BioID database. To verify the results of the registration we use the high-quality FRCG and the low-quality FRGC, where we search for landmarks in the region determined by the face detection algorithm.
The face recognition is performed on the low- and high-quality FRGC datasets. The datasets are randomly split into two subsets, each consisting of approximately half of the images of each person. One subset is used for training and the other for testing. We train and test the face recognition algorithm on the same resolution. The same holds for other parameters, if manual landmarks are used for training then they are also used for testing. The results of the face recognition are measured in Equal Error Rate (EER), at the point of operation where False Accept Rate (FAR) is equal to the False Reject Rate (FRR). To get more accurate results, we repeat the experiments 20 times, randomly splitting the datasets so other subsets are used in the training and test set. The EER in our case is calculated from the total set of matching and non-matching scores of all experiments [17].

3.4.2 Experiments

We conducted several experiments, investigating parts of the system and the entire system.

3.4.2.1 Face Recognition

In Experiment 1 we use manual landmarks provided with the dataset, to avoid the effects of incorrect registration. By doing this experiment we can determine a minimum resolution which still gives good results and we can also verify if the claims made in the literature are valid. We also look at the effect of the components of PCA and LDA under various resolutions. Our hypothesis is that adding more dimensionality will mainly have a positive effect on results of the high resolution images, because then more discriminating details of the face can be used in classification.

3.4.2.2 Face Registration

In Experiment 2 we investigate the results of our landmark finder under different resolutions. Our first hypothesis is that the resolution needed for accurate face registration is higher than for face recognition. Our second hypothesis is that at a certain test resolution the landmark finder trained at the same resolution gives the best results. We first compare the automatically found landmarks directly with the manually labelled landmarks. We perform some experiments by training and

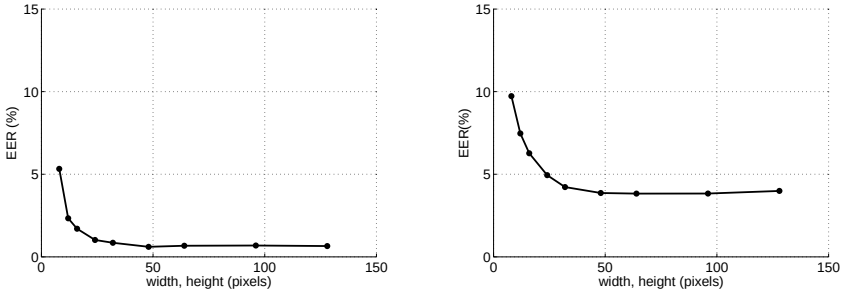
testing the landmark finder on different resolutions. Finally we investigate the effect of face registration on the face recognition by performing registration on all resolutions, while doing the recognition on the highest resolution.

3.4.2.3 Face Registration and Recognition

To study the effect of resolution on the face registration and face recognition we use in Experiment 3 the automatically found landmarks to register the face images and then perform the recognition. The results of the registration and recognition on the same resolution are calculated to determine the effect of resolution on the overall system.

3.4.2.4 Face Recognition using erroneous landmarks

Because we only use one landmark finding technique, we decided to investigate what will happen at different resolutions if registration errors are made. Our landmark finding algorithm may perform optimal for a certain resolution, but that doesn't mean other algorithms will. This experiment allows us to predict the EER of face recognition algorithm based on the RMS error made by the registration. In the next section we present the details and the results of the experiments.



(a) High Quality FRGC (b) Low Quality FRGC

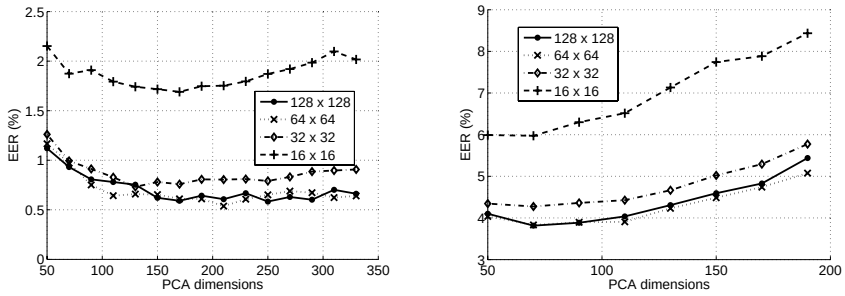
Figure 3.2: Results in Face Recognition at different resolutions - Face recognition performed at different resolutions with manual landmarks for registration

3.4.3 Results

3.4.3.1 Face Recognition (Experiment 1)

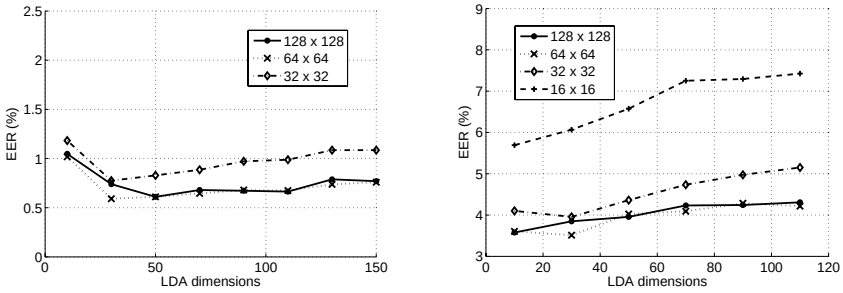
We first perform experiments using the manual labelling given by the FRGC dataset. The EER is calculated for every resolution on the low- and high-quality FRGC database and is shown in Figure 3.2, where this is done for the low- and

high-quality FRGC database. The EERs remain almost constant downto a resolution of 32×32 pixels, below this resolution the EER increases rapidly. This seems to be in accordance with other papers where still good results are mentioned on low resolution face images. Our results on face recognition show that training and testing on low resolutions still give good results. This also means that if the lowest resolution of the face images for a certain system is given, training on this resolution gives good results for all the above resolutions downto 32×32 pixels. For these experiments on the high- and low quality FRGC we use respectively 150 and 90 PCA dimensions and 50 LDA dimensions.
To investigate the influence of the dimensionality at various resolutions, two other experiments are performed. In the first experiment we increase the number of PCA components beginning with 50 PCA components with steps of 20, and we use 50 LDA components (Figure 3.3). In the second experiment we use 270 PCA components for high quality FRGC and 110 PCA components for low quality FRGC and we increase the number of LDA components beginning with 10 components using steps of 20 (see Figure 3.4). These experiments are done on both FRGC datasets for the resolutions 128×128, 64×64, 32×32 and 16×16. Figure 3.3 and 3.4 show that the results on the resolutions 64×64 and 128×128 are almost the same and that the results for 32×32 are slightly worse. Using more than 90 PCA components on the high quality FRGC will give better results for high resolutions, beyond that the results seem to remain stable on this dataset. On the low quality FRGC dataset using above 110 PCA components will worsen the results. Using more LDA components than the 50 components we already use in the first experiment also seems to worsen the results. In most cases around 30 LDA components seems to be the optimal choice. Figure 3.3 and 3.4 also show that the number of components depends more on the database that is used than on the different resolutions.



(a) High Quality FRGC (b) Low Quality FRGC

Figure 3.3: Results in Face Recognition with using different PCA components - EER as a function of the number of PCA components while using 50 LDA components



(a) High Quality FRGC (b) Low Quality FRGC

Figure 3.4: Results in Face Recognition with using different LDA components - EER as a function of the number of LDA components while using for the high- and low-quality FRGC respectively 270 and 110 PCA components

3.4.3.2 Face Registration (Experiment 2)

In this experiment we search for the location of the landmarks in the different resolution face images. To measure the performance of the landmark finding we use the RMS error which compares manually labelled landmarks with the automatically found landmarks after the alignment. The exact calculation of the RMS error is given below, because a straightforward comparison cannot be used due to the difference in scale in the face images.

1. Translate, scale and rotate the groundtruth data so that the eye landmarks are on a horizontal line at a 100-pixels distance form each other.

2. Align the shape found to the corresponding groundtruth shape.

3. Calculate the Euclidian distance between each landmark and its groundtruth equivalent.

4. Remove the bias caused by the different labelling policies in the databases, i.e. tip of the nose (BioID) versus a point between the nostrils (FRGC).

5. Calculate the RMS value of the remaining difference between the found shape and the groundtruth shape, which is now given as a percentage of the inter-eye distance.

In the FRGC database the center of the mouth is labelled, while our methods label the mouth corners. Therefore, prior to calculating the error an estimate of the center of the mouth was obtained by computing the midpoint of the mouth corners. During our experiment the training of MLLL is done using the highest resolution images of the BIOID and for testing we used the high-quality FRGC database. The results are shown in Table 3.1. Because the resolution of 32×32 gives the best results, we have studied the effect of training on other resolutions.

Resolution	right eye	left eye	nose	mouth
128×128	3.0	3.2	4.4	2.8
64×64	2.6	2.7	3.7	2.5
32×32	**2.1**	**2.2**	**3.2**	**2.2**
16×16	3.1	3.2	4.6	2.9
8×8	6.4	6.8	8.2	5.6

Table 3.1: RMS error on High Quality FRGC for different facial features

The results of these experiments are shown in Table 3.2 and 3.3 on both the high- and low-quality FRGC, the values in the table are the mean of the RMS errors of the eyes, nose and mouth. Tables 3.2 and 3.3 show that training on the highest resolution still gives overall better results than training on lower resolutions. Tables 3.2 and 3.3 also show that the results of training on 128×128 pixels and the highest resolution are almost the same. The reason is that the resolution of the face images of this dataset is around 150×150 pixels and thus close to 128×128 pixels. It seems that MLLL performs best at the resolution of 32×32 pixels, nearly for almost all training resolutions.

Test Resolution	Highest	128×128	64×64	32×32
128×128	**3.4**	3.6	5.1	7.7
64×64	**2.9**	3.0	4.0	6.9
32×32	**2.4**	2.5	2.7	4.2
16×16	**3.5**	3.5	3.4	3.4
8×8	**6.7**	6.5	6.8	5.9

Table 3.2: RMS error on High Quality FRGC trained on different resolutions (columns)

We investigate the effects of this registration methode on the face recognition, after performing face alignment based on the landmarks obtained by the landmark finder trained on the highest resolution. Because here we are only interested in the performance of the registration under different resolutions, we performed landmark finding under the different resolution, while doing face recognition under the highest resolution of 128×128. The results are shown in Figure 3.5 by the

Test Resolution	Highest	128×128	64×64	32×32
128×128	**4.3**	4.4	4.7	6.9
64×64	**4.1**	4.2	4.4	6.0
32×32	**4.0**	4.2	4.0	4.5
16×16	**4.3**	4.4	4.2	4.3
8×8	**5.7**	5.6	5.5	5.2

Table 3.3: RMS error on Low Quality FRGC trained on different resolutions (columns)

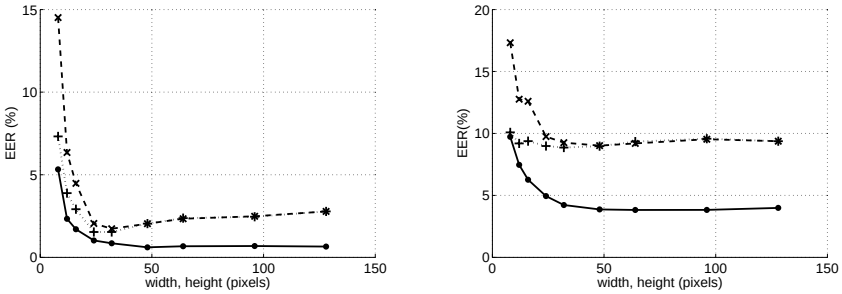
dotted line. Optimal results in EER are also reached at 32×32. It appears that the RMS error gives a good indication of the EER, because the RMS error and EER follow the same trend.

3.4.3.3 Face Recognition and Registration (Experiment 3)

In this experiment we investigate the effects of the entire system under different resolutions. We first perform face registration and then the face recognition on the same resolution. The results of face registration and recognition under various resolutions are shown in EER in Figure 3.5 by the dashed line. The other lines in Figure 3.5 are the results of only face recognition using the manual landmarks (solid line), and the results of only face registration doing face recognition on resolution of 128×128 pixels (dotted line). Figure 3.5 shows that our landmark finding algorithms works best at 32×32 and also for the whole face recognition system the performance is best for this resolution. For resolutions above 32×32 pixels the influence of the registration on the EER is significant, while for lower resolutions the EER is dominated by the poor performance of the face recognition. The difference obtained in EER between manual and automatic landmarks is rather large, so there is much to gain by improving the face registration.

3.4.3.4 Face Recognition by using erroneous landmarks (Experiment 4)

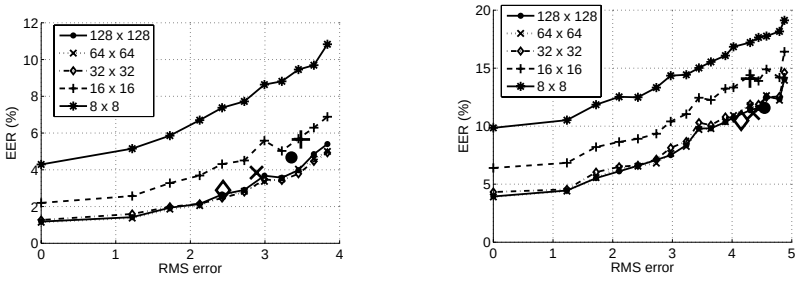
In this experiment we added Gaussian noise to the manually labelled landmarks and, based upon these landmarks, the face registration and recognition is performed. For the noisy landmarks we calculated the RMS error of the landmarks and the EER of the face recognition algorithm. This experiment is performed on both FRGC datasets and the results are shown in Figure 3.6. It shows that the EERs of resolutions 32×32 up to 128×128 remain almost the same for all RMS errors. Results in RMS error and corresponding EER of our face registration and recognition for the resolutions 16×16, 32×32, 64×64 and 128×128 are shown as



(a) High Quality FRGC (b) Low Quality FRGC

Figure 3.5: Results in Face Registration and Recognition at different resolutions - 1. manual landmarks (solid), 2. landmark finding on all resolution and recognition at 128×128 (dotted), 3. landmark finding and recognition on all resolutons (dashed)

the large symbols (respectively, plus, diamonds, cross and dot) in the Figure 3.6. These results correspond well with the erroneous landmark results, which indicates that with this graph we can roughly predict the results on face recognition if we know the RMS error of the registration.



(a) High Quality FRGC (b) Low Quality FRGC

Figure 3.6: Results in Face Recognition by simulating errors in Face Registration - EER under different RMS errors using erroneous landmarks, the large symbols are the locations using automatically obtained landmarks

3.5 Conclusion

In this chapter, we investigate the effect of image resolution on the result of a face recognition system. The results confirm that face recognition algorithms using a PCA/LDA-based system are not very sensitive to resolution and still give good

results at resolutions as low as 32×32 pixels. This also means training on 32×32 pixels will give good results if the face recognition system has to deal with various resolutions above 32×32 pixels. We also show that the optimal choice for the number of PCA and LDA components depends on the resolution, but much more on the dataset that is used. Increasing the amount of PCA and LDA components can help in cases where the resolution and quality of the images is high, however at the risk of overtraining.

We have also investigated the relation between face registration and resolution. Our registration algorithm performs best on the upscaled image with a resolution of 32×32 pixels. The landmark finding was not improved by training on the same resolution as used for testing. Other registration methods may behave differently under various resolutions. We also show that there is much to be gained by accurate registration. This can be seen in the difference in results of the face recognition between manually and automatically found landmarks. This confirms that accurate registration is of vital importance for face recognition. Our entire face recognition system works best at the resolution of 32×32 pixels.

Conclusion Part I

In camera surveillance, the resolution of the faces in the images is usually low. For this reason, we would like to know the effect of low resolution on the different components of the face recognition system. Chapter 3 shows that the recognition results of the subspace based face recognition methods are hardly effected by the resolution until around 32×32 pixels. Our face registration method is also able to perform reasonably at this resolution. Although camera surveillance footage often contains faces with lower resolution than 32×32 pixels, the use of these face images becomes questionable because humans can not verify the results obtained using these images either. With techniques like face hallucinating or super-resolution, we might be able to further enhance the image. We observed however, that a minimum resolution of around 32×32 pixels is usually achieved in the scenarios (Section 1.1.4) on which we focussed our attention. For this reason, we decided to focus our efforts on other problems already visible in the experiment performed in Chapter 3.

The experiments in Chapter 3 already show two other important issues for the camera surveillance applications. The first issue is the difference between registration using manually labelled landmarks and the automatically found landmarks. This difference is even larger for the Low Quality (uncontrolled conditions) images. The last experiment, which simulates erroneous registration, indicates that much can be gained by improving the registration. The second issue is the difference between "High Quality" (controlled) and "Low Quality" (uncontrolled), which answers the question: What is the effect of illumination on the different components of the face recognition system? We observe that for uncontrolled conditions, face recognition methods perform far worse than for controlled conditions. The face recognition component is most sensitive to varying illumination conditions, making face intensity normalization methods necessary to improve the performance.

Part II

REGISTRATION

The second specific research question is: Which measures can be taken to improve the face recognition system for low resolution facial images? In the previous part, we showed that there still is a large difference between manual and automatic face registration. In this part, we focus on improving face registration for camera surveillance. In order to improve face registration, the number of landmarks is usually increased. This approach fails for low resolution faces obtained in camera surveillance. The reason is that the facial images contain not enough information to correctly locate multiple landmarks. In order to solve this problem, we decided to find the registration parameters based on the entire image. In order to find the best registration parameters, we maximize the similarity score from face recognition methods by varying the registration parameters.
This part contains two chapters about face registration:

- Chapter 4 is based on papers [22] and [26]. In this chapter, we perform face registration by maximizing the scores of several face recognition methods. These face recognition methods are all holistic face recognition methods. We show that our registration method on high resolutions outperforms the landmark based registration. We observe that on lower resolution, we still achieve similar results as landmark based registration on high resolutions. We discovered that our registration method sometimes finds a local instead of a global maximum score for the registration and we developed better search strategies to overcome this problem.

- Chapter 5 is based on [24]. In Chapter 4, we found the registration by maximizing the similarity scores from face recognition methods. In this chapter, we used a matching criterion that includes the probability that a face might be misaligned, instead of using the output of the face recognition methods which only measures similarity between faces of different individuals. A second search method is also introduced and more robust features for registration are used. To evaluate the new improvements, a face recognition experiment is performed on the FRGCv2 database. We show clear improvements in comparison with our earlier work and landmark based methods on high resolution images. Our face registration method performs well on low resolution facial images.

4

AUTOMATIC FACE ALIGNMENT BY MAXIMIZING THE SIMILARITY SCORE

4.1 Introduction

Several papers have shown that correct registration is essential for good face recognition performance [95],[17]. The performance of popular face recognition algorithms, for instance PCA, LDA and ICA, depend on accurate face registration. We propose a new method for face registration which searches for the optimal face alignment by maximizing the score of a face recognition algorithm. Our new method outperforms the landmarks methods described in [18]. In this chapter, we investigate the practical usability of the new face registration method for face verification.
In practice, we need to locate the face region using a face detection algorithm. Using this region, we register the face image to a user template in the database and then recognize the face. Our face registration algorithm, first described in [22], finds an optimal face alignment from the located face region. In this chapter, we investigate different practical aspects of our face registration method. We test our face registration method with different face classifiers as evaluation criteria in the search procedure. We test our method under circumstances where lighting is controlled and uncontrolled, and we also lower the resolution of the face images. We investigate if our method works with automatically registered training images,

so it becomes fully automatic. Finally, we look at the mistakes of our registration method and introduce some solutions to overcome these problems.
In the literature, face registration is usually achieved by finding landmarks in a face image. An approach which is similar to our face registration algorithm is described in [68] and [79], which uses a form of robust correlation to find the alignment to a user template. Recently, Wang et al [131] improve the face identification on the the FERET database by calculating the similarity score of different alignments and selecting the maximal score. The main differences with their approach are the assumption of a face identification problem and using the manually labelled eye coordinates as start points.
In section 4.2 we firstly explain our method. Secondly, we specify our search procedure and finally we describe the face recognition algorithms. In section 4.3, we describe the experimental setup we used to evaluate our method. Section 4.4 describes the various experiments carried out using our registration method. The final section gives a conclusion about this face registration method.

4.2 Matching Score based Face Registration

We have developed a new face registration method, namely Matching Score based Face Registration (MSFR). This method searches for the optimal alignment between the probe image and a user template in the database. To evaluate the alignment, we use the output of a face classifier. This output is also called the matching score in the case of a genuine user or the similarity score in the case of a unknown user. We assume that the similarity score becomes better if the alignment of face image to the user template improves. To verify this assumption, we have performed several experiments in which we vary one of the registration parameters (translation, rotation and scale) of a manually registered face image. The resulting similarity score given in Figure 4.1, where four graphs are shown containing the matching score for a single face image while varying one of the registration parameters. These parameters are varied relative to a registration based on manually labelled landmarks. In the graphs this corresponds with scale = 1, angle = 0 and translation in x- and y-direction = 0. Figure 4.1 shows that the manual registration is rather good although the matching score can be improve by using a slightly different translation in x-direction. Our second assumption is that the optimal alignment of the genuine user's face image gives a better similarity score than the optimal alignment of an imposter's face image.

4.2.1 Face Registration

Based on the assumptions described above, we have developed the following method. The region of the face is found by a face detection algorithm, in our case the face detector first described by Viola and Jones [127]. Using an affine transformation $T_{\boldsymbol{\theta}}$ on the pixel p of the probe image I_p, given by the region found by the face detection algorithm, we vary the multiple registration parameters $\vec{\theta}$ searching for the optimal alignment. The geometric transformation function is :

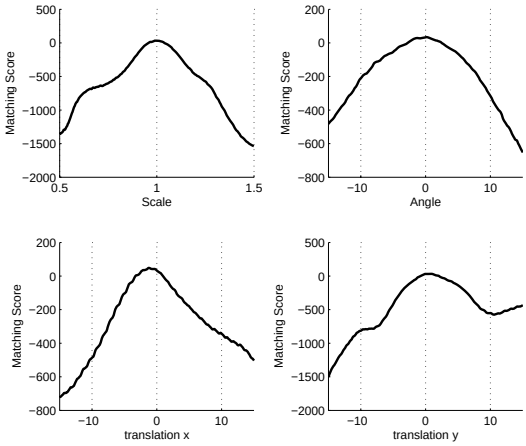


Figure 4.1: Influence on Matching Score when varying the registration parameters - Best Matching Score is achieve near manual registration, maximizing the Matching Score will achieve good registration results

$$\begin{aligned} T_{\boldsymbol{\theta}}(x,y) &= (\theta_1 cos\theta_2 x - \theta_1 sin\theta_2 y + \theta_3, \\ &\quad \theta_1 cos\theta_2 x + \theta_1 sin\theta_2 y + \theta_4) \end{aligned} \tag{4.1}$$

In the transformed image $I_p(T_{\boldsymbol{\theta}}(p))$, the pixel values are calculated using bilinear interpolation. The optimal alignment parameters to person i in the database are given by:

$$\boldsymbol{\theta}_{max} = \arg\max_{\boldsymbol{\theta}\in\Theta} S(I_p(T_{\boldsymbol{\theta}}(p)), i) \tag{4.2}$$

Of course, the similarity score $S(I_p(T_{\boldsymbol{\theta}_{max}}(p)), i)$ can also be used to verify the person's identity. It is also possible to use one face recognition algorithm to find the optimal alignment parameters but another face recognition algorithm to classify the face. A schematic representation of the entire system is given in Figure 4.2.

4.2.2 Search for Maximum Alignment

To search for the maximum similarity score, we use a search algorithm called the downhill simplex method [81]. This search algorithm finds four parameters $\vec{\theta}$ which maximize the similarity score. The starting point of the search algorithm is the region given by the face detection algorithm, where $\boldsymbol{\theta}_0 = (1, 0°, 0, 0)$. For the downhill simplex method, we need to determine a simplex (geometrical figure in N dimensions consisting of $N+1$ points). This is created from the starting point

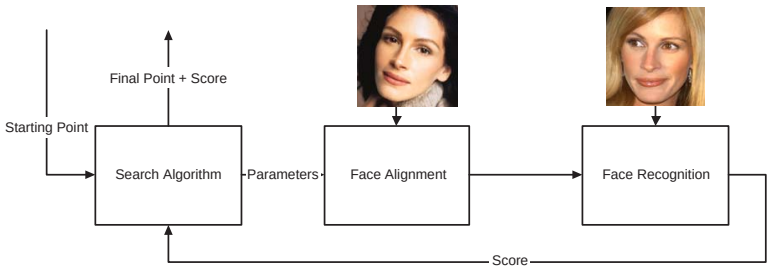


Figure 4.2: Scheme of matching score face registration - Using a search algorithm, we estimate the registration parameters, which give the best score as final point. The starting point is the result obtained from the face detection

parameters and four points for which we varied a single parameter. The other four points of the simplex are: $\boldsymbol{\theta}_1 = (1.2, 0°, 0, 0)$, $\boldsymbol{\theta}_2 = (1, 5°, 0, 0)$, $\boldsymbol{\theta}_3 = (1, 0°, 5, 0)$ and $\boldsymbol{\theta}4 = (1, 0°, 0, 5)$. We have also experimented with other simplexes to start the search algorithm. Details will be given later on in this chapter. We have also experimented with the search algorithm of Powell-Brent [30],[92], but it performs worse for this search problem.

4.2.3 Face Recognition Algorithms

Face recognition involves performing several steps to be able to recognize a face in an image. Using an aligned image $I_p(T_{\boldsymbol{\theta}}(p))$ given by the search algorithm, we select a region of interest (ROI) and we normalize the image inside the ROI to zero mean, unit variance. After that, the pixels in the ROI are vectorized and the resulting vectors are then used in our face recognition algorithm.
We use four algorithms to calculate the similarity score, these algorithms are based on PCA [123] some in combination with LDA [16]. In this chapter, we used a fixed number of PCA and LDA dimensions, 100 and 50 respectively. The first algorithm is PCA in combination with the Euclidean distance (eucl), where we calculate the Euclidean distance between the probe image with the template in the database and use this as similarity score. The second algorithm is PCA in combination with the Mahalanobis distance (mah), where we use the Mahalanobis distance instead of the Euclidean distance. In the third algorithm, we perform feature reduction using PCA and LDA and use the log-likelihood ratio proposed in [126] to calculate the similarity score. For a certain class i, the similarity score S is calculated by:

$$\begin{aligned} S_{\boldsymbol{y},i} &= -(\boldsymbol{y} - \boldsymbol{\mu}_{W,i})^T \Sigma_W^{-1} (\boldsymbol{y} - \boldsymbol{\mu}_{W,i}) \\ &\quad + \boldsymbol{y}^T \Sigma_T^{-1} \boldsymbol{y} - \log|\Sigma_W| + \log|\Sigma_T| \end{aligned} \tag{4.3}$$

Where $\boldsymbol{y}$ is a vector which is a representation of the face image after feature reduction, Σ_T is the total covariance matrix, Σ_W is the within class covariance

matrix and $\boldsymbol{\mu}_{W,i}$ is the ith class average. The final algorithm uses the numerator of the likelihood ratio, which is given by:

$$S_{\boldsymbol{y},i} = -(\boldsymbol{y} - \boldsymbol{\mu}_{W,i})^T \Sigma_W^{-1} (\boldsymbol{y} - \boldsymbol{\mu}_{W,i}) - \log |\Sigma_W|$$

The reason behind using only the numerator is that we register to a certain user template. This means that it is not necessary to maximize with respect to the background distribution, given by the denominator in the likelihood ratio. We call this final method the within ratio. This method is only intended for finding a maximal alignment to a user template and not for face recognition. After the face registration, a final face verification is always performed using the likelihood ratio.

4.3 Experimental Setup

In our experiments, we use the Face Recognition Grand Challenge (FRGC) version 1 database [83]. We only used face images in which the face was correctly found by the face detection algorithm of Viola-Jones [127], because we are not interested in the mistakes made during face detection. The face images are correctly found when the eyes, nose and mouth coordinates lie inside the face region and the width and height of this region are less then four times the distance between the eyes. The FRGC version 1 database contains 275 individuals, from which we use a set of 3761 face images taken under controlled conditions and a set of 1811 face images taken under uncontrolled conditions. In our experiments we randomly split these sets into two subsets, each consisting of approximately half of the images of each person. One subset is used for training and enrollment and the other is used for testing. The same random split is used for all experiments.
We compare our face registration method with the best landmark based face registration methods in [18], namely MLLL + BILBO. The results of the face registration are measured on the performance in face verification by calculating Equal Error Rate (EER): this is the point of operation where the False Accept Rate (FAR) is equal to the False Reject Rate (FRR). To measure the accuracy of registration we use the RMS error. We calculated the location of the eyes, nose and mouth in the original image based on the alignment found by MSFR. The RMS error is then calculated between these positions and the manually labelled landmarks given by the FRGC database and we normalize to a distance of 100 pixels between the eyes.

4.4 Experiments

Since our earlier paper [22], we have performed more extensive experiments. We have done several experiments to gain a better understanding of our method. In our first experiment we report the results of the different recognition algorithms on

	FRGC Conrolled vs Controlled			FRGC Uncontrolled vs Uncontrolled		
Registration Method	EER [%]	RMS error users	RMS error impostors	EER [%]	RMS error users	RMS error impostors
Manually labelled	0.59	-	-	1.7	-	-
MLLL+BILBO	3.6	7.9	7.9	9.7	10.2	10.2
PCA eucl	1.8	3.4	9.5	3.2	4.3	10.3
PCA mah	1.3	3.0	5.7	**1.5**	2.9	4.2
Likelihood ratio	**1.01**	3.2	9.4	2.3	4.0	7.9
Within ratio	**1.07**	3.2	8.7	2.1	3.8	7.2

Table 4.1: Results of the face verification using various registration algorithms

the FRGC database. In the second experiment, we use a lower resolution making the algorithm faster and applicable to video surveillance environments. The third experiment investigated if this face registration algorithm can be trained on face images which are registered using automatically obtained landmarks. The final experiments try to address some failures of the search algorithm by performing the search several times and adding registration noise to the training data.

4.4.1 Comparison between recognition algorithms

Searching for the best alignment requires a recognition algorithm. In this section, we describe our experiments using the various recognition algorithms. We compare the results with both manually labelled landmarks and automatically obtained landmarks. For our experiment, we use the experimental setup described in section 4.3. We use face images with a resolution of 128×128 pixels. Training the face classifier is achieved using a training set which is aligned using the manually labelled landmarks. Face recognition applied to images registered using MLLL + BILBO, however, is train also on images which were registered using MLLL + BILBO [18]. After registering the face, we recognize the registered faces using the Likelihood ratio classifier.

In Table 4.1, we compare the results of the various registrations in EER and RMS error. The columns for the EER show that the MSFR outperforms the landmark registration. By comparing the various classifiers of MSFR, it becomes clear that the likelihood ratio and the within ratio perform best on the controlled images of FRGCv1, although the performance of PCA with Mahalanobis distance is also very good. On the uncontrolled image of FRGCv1, PCA with Mahalanobis distance performs best, even better than the manually labelled landmarks. In RMS error, the various MSFR methods are more accurate than MLLL+BILBO when it comes to registration to the genuine user. If we look at registration of an imposter, the RMS error of most of the MSFR is higher than the RMS error of MLLL + BILBO. This does not need to have any effect on the EER,

because poorly registered images usually do not improve the similarity score. In the case of PCA with Mahalanobis distance, the RMS error of the impostor is still lower than that of MLLL+BILBO. Figure 4.3 shows the FAR and FRR curves

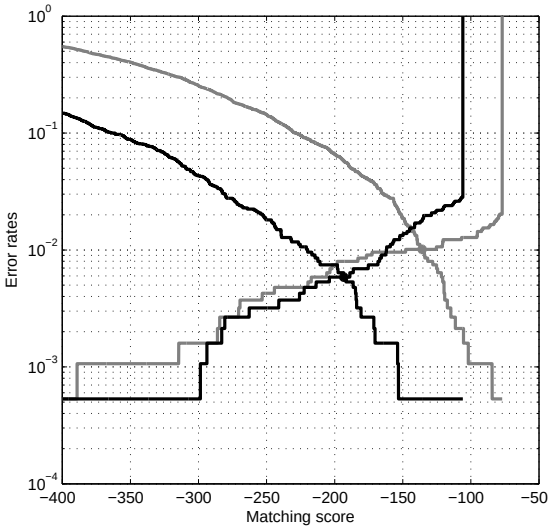


Figure 4.3: Registration Effects on similarity scores - FAR and FRR curves: the grey line is the Matching Score Face Registration and the black line is the manual registration

of the manually labelled landmarks and the MSFR with the likelihood ratio on the controlled images of the FRGC. Both the matching and non-matching scores increase when using MSFR, which means that for genuine users, better alignments than manually labelled landmarks can be found using MSFR.

4.4.2 Lowering resolution

In [22] we report that our method takes about 20-30 seconds to register and classify a face image using an Intel Pentium 2.80 GHz. Currently, it takes about 5-10 seconds on the same computer for a face image of 128×128 pixels, because we optimized some of our source code. In [21] we investigated the effect of the image resolution on face recognition. It turns out that the EER does not increase much on face recognition by lowering the resolution to 32×32. In practice, we do not always have high resolutions face images, so we performed a experiment at a resolution of 32×32 pixels, which also leads to a decrease in computation time. The results of the recognition are stated in Table 4.2 together with the results of using the normal resolution of 128×128 pixels.

Although Table 4.2 shows that the EER increases somewhat by using face images of 32×32 pixels, these results are still acceptable and better than face registration

	FRGC Conrolled vs Controlled		FRGC Uncontrolled vs Uncontrolled	
Registration Methods	resolution 128×128	resolution 32×32	resolution 128×128	resolution 32×32
PCA eucl	1.8	2.4	3.2	5.5
PCA mah	1.3	2.3	1.4	2.8
Likelihood ratio	1.01	1.3	2.3	3.8
Within ratio	1.07	1.7	2.1	3.6

Table 4.2: The EERs by using face images with a resolution of 32×32

Registration Methods	EER [%] manual	EER [%] automatic
PCA eucl	1.8	2.0
PCA mah	1.3	1.5
Likelihood ratio	1.01	1.7
Within ratio	1.07	1.8

Table 4.3: The EERs when the training en enrollment are registered using MLLL + BILBO on images from the controlled set of FRGCv1

using MLLL + BILBO. It takes about 2-5 seconds to register and classify a face image of 32×32 pixels on a Intel Pentium 2.80 GHz. More improvements in the operation time of our method can be realised, because we have not payed much attention to this subject yet.

4.4.3 Training using automatically obtained landmarks

Until now, we have assumed that for training and enrollment of the face registration we can use a set of manually labelled face images. In practice, this usually is not the case, especially for the enrollment of a new user. This is the reason we performed an experiment where we trained and enrolled images which have been aligned using the landmarks given by MLLL + BILBO. The results of this experiment are given in Table 4.3. Although the results we report in table 4.3 show increased error rates, the performance is still much better than using MLLL + BILBO for face registration. This shows that some small mistakes in the registration of training set do not have a large influences on recognition results.

4.4.4 Improving maximization

In this chapter, we already reported a large improvement in the results of face registration. But after some analysis of our method, we found that a correct registration is not always found by simply running the downhill simplex search algorithm. The main reason is that the search algorithm can find a local maximum far away from the global maximum. In figure 4.4, we show the incorrectly found registration results by the likelihood ratio classifier. These results can easily be determined by considering the RMS error of face images. The faces depicted in figure 4.4 all have a RMS error bigger than 11 pixels, except for the bottom right face image. The main reason for these errors is that in these cases, the search algorithm searches in the wrong direction and gets stuck in a local maximum. To

Figure 4.4: Mistakes in Registration - Face images which have been incorrectly registered, the bottom-right image is an example of a correct alignment

correct these outliers, we have developed two strategies to address this problem. Firstly, we use the downhill simplex method several times but start with a different simplex in the search space. Secondly, we change the search space by training on a database with some registration noise.

4.4.4.1 Using a different start simplex

The first strategy is based on the idea that if we start searching from another side in the search space, we will probably never come across the same local maximum. For this experiment, we have defined two new start simplexes; the first simplex consists of the points: $\boldsymbol{\theta}_0 = (1, 0°, 0, 0)$, $\boldsymbol{\theta}_1 = (0.8, 0°, 0, 0)$, $\boldsymbol{\theta}_2 = (1, -5°, 0, 0)$, $\boldsymbol{\theta}_3 = (1, 0°, -5, 0)$ and $\boldsymbol{\theta}_4 = (1, 0°, 0, -5)$, so that we start from the opposite side of the search space. For the second simplex, we start at the points: $\boldsymbol{\theta}_0 = (0.9, -2.5°, -2.5, -2.5)$, $\boldsymbol{\theta}_1 = (1.1, -2.5°, -2.5, -2.5)$, $\boldsymbol{\theta}_1 = (0.9, 2.5°, -2.5, -2.5)$, $\boldsymbol{\theta}_2 = (0.9, 2.5°, -2.5, -2.5)$, $\boldsymbol{\theta}_3 = (0.9, -2.5°, -2.5, 2.5)$ and $\boldsymbol{\theta}_4 = (0.9, -2.5°, -2.5, 2.5)$,

Registration Methods	start position 1	start position 2	start position 3	combining 1,2,3
PCA eucl	1.8	7.4	3.5	3.6
PCA mah	1.3	1.6	5.2	**0.64**
Likelihood ratio	1.01	2.6	6.7	**0.59**
Within ratio	1.07	2.4	6.5	**0.64**

Table 4.4: The EERs when searching from different positions in the search space on images from the controlled set of FRGCv1

which gives us a search area around the located face region. In Table 4.4 we present the results of the three different start positions. The EERs 2 and 3 in table 4.4 are the new start positions, while EER 1 gives the start position which has been used throughout the entire chapter. Also, the results of combining these outcomes of registration by using the maximum similarity score of the three different start positions is given in table 4.4, this procedure is done for both genuine users and impostors. These combinations give results which are similar to registration using manually labelled landmarks. The EERs of the other start positions are not particularly good, but using other starting points results in different failures. By using the similarity score to evaluate the final outcomes of the different starting points, the local maxima are discarded.

4.4.4.2 Adding noise to train our registration method

The second strategy is based on changing the search space. This is done by adding Gaussian noise to the manually labelled landmarks of the training examples. By adding noise to the training, we also hope that we can model the registration error better. The results of this experiment are given in Table 4.5 where the t is the standard deviation of the noise in pixels, normalized to 100 pixels between the eyes. In Table 4.5 we show that by combining the results of adding registration noise to the training set we reach the same result as with manually labelled landmarks. Another observation is that by adding a little registration noise to the training, the EER seems to decrease anyway. This is because a reduction in the number of outliers. We suspect that the registration noise makes the search space smoother in the areas further away from the optimal registration. By adding too much noise, the EER increases. This can be observed for $t = 5$ in table 4.5.

4.5 Conclusion

We present a system for face registration which uses the output of the face recognition classifier to find an optimal registration. We search for an optimal registration

Registration Methods	noise						combining $t = 0, 1, 2$
t	0	1	2	3	4	5	
PCA eucl	1.8	1.5	1.5	1.8	1.3	1.5	1.9
PCA mah	1.3	0.91	0.85	0.80	0.69	1.2	**0.80**
Likelihood ratio	1.01	0.69	0.75	0.91	0.91	1.3	**0.59**
Within ratio	1.07	0.64	0.75	0.91	0.96	1.2	**0.64**

Table 4.5: The EERs when adding noise to the registration of the training on controlled face image of FRGCv1

by varying the face alignment parameters. Our new face registration method performs better than the landmark based methods of [18]. The experiments show that our method performs well with both face images taken under controlled as well as uncontrolled conditions. The operating speed of the method has been improved and we show that lowering the resolution improves the speed even more while still obtaining good performance. Our face registration method also works with an automatically registered training set and achieves good results despite registration errors in the training set. This makes our face registration method very useful in practise when dealing with a face verification problem. By using multiple searches, the results of our face registration method are equivalent to the results obtained with registration using manually labelled landmarks. This kind of performance has not yet been achieved by any fully automatic face registration method known to us.

5

SUBSPACE-BASED HOLISTIC REGISTRATION FOR LOW RESOLUTION FACIAL IMAGES

5.1 Introduction

Face recognition in the context of camera surveillance is still a challenging problem. For reliable face recognition, it is crucial that an acquired facial image is registered to a reference coordinate system. Most conventional registration methods are based on landmarks. To locate these landmarks accurately, high resolution images are needed. For those methods, it is problematic to register low[1] resolution facial images as obtained in video surveillance. Face registration on low resolution images is in these cases often omitted and the region found by the face detection is directly used for face recognition [2; 12]. In our opinion, accurate face registration can contribute to better recognition performance on low resolution images. Therefore, we developed a Subspace-based Holistic Registration (SHR) method, which uses the entire face region to correct for translation, rotation and scale transformation of the face, which enables us to accurately register low resolution facial images. The face registration is performed after a frontal face detector, which

[1] In the Face Recognition Vendor Test [89], low resolution face images are defined to contain an interocular distance of 75 pixels, we used even lower resolutions with interocular distances of 50 pixels and lower. High resolution face images have an interocular distance of more than 100 pixels

detects a face at a certain scale and rotation variations, limiting the search for the final registration parameters.

As already pointed out above, registration methods can be divided into two categories: landmark-based registration, using landmarks to register the face image, and holistic registration, using the entire image for registration. Of the latter only a few methods have been reported.

In the first category, the object detection method of Viola and Jones [127], originally proposed for face detection, is a popular approach to locating landmarks [32; 34; 41]. The advantages of this method are that it is fast and robust in comparison with other landmark methods. Many papers report good results especially in uncontrolled scenarios. However, occasionally landmarks are not found by this method. In [80], a probabilistic approach using Principal Component Analysis (PCA) is used to locate the landmarks. Subspace methods for facial feature detection are also used in [15; 18; 46]. Some landmarking techniques are not only based on texture, but also use geometric relations between landmarks, for instance [6; 98; 108; 133]. These methods usually require more landmarks and high resolution facial images. A well-known example of such a method is Elastic Bunch Graphs [133]. Elastic Bunch Graphs are used to determine the relation between different landmarks. The relation between the landmarks and the scores of Gabor Jets are combined to register and recognize the face. Active Shape Models [37] and Active Appearance Models [38] can also be used to perform a fine registration of a face, by using both texture and the relation between the landmarks. Both methods, however, need a good initialization to find an accurate registration, which can be provided by, for instance, the Viola and Jones landmark finding method.

In the second category of registration, there are correlation-based registration methods that are invariant to translation. The MACE filter originally described in [77] and used in face recognition in [100; 102], is invariant for translations. In [65], a face registration method using super resolution is described that performs correlation to compare the original image with a reconstructed image obtained using super resolution, correcting for translation and scale variations. The method described in [68] and [79] is a correlation based method that finds a rigid transformation to align the facial images, which is done using robust correlation to a user template.

Another way of evaluating the registration quality is by using the similarity score determined by a face recognition algorithm. In [131], the manually labelled eye coordinates are used as a starting point from which the eye coordinates are varied to obtain different registrations. The registration that resulted in the best similarity score is selected. This experiment was performed using several different face recognition algorithms. In [115], we performed a similar experiment and in addition showed that small changes in the registration parameters can have a huge effect on the similarity scores of face recognition algorithms. In [22] and [26], we proposed a matching score based face registration approach, which searches for the optimal alignment by maximizing the similarity score of several holistic face recognition algorithms, e.g. PCA Mahalanobis distance. In [76], the PCA Mahalanobis distance is used to find the registration parameters for low resolution images using a different search strategy as in [26], where the focus of the paper is

face hallucination. In [64], this face registration method is extended especially for the purpose of face hallucination. We performed no experiment using face hallucination, because our focus is on face registration and its effect on the recognition. In this paper, we extended the work in [22; 26], by developing Subspace-based Holistic Registration (SHR) method. The novelty of this method is that we use a probabilistic framework designed to evaluate the registration of faces, instead of maximizing the score of a face recognition method, which might not be suited for comparing unregistered face images.

5.2 Face Registration Method

5.2.1 Subspace-based Holistic Registration

Face registration is performed to correct for variations that occur when the face region is selected from an image. We assume that the face detection obtains frontal faces from a camera and that we have to correct for in plane rotations of these faces. The exact positions of the camera and the face are usually unknown, making a correction for scale and translation necessary as well. A Procrustes transformation denoted by $T_{\boldsymbol{\theta}}$ corrects for these variations, allowing us to scale by a factor s, rotate with an angle α and translate over a vector $\mathbf{u}$ an image. The optimal face registration is assumed to be found if there is a maximum similarity between the transformed input image (probe image) and the gallery images. In SHR, we try to find the best registration parameters $\boldsymbol{\theta} = \{\mathbf{u}, \alpha, s\}$, by maximizing a similarity function $S(T_{\boldsymbol{\theta}}H, K|\Omega)$. Here H denotes the probe image, which is transformed by $T_{\boldsymbol{\theta}}$, K denotes a registered reference object (gallery image) and Ω denotes a model of the reference object (faces). The equation for finding the best registration parameters $\hat{\boldsymbol{\theta}}$ is:

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} S(T_{\boldsymbol{\theta}}H, K|\Omega) \tag{5.1}$$

An important issue is how to measure the similarity between probe and gallery image. In our previous work, we used similarity scores from well-known face recognition algorithms for this purpose. However, these scores are usually optimal for face recognition, measuring the similarity between faces of different individuals in a face space. In this paper, we argue that the correct quantifier for the face registration should also include the probability that the face might be misaligned, measuring also the error outside face space. We thus use the probability that the aligned image $T_{\boldsymbol{\theta}}H$ belongs to the object class Ω of the gallery image K. Let V be an operator that vectorizes the features in H and K using a set of predefined locations $\{\mathbf{p}_n\}_{n=1}^{N}$ in the images. We adopt a Gaussian model of which VK is the mean and Σ_Ω the covariance matrix

$$S(T_{\boldsymbol{\theta}}H, K|\Omega) = \mathcal{N}(VT_{\boldsymbol{\theta}}H|VK, \Sigma_\Omega) \tag{5.2}$$

Our goal is to optimize $S(T_{\boldsymbol{\theta}}H, K|\Omega)$ as function of the registration parameters $\boldsymbol{\theta}$. For notational compactness, we define $\mathbf{x} = VT_{\boldsymbol{\theta}}H$ and $\bar{\mathbf{x}} = VK$ and

$$P(\mathbf{x}|\Omega) \stackrel{\text{def}}{=} \mathcal{N}(VT_{\boldsymbol{\theta}}H|VK, \Sigma_\Omega) \quad (5.3)$$

$$= \frac{\exp\Big(-\frac{1}{2}(\mathbf{x}-\bar{\mathbf{x}})^T\Sigma_\Omega^{-1}(\mathbf{x}-\bar{\mathbf{x}})\Big)}{(2\pi)^{N/2}\cdot|\Sigma_\Omega|^{1/2}} \quad (5.4)$$

The training samples $\mathbf{x}$ to determine both the mean $\bar{\mathbf{x}}$ and covariance matrix Σ_Ω are correctly aligned images. Notice that K needs to be a registered image in order to find the registration parameters $\boldsymbol{\theta}$ for H. The exact estimation of the covariance matrix Σ_Ω is not possible with a limited number of training samples. As a consequence, the estimate of Σ_Ω is often singular, so that Σ_Ω^{-1} cannot be computed and even if Σ_Ω^{-1} can be calculated, the results will be inaccurate. Furthermore, the computational costs of evaluating Equation 5.4 are large, due to the high dimensionality of Σ_Ω and $\mathbf{x}$. For these reasons, we use Principal Component Analysis (PCA) to reduce the dimensionality. We obtain a subspace by solving the eigenvalue problem:

$$\Lambda = \Phi^T\Sigma_\Omega\Phi \quad (5.5)$$

where Λ are the eigenvalues and Φ are the eigenvectors of the covariance matrix Σ_Ω. We can obtain a reduced feature vector $\mathbf{y} = \Phi^T\tilde{\mathbf{x}}$, where $\tilde{\mathbf{x}} = \mathbf{x} - \bar{\mathbf{x}}$. The principal subspace $F = \{\Phi_i\}_{i=1}^M$, which reduces the feature vector from N to M dimensions, has an orthogonal complement $\bar{F} = \{\Phi_i\}_{i=M+1}^N$, which contains the variations that are not modelled by PCA. Using only similarities in the principal subspace as in our previous work [26], results in the Mahalanobis distance. However, if we optimize the alignment only for the principal subspace F, we might walk further away in the orthogonal complement $\bar{F}$, ignoring details not included in our model but which indeed might be important for the registration. To overcome this problem, we use a distance measure, proposed in [80].

$$\epsilon^2(\mathbf{x}) = \sum_{i=M+1}^{N} y_i^2 = ||\tilde{\mathbf{x}}||^2 - \sum_{i=1}^{M} y_i^2 \quad (5.6)$$

$$\hat{d}(\mathbf{x}) = \sum_{i=1}^{M}\frac{y_i^2}{\lambda_i} + \frac{\epsilon^2(\mathbf{x})}{\rho} \quad (5.7)$$

where λ_i are the eigenvalues in F and $\rho = \frac{1}{N-M}\sum_{i=M+1}^{N}\lambda_i$ which is the average eigenvalue in $\bar{F}$. This distance measure consist of two parts, the first $\sum_{i=1}^{M}\frac{y_i^2}{\lambda_i}$ is called "distance-in-feature-space" (DIFS) and the second $\frac{\epsilon^2(\mathbf{x})}{\rho}$ is called "distance-from-feature-space" (DFFS). In our experiments, we compare the results of using only DIFS for face registration, which is used in [26; 76], and using both DIFS and DFFS (see section 5.4.1). We show that using both distances result in a better performance than using DIFS

In Figure 5.1, we give a schematic representation of the components needed for

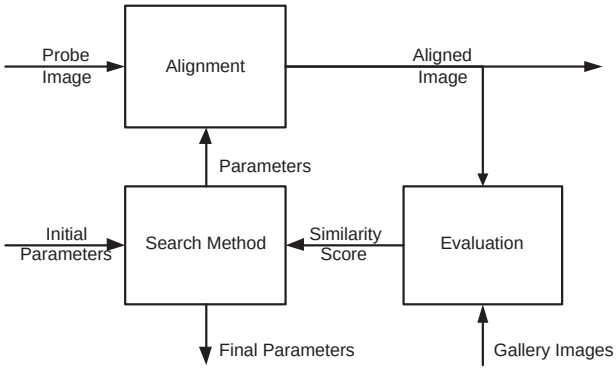


Figure 5.1: Schematic representation of SHR - Flow diagram of our iterative method

SHR and the interaction between them. We use an iterative search method to find the optimal similarity between probe image and gallery images. The initial registration parameters are given by a face detection algorithm, for instance the method of Viola and Jones [127]. The alignment registers the probe image based on the specified parameters. We will discuss the components in Figure 5.1 in the following sections: evaluation (Section 5.2.2), the alignment (Section 5.2.3) and the search methods (Section 5.2.4).

5.2.2 Evaluation

Two important issues in the evaluation function are the model and the features. The model can be either user independent as explained in the previous section or user specific. This we will discuss in the first paragraph below. As features, we propose edge images, instead of grey level images, which reduce the number of local minima in the evaluation. This will be explained in the second paragraph.

5.2.2.1 Evaluation to a user specific face model

Instead of registration to a mean face model, which may differ substantially from individual faces, registration to a user specific model, if available may improve registration results. For user specific face registration, we need a user template to register a probe image. For face identification, user specific registration has the drawback that we have to register the probe to every user template in the database.
For user specific registration, we define the similarity measure $S(T_{\boldsymbol{\theta}}H, K_c|\Omega_c)$, where Ω_c models registered facial images of user c. The user specific model consists of a user template K_c and the covariance matrix Σ_{Ω_c}. For the covariance matrix Σ_{Ω_c}, we use a within class covariance matrix that models the variations among face images of the same person for all users, because we often do not have enough images to estimate a user specific covariance matrix. The similarity function for

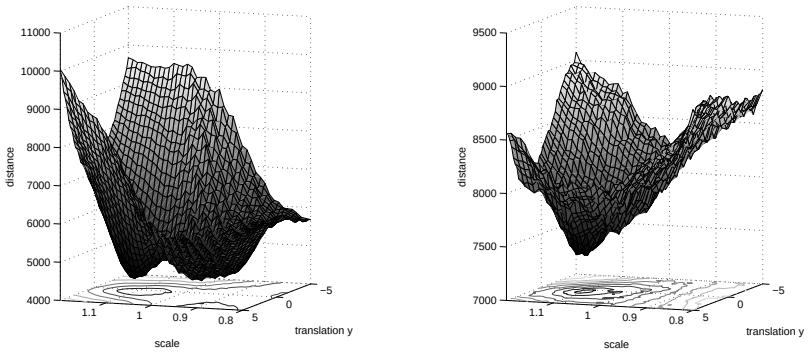


Figure 5.2: Search space for grey level image and edge image - A 2D search space based on the grey level image (left) and edge image (right), for scale (left-right) and translation in y direction (front-back), showing a local minimum in the left score landscape

the user specific model is

$$S(T_{\boldsymbol{\theta}}(H), K_c; \Omega_c) \quad = \quad \mathcal{N}(VT_{\boldsymbol{\theta}}H; VK_c, \Sigma_{\Omega_c}) \tag{5.8}$$

5.2.2.2 Using edge images to avoid local minima

Using grey level images for registration often leads to local minima in the search space. Better registration results can be obtained by using edge images, which is for instance shown in [39] for Active Appearance Models. In image registration, regions containing large variations (structure) contribute more to registration than homogeneous regions. By applying edge filters, the regions that contain structure will be highlighted and the homogeneous regions will be suppressed. In our case, the use of edge filters results in a search space with fewer local minima. In Figure 5.2, a 2D search space is shown where we varied the scale and translation in y-direction of a grey level image and an edge image. The edge image (right) shows a single clear minimum, while the grey level image has a global minimum at the same place, but also a large local minimum in the right corner. In order to calculate the edges in the image, we take the derivatives in the x and y directions in the images. Because images usually contain noise, we use the Gaussian kernels G_x and G_y:

$$\begin{aligned} G_x(x,y) &= \frac{-x}{2\pi\sigma^4}\exp(-\frac{x^2+y^2}{2\sigma^2}) \\ G_y(x,y) &= \frac{-y}{2\pi\sigma^4}\exp(-\frac{x^2+y^2}{2\sigma^2}) \end{aligned} \tag{5.9}$$

The derivatives H_x and H_y of the images are calculated by convolution. We refer to these as 'edge images'. If we use both edge images in the feature vector

instead of the grey level image, this doubles the length of the feature vector, resulting in increased computation time. An alternative is to combine the two edge images as follows into a 'magnitude image'

$$H_{\text{mag}} = \sqrt{H_x^2 + H_y^2} \tag{5.10}$$

The default features used in this paper are the 'edge images' and a comparison between the features is performed in Section 5.4.1.

5.2.3 Alignment

We use a Procrustes transformation to align the probe image H to the gallery images, which is common practice in face recognition, preserving the distance ratios. Given the pixel location $\mathbf{p} = (x, y)^T$, we can define a transformation $U_{\boldsymbol{\theta}}\mathbf{p}$ on the pixel location as follows:

$$U_{\boldsymbol{\theta}}\mathbf{p} = sR(\alpha)\mathbf{p} + \mathbf{u} \tag{5.11}$$

$R(\alpha)$ is the rotation matrix. The transformation of the image is defined as:

$$T_{\boldsymbol{\theta}}H(\mathbf{p}) = H(U_{\boldsymbol{\theta}}^{-1}\mathbf{p}) \tag{5.12}$$

This allows us to obtain an aligned image $T_{\boldsymbol{\theta}}H(\mathbf{p})$ by backward mapping and interpolation. Most landmark based methods also perform this transformation based on the found landmarks in order to obtain a registered face image [108].

5.2.4 Search Methods

In Equation 5.1, we have to maximize the similarity score to find the best alignment parameters $\boldsymbol{\theta}$. Ideally, an iterative search method should be able to find the optimal solution using a small number of evaluations, making it possible to register the probe image almost real-time. The search method also has to be robust against local minima. Confirmed by our observations, we assume reasonably smooth search landscapes. We applied two different search methods the first is the downhill simplex method [81] that we also used in [22] and [26], and the second is a gradient based method.

5.2.4.1 Downhill Simplex search method

This method is able to maximize a similarity function using around 100 evaluations. A good initialization of the downhill simplex method is necessary to be robust against local minima. This was also observed in [26], where we used several initializations to reduce outliers. To initialize the downhill simplex method, we need to create a simplex $\Theta \in \Re^{N\times(N+1)}$ (geometric shape in N dimensions, consisting of $N + 1$ points). To obtain the four registration parameters, this means

that we have to select five starting points. The first starting point is given by the initial parameter vector $\boldsymbol{\theta}_0$. The other starting points are given by

$$\Theta = \begin{bmatrix} \boldsymbol{\theta}_0 & \boldsymbol{\theta}_0 \pm \Delta s & \boldsymbol{\theta}_0 \pm \Delta\alpha & \boldsymbol{\theta}_0 \pm \Delta x & \boldsymbol{\theta}_0 \pm \Delta y \end{bmatrix} \tag{5.13}$$

where Δ is the maximum expected offset for a single registration parameter in positive or negative direction, where we use the offset which gives the best similarity score. The downhill simplex methods is however able to find optimal registration parameters that lay outside the maximum expected offsets. This search method maximizes the similarity function by replacing those registration parameters in the simplex that give the worst similarity score by a better set using some simple heuristics.

5.2.4.2 Gradient based search method

In Equation 5.1, we find the best alignment parameters $\hat{\boldsymbol{\theta}}$ by maximizing the similarity score. We start with the initial registration parameters $\boldsymbol{\theta}_0$, improving these parameters means that we have to determine an offset to the optimal alignment called $\boldsymbol{\delta}$ [11; 52]. We achieve this by expanding the image using a first order Taylor expansion:

$$T_{\boldsymbol{\theta}_k+\boldsymbol{\delta}_k} H \simeq T_{\boldsymbol{\theta}_k} H + \mathbf{M}_{\boldsymbol{\theta}_k} \boldsymbol{\delta}_k \tag{5.14}$$

In this case, $\mathbf{M}_{\boldsymbol{\theta}}$ is the Jacobian matrix of $\mathbf{H}$ with respect to the parameters $\boldsymbol{\theta}$, given in [52] for a transformation with translation, rotation and scale . By setting the derivative of Equation 5.2 with respect to $\boldsymbol{\delta}$ to zero, we can determine the offset from the original parameters:

$$\frac{\partial}{\partial \boldsymbol{\delta}_k} S(T_{\boldsymbol{\theta}_k} H + \mathbf{M}_{\boldsymbol{\theta}_k} \boldsymbol{\delta}_k, K|\Omega) = 0 \tag{5.15}$$

In Appendix 5.6, it is shown how this equation is solved and how updated parameters $\boldsymbol{\theta}_{k+1} = \boldsymbol{\theta}_k + \boldsymbol{\delta}_k$ are obtained analytically. This procedure is repeated until convergence has been reached

5.3 Experiments

In this section, we describe experiments to evaluate the performance of SHR. The main purpose of SHR is to improve the face recognition performance, particularly at low resolutions. The goal of the experiments, therefore, is to demonstrate and quantify the improvement of face recognition performance if SHR is used for face registration. We will present results of the following comparisons:

- Comparison with earlier versions of SHR [26]. These experiments are included to illustrate the positive effect of the new evaluation criteria given in Equation 5.7 and of the features discussed in (Section 5.2.2.2).

- Comparison with landmark based registration based on automatically detected landmarks as well as on manual landmarks

- Comparison between user independent and user specific registration.

- Comparison between two search methods (Section 5.2.4) in both performance and computation time.

- Comparison of SHR performed on lower resolutions.

5.3.1 Experimental Setup

5.3.1.1 Face Database

To perform the experiments, we use the Face Recognition Grand Challenge version 2 (FRGCv2) database [90], on which we perform the one-to-one controlled versus controlled experiments. We train both face registration (landmark methods and SHR) and face recognition methods on the training set defined in the FRGCv2. We calculated all the similarity scores, which resulted in the Receiver Operating Characteristic (ROC) of the entire set and the ROC of the three masks defined by the FRGCv2 database. Mask I compares the images that are recorded within a semester, for Mask II this is within a year, while Mask III compares images that are recorded between semesters. To compare the different settings of SHR, we use a random subset to reduce computational costs of the face recognition. We still register every gallery and probe image but instead of computing all the scores, we calculate for every probe image one genuine and one impostor score from a randomly chosen image in the gallery. The same random images are used for all the experiments. We show in Table 5.1, that the recognition results of the random subset are comparable to the results on the entire set.

5.3.1.2 Face Detection

Face registration depends on the input of a Face Detection method. We used the OpenCV implementation [61; 74] of the Viola and Jones algorithm [127] to find the faces. We used the pretrained model called 'haarcascade_frontalface_default.xml'. In order to avoid misdetections, we included some simple heuristics[1] based on the manually labelled landmarks to determine if the face regions were correctly found. Facial images in which the face is not correctly found are removed from all experiments.

[1] All landmarks have to be inside the face region and the width and height of this region is less than four times the distance between the eyes

	Mask I	Mask II	Mask III	Entire Set	Random Subset
PCA Mah	54.0%	48.8%	42.9%	50.3%	52.2%
PCA MahCos	72.4%	67.2%	61.8%	68.2%	69.8%
Adaboost	91.4%	88.3%	84.9%	88.9%	89.5%
PCA LDA	92.1%	90.4%	88.6%	90.8%	91.0%

Table 5.1: The verification rates at FAR = 0.1% of several Face Recognition Methods which allow us to compare the registration methods, these verification rates are achieved using manually registered images

5.3.1.3 Low Resolution

SHR is developed for low resolution images. Because there are no large low resolution face databases, we used the FRGCv2 database and created low resolution facial images by lowpass filtering and subsequent downsampling. Using low resolution facial images makes the comparison of the performance of our face recognition methods with the state of the art difficult, because these are primarily focussed on high resolution facial images. Also, landmark based registration methods work poorly on these resolutions. For this reason, we performed the landmark finding on high resolutions images, thus given them an advantage over SHR.

5.3.1.4 Face Recognition

We measured the performance of face registration by its effect on face recognition. In [130], a similar comparison is performed on the FRGC database, where the baseline PCA and PCA-LDA face recognition methods are used. We decided to use not only holistic but also feature based methods, in order to demonstrate that different face recognition methods benefit from improved registration. We used our own implementation of the following face recognition methods:

- PCA Mahalanobis distance (baseline) [88]
- PCA Mahalanobis Cosine distance [88]
- Adaboost with Local Binary Patterns (LBP) [137]
- PCA LDA likelihood ratio [126]

In Table 5.1, we show the face recognition results with an interocular distance (distance between centers of the eyes) of 50 pixels using registration with manually labelled landmarks, showing the capacity of the face recognition methods if the registration is almost perfect. This is confirmed by [130], where their registration method is not able to perform better than manually registered images. From the results in Table 5.1, we observe that of the selected face classifiers, the PCA-LDA

likelihood ratio performs best, closely followed by Adaboost with LBP. SHR is developed for low resolution images using an interocular distance of 50 pixels instead of the available 350 pixels, this makes comparison with other results published on these databases difficult. In Figure 5.3, we attempt to show the relation between

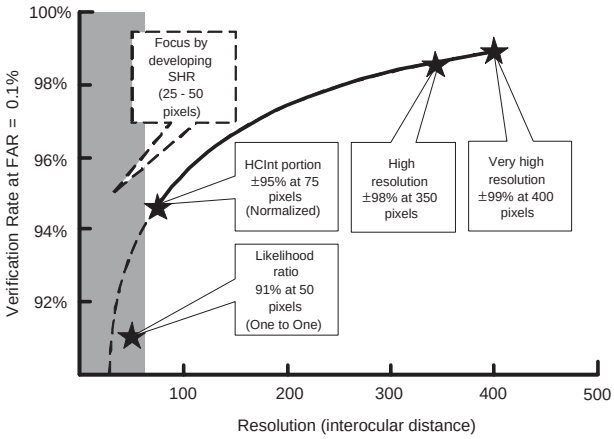


Figure 5.3: Expected results for low resolution face recognition - Best verification rates reported during the FRVT 2006, where we show that our focus is at even lower resolutions expecting a slightly lower verification rat

resolution and verification rate. Below approximately 50 pixel interocular distance, we expect that the verification rate decreases rapidly. At least part of this decrease is caused by failing registration at low resolutions, which we address in this paper. The area of interest for camera surveillance is the shadowed area in figure 5.3 and the stars mark the published results. In [89], an experiment is performed on a low resolution database called HCInt portion of the FRVT 2002 (not available to us), which uses an interocular distance of 75 pixels. The best verification rate reported on the HCInt portion are ± 95% at FAR = 0.1% for a gallery normalized experiments. Our best face recognition method gave a verification rate of 91% at FAR = 0.1% for an interocular distance of 50 pixels with an one-to-one experiment, which is more difficult than a gallery normalized experiment. This matches the expectations we have of good results that can be obtained using face recognition on facial images with an interocular distance of 50 pixels. In [91], a verification rate of ± 67% at FAR = 0.1% was reported for the PCA Mahalanobis distance classifier on the high resolution experiments. For the same classifier, we obtained a verification rate of 50.3% at FAR = 0.1% for an interocular distance of 50 pixels. This once again illustrates the drop in verification rates for low resolutions.

5.3.1.5 Landmark Methods for Comparison

We compared SHR to two landmark registration methods. The first method is the Viola and Jones detector [127] trained to find facial landmarks. The second method

is called MLLL (Most Likely Landmark Locator) [18], which finds the landmarks by maximizing the likelihood ratio using PCA and LDA. This algorithm is run in combination with BILBO, which is a subspace based method to correct for outliers. We have trained both methods on the FRGCv2 database and evaluated them using high resolution images. Both the Viola-Jones and MLLL+BILBO find four landmarks (eyes, nose and mouth). Based on the found landmarks, we calculate the Procrustes transformation to align the images.

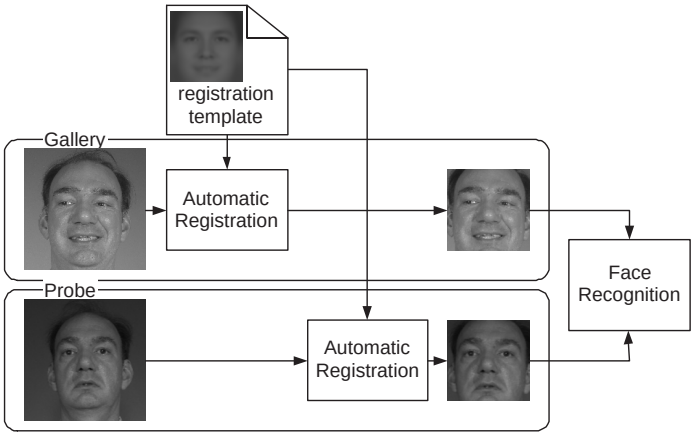


Figure 5.4: Protocol for user independent registration - Schematic representation of user independent registration using the same template for the gallery and probe image

5.3.2 Experimental Settings

In this section, we introduce the default experimental settings, unless other setting are explicitly mentioned, these settings are used in the experiments. We use the user independent registration, with edge images as features and the downhill simplex search method to find the registration parameters. The number of subspace components is set to 300, which is a good compromise between speed and accuracy. For the edge images, we use kernels of 17×17 pixels with $\sigma = 2$, which, according to our observations, gives good results on several databases. The maximum expected offsets for scale, rotation and translation needed to create the initial simplex are respectively 0.2, 5 degrees and 5 pixels. The downhill simplex method can also find the optimal registration parameters outside the maximum expected offsets. The gradient based search method is not limited in the registration parameter search either. In the case of user independent registration, both gallery image and probe image are registered to the same user independent registration template (depicted in Figure 5.4). The registration template is the mean face obtained from the training set. For user specific registration, we register to a single gallery image. Our subspace model is based on registered facial images, therefore,

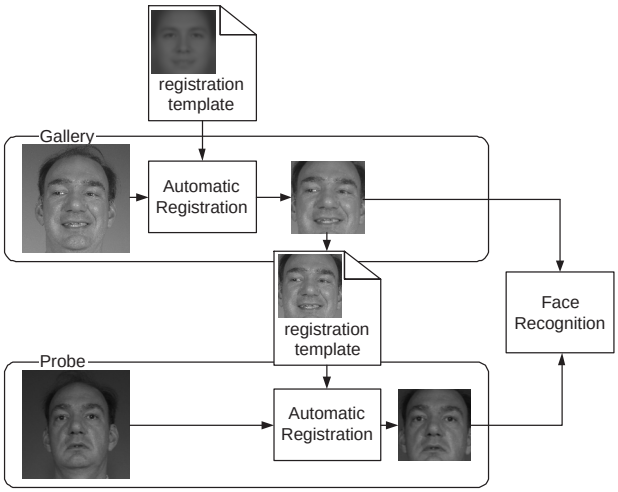


Figure 5.5: Protocol for user specific registration - Schematic representation of user specific registration, where the template is an automatically registered gallery image

we need a correctly registered template. Furthermore, face recognition methods assume that both gallery and probe images are correctly registered, making proper registration of the gallery image important for user specific registration. To obtain a registered gallery image, we perform the user independent registration with the mean face as registration template (see Figure 5.5). Although in our experiments we use a single image as registration template, it is also possible to use multiple images to build a user specific template. In this case, registration among gallery images can also be applied to improve the accuracy of the alignment of the gallery images.

5.4 Results

5.4.1 Comparison with Earlier Work

In Sections 5.2.1 and 5.2.2.2, we introduce a new evaluation criterion instead of the PCA Mahalanobis distance [26; 76] and new edge features for registration. In this section, we compare the effects of these changes separately. Figure 5.6 shows the effects which the new evaluation criteria (Bayesian Framework) and the new features have on the face recognition results, which are depicted using a ROC. After performing the registration with the different settings, we used the PCA-LDA likelihood ratio method for the recognition. In Figure 5.6, the ROC of the Bayesian Framework (grey values) shows that for FAR $> 50\%$ the verification rate decreases quickly, and for FAR $< 50\%$ the distance to the Bayesian Framework (edge images) remains constant. This behaviour is caused by incorrect

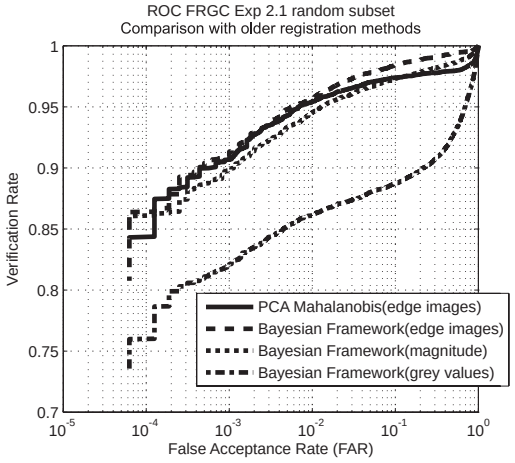


Figure 5.6: ROCs of different changes in our registration method - Comparing the effects of our new evaluation criteria and new features, this shows that the Bayesian Framework with edge images achieves the best results

registration, due to local minima in the search space, an example was shown in Figure 5.2. Comparing the performance of the Bayesian Framework (edge images) to the Bayesian Framework (magnitude images), we observe that edge images are slightly better. For this reason, we use the edge images in the remaining part of the paper. In Figure 5.6, we also show that the verification rate of the PCA Mahalanobis (edge images) distance drops rapidly to 98 % when FAR decreases from 100 %. This is caused by failures to find a correct registration. Figure 5.6 shows that the Bayesian Framework (edge images) containing the Distance From Features Space has made SHR more robust against these failures, resulting in a higher overall recognition performance.

5.4.2 Subspace-based Holistic Registration versus Landmark based Face Registration

In this experiment, we registered every face image using two landmark based face registration methods, SHR (user-independent face model) and the manually labelled landmark given by FRGCv2 database. For each face recognition method, we had to train the recognition methods on face images, which were registered by the specific registration method. This made the recognition method more robust against the specific variations. For SHR, we used the manual registration of the training set to train the face recognition methods. The results of our face recognition experiments using PCA-LDA likelihood ratio face recognition method are shown in Figure 5.7. Note that these results are obtained for verification at 50 pixels interoccular distance. Our focus is on the registration, which means that the relative results to manual registration are important. Other papers on

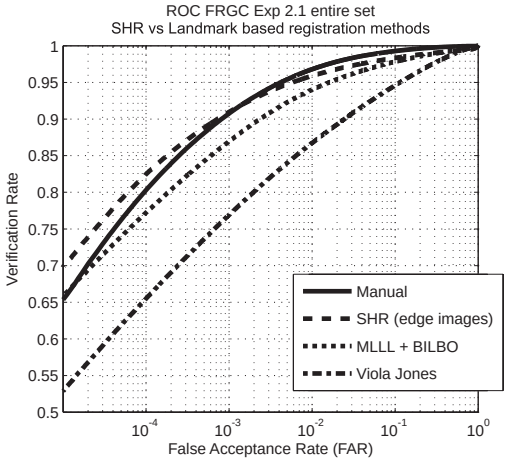


Figure 5.7: ROCs of different registration methods - Comparison of face recognition (PCA-LDA likelihood ratio) with several registration methods on FRGC experiment 2.1 using the entire set. SHR outperforms the results of face recognition with landmark based methods

face registration like [18; 130] do not achieve better recognition results than manual registration on the FRGC. In Figure 5.7, we observe that the performance of SHR is better than manual registration at FAR $\leq 0.1\%$. SHR also outperformed the automatic landmark based registration algorithms, which used high resolution images to obtain a registration. In Figure 5.7, the best landmark based registration method is MLLL+BILBO, which performed better than the Viola-Jones landmark method. In the case of the Viola-Jones landmark method, we removed 997 of the 15982 images from the query set of experiment 2.1, because 3 or less landmarks where found in these images which often resulted in poor alignments. We also experimented with the Viola-Jones method at an interocular distance of 50 pixels. In this case it failed to find the 4 landmarks for 10734 of the 15982 face images. In Table 5.2, we present the verification rates of all registration methods and the gain or loss in the recognition results by using automatic face registration methods instead of the manual face registration. Again all face recognition results were obtained at 50 pixels interocular distance. We observe that SHR improved the performance of all the face recognition methods in comparison with automatic landmark registration, which indicates that it is not dependent on the choice of the face recognition method. Some face recognition methods seem to be more robust against registration variations, for example Adaboost, but still more accurate registration improves the final recognition performance. In Table 5.2, the performance of the user-independent SHR is for most recognition methods similar or better than manual registration. To understand why SHR sometimes performs better than manually registered images, we first determined the difference in found registration parameters between manual and automatic registration,

Face Classifier	FAR	Viola-Jones (high resolution)	MLLL+BILBO (high resolution)	SHR (low resolution)	Manual
PCA Mah	1%	57.3% (-8.9%)	67.4% (+1.3%)	68.1% (+2.0%)	66.2%
	0.1%	44.5% (-5.8%)	52.9% (+2.6%)	54.0% (+3.3%)	50.3%
	0.01%	34.0% (-3.4%)	40.9% (+3.4%)	42.2% (+4.7%)	37.5%
PCA MahCos	1%	73.2% (-13.8%)	85.2% (-1.7%)	87.9% (+2.0%)	87.0%
	0.1%	57.4% (-10.8%)	68.2% (-0.0%)	71.9% (+3.3%)	68.2%
	0.01%	39.7% (-9.3%)	47.4% (+4.1%)	50.7% (+4.7%)	43.3%
Likelihood ratio	1%	86.7% (-10.1%)	94.0% (-2.8%)	95.9% (-0.9%)	96.8%
	0.1%	76.9% (-13.9%)	86.9% (-4.7%)	91.0% (+0.2%)	90.8%
	0.01%	65.5% (-14.8%)	77.2% (-3.1%)	82.5% (+2.2%)	80.3%
Adaboost	1%	86.5% (-8.4%)	93.5% (-1.4%)	94.1% (-0.8%)	95.0%
	0.1%	78.3% (-10.5%)	87.1% (-1.7%)	87.9% (-1.0%)	88.9%
	0.01%	69.8% (-11.1%)	78.9% (-2.0%)	80.1% (-0.8%)	80.9%

Table 5.2: Verification rate at FAR = {1%, 0.1%, 0.01%} and in parenthesis the relative contribution that automatic registration has in comparison with manual registration on FRGC Exp 2.1, comparing all registration methods using all face classifiers. The best automatic registration is achieved using user independent SHR using low resolutions, this often performs even better than manual registration

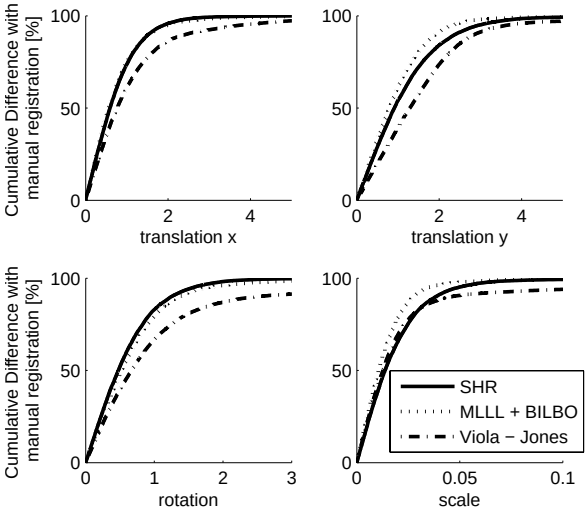


Figure 5.8: Differences in Registration Parameters from Manual Registration - Cumulative differences of registration parameters compared with manual registration, showing that MLLL+BILBO produces very accurate landmarks and that SHR and Manual differ especially in scale and translation in y-direction

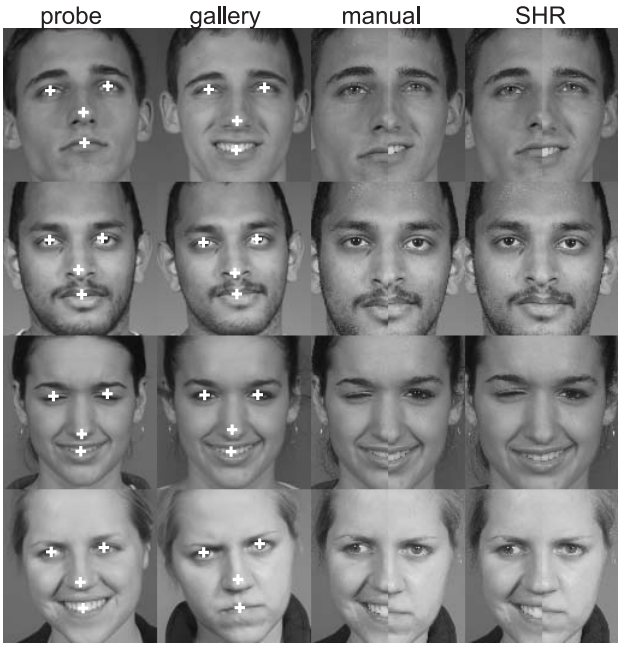


Figure 5.9: Examples of misalignments using manual registration - Examples of registration, the first and second column contain the face detection regions of probe and gallery images together with the manual landmarks. The third column, we present half of the probe image and other half of gallery image to compare the final alignment of manual registration. For the fourth column, we performed the same procedure as in the third column but with user independent SHR

which is shown in Figure 5.8. We observe that the results of MLLL+BILBO, which finds landmarks very accurately, are closer to manual landmarks in scale and y-translation. Both SHR and MLLL+BILBO have similar results in rotation and x-translation, but SHR finds different scale and y-translations. In Figure 5.9, a few examples of facial images with large differences in scale and y-translation between registration with manual landmark (third column) and SHR user independent (fourth column) are shown, together with the input for the registration determined by the face detection of the probe image (first column) and gallery image (second column). The white marks on the face are the manually labelled landmark locations. We pictured half of the registered probe image (left) and the other half registered gallery image (right) to show the alignment between the images. In the first row of Figure 5.9, we show a probe image with the head tilted up and a gallery image without tilt, because of the tilt of the head the relative positions of the landmarks change. We observe that the eyes, nose and mouth in the probe and gallery image are on almost the same line using manual registration, but there is a big difference in scale. On the other hand, SHR aligned both images on the same scale, this places the nose of the probe image higher but gives a better match with the mouth. In the second and third row, a slightly different definition of the landmark location is used (especially the nose), resulting in misalignments for manual registration, where the two halves in the third column do not overlap in the nose and mouth regions because of scaling differences. Another difficulty in the third images are the landmark locations of closed eyes, which is done correctly in this case, positioning the eyes somewhat above the closed eyebrows, but this is often not the case. In the last column of Figure 5.9, we observe that expressions can also change the ratio between landmark especially in the mouth area. The nose in the probe image is located higher than the nose in the gallery image using manual registration.

5.4.3 User independent versus User specific

In this section, we compare user specific registration to the user independent registration. In Figures 5.4 and 5.5, we show the two scenarios to obtain the user independent and user specific templates. In Figure 5.10, we show ROCs of the user independent and user specific face registration using the edge images. We observe that the performance consistently improves by using user specific registration. Figure 5.10 also shows that user specific registration performs slightly better than manual registration, which indicates that SHR gives more stable registration than the landmarks located by humans.

5.4.4 Comparing Search Algorithms

The two search methods, described in Section 5.2.4, were compared using a similar experiment as performed in the previous section. In all other experiments, the downhill simplex search method is used. It costs our matlab implementation on AMD opteron 275 around the 2.7 seconds to perform a registration for a single image, while the obtained matlab implementation of MLLL + BILBO [18] takes

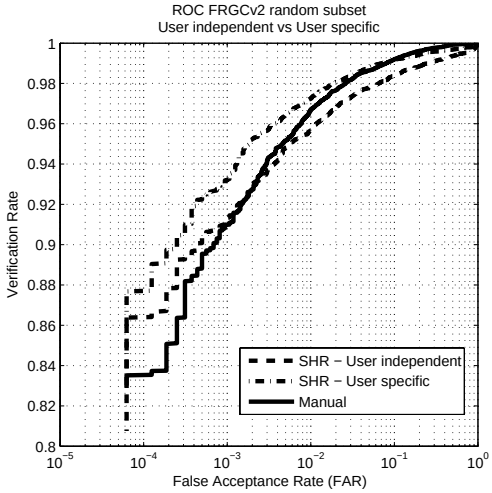


Figure 5.10: ROCs of user independent en user specific registration - Comparison of user independent and user specific face registration. User specific registration obtains better results than user independent registration and manual registration

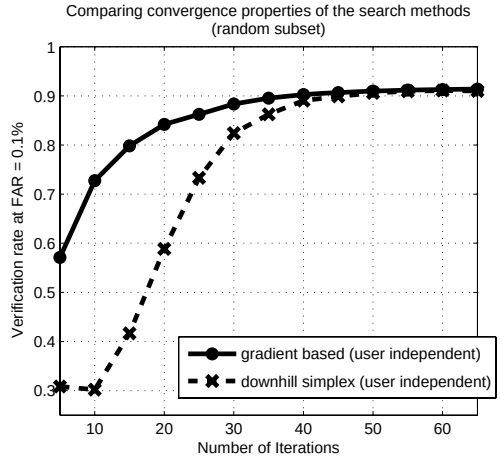


Figure 5.11: Convergence of Search Methods in number of iterations - Comparison of the search algorithms showing the verification rates of the likelihood ratio at different number of iterations. It takes the gradient-based method 3 times more computation time to calculate the same number of iterations as the downhill simplex method

around 7 seconds for a single image. The Viola and Jones landmark implementation in C++ performs almost real-time registration. Note that we spent not much effort in optimizing our code, because our main focus is on improving the accuracy. However, we can imagine that computation time in practical scenarios can be an issue. For this reason, we show a trade off between computation time measured in the number of iteration and accuracy measured in the verification rate, see Figure 5.11. Although the average search time of the gradient based method is larger, Figure 5.11 shows that it is able to find a good solution within a smaller number of iterations. This makes the difference between both search method in computation and accuracy very small.

5.4.5 Lower resolutions

In video surveillance, the resolution of the facial images is often below the interocular distance of 50 pixels used in previous section. To simulate this, we downsampled the images even more. In this section, we ran experiments using several lower resolutions to test the performance of SHR. After finding the alignment parameters for these resolutions, we use the alignment to register the facial images using an interocular distance of 50 pixels. This allows us to show the effects of low resolution on the registration, while ignoring the effects of low resolution on the face recognition. In Figure 5.12, we show the results on user independent

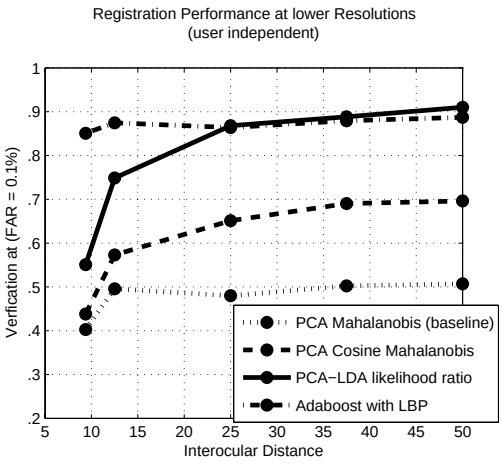


Figure 5.12: Registration performance by lower resolutions - Registration performance by varying the resolution used in SHR, the found registration parameters are then used to align facial images with an interocular distance of 50 pixels, showing only the performance of SHR at low resolution, which is still good at an an interocular distance of 25 pixels

registration for all the face recognition methods. We expect that registration performance decreases for lower resolutions. The registration results start becoming

worse at an interocular distance smaller than 25 pixels. Some methods like Adaboost are less sensitive for the registration errors caused by the lower resolutions than for instance PCA-LDA likelihood ratio.

5.5 Conclusion

We presented a novel subspace-base holistic registration (SHR) method, which is developed to perform registration on low resolution face images. In contrast to most landmark based registration methods, which can only perform accurate registration on high resolutions. SHR is able to use a user independent face model or a user specific face model to register face images. For the user specific registration, we defined two scenarios to register the gallery images. We show that by using edges as features for the registration, we obtain better results than using the grey levels of the image. The search for the best registration parameters is iterative and we proposed two search methods namely the downhill simplex method and a gradient-based method.
To evaluate the face registration, we measured the effects it has on the results of face recognition. We used the FRGCv2 database to perform our face registration experiments. We compared SHR with two landmark based registration methods, working on high resolution facial images. Nevertheless, the recognition results of SHR were better than those of the landmark based methods. User independent SHR gives a similar performance in face recognition results than registration with manually labelled landmarks. User-specific SHR performs better than the user-independent SHR and manual registration. One of the advantages over the landmark based methods is that SHR is able to register low resolution face images with an interocular distance as low as 25 pixels. The results at this resolution make SHR suitable for use in video surveillance.

5.6 Appendix: Gradient based search method

In this appendix, we discuss the gradient based search method in more detail. In Equation 5.14, we use first order Taylor series to rewrite the probe image into $T_{\boldsymbol{\theta}_k} H + \mathbf{M}_{\boldsymbol{\theta}_k} \boldsymbol{\delta}_k$. This allows us to find the maximum by taking the derivative of the similarity function, which in our case is the same as minimizing the distance $\hat{d}(\mathbf{x})$ in Equation 5.7. We write the probe image $T_{\boldsymbol{\theta}_k} H + \mathbf{M}_{\boldsymbol{\theta}_k} \boldsymbol{\delta}_k$ in terms of a feature vector $\mathbf{x}_k + \mathbf{M}_{\boldsymbol{\theta}_k} \boldsymbol{\delta}_k$. From [52], we know that the Jacobian matrix $\mathbf{M}_{\boldsymbol{\theta}}$ for transformation of scale, rotation and translation is defined as follows:

$$\mathbf{M}_{\boldsymbol{\theta}_k} = \begin{bmatrix} \nabla_{\mathbf{p}} T_{\boldsymbol{\theta}_k} H(\mathbf{p}_1)^T \Gamma(\mathbf{p}_1) \\ \nabla_{\mathbf{p}} T_{\boldsymbol{\theta}_k} H(\mathbf{p}_2)^T \Gamma(\mathbf{p}_2) \\ \cdots \\ \nabla_{\mathbf{p}} T_{\boldsymbol{\theta}_k} H(\mathbf{p}_N)^T \Gamma(\mathbf{p}_N) \end{bmatrix} \Sigma(\boldsymbol{\theta}) \tag{5.16}$$

$$\Gamma(\mathbf{p}) = \begin{bmatrix} 1 & 0 & -y & x \\ 0 & 1 & x & y \end{bmatrix} \tag{5.17}$$

$$\Sigma(\boldsymbol{\theta}) = \begin{bmatrix} \frac{1}{s}R(\alpha) & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & 1 & 0 \\ \mathbf{0} & 0 & \frac{1}{s} \end{bmatrix} \tag{5.18}$$

In this case, $\mathbf{p} = (x, y)^T$ is the pixel location, where $\nabla_{\mathbf{p}}$ gives the gradients in x and y direction. For clarity we rewrite the distance, in (Equation 5.7):

$$\hat{d}(\mathbf{x}) = \mathbf{y}^T \Lambda^{-1} \mathbf{y} + \frac{||\tilde{\mathbf{x}}||^2 - ||\mathbf{y}||^2}{\rho} \tag{5.19}$$

$$\begin{aligned} \hat{d}(\mathbf{x}) = (\Phi^T \mathbf{x} - \Phi^T \bar{\mathbf{x}})^T \Lambda^{-1} (\Phi^T \mathbf{x} - \Phi^T \bar{\mathbf{x}}) + \\ \frac{||\mathbf{x} - \bar{\mathbf{x}}||^2 - ||\Phi^T \mathbf{x} - \Phi^T \bar{\mathbf{x}}||^2}{\rho} \end{aligned} \tag{5.20}$$

We have to substitute $\mathbf{x}$ by $\mathbf{x}_k + \mathbf{M}_{\boldsymbol{\theta}_k} \boldsymbol{\delta}_k$, which results in:

$$\begin{aligned} \hat{d}(\mathbf{x}_k + \mathbf{M}_{\boldsymbol{\theta}_k} \boldsymbol{\delta}_k) &= (\mathbf{y}_k + \Phi^T \mathbf{M}_{\boldsymbol{\theta}_k} \boldsymbol{\delta}_k)^T \Lambda^{-1} (\mathbf{y}_k + \Phi^T \mathbf{M}_{\boldsymbol{\theta}_k} \boldsymbol{\delta}_k) + \\ &\quad \frac{||\tilde{\mathbf{x}}_\mathbf{k} + \mathbf{M}_{\boldsymbol{\theta}_k} \boldsymbol{\delta}_k||^2 - ||\mathbf{y}_k + \Phi^T \mathbf{M}_{\boldsymbol{\theta}_k} \boldsymbol{\delta}_k||^2}{\rho} \end{aligned} \tag{5.21}$$

We take the derivative of the distance function with respect to $\boldsymbol{\delta}_k$ and set it equal to zero. This gives us the following equation where for clarity $A = \Phi^T \mathbf{M}_{\boldsymbol{\theta}_k}$. Note that Λ^{-1} is a diagonal matrix.

$$A^T \Lambda^{-1} (\mathbf{y}_k + A\boldsymbol{\delta}_k) + \frac{1}{\rho} \mathbf{M}_{\boldsymbol{\theta}_k}^T (\tilde{\mathbf{x}}_k + \mathbf{M}_{\boldsymbol{\theta}_k} \boldsymbol{\delta}_k) - \frac{1}{\rho} A^T (\mathbf{y}_k + A\boldsymbol{\delta}_k) = 0 \tag{5.22}$$

This give us the follow linearly solvable equation for $\boldsymbol{\delta}_k$:

$$(A^T \Lambda^{-1} A + \frac{1}{\rho} \mathbf{M}_{\boldsymbol{\theta}_k}^T \mathbf{M}_{\boldsymbol{\theta}_k} - \frac{1}{\rho} A^T A) \boldsymbol{\delta}_k = (A^T \Lambda^{-1} \mathbf{y}_k + \frac{1}{\rho} \mathbf{M}_{\boldsymbol{\theta}_k}^T \tilde{\mathbf{x}}_k + \frac{1}{\rho} A^T \mathbf{y}_k) \tag{5.23}$$

Using this function, we can determine the new registration parameters $\boldsymbol{\theta}_{k+1} = \boldsymbol{\theta}_k + \boldsymbol{\delta}_k$. We repeat this gradient-based search method multiple times to find the final registration parameters.

Conclusion Part II

In Chapter 4 and Chapter 5, we described new face registration methods. These are the measures that we have taken to improve the face recognition system for low resolution facial images. In order to answer our last specific research question: how much improvement of the face recognition performance is obtained with the measures mentioned before? We have observed that these registration methods perform well for both controlled and uncontrolled conditions. These registration methods achieve the same face recognition performance as face images registered using manually labelled landmarks. In comparison with manually labelled landmarks there is no improvement, but this registration is often seen as almost perfect and we have to learn our methods based on this type of registration. In comparison with the automatic landmark based methods, the registration has achieved a large improvement. This makes the registration methods also suitable for applications not limited by low resolution like for example access control.
Our registration methods perform a search operation in order to find the best registration parameters. This search operation is sometimes more time consuming than other registration methods. There are, however, many accepted methods that perform similar or even more complex search operations like the Active Appearance Models. An advantage is that in video recordings, we can use the registration parameters of the previous frame, to register the face in the next frame. This can be useful in case of camera surveillance, where video recordings are often available.
In camera surveillance, there are, however, still challenges for the face registration. In most entrance and compulsory scenarios (Section 1.1.4), frontal or almost frontal face images are available. However, in camera surveillance, people do not always cooperate, which can result in face images under pose. Although the registration methods can deal with small variations in poses, these face registration methods fail for large pose variations. In order to deal with this, pose registration parameters must be estimated in addition to scaling, rotation and translation. Furthermore, 3D models of the face are necessary in order to model the appearance of the face for the different poses. In the next part, we already investigate the use of 3D models on 2D face images. Although we use the 3D models in the face intensity normalization component, the estimation of the surface can also be used for pose estimation.

Part III

ILLUMINATION

In camera surveillance, faces are recorded under uncontrolled conditions. In these scenes, the illumination in facial images can not be controlled. This differs from other applications, like access control where the acquisition is usually performed in a more controlled environment. The gallery images in camera surveillance are often recorded under controlled conditions, for instance if passport photos are used or if a criminal has been caught by the police and a mugshot is taken. Comparing facial images which are recorded under uncontrolled conditions is difficult, because the illumination in the faces causes large variations in the appearance. For this reason, we defined the following research question: Which measures can be taken to improve the face recognition system for images captured under uncontrolled illumination conditions? The uncontrolled illumination conditions mainly affect the face comparison component. It sometimes has a small influence on the face detection and registration, although the face detection and registration are in most cases very robust to illumination variations. For this reason, we investigate if face intensity normalization methods can improve the face comparison component. Several face intensity normalization methods have been developed to correct the illumination in faces, but they are often tested on datasets recorded in laboratory conditions. We also investigated the performance on images taken under uncontrolled conditions.
This part contains four chapters which are based on our publications on the face intensity normalization component. In the first chapter, we describe a face intensity normalization method, which is extended in the next chapters. To extend this method, we gradually used more advantaced reflectance and face shape models. We will discuss the major difference between the chapters in short:

- Chapter 6 is based on [23]. In this chapter we explain a new illumination correction method for facial images. This method is able to correct for a single light source and uses an additional error term to correct for shadows and reflections. A subspace model of the face shape is developed from 3D range images in order to model the facial appearance. By finding an estimate of the face shape, we are able to render a facial image under frontal illumination conditions. These images are used in the experiments, where the correction method is tested on both uncontrolled illumination conditions and illumination conditions created in a laboratory.
- Chapter 7 is based on [27]. In this chapter we discuss some extensions on the previous method. Instead of modelling a single light source, we model both an ambient and a diffuse light source. This allows us to explain reflectance in shadow areas. Another difference is that we used a surface and albedo model instead of the face shape model. With the surface model, we can apply geometrical constraints to the face shape which improves the face shape estimation.
- Chapter 8 is based [28], in which we combine face intensity normalization methods. In Section 2.4, we explain that there are two categories of Face Intensity Normalization methods. The methods in the first category perform normalization based on local regions, while methods in the second category use a physical model to determined the illumination condition in the entire face images. The methods from both categories have advantages and disadvantages. We have combined these methods using score and decision level fusion to improve the overall recognition results.
- Chapter 9 is based on [25]. In this chapter, we extend the method explained in Chapter 7. The major improvement is that we are able to model multiple light sources, using a virtual illumination grid. We have also coupled the surface and albedo models, instead of using two separate subspace models. Experiment 4 of the Face Recognition Grand Challenge version 2 database is used to compare several face intensity normalization methods.

6

MODEL-BASED RECONSTRUCTION FOR ILLUMINATION VARIATION IN FACE IMAGES

6.1 Introduction

In video surveillance applications, face recognition has been a difficult problem due to variations in the face, like illumination, pose and facial expressions. In this chapter, we introduce a method for correcting for the illumination variations. Correcting for these variations is necessary because they are often larger than the variation due to changes of the person's identity. We propose a method that is able to correct the illumination in a single image.
Several approaches have already been proposed in the literature, that make face images invariant for illumination. These approaches can be divided into two categories. The first category works by applying a preprocessing step to the images, like Histogram Equalization [105] or Local Binary Patterns [55]. These methods are direct and simple but often lack a theoretical explanation. The second category works using physical models of the illumination mechanism and its interactions with the object surface. The Illumination Cone [48] falls in this category, which estimates the shape from at least three images under different illumination conditions. An other method is the Quotient Image [107] which estimates illumination of a single image allowing them to compute a quotient image. This method does not model shadows and reflection. The method proposed in [110] can

deal with shadows and reflections by adding an error term. In the 3D morphable model described in [20], the Phong model is applied to simulate the reflection, also shadows are properly modeled by the 3D shape of the face. In our method, we choose to model shadows and reflection by adding an error term as in [110], making our method computationally efficient, but we use a shape model derived from 3D range maps, allowing us to estimate the shape from a single image.
This chapter is organized as follows, in Section 6.2 we describe our method to correct illumination variation in face images. In Section 7.3 we describe the experiments and the results of the face recognition algorithm on the reconstructed face images. In Section 6.4 we present some conclusions.

6.2 Method

6.2.1 Lambertian model

In order to solve the illumination problem in face recognition for a single image, we begin with some simple assumptions. Our first assumption is that we have a single light source at infinite distance, making the problem easier and computationally tractable. The direction and intensity of the light source is not known to us. Recovering an unknown shape for a 2D image without further knowledge is impossible, because many 3D shapes result in the same 2D image. For this reason we use a 3D face shape model which allows us to estimate the face shape in the face image. In this chapter, we assume that the illumination of faces behaves to a large extent according to the Lambertian equation. We introduce an error term to model shadows and reflections which are normally not included in the Lambertian equation [110]. This gives the following Lambertian equation:

$$b(x) = c(x)\mathbf{n}(x)^T \mathbf{s}i + e(x;\mathbf{s}) \tag{6.1}$$

x is the pixel position and $b \in \mathcal{R}$ is the pixel intensity in the image. The pixel intensity is determined by the dot product of the shape and the illumination. The surface normals $\mathbf{n} \in \mathcal{R}^3$, which define the direction of the reflection, and the albedo of the surface given by $c \in \mathcal{R}$ together are the shape $\mathbf{h}(x) = c(x)\mathbf{n}(x)^T$. The direction of the light, which is a normalized vector given by $\mathbf{s} \in \mathcal{R}^3$, and the intensity of the light, which is given by $i \in \mathcal{R}$, describe the light conditions $\mathbf{v} = \mathbf{s}i$. The error term $e \in \mathcal{R}$ that we introduce allows us to handle reflection and shadows which are not modeled in the Lambertian equation. Instead of writing these term for every pixel position x, we can also vectorize the image giving us the following equation:

$$\mathbf{b} = H\mathbf{v} + \mathbf{e}(\mathbf{s}) \tag{6.2}$$

In our case we have M pixel positions, which gives us the vectorized face image $\mathbf{b} \in \mathcal{R}^{M\times 1}$, a matrix which contains the face shape $H \in \mathcal{R}^{M\times 3}$ and the error term $\mathbf{e}(\mathbf{s}) \in \mathcal{R}^{M\times 1}$.

6.2.2 Overview of our correction method

In this chapter, we want to correct the illumination in a single image, where we only have the pixel intensities $b(x)$ of the Lambertian equation. The other terms; the shape $\mathbf{h}(x)$ and light $\mathbf{v}$ have to be estimated by our method. Our method uses a model of the face shape, which allows us to calculate the parameters of the face shape given that we know the light parameters. Because the lighting is unknown, we search for the light parameters which give the optimal face shape parameters. For different light directions, we estimate a light intensity and calculate the face shape. We evaluate all the face shapes calculated under different light directions. Using kernel regression, we are then able to determine the final face shape from the different face shapes. The different steps of our method are given below in pseudo-code:

- Learning a model of the face shape (offline)
- For different light direction $\mathbf{s}_j$
  - Calculate a shadow and reflection term $e_j(x; \mathbf{s}_j)$
  - Estimate the light intensity i_j which gives us light source $\mathbf{v}_j$
  - Fit the face shape model to the face image which gives us the shape $\mathbf{h}_j(x)$
  - Evaluate the shape $\mathbf{h}_j(x)$ which gives us a distance measure d_j
- Calculate the final shape $\mathbf{h}(x)$ using kernel regression
- Refine the albedo of the surface $c(x)$ to obtain more details

Throughout this chapter, j is a index for the different light direction, where we calculate for J light directions error terms $e_j(x; \mathbf{s}_j)$, light intensities i_j, face shapes $\mathbf{h}_j(x)$ and distance measures d_j. Using the different shapes and distances, we can obtain one final shape which we refine. With this final shape, we can calculate a face image under frontal illumination conditions, using the following equation:

$$b_{\text{frontal}}(x) \quad = \quad \mathbf{h}(x)^T \mathbf{v}_{\text{frontal}} \tag{6.3}$$

In the next sections, we will discuss the different steps described in our pseudo code in more detail.

6.2.3 Learning the Face Shape Model

A set of face images and 3D range maps provide us with the means to calculate the face shape H for each face image (See Section 6.3.2). Because we have multiple face shapes we can determine a mean shape $\overline{H}$ and variations from this mean shape. To calculate the variations, we vectorized all the L obtained shapes into the data matrix $X \in \mathcal{R}^{3M \times L}$ from which we compute the covariance matrix Σ. Using Principal Component Analysis (PCA), we can decompose Σ as $\Sigma = \Phi \Lambda \Phi^T$, where the columns of Φ are the eigenvectors and the matrix Λ contains the eigenvalues on the diagonal. The columns of $\Phi \in \mathcal{R}^{3M \times K}$, are converted to $M \times 3$ matrices, which we denote by $T_k \in \mathcal{R}^{M \times 3}$. These matrices $\{T_k\}_{k=1}^{K}$ contain the most important

variations of the face shape. In Figure 6.1, we calculated the surface from the shape and on this surface we have drawn a reconstructed image under frontal illumination. The left image on top is the mean shape and the other images from left to right are the first 5 most important variations in the shape model.

Figure 6.1: Eigenshapes determine with PCA - The mean shape and first five deviation of the shape model, on the face shape we draw the face image under frontal illumination

6.2.4 Shadow and Reflection Term

In this chapter we use the Lambertian equation as basis, which is unable to deal with shadows and reflection. For this reason, we added the error term $e(x;\mathbf{s})$ in the equation 9.1. We assume that the error term depends on the light direction $\mathbf{s}$, which is the determining factor for shadows and reflections. We use a 2D face database with labelled illumination conditions to learn the error term for each light direction. Using the kernel regression method described in [110], we can calculate for a light direction $\mathbf{s}_j$ (in Section 6.2.2) the mean error $\mu_e(x,\mathbf{s}_j)$ and the variance $\sigma_e^2(x,\mathbf{s}_j)$. For the error term in equation 9.1, we take $e_j(x;\mathbf{s}_j) = \mu_e(x,\mathbf{s}_j)$. The variance $\sigma_e^2(x,\mathbf{s}_j)$ will be used later in the Section 6.2.9. Although we use a face-independent mean term for the shadows and reflections, this estimation is better than ignoring the error term.

6.2.5 Light Intensity

We want to estimate a face shape using the face image and our face shape model, where we know the light direction $\mathbf{s}_j$ and the error term $\mathbf{e}_j(\mathbf{s}_j)$. To estimate this shape we have to know the light intensity i which allows us to calculate the light conditions $\mathbf{v} = \mathbf{s}i$. To calculate the intensity, we replace the unknown shape H with the mean shape $\overline{H}$. Using a linear solver, we are able to solve the following

equation, where the light intensity i_j is the unknown:

$$i_j = \arg\min_{i_j} \|(\overline{H}\mathbf{s}_j)i_j - (\mathbf{b} - \mathbf{e}_j(\mathbf{s}_j)\|^2 \tag{6.4}$$

Because this is a overcomplete system, we can use the mean face shape $\overline{H}$ to estimate the light intensity i_j, which still gives a very accurate estimation. However, this might normalize the different intensities in the skin color of different persons.

6.2.6 Estimation of the Face Shape

In the previous sections, we calculated for a light direction $\mathbf{s}_j$, the error term $\mathbf{e}_j(\mathbf{s}_j)$ and light condition $\mathbf{v}_j$. Using these terms we are now able to obtain the face shape H_j. In this section, we will calculate the shape using the face shape model obtained in Section 6.2.3. Using the face shape model, we can replace H_j with:

$$H_j = \overline{H} + \sum_{k=1}^{K} T_k y_{j,k} \tag{6.5}$$

This is the mean shape and the K most important variations. Our approach has much in common with 3D morphable models [20], however we apply it solely for shape recovery. We can now rewrite the Lambertian equation as follows 9.1:

$$\overline{H}\mathbf{v}_j + \sum_{k=1}^{K} T_k \mathbf{v}_j y_{j,k} = \mathbf{b} - \mathbf{e}_j(\mathbf{s}_j) \tag{6.6}$$

$$\sum_{k=1}^{K} T_k \mathbf{v}_j y_{j,k} = \mathbf{b} - \mathbf{e}_j(\mathbf{s}_j) - \overline{H}\mathbf{v}_j \tag{6.7}$$

Instead of calculating the shape H_j, we are now able to calculate the variations $\mathbf{y}_j \in \mathcal{R}^K$ of the shape using a linear solver. In this case, we write $T_k\mathbf{v}_j = A_k$ and $\mathbf{b} - \mathbf{e}_j(\mathbf{s}_j) - \overline{H}\mathbf{v}_j = \mathbf{c}$ giving us the following linear solvable equation:

$$\mathbf{y}_j = \arg\min_{\mathbf{y}_j} \|A\mathbf{y}_j - \mathbf{c}\|^2 \tag{6.8}$$

To calculate the shape H_j from the parameters $\mathbf{y}_j$ obtained in Equation 6.8, we use Equation 6.5. In this case, we are able to estimate the shape given the direction of the illumination $\mathbf{s}_j$. In the next section, we explain how we evaluated the obtained face shape.

6.2.7 Evaluation of the Face Shape

To evaluate the face shape we made two observations: First, the found variations will be small when the light direction is similar to the light direction in the face image. Second, when the light directions are similar, the found variations of the face shape create a reconstructed image which matches the input image. These

criteria are similar to the criteria used in the active shape models [40]. Although in their case, they can compare their output directly with the image, we have to convert our found shape H_j belong with the light direction $\mathbf{s}_j$ to a reconstructed image $\mathbf{b}_j$, which is calculated as following:

$$\mathbf{b}_j = H_j \mathbf{v}_j + \mathbf{e}_j(\mathbf{s}_j) \tag{6.9}$$

Because we calculate for different light directions $\mathbf{s}_j$ new reconstructed images $\mathbf{b}_j$, we can compare these images with the original image $\mathbf{b}$. We do this by calculating the sum of the square differences between both images as follows:

$$\epsilon^2 = (\mathbf{b} - \mathbf{b}_j)^T (\mathbf{b} - \mathbf{b}_j) \tag{6.10}$$

We know that the reconstructed image $\mathbf{b}_j$ contains the variations $\mathbf{y}_j$ to the face shape model, because $\mathbf{v}_j$ and $\mathbf{e}_j(\mathbf{s_j})$ are constants. Our face shape model has only K variations, making it impossible to explain all deviations in the face images. In [40], the distance measure how well the model fits is given by:

$$d_j = \sum_{k=1}^{K} \frac{y_k^2}{\lambda_k} + \frac{\epsilon^2}{\rho} \tag{6.11}$$

We use the estimation of [80] for $\rho = \frac{1}{3M-K} \sum_{k=K+1}^{3M} \lambda_k$. Using this distance measure, we can easily evaluate the quality of the found shape for a certain light direction.

6.2.8 Calculate final shape using kernel regression

In Section 6.2.6, we calculated the face shape parameters $\{\mathbf{y}_j\}_{j=1}^{J}$ for different light directions $\mathbf{s}_j$. In Section 6.2.7, we evaluate the distance measures $\{d_j\}_{j=1}^{J}$, which determine the quality of the different face shape parameters. The light directions $\mathbf{s}_j$ are the same as the light directions used in the face database with labelled illumination conditions. The main reason that we use the same light directions is because these light directions cover a complete grid of light directions. This is ideal for applying kernel regression [7]. Using the obtained face shape parameters $\{\mathbf{y}_j\}_{j=1}^{J}$ and the distances $\{d_j\}_{j=1}^{J}$ our regression method can be seen in the following equations:

$$\mathbf{y} = \sum_{j=1}^{J} w_j \mathbf{y}_j / (\sum_{j=1}^{J} w_j) \tag{6.12}$$

$$w_j = \exp[-\frac{1}{2}(d_j)/\sigma)^2] \tag{6.13}$$

In the above equation, σ is determined so that 5 percent of the distances lie within $1 \times \sigma$. The final shape parameters obtained from $\mathbf{y}$ give us the final shape H using Equation 6.5. We can calculate the light conditions $\mathbf{v}$ using a linear solver, which allows us to determine the final light conditions.

6.2.9 Refinement

The obtained shape can be divided into two parts, the first part are the surface normals $\mathbf{n}(x)$, the second part is the albedo of the surface denoted by $c(x)$. After estimating the final shape H, we observed that the reconstructed frontal illuminated image from shapes does not contain all details present in the original face image. Partially, this phenomenon can be explained by the fact that we do a dimension reduction. Another part can be explained by the fact that the kernel regression performs an interpolation. To recover these details, we recalculate only the albedo of the surface $c(x)$. Because the albedo of the surface contains most of the details, while the surface normals contain the larger structures of the face shape. To calculate the albedo $c(x)$ we use a MAP estimate give by the following equation:

$$c(x)_{MAP} = \arg\max_{c} P(b(x)|c(x))P(c(x)) \tag{6.14}$$

As can be seen from Equation 6.14, we estimate the albedo for every pixel. For clarity, we will drop the "(x)" in the following equations and replace the surface normals and final light condition $\mathbf{n}(x)^T \mathbf{s}i$ with the constant q. The albedo of the surface $c(x)$ can be estimated by the following equation, where we assume two Gaussian distributions:

$$c_{MAP} = \arg\max_{c} \mathcal{N}(cq + \mu_e(\mathbf{s}), \sigma_e^2(\mathbf{s})) \times \mathcal{N}(\mu_c, \sigma_c^2) \tag{6.15}$$

$$\arg\min_{c} L = \left(\frac{i - cq - \mu_e(\mathbf{s})}{\sigma_e(\mathbf{s})}\right)^2 + \left(\frac{c - \mu_c}{\sigma_c}\right)^2 \tag{6.16}$$

The mean and variance of the error term are calculated with kernel regression, which is described in Section 6.2.4. The log probabilities L are given in Equation 6.16. We find the minimum by taking the derivative and make it equal to zero. The new albedo c_{MAP} contains more details than the albedo obtained using the PCA model. We have observed that the details are very important in the face recognition processes.

6.3 Experiments and Results

The purpose of the illumination correction is to improve the recognition rate of the face classifier. To evaluate this we performed two experiments to test if our correction algorithm indeed improves the recognition results. The first experiment is done on the Yale B databases to see if our algorithm can deal with different illumination conditions. The second experiment is performed on the FRGCv1 database where we use controlled face images for enrollment and the uncontrolled face images for the classification.

6.3.1 Face databases for Training

In the Section 6.2, we described our method which uses two different face databases for training. In this chapter, we use publicly available face databases. For the shad-

ows and reflections, we used the Yale B [48] and extended Yale B [72] database (for short just the Yale B databases), these databases contain face images illuminated under different labelled light directions. To make a shape model, we needed a face database which contains both images and 3D range maps. The Face Recognition Grand Challenge (FRGC) [83] database is a publicly available database with a subset which contains 3D range maps. To make the shape model we used a subset of the entire 3D FRGC database namely the Spring 2003 3D range maps and face images. These face images contain almost frontal illumination and no shadows, making this subset of the database ideal for the calculation of the shape model. We performed some simple spike removal and filling techniques to obtain better surfaces from which we calculated the surface normals. Using these surface normals we are able to derive the shape model.

6.3.2 Determine albedo of the Shape

From the range maps, we are able to calculate the surface normals $\mathbf{n}(x)$ for each pixel in the image. In this case, we obtained for every pixel the image intensity $b(x)$ and the surface normals $\mathbf{n}(x)$. Looking at the Lambertian equation 9.1, we need for the shape $h(x)$ also the albedo of the surface $c(x)$. Because the face images and 3D range maps where taken under almost frontal lighting, we decided to ignore the $e(x)$ term. In this case we only have to obtain the light conditions, to be able to calculate the albedo of the surface $c(x)$. To estimate the light conditions $\mathbf{v}$ we determined the mean albedo of the surface $\mu_c(x)$ from the Yale B databases, which allows us to estimate $\mathbf{v}$ as follows:

$$\begin{aligned} \mathbf{g}(x) &= \mu_c(x)\mathbf{n}(x) \\ \mathbf{v} &= \arg\min_{\mathbf{v}} \|G\mathbf{v} - \mathbf{b}\|^2 \end{aligned} \tag{6.17}$$

In Equation 6.17, we first calculate the temporary shape $\mathbf{g}(x)$ to estimate the light conditions $\mathbf{v}$. We represent the temporary shape in the matrix $G \in \mathcal{R}^{M\times 3}$. Using a linear solver we are able to calculate $\mathbf{v}$, see Equation 6.17. Using the light conditions $\mathbf{v}$ we can calculate the albedo of the surface with the following equation:

$$c(x) = \frac{b(x)}{\mathbf{n}(x)^T\mathbf{v}} \tag{6.18}$$

We calculate for every pixel the albedo of the surface $c(x)$, sometimes these values become very large because the surface normals are perpendicular with the light source. In those case we use a filling algorithm to correct for these mistakes.

6.3.3 Face Recognition

The illumination correction is a preprocessing step to improve the face recognition. For this reason, we performed a recognition experiment on two face databases to see if the correction indeed helps improving the recognition results. For the face

recognition, we performed a feature reduction by subsequently using a PCA [123] and LDA [16] transformation to the corrected face images, using 200 PCA and 50 LDA components. After feature reduction, we use the likelihood ratio, described in [126] to obtain the similarity scores. In the next sections, we discuss how this classifier is used to obtain the results on the different databases.

6.3.4 Yale B database

The Yale B databases are in our case used to train the illumination model. These databases also allow us to evaluate our algorithm under different illumination conditions. Because we trained our illumination model on the Yale B databases, we performed a leave one person out experiment, with all the face images with the azimuth and elevation angle below the ± 90 degrees. In Figure 6.2, we show the original and reconstructed face images of a person in the Yale B database, where we used both our method and the method describe in [110] for correction. In the case of big shadows, like the eyes in the most right image in Figure 6.2, the method in [110] corrects by filling in the mean face. Our method however is bound to the shape model which also has to explain dependencies between pixels, which allows us to correct more user-specific
We used all the corrected face images obtained from the Yale B databases for a

Figure 6.2: Reconstructed face images of the Yale B database - Face images from the Yale B database, the first row contains uncorrected images, the second row contains the correction of [110] and the last row is corrected using our method.

recognition experiment. We trained on thirty persons and used the face images of the remaining eight persons for enrollment and testing. We compared each image with the other images taken under all the different illumination conditions, where we left out the images which are taken under similar illumination conditions. We repeated this experiment until we obtained the similarity scores for every person with all the other persons in the database. The Receiver operating characteristic (ROC) of this experiment is shown in Figure 6.3. For this experiment, we clearly

see that our method (EER: 12.83%) performs better than the method in [110] (EER: 17.84%) and the uncorrected images (EER: 20.82%).

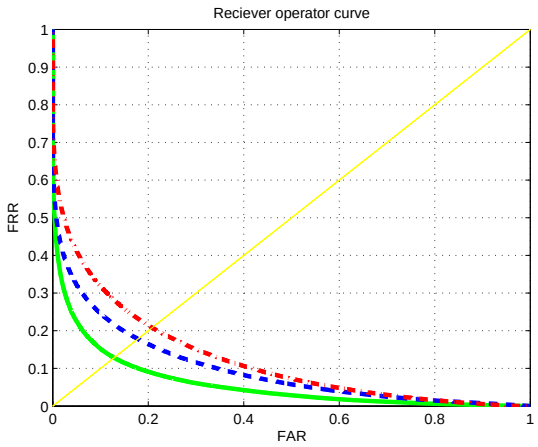


Figure 6.3: ROCs of illumination correction methods on the Yale B database - The ROC of the face recognition experiment on the Yale B databases, where the solid line is our correction method, the dashed line is the method describe in [110] and the dotted-dashed line is for the uncorrected images

6.3.5 FRGCv1 database

The main purpose of our method is to solve the surveillance problem where we have face images taken under controlled conditions but we also want to recognize these persons under uncontrolled conditions. In our case, the main focus is the illumination correction of these face images. The Face Recognition Grand Challenge version 1 contains frontal face images taken under both controlled and uncontrolled conditions which allows us to setup an experiment using this database. We corrected the illumination effects on all images in the FRGCv1 database, both the controlled and uncontrolled conditions. Examples of some reconstructed face images from the FRGCv1 database are shown in Figure 6.4.
For our recognition experiment, we randomly divided the uncontrolled and controlled face images into two parts, each containing approximately half of the face images. We used the first halves of both sets to train our face classifier, the second half of the controlled images are used for the enrollment of the one user template for every user and the second half of the uncontrolled images are used as probe images. We repeated this experiment 20 times randomly splitting the database to remove statistical fluctuations. The ROC curves of our experiment are shown in Figure 6.5, where our method obtained a EER of 4.35 %, while using uncorrected images we obtain a EER of 4.80 %. We also show the results of [110], but their method is not able to deal with different light intensities because it has

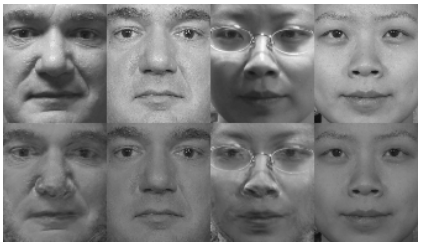


Figure 6.4: Reconstructed face images of the FRGCv1 database - Face images from the FRGCv1 database, the first row contains uncorrected images, the second row contains the corrected images, with for both persons a controlled and uncontrolled images.

been trained on the Yale B databases which only contains one intensity. Although our method improves the recognition rates in this experiment, the improvement is smaller than the one reported on the Yale B database. A reason is that the FRGC database contains other challenges like small poses, expressions, motion blur, while the Yale B database only focusses on illumination problems.

6.4 Conclusion

In this chapter, we propose a novel approach to correct face images for unknown illumination conditions. By fitting a face shape model under different light directions on the face images, we are able to estimate the face shape from which we can reconstruct a face image under frontal illumination. To test if these reconstructed face images improve the recognition rates we setup two experiments. In our first experiment, we achieve better recognition rates on the reconstructed face images acquired in a laboratory under different kinds of light conditions. The second experiment shows our method also improves the recognition results under uncontrolled illumination, making the algorithm suitable for surveillance applications.

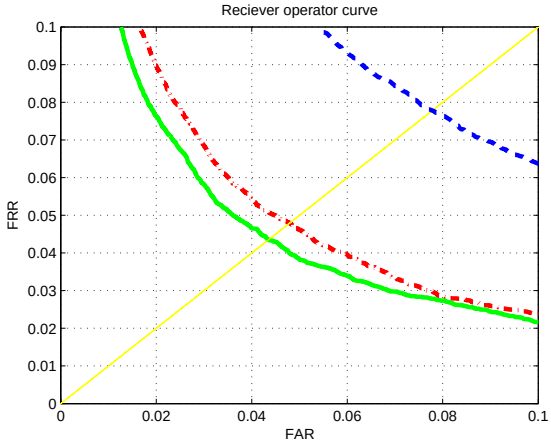


Figure 6.5: ROCs of illumination correction methods on the FRCGv1 database - The ROC of the face recognition experiment on the FRCGv1 database, where the solid line is our correction method, the dashed line is the method describe in [110] and the dotted-dashed line are the uncorrected images

7

MODEL-BASED ILLUMINATION CORRECTION FOR FACE IMAGES IN UNCONTROLLED SCENARIOS

7.1 Introduction

One of the major problems with face recognition in uncontrolled scenarios is the illumination variation, which is often larger than the variations between individuals. We want to correct for these illumination variations in a single face image. In the literature, several methods have been proposed to make face images invariant to illumination. These methods can be divided into two categories: The first category contains methods that perform preprocessing based on the local regions, like Histogram Equalization [105] or (Simplified) Local Binary Patterns [55; 121]. These methods are direct and simple, but fail to model the global illumination conditions. The methods in second category estimate a global physical model of the illumination mechanism and its interaction with the facial surface. One of the earlier methods in this category is the Quotient Image [107], which estimates illumination in a single image allowing the computation of a quotient image. More recent correction methods [110; 139] are also able to deal with shadows and reflections using an addition error term. In our experience, these methods work on images with illumination conditions created in a laboratory, but fail in uncontrolled scenarios. In [20], 3D morphable models are used to simulate the illumination conditions in a single images, calculating shadows and reflections properly. The disadvantage

of this method is the computational cost for a single image. In [5], a illumination normalization is proposed for uncontrolled conditions which requires a color image together a 3D range image.
We developed a new method for illumination correction in [23] which used only a single grey level image, but this method improved the recognition for face images taken under uncontrolled conditions. During these experiments, we discovered that both our method and [110] have problems modelling shadow areas which still contain some reflection. This often occurs in face images taken under uncontrolled conditions. Furthermore, we observed that the found surface normals were not restricted by the geometrical constrains. In this chapter, we tried to solve these issues by improving our previous method.

7.2 Illumination Correction Method

7.2.1 Phong Model

To model the shadow areas that contain some reflections, we use the Phong model, which explains these areas using the ambient reflection term. In our previous work and in [110], the Lambertian model with a summed error term was used to model the shadows. This however fails when both the intensities of the light source on the face and the reflections in the shadow areas vary. The Phong model in combination with a shadow expectation is able to model these effects. If we assume a single diffuse light source l, the Phong model is given by the following Equation:

$$b(\mathbf{p}) = \quad c_a(\mathbf{p})i_a + c_d(\mathbf{p})\mathbf{n}(\mathbf{p})^T\mathbf{s}_l i_d + \text{specular reflections} \tag{7.1}$$

The image $b(\mathbf{p})$ at location $\mathbf{p}$ can be modelled using three parts namely: the ambient, diffuse and specular reflections. The ambient reflections exist of the albedo $c_a(\mathbf{p})$ and the intensity of the ambient light i_a. The ambient reflections are still visible if there is no diffuse light, for instance in shadow areas which are not entirely dark. The diffuse reflections are similar to the Lambertian model, where the surface normals $\mathbf{n} \in \mathcal{R}^3$ define the direction of the reflection and together with the albedo $c_d(\mathbf{p})$ give the shape $\mathbf{h}(\mathbf{p}) = c_d(\mathbf{p})\mathbf{n}(\mathbf{p})^T$. The diffuse light can be modelled by a normalized vector $\mathbf{s} \in \mathcal{R}^3$, which gives the light direction and the intensity of the diffuse light i_d. The final term contains the specular reflections, which explain the highlights in the image, but because this phenomenon is usually only present in a very small part of the image we will ignore this term.
The shadow can be modelled as a hard binary decision. If a light source can not reach a certain region, it makes a shadow. This holds expect for areas which contain the transition between light and shadow. Using a 3D range map of a face, we can compute the shadow area given a certain light direction using a ray tracer. Computing these shadow areas for multiple images, allows us to calculate an expectation of shadow $e_l(\mathbf{p})$ on the position $\mathbf{p}$ for the light directions $\mathbf{s}_l$. This

gives us a user independent shadow model given the light direction.

$$b(\mathbf{p}) = c(\mathbf{p})i_a + c(\mathbf{p})\mathbf{n}(\mathbf{p})^T\mathbf{s}_l i_d e_l(\mathbf{p}) \quad (7.2)$$

$$\mathbf{b} = \mathbf{c}i_a + H\mathbf{s}_l i_d \star \mathbf{e}_l \quad (7.3)$$

In Equation 7.2, we simplified the Phong model and we added the expectation term $e_l(\mathbf{p})$ to model shadows. We also use the same albedo term for ambient and diffuse illumination, which is common practice [20]. In Equation 7.3, we vectorized all the terms, where the $\star$ denotes the Cartesian product. Our goal is to find the face shape and the light conditions given only a single image.

7.2.2 Search strategy for light conditions and face shape

An method to estimate both the face shape and the light conditions is to vary one of the variables and calculate the others. In our case, we chose to vary the light direction allowing us to calculate the other variables. After obtaining the other variable, e.g. light intensity, surface and albedo, we use an evaluation criteria to see which light direction gives the best estimates. The pseudo-code of our correction method is given below:

- For a grid of light directions $\mathbf{s}_l$
  - Estimate the light intensities i_a and i_d
  - Estimate the initial face shape
  - Estimate the surface using geometrical constrains and a 3D surface model
  - Computing the albedo and its variations
  - Evaluation of the found parameters
- Refine the search to find the best light direction.
- Reconstruct a face images under frontal illumination.

We start with a grid where we vary the azimuth and elevation of the light direction with 20 degrees. The grid allows us to locate the global minimum, from there we can refine the search using the downhill simplex search method [81] to find the light direction with an accuracy of ± 2 degrees. Using the found parameters like light conditions and face shape, we can reconstruct a face image under frontal illumination, which can be used in face recognition. In the next sections, we will discuss the different components mentioned in the pseudo-code.

7.2.3 Estimate the light intensities

Given the light direction $\mathbf{s}_l$ and the shadow expectation $\mathbf{e}_l(\mathbf{p})$, we can estimate the light intensities using the mean face shape $\overline{\mathbf{h}}(\mathbf{p})$ and mean albedo $\overline{c}(\mathbf{p})$. The mean face shape and albedo are determined using a set of face images together

with there 3D range maps. This gives us the following linearly solvable equation, allow us to obtain the light intensities $\{i_a, i_d\}$:

$$\{i_a, i_d\} = \arg \min_{\{i_a, i_d\}} \sum_{\mathbf{p}} \|b(\mathbf{p}) - \overline{c}(\mathbf{p}) i_a - \overline{\mathbf{h}}^T \mathbf{s}_l i_d e_l(\mathbf{p})\|^2 \quad (7.4)$$

Because this is an over-determined system, we can use the mean face shape and mean albedo to estimate the light intensities, which still gives a very accurate estimation. However, this might normalize the difference in intensity of the skin color. If the light intensities are negative, we skip the rest of the computations.

7.2.4 Estimate the initial face shape

To estimate the initial face shape given the light conditions $\{\mathbf{s}_l, \mathbf{e}_l(\mathbf{p}), i_a, i_d\}$, we use the following two assumptions: Firstly, the Phong model must hold, which gives us the following equations:

$$b(\mathbf{p}) = c(\mathbf{p}) i_a(\mathbf{p}) + h_x(\mathbf{p}) s_{x,l} i_d e_l(\mathbf{p}) + h_y(\mathbf{p}) s_{y,l} i_d e_l(\mathbf{p}) + h_z(\mathbf{p}) s_{z,l} i_d e_l(\mathbf{p}) \quad (7.5)$$

Secondly, the face shape should be similar to the mean face shape. This can be measure by taking the Mahalanobis distance between the face shape $\mathbf{h}(\mathbf{p})$ and the mean face shape $\overline{\mathbf{h}}(\mathbf{p})$. Using Lagrange multipliers, we can minimize the distance with Equation 7.5 as a constrain. This allows us to find an initial face shape $\hat{\mathbf{h}}(\mathbf{p})$, which we will improve in the next steps using a surface model together with geometrical constrains.

7.2.5 Estimate surface using geometrical constrains and a 3D surface model

Given an estimate of the face shape $\hat{\mathbf{h}}$, we want to determine the surface $\mathbf{z}$, which is a depth map of the face image. Given a set of 3D range images of faces, we can calculate depth maps $\{\mathbf{z}^t\}_{t=1}^T$ and we can obtain the mean surface $\overline{\mathbf{z}}$ and a covariance matrix $\Sigma_{\mathbf{z}}$. Using Principal Component Analysis (PCA), we computer the subspace by solving the eigenvalue problem:

$$\Lambda_{\mathbf{z}} = \Phi^T \Sigma_{\mathbf{z}} \Phi \qquad \hat{\mathbf{z}} = \overline{\mathbf{z}} + \sum_{k=0}^{K} \Phi_k u_{\mathbf{z}}(k) \quad (7.6)$$

where $\Lambda_{\mathbf{z}}$ are the eigenvalues and Φ are the eigenvectors of the covariance matrix $\Sigma_{\mathbf{z}}$, which allows to express the surface in variations $\mathbf{u}_{\mathbf{z}}$ for the mean surface $\overline{\mathbf{z}}$. We also know that $h_{xz}(\mathbf{p}) = \frac{h_x(\mathbf{p})}{h_z(\mathbf{p})} = \nabla_x z(p)$ and $h_{yz}(\mathbf{p}) = \frac{h_y(\mathbf{p})}{h_z(\mathbf{p})} = \nabla_y z(\mathbf{p})$ holds, where ∇_x and ∇_y denote the gradient in x and y direction. This allows us to calculate the variations of the surface $\mathbf{u}_{\mathbf{z}}$ using the following equation:

$$\mathbf{u}_{\mathbf{z}} = \arg\min_{\mathbf{u}_{\mathbf{z}}} \; ||\nabla_x \overline{\mathbf{z}} + \nabla_x \Phi \mathbf{u}_{\mathbf{z}} - \hat{\mathbf{h}}_{xz}||^2 + ||\nabla_y \overline{\mathbf{z}} + \nabla_y \Phi \mathbf{u}_{\mathbf{z}} - \hat{\mathbf{h}}_{yz}||^2 \quad (7.7)$$

The surface $\hat{\mathbf{z}}$ can be found using Equation 9.3 and from this surface we can also find the surface normals $\mathbf{n}(\mathbf{p})$. In this case, the surface normals are restricted by geometrical constrains. Using only the geometrical constrains does not have to be sufficient to determine the face surface, therefore, we use the surface model to ensure the convergence.

7.2.6 Computing the albedo and its variations

In the previous sections, we obtained the surface normals $\mathbf{n}(\mathbf{p})$ and the illumination conditions $\{\mathbf{s}_l, \mathbf{e}_l(\mathbf{p}), i_a, i_d\}$. This allows us to calculate the albedo $\mathbf{c}$ from Equation 7.2. In order to find out whether the albedo is correct, we also create a PCA model of the albedo. Given a set of face images together with their 3D range maps, we estimated the albedo, see [23]. Vectorizing the albedo $\{\mathbf{c}^t\}_{t=1}^T$ allows us to calculate a PCA model and find the variations $\mathbf{u_c}$, which is also used for the surface model. Using the variations $\mathbf{u_c}$, we calculated also a projection of albedo $\hat{\mathbf{c}}$ to PCA model. The projection $\hat{\mathbf{c}}$ does not contain all details necessary for the face recognition. For this reason, we use the albedo $\hat{\mathbf{c}}$ from the PCA model in the evaluation criteria, while we use the albedo $\mathbf{c}$ obtained from Equation 7.2 in the reconstructed image.

7.2.7 Evaluation of the found parameters

Because we calculate the face shape for multiple light directions, we have to determine which light direction results in the best face shape. Furthermore, the downhill simplex algorithms also needs an evaluation criteria to be able to find the light direction more accurately. Using the found light conditions and face shape, we can reconstruct an image $\mathbf{b}_r$ which should be similar to the original image. This can be measured using the sum of the square differences between the pixels values. Minimizing this may cause overfitting of our models at certain light directions. For this reason, we use the maximum a posterior probability estimator given by $P(\mathbf{u_c}, \mathbf{u_z}|\mathbf{b})$, which can be minimized by the following equations, see [20]:

$$E = \frac{1}{\sigma_b} \sum_p \|b(\mathbf{p}) - b_r(\mathbf{p})\|^2 + \sum_{k=1}^{K} \frac{u_{\mathbf{z}}^2(k)}{\lambda_{\mathbf{z}}(k)} + \sum_{j=1}^{J} \frac{u_{\mathbf{c}}^2(j)}{\lambda_{\mathbf{c}}(j)} \tag{7.8}$$

In this case, σ_b controls the relative weight of the prior probability, which is the most important factor to minimize. $\lambda_{\mathbf{z}}$ and $\lambda_{\mathbf{c}}$ are the eigenvalues of the surface and albedo. The light directions that minimizes Equation 7.8, give us the parameters from which we can reconstruct a face image with frontal illumination.

7.3 Experiments and Results

We correct for illumination by estimating both the illumination conditions and the face surface. In this section, we will show some of the estimate surfaces together with their corrected images. The main purpose of the illumination correction is

to improve the performance of the face recognition method. Our goal is therefore to demonstrate that our face recognition method indeed benefits from the improvement in the illumination correction. For this purpose, we use the FRGCv1 database where we have controlled face images in the enrollment and uncontrolled face images as probe images.

7.3.1 3D Database to train the Illumination Correction Models

For our method, a database is needed that contains both face images and 3D range maps to compute the surface, shape, shadow and albedo models. In this case, we used the Spring 2003 subset of Face Recognition Grand Challenge (FRGC) database, which contains face images together with their 3D range maps. These face images contain almost frontal illumination and no shadows, making this subset of the database ideal to compute the surface and albedo. The exact method to retrieve the albedo is describe in [23].

7.3.2 Recognition Experiment on FRGCv1 database

Figure 7.1: Reconstructed frontal illuminated face images and surfaces - First row contains the original images from the FRGCv1 database, second and third row show the resulting surface, the fourth row depicts the reconstructed frontal illumination

The FRGCv1 database contains frontal face images taken under both controlled and uncontrolled illumination conditions as is shown in Figure 7.1. The first image in Figure 7.1 is taken under controlled conditions, while the other images are taken under uncontrolled conditions. For the first person, we show that our method is able to correct for different unknown illumination conditions. In case of the last image, we observe more highlighted areas directly under the eyes, this is caused by the reflection of the glasses which are not modelled by our method.

In order to test if this illumination correction method improves the performance in face recognition, we performed the following experiment to see if illumination conditions are removed in the images taken under uncontrolled conditions. In this case, we use the images with uncontrolled illumination as probe image and make one user template for every person with the images taken under controlled conditions. To train our face recognition method, we randomly divided both the controlled and uncontrolled set of the FRGCv1 database into two parts, each containing approximately half of the face images. The first halves are used to train the face recognition method. The second half of the controlled set is used to compute the user templates, while the second half of the uncontrolled set is used as probe images. We repeat this experiment 20 times using different images in both halves to become invariant against statistical fluctuations.

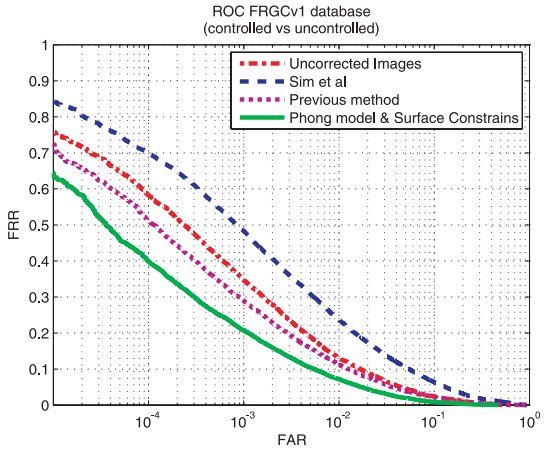


Figure 7.2: ROCs of illumination correction methods on the FRCGv1 database - ROC on the FRGCv1 database with a comparison to our previous work [23] and to work of Sim et al [110]

The Receiver operating characteristic (ROC) in Figure 7.2 is obtained using the PCA-LDA likelihood ratio [126] for face recognition. The first three lines in Figure 7.2 are also stated in our previous work [23], where the last line depicts the improvements obtained using the method described in this chapter. The ROC curves in Figure 7.2 shows only the improvements due to illumination correction. In the case of our previous method, the False Reject Rate (FRR) becomes significantly better at a False Accept Rate (FAR) smaller than 1%, while the last line is overall better. Most other illumination correction methods like [5; 139] evaluated their method only on a database create in a laboratory or do not perform a recognition experiment, which makes the comparison with other methods difficult.

7.4 Discussion

In figure 7.1, we observe that the phenomenon, where the shadow areas are still not completely dark, often occurs in uncontrolled illumination conditions. Improving our model on this point gave also improvements in the recognition results, which was the main purpose of our illumination correction. We choose to ignore other illumination effects like specular reflections, because we expected a small performance gain in face recognition and a large increase in computation time.
The second improvement is a restriction to the face shape by computing the surface instead of the surface normals, which also slightly improved the face recognition results. Another benefit is that we obtained an estimation of the surface of the face, which might be handy in other applications. In our research, the focus has not been on the quality of the estimated surfaces. Although we expect that this can be an interesting result to improve for instance 3D face acquisition and recognition.

7.5 Conclusion

We present two major improvements for our illumination correction method for face images, where the purpose of our method is to improve the recognition results of images taken under uncontrolled illumination conditions. The first improvement uses a better illumination model, which allows us to model the ambient light in the shadow areas. The second improvement computes an estimate of the surface given a single face image. This surface gives us a more accurate face shape and might also be useful in other applications. Because of both improvements, the performance in face recognition becomes significantly better for face images with uncontrolled illumination conditions.

8

COMBINING ILLUMINATION NORMALIZATION METHODS

8.1 Introduction

One of the major problems with face recognition under uncontrolled conditions is the illumination variation, which is often larger than the variations between individuals. Using illumination normalization methods, we want to correct the illumination variations in a single face image. In the literature, several methods have been proposed to make face images invariant for illumination. These methods can be divided into two categories. The first category normalizes the face image by applying a preprocessing step on the pixel values using information from the local region around that pixel. Examples of these approaches are Histogram Equalization [105] or (Simplified) Local Binary Patterns [55],[121]. These approaches are direct and simple, but fail to model the global illumination conditions. The second category estimates a global physical model of the illumination mechanism and its interaction with the facial surface. In this category falls for instance the Quotient Image [107], Spherical harmonics [14], 3D morphable models [20]. These methods are able to estimate the global illumination condition, but are also more complicated and require training to model the illumination conditions.
In practise, we have observed that both categories of illumination normalization algorithms have their advantages and disadvantages. The methods in the first category have problems with regions which are not illuminated because of hard

shadows. These shadow regions have a large signal to noise ratio which makes the correction for local methods almost impossible. However, these local methods work well on the illuminated parts of the image and on lightly shadowed (soft shadows) areas. The second category is able to reconstruct the parts with hard shadows, using statistical models. But our current implementation of a global method does not model face variations like glasses and expressions. To summarize, the local methods work well on illuminated part of the image, also the parts which are not modelled by the global methods. The global methods are however able to reconstruct parts which contain hard shadows, which is not possible using a local method. By combining methods from both categories, we aim to improve the performance in face recognition under different and uncontrolled illumination conditions.
Combining the two different illumination correction methods can be done on three different levels, namely at the feature level, the score level and the decision level. We will concentrate on the last two levels of fusion because of its simplicity. To achieve this, the preprocessed images are individually classified and the scores are fused using score-level [44], decision-level fusion [122] and a combination of both these methods named hybrid fusion [120].
This chapter is organized as follows, in Section 2 we describe the two illumination correction algorithms. Section 3 explains how we combine these algorithms for face recognition. In Section 4, we show the experiments and results and Section 5 gives the conclusions.

8.2 Illumination normalization

For illumination normalization, we use two methods which come from different categories. The method from the first category is the Local Binary Patterns [86] preprocessing, where different papers [55],[121] claim its invariance to illumination conditions. In the second category, we use the illumination correction approach in [23], which is able to correct illumination variations using a single 2D facial image as its input. If we compare this method with [107] or [110], it is more advanced using a 3D shape model and a shadow and reflection model, but in comparison with 3D morphable models [20] it is still computational efficient. In the following subsections, we describe both methods in more detail.

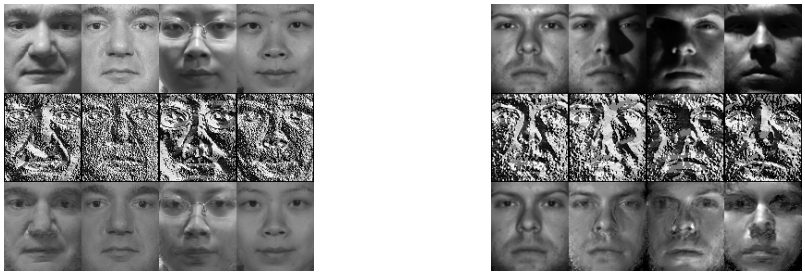
8.2.1 Local Binary Patterns

Local Binary Patterns method (LBP) is proposed in [86] and are often used as features in face recognition. The standard LBP give the 3×3-neighbors the value 0 if they are smaller than the center pixel value and 1 otherwise. This result in a 8 bit string, which represents the pattern at the center point. We can also obtain from it a decimal representation between 0 and 255. LBP is a relative measure which makes it invariant against monotonic gray-scale transformations. A big range of illumination changes consist of monotonic gray-scale transformations in large regions in the image.

There are also extensions to the original LBP, which use a bigger radius and more spacing for the neighbors or use a different weighting scheme for the different bits. In this chapter, we use the simplest LBP as preprocessing to obtain the illumination invariant images, i.e. based on a 3×3-neighborhood.

8.2.2 Model-based Face Illumination Correction

The method from the second category is described in Chapter 6 and published in [23]. It uses a global method based on the Lambertian Reflectance model and subspace methods to correct for illumination in facial images. Examples of images corrected with this method are shown in Figure 8.1. This method reconstructs a frontal illuminated facial image based on the input image, instead of creating a illumination free feature vector.



(a) FRGCv1 database (b) Yale B databases

Figure 8.1: Examples of corrected face images - Face Images from the Yale B database and FRGCv1 database, upper without correction, middle preprocessed with LBP, lower preprocessed with Model-based Illumination Correction

8.3 Fusion to improve recognition

In the previous section, we proposed two methods to obtain illumination invariant face images. The resulting images of these methods can be seen in Figure 8.1. In this section, we combine these methods to improve the face recognition under different illumination conditions. We train two face classifiers, one with the LBP and one with Model-based Face Illumination correction. The face classifier (log-likelihood ratio after a feature reduction using a PCA and LDA transformation) gives us a similarity score, which we are able to fuse. For fusion, we use the following methods: SUM rule score-level fusion [44], OR rule decision-level fusion [122] and hybrid fusion [120].

In the case of score-level fusion, we can take the joint likelihood ratio, which in our case means that we can sum the scores obtained from the log-likelihoods ratios. This gives us the advantage that we do not have to estimate the different density functions or perform a normalization step to the similarity scores. This method of fusion, we denote as SUM rule fusion.
Although theoretically, score-level fusion should achieve the optimal performance, it is not very robust to outliers. For this reason, we also use decision-level fusion with the OR rule to combine the reciever operation characteristics (ROC). The ROC is determined from the similarity scores of the face classifiers and can be obtained by varying the threshold, thus producing a different false reject rate β and false accept rate $\alpha = 1 - p_r$. This specific pair (α, β) is called an operation point, which corresponds to a threshold t in the similarity scores. In the case of fusion, there can be N classifiers and each is characterized by its ROC, $p_{r,i}(\beta_i), i = 1, ..., N$. By assuming that our classifiers are independent, the final performance of the OR rule can be estimated as $\beta = \prod_{i=1}^{N} \beta_i$ and $p_r(\beta) = \prod_{i=1}^{N} p_{r,i}(\beta_i)$. By searching for the optimal operation points, the fusion with the OR rule can be formulated as:

$$\hat{p}_r(\beta) = \max_{\beta_i | \prod_{i=1}^{N} \beta_i = \beta} \{ \prod_{i=1}^{N} p_{r,i}(\beta_i) \} \tag{8.1}$$

We can prove that the estimated $\hat{p}_r(\beta)$ is never smaller than any of the components $p_{r,i}(\beta), i = 1, ..., N$ at the same β. Because we do not have the ROC $\hat{p}_r(\beta)$ in analytical form, we estimate a ROC from evaluation data. The ROC $\hat{p}_r(\beta)$ is therefor characterized by discrete values and can be solved numerically [122].
The hybrid fusion is a combination between score-level fusion and decision-level fusion. In hybrid fusion, we first perform the SUM rule score-level fusion to combine the ROCs of both classifiers. The ROCs of both classifiers together with ROC given by the SUM rule are then fused using OR rule decision-level fusion. Using hybrid fusion, we hope to combine the advantages of score-level fusion and the decision-level fusion.

8.4 Experiments and Results

The purpose of the illumination corrections is to improve the verification rates in face recognition. We performed a recognition experiment on the Yale B databases (the Yale B [48] and extended Yale B [72] database), which contain face images under different labelled illumination conditions created in a laboratory. This experiment tests the ability to make illumination invariant images under all kinds of illumination conditions (also with hard shadows). We also perform an experiment on the Face Recognition Grand Challenge version 1 (FRGCv1) database [90] which contain face images taken under controlled and uncontrolled conditions. This experiment allows us test the ability to correct the uncontrolled illumination conditions and compare them to the controlled images in the gallery.

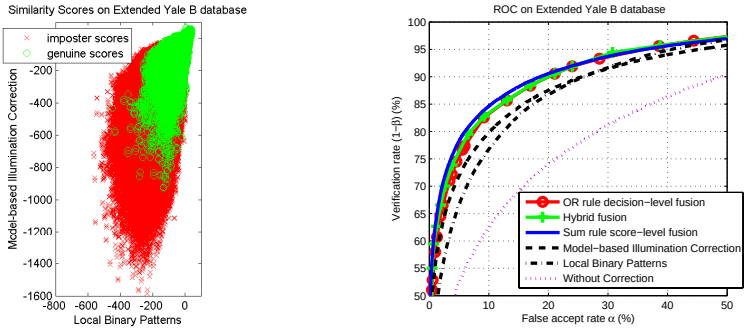


Figure 8.2: Score Plot and ROC on the Extended Yale B database - Score Plot and ROC of the two illumination correction methods on the Extended Yale B database, ROCs also contains the fusion results and the results without illumination correction

8.4.1 The Yale B databases

The Yale B databases are created to model and test the effects of illumination on face images. In our correction algorithm, we use the Yale B databases to obtain our error model for shadows and reflections. Because we trained our error model on the Yale B databases, we performed a leave one person out experiment for the Model-based Illumination Correction.
In our face recognition experiment, we correct all the face images with both correction methods. We use only the face images with a azimuth and elevation angle below ± 90 degrees in this experiment. For face recognition, we trained on the face images of thirty persons and performed a one-to-one verification experiment on the remaining eight persons, leaving out the face images taken under similar illumination conditions. We repeated this experiment until we compared the face images of all person in the Yale B databases with each other. To train the fusion methods, we used the scores from the Yale B database (10 persons). For testing, we used the scores of Extended Yale B database (29 persons). In Figure 8.2, we show all the results on the Extended Yale B database.
The experiment which we perform on the Yale B databases is a difficult experiment, because sometimes two images illuminated from opposite positions are compared. We observe from Figure 8.2(b) that the Model-based Illumination Correction works better at a FAR $< 25\%$ than the Local Binary Patterns. By fusing the two methods, we can improve the recognition results significantly. In Figure 8.2(a), we observe that a diagonal line is probably the best separation between imposter scores and genuine scores. This explains why the SUM rule performs slightly better than the OR rule and hybrid fusion (see Figure 8.2(b)).

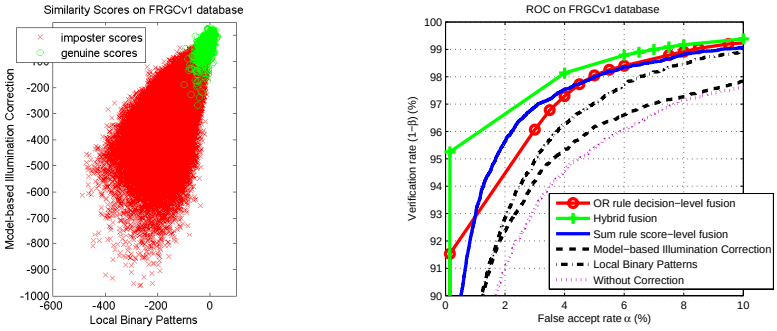


Figure 8.3: Score Plot and ROC on FRGCv1 database - Score Plot and ROC of the two illumination correction methods on the FRGCv1 database, ROCs also contains the fusion results and the results without illumination correction

8.4.2 The FRGCv1 database

The FRGCv1 database contains frontal face images taken under both controlled and uncontrolled conditions. In our experiment, we randomly divided the uncontrolled and controlled face images into two parts, each containing approximately half of the face images. We used the first halves of both sets to train our face classifiers and the fusion methods, the second half of the controlled images are used for the enrollment of the one user template for every person and the second half of the uncontrolled images are used as probe images. We repeat this experiment 20 times using different random splits of the database to become invariant for statistical fluctuations. This experiment simulates a video surveillance scenario, where we usually have a gallery of high quality images, but the probe images are obtained under uncontrolled conditions. Both our illumination correction algorithms preprocess all the images, also the controlled images. The recognition results are shown in Figure 8.3.
Although, we have in this experiment less extreme illumination conditions, there are also other challenges in the FRGCv1 database beside illumination, like expressions and out of focus images. From Figure 8.3(b), we observe that the Local Binary Patterns work better on this database than the Model-based Illumination Correction. The main reason for this difference is that Model-based Illumination Correction has a larger amount of outliers, due to glasses and expressions (see also Figure 8.3(a) where relatively many genuine scores (circles) have larger negative values for the Model-based Illumination Correction). Using the simple SUM rule to combine the face classifiers already improves the overall recognition results. In figure 8.3(b), we observe that the recognition results of the OR rule are similar to the sum rule, and the hybrid fusion clearly outperforms the other fusion methods on this database.

8.5 Conclusions

We combine two different methods to correct for illumination in face images and obtain better results in face recognition. We show that both methods are able to correct for illumination in face images. The Local Binary Patterns method corrects pixel values based on the local neighborhood in the image. This method shows good results on uncontrolled images, but cannot recover large regions with hard shadows. The Model-based Illumination Correction shows that it can deal with these shadow regions, but it has problems in uncontrolled conditions which contain unmodelled effects, like glasses and expressions. Because both methods have different strengths and weaknesses, we combine the illumination normalization methods using fusion. We use three different fusion methods: SUM rule score-level fusion, OR rule decision-level fusion and hybrid fusion. The performance of the simple SUM rule fusion already improves the results significantly and works best for large variations in illumination. The performance of the OR-rule is in both experiments slightly worse than the SUM rule. The hybrid fusion, which tries to combine the advantages of both fusion methods, gives the largest improvement in performance when we correct for uncontrolled illumination conditions occurring in a video surveillance environment.

9

VIRTUAL ILLUMINATION GRID FOR CORRECTION OF UNCONTROLLED ILLUMINATION IN FACIAL IMAGES

9.1 Introduction

One of the major problems of face recognition in uncontrolled conditions is the variation caused by illumination. Our contribution is an illumination correction method that is capable of handling multiple light sources. The purpose is to correct for multiple light sources in a single facial image. Correcting for illumination effects in images taken under uncontrolled illumination conditions is more challenging than the standard experiments for face illumination, which address removing illumination from facial images recorded in laboratory conditions (Yale B database, CMU-PIE database), illuminated with a single light source. In order to correct for uncontrolled illumination conditions, we try to reconstruct the illumination conditions. The requirements of the fontal illuminated facial image are that it removes the illumination variations without introducing artifacts, while preserving the identity information for recognition. The illumination correction method is used as an independent preprocessing method. The advantage is that the generated frontal illuminated facial image can be the input of various face recognition methods and allows us to use a single gallery image in face recognition.
Several methods have been proposed to correct for illumination variations in facial

images. We will categorize them based on two criteria. The first criterion is the complexity of the reflectance model. We discriminate between reflectance models that use a weak assumption about illumination, the Lambertian reflectance model, a reflectance function based on the Lambertian assumption (for example Spherical Harmonics) and advanced reflectance models (Phong, Torrance-Sparrow). The second criterion is the complexity of the face model. There are methods, which use no model of the face, other methods make models of the appearance, some use an implicit model of the 3D surface and texture, while other methods have applied an explicit model of the 3D surface and texture. In Figure 9.1, we categorize illumination correction methods based on these criteria and divide them into four groups.

The first group (bottom-left oval of Figure 9.1) contains methods that do not need a face model and do not assume an explicit reflectance model. These methods usually perform preprocessing based on the local regions, for example Histogram Equalization [105] or (Simplified) Local Binary Patterns [55; 121]. Other methods like Gross et al [50] and Tan et al [119] use the local region around the pixel to perform illumination correction based on some properties of the reflectance. The Self Quotient Image [128] uses the local region for correction based on the Lambertian reflectance model without the need of a model of the face.

There are also methods that learn the behaviour of reflectance and so become invariant for illumination in the face. These methods use only the appearance learned for a bootstrap database with different labeled illumination conditions. Note that they do not assume a reflectance model (center-left oval of Figure 9.1). Tensorfaces [125] can be trained to handle multiple variations like expression, illumination and pose using multidimensional subspace models. This is extended in [71] with a model which also include 3D shape parameters, but the illumination is still modelled using subspace models. In [33], a subspace model is used to compute Intrinsic Images, separating the image in reflectance and illuminance. Subspace models that learn the behaviour of the reflectance on faces also allow correction for this reflectance. However, these methods usually depend heavily on a bootstrap database with varying illumination conditions. This makes it difficult to predict whether these methods are robust to unseen conditions, which is mostly the case with uncontrolled conditions.

In face illumination correction, many correction methods use both assumptions on the illumination reflectance as well as implicitly taking into account the 3D surface and texture, usually by estimation of surface normals and albedo (center-right oval of Figure 9.1). An example is the Quotient Image [107] which estimates illumination using the Lambertian reflectance model. This method computes a quotient image based on the assumption that faces have a similar surface. A estimate of the surface is then obtained from a bootstrap set of faces. In [48], an illumination cone can be determined from three images illuminated with independent light sources. Sim et al [110] proposed a method based on the Lambertian reflectance model which corrects for illumination in a single facial image. This method uses a large bootstrap database containing many illumination variations in order to correct for these conditions. Spherical Harmonics are proposed in both [14] and [94] which give an approximation of a 9D linear subspace under all possible Lambertian il-

luminations. Zhang et al [139],[140] proposed a method to obtain the Spherical Harmonics for a single image illuminated under unknown illumination by using a bootstrap database to model shadows and reflections. In [72], a configuration of nine points of light (9PL) is determined to construct a linear subspace for face recognition. In face recognition, nine gallery images illuminated or rendered under predefined conditions are necessary to perform face recognition, which requires specialized data acquisition of the gallery images. Zhou et al [144],[143] span a linear subspace using object-specific albedo-shape matrices. They find an illumination free identity vector by optimizing both identity vector and illumination conditions to best resemble the input image. This gives them an illumination free identity vector instead of an estimate of the 3D surface and albedo, from which we render the frontal illuminated facial image. Because Zhou et al's method is the closest to ours, we have chosen to point out other differences with Zhou et al's method throughout the text.
The last group in Figure 9.1 (top-right oval) differs from other methods because it uses 3D face models. The 3D morphable models [20] are among the first to use the 3D information of faces, where PCA models are used for both the 3D shape and texture. Using the Phong reflectance model for illumination, a parameter optimization method is used to render a facial image which is close to the input image. In [111], Smith et al propose a statistical model for normal maps to accurately estimate the surface and in [112] the same authors use a subspace model of the depth map as a geometrical constraint. In [23], we use a shape model in combination with the Lambertian reflectance model to correct for illumination of a single light source. In [27], we have improved this method by modelling both ambient and diffuse illumination and estimated the depth maps. We observe that most illumination correction methods have difficulties in modelling multiple light sources, especially when they cause some reflectance in shadow areas.

This paper is organized as follows: In Section 9.2, we describe the illumination correction method that is able to deal with multiple light sources. We first introduce the necessary reflectance and face models which then allows us to calculate a reconstruction of the face shape iteratively. In Section 9.3, we describe experiments and the results and fuse our global illumination correction with a local illumination correction method. We will discuss the results obtained with our method in Section 9.4 and finally provide conclusions in Section 9.5.

9.2 Method

9.2.1 Reflectance model

In order to correct for the illumination in a single facial image, we use a reflectance model for the behaviour of illumination. Because of our focus on uncontrolled illumination conditions, we assume that faces are illuminated by multiple light sources. We use the Lambertian reflectance model, which gives a good approximation of the reflectance behaviour on the surface of faces [48]. The image intensity $b \in \mathcal{R}$ at a certain position $\mathbf{p} = \{x, y\}$ in the image can be described by the following

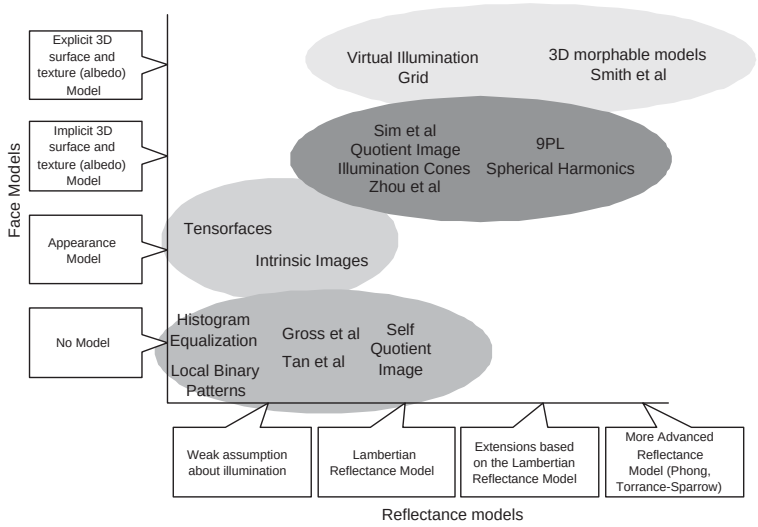


Figure 9.1: The categorization (in face and reflectance models) of the illumination correction methods: separating four major groups based on two criteria

equation:

$$b(\mathbf{p}) = \rho(\mathbf{p}) \sum_{l} \max(0, \mathbf{n}(\mathbf{p})^T \mathbf{s}_l i_l) \tag{9.1}$$

where the face shape $\mathbf{h}(\mathbf{p}) = \rho(\mathbf{p})\mathbf{n}(\mathbf{p})^T$ consists of the surface normals $\mathbf{n} \in \mathcal{R}^3$, the albedo of the surface given by $\rho \in \mathcal{R}$ and the max operation allows us to model attached shadows. A normalized vector $\mathbf{s} \in \mathcal{R}^3$, defines the direction of the illumination. The intensity of the light is given by $i \in \mathcal{R}$. Instead of finding multiple light directions in a continuous domain, we use L discrete directions, assuming that a light source in continuous direction can be created using multiple light sources in discrete directions. The Lambertian reflectance model in Equation 9.1 in this form cannot model cast shadows on the face surface. There are two kinds of shadows on faces. The first kind is called "attached shadows". In this case, the Lambertian reflectance model does not hold because the normal is not directly facing the light source. This results in a negative image intensity, which can be easily detected and corrected by replacing the negative value by zero. The second kind of shadow is due to the geometry of the face that blocks the light source, these are called "cast shadows". These shadows are harder to calculate because we need to perform ray tracing. Shadows can be seen as hard binary decisions. This definition holds with the exception of areas, which contain the transition between light and shadow areas. We propose to model shadows in the Lambertian reflectance model using a weight $e_l(\mathbf{p})$, which is linked to the light direction. This weight is in fact the expectation $e_l \in [0, 1]$ that a shadow occurs

at position $\mathbf{p}$ given a certain light direction l:

$$\hat{b}(\mathbf{p}) = \rho(\mathbf{p})\mathbf{n}(\mathbf{p})^T \sum_l \mathbf{s}_l i_l e_l(\mathbf{p}) \quad (9.2)$$

The illumination conditions for a certain position $\mathbf{p}$ can then be described by $\mathbf{v}(\mathbf{p}) = \sum_l \mathbf{s}_l i_l e_l(\mathbf{p})$. In case of an attached shadow, $e_l = 0$, thus making the max operation in Equation 9.1 unnecessary. This user-independent expectation can be used as weight, giving smooth values in the areas which contain the transition between light and shadow, while we have a hard binary decision in area that certainly contain shadows. This expectation is determined from a training set of multiple surfaces, where we calculate for a single surface a binary decision that a shadow occurs at position $\mathbf{p}$ give a light direction l using a ray tracer. The expectation is obtained by taking the mean over all the binary values. We determine the expectation at all positions $\mathbf{p}$ and for all L light directions in the grid. Apart from the expectation, we also determine the variations $\sigma_l^2(\mathbf{p})$ at every position, which we use for the albedo estimate, described in Section 9.2.7.
The goal of our correction method is to find the illumination conditions $\mathbf{v}(\mathbf{p})$ and the face shape $\mathbf{h}(\mathbf{p})$ that best explain our input image $b(\mathbf{p})$. This method minimizes the distance between the input image $b(\mathbf{p})$ and an estimate based on the models $\hat{b}(\mathbf{p})$, in our case obtained from Equation 9.2. Note that multiple combinations of light conditions and face shapes can result in the same image based on Equation 9.2. For this reason, it is necessary to use domain specific knowledge to constrain the shape from shading problem. This domain specific knowledge is enforced using subspace models of the face shape (Section 9.2.2). The subspace models also allow us to estimate the face shape (surface and albedo). By obtaining the face shape, we can easily compute facial images under frontal illumination by replacing the $\mathbf{v}(\mathbf{p})$ for $\mathbf{v}_{frontal}(\mathbf{p})$. Of course, it is also possible for us to illuminate the face with other illumination conditions using our virtual illumination grid.

9.2.2 Face Shape and Albedo Models

The reflectance of the light depends on the object that is illuminated. Faces are difficult to model, but multiple techniques have been proposed in the literature to achieve this. Zhou et al [143] use a linear subspace of object-specific albedo-shape matrices as illumination free term containing both albedo and surface normals. In VIG, we have chosen to split the illumination free terms in order to create two separate linear subspaces. This has the advantage of allowing us to perform an estimate of the surface (Section 9.2.6). For the subspace models, we use a vectorized representation of albedo $\boldsymbol{\rho}$ and surface $\mathbf{z}$ (depth map of the face), instead of a notation for every position $\mathbf{p}$ in an image. From a set of surfaces $\{\mathbf{z}^m\}_{m=1}^M$, we obtain the mean surface $\overline{\mathbf{z}}$ and a covariance matrix $\Sigma_{\mathbf{z}}$. This allows us to compute a subspace by solving the eigenvalue problem, obtaining the eigenvalues $\Lambda_{\mathbf{z}}$ and the eigenvectors $\Phi_{\mathbf{z}} = [\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_N]^T$ of the covariance matrix $\Sigma_{\mathbf{z}}$.

$$\hat{\mathbf{z}} = \overline{\mathbf{z}} + \Phi_{\mathbf{z}}\mathbf{u}_{\mathbf{z}} \quad (9.3)$$

Given that we are able to obtain the variations of the surface $\mathbf{u_z}$, we can estimate a surface $\hat{\mathbf{z}}$ using Equation 9.3. A similar PCA model is obtained for the albedo. In order to take into account the correlation between depth maps and albedo, which is probably present, we combine both PCA models, creating the following concatenated vector:

$$\mathbf{y_t} = \begin{pmatrix} W_\mathbf{z}\mathbf{u_z} \\ \mathbf{u}_{\boldsymbol{\rho}} \end{pmatrix} \tag{9.4}$$

where $W_\mathbf{z}$ is a diagonal weighting matrix, allowing difference in weight between depth maps and albedo. This method of combining both PCA models is similar to the approach used in the Active Appearance Model [38]. Because albedo and depth maps cannot be compared directly, we measure the effects in appearance of changing $\mathbf{u_z}$ and $\mathbf{u_c}$ using the RMS error. This gives us a relative weight between albedo en depth maps changes. We apply PCA to the vector $\mathbf{y_t}$, giving us a model which contains both depth maps and albedo:

$$\mathbf{y_t} = \Phi_\mathbf{t}\mathbf{u_t} \tag{9.5}$$

In this case, $\Phi_\mathbf{t} = [\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_K]^T$ are the eigenvectors and $\mathbf{y_t}$ is already zero mean allowing us to control the depth map and shape in the following way:

$$\hat{\mathbf{z}} = \overline{\mathbf{z}} + \Phi_\mathbf{z} W_\mathbf{z}^{-1} \Phi_\mathbf{tz} \mathbf{u_t} \tag{9.6}$$

$$\hat{\boldsymbol{\rho}} = \overline{\boldsymbol{\rho}} + \Phi_{\boldsymbol{\rho}} \Phi_{\mathbf{t}\boldsymbol{\rho}} \mathbf{u_t} \tag{9.7}$$

where

$$\Phi_\mathbf{t} = \begin{pmatrix} \Phi_\mathbf{tz} \\ \Phi_{\mathbf{t}\boldsymbol{\rho}} \end{pmatrix} \tag{9.8}$$

Several papers already have used separate PCA models for texture and shape in order to correct for illumination [20]. However, they usually do not take into account the correlation between the depth map and the albedo. Using this correlation has the advantage of prohibiting improbable combinations of depth maps and albedo to explain appearances in images.

9.2.3 Illumination Correction Method

Given a single image, we want to estimate the face shape and the illumination conditions, using both the Lambertian Reflectance model (Equation 9.2) and the PCA models of surface and albedo (Equations 9.6 and 9.7). Of course, we can perform an exhaustive search, as in the 3D morphable models where optimized code can perform such an operation with 4.5 minutes [19]. Instead, we chose an iterative scheme, to first estimate the illumination conditions and then find the model parameters of the depth map and surface. Using the found depth map and albedo, we can then improve the accuracy of the illumination conditions, which in turn improves the depth map and albedo. We repeat these steps several times

(average of 5 iterations). Because the most time consuming computations are linear, this method can be computed much faster than the 3D morphable model. The pseudo-code of our correction method is given below:

- Repeat
  - Estimate illumination conditions (Section 9.2.4)
  - Estimate crude face shape (Section 9.2.5)
  - Estimate the surface parameters (Section 9.2.6)
  - Estimate the albedo parameters (Section 9.2.7)
- Until convergence (based on evaluation of the obtained illumination conditions, surface and albedo (Section 9.2.8))
- Refinement of the albedo (Section 9.2.9)

We will discuss the different components in the following sections.

9.2.4 Estimation of the illumination conditions

Given the image, we want to estimate the illumination conditions in the image. Because we assume a grid of discrete light directions, we have to determine the intensity of the light for every point in the grid in order to calculate the global illumination conditions. To obtain the illumination conditions, we use an estimate of the face shape $\tilde{\mathbf{h}}(\mathbf{p})$. In the first iteration, we use the mean face shape $\overline{\mathbf{h}}(\mathbf{p})$ to obtain the light intensities, while in the next iterations, we use the estimated face shape from the previous iteration. The light intensities are calculated as follows

$$\begin{aligned} \mathbf{i} &= \arg\min_{\mathbf{i}} \sum_{\mathbf{p}} \| \sum_{l=0}^{L} \tilde{\mathbf{h}}(\mathbf{p})^T \mathbf{s}_l i_l e_l(\mathbf{p}) - b(\mathbf{p}) \|^2 \\ &\quad \text{where } i_l \geq 0 \end{aligned} \tag{9.9}$$

This can be solved using a constrained linear least square solver where the light intensity cannot be negative. Because Equation 9.9 is an overcomplete system, even using a relatively poor estimate of the face shape $\tilde{\mathbf{h}}$, still gives acceptable results. The accuracy of the illumination conditions depends also on the configuration of the light sources in the grid. We have experimented with two grids by varying the azimuth and elevation angle by 10 or 20 degrees from -80 to 80 degrees. We have observed that the results of both grids are very similar, while the computations with a grid of 10 degrees take much longer. For this reason, we decide to use the grid of 20 degrees for all experiments. In [72], a configuration of nine points of light (9PL) at predefined locations is used to obtain an illumination free representations. In VIG, however, we want to obtain a reconstruction of the surface and the albedo, for which the large illumination grid is necessary to accurately model cast shadows. With the 9PL methods, accurate modelling of cast shadows is not possible, which will result in artifacts in the reconstruction.

9.2.5 Estimation of the crude face shape

Given the image and the illumination conditions, the goal is to obtain the face shape. A crude estimate of the face shape is necessary in order to improve the face shape using the linear subspace models. To obtain the crude estimate, we use the following two assumptions. The first assumption is that the Lambertian reflectance model (Equation 9.11) holds. The second assumption is that the face shape should be similar to the mean face shape. This can be measured by taking a distance between the face shape $\mathbf{h}(\mathbf{p})$ and the mean face shape $\overline{\mathbf{h}}(\mathbf{p})$. In our case, we have taken the Mahalanobis distance between two vectors, where $\Sigma_{\mathbf{h}(\mathbf{p})}$ is the covariance matrix of the face shape at location $\mathbf{p}$ obtain from a training set.

$$\hat{\mathbf{h}}(\mathbf{p}) = \arg\min_{\mathbf{h}(\mathbf{p})}(\mathbf{h}(\mathbf{p}) - \overline{\mathbf{h}}(\mathbf{p}))^T \Sigma^{-1}_{\mathbf{h}(\mathbf{p})}(\mathbf{h}(\mathbf{p}) - \overline{\mathbf{h}}(\mathbf{p})) \quad (9.10)$$

$$\text{where } b(\mathbf{p}) = \mathbf{h}(\mathbf{p})^T \mathbf{v}(\mathbf{p}) \quad (9.11)$$

We can minimize the Mahalanobis distance (Equation 9.10) with the Lambertian reflectance model (Equation 9.11) as a constraint using Lagrange multipliers. This gives us a crude estimate of face shape $\hat{\mathbf{h}}$, which we will improve in the following sections.

9.2.6 Estimation of the surface

Given an estimate of the face shape $\hat{\mathbf{h}}(\mathbf{p})$, we further improve this estimate by applying geometrical constrains. By calculating a depth map from the crudely estimated face shape, we can automatically enforce the geometrical constrains. The PCA model allows us to introduce domain specific information, which ensures convergence of this shape from shading problem. In the method of Zhou et al [143], integrability and symmetry constrains are used in generalized photometric stereo to recover the shape and albedo, using multiple face images under different illumination conditions. For Zhou et al to become invariant to illumination in a single image, no geometrical constraints are used, while our method allows us to estimate the surface and thus enforce the integrability constraints even for a single image. We know that the gradient of the surface in x and y direction is equal to $\nabla_x z(p) = \frac{h_x(\mathbf{p})}{h_z(\mathbf{p})} = h_{xz}(\mathbf{p})$ and $\nabla_y z(\mathbf{p}) = \frac{h_y(\mathbf{p})}{h_z(\mathbf{p})} = h_{yz}(\mathbf{p})$. Instead of calculating the depth map directly, we estimate the variations $\mathbf{u_z}$, using the following equation:

$$\begin{aligned}\mathbf{u_z} = \arg\min_{\mathbf{u_z}} \quad & ||\nabla_x \overline{\mathbf{z}} + \nabla_x \Phi \mathbf{u_z} - \hat{\mathbf{h}}_{xz}||^2 + \\ & ||\nabla_y \overline{\mathbf{z}} + \nabla_y \Phi \mathbf{u_z} - \hat{\mathbf{h}}_{yz}||^2 \end{aligned} \quad (9.12)$$

A similar procedure to obtain the face surface is performed in [112] to find the variations from the surface model. This can be solved using a linear least square solver. The final depth map can be computed using Equation 9.3. In our case, we combine the depth map and albedo models, to calculate the final depth map taking into account the correlation between the albedo, see Section 9.2.2. A

practical problem in calculating the depth maps is that vectors which are almost perpendicular to the viewer direction sometimes cause large spikes. In order to deal with this problem, we remove these locations. In this case, we are still able to calculate the variations of the surface because PCA can also be applied on an incomplete set of locations.

9.2.7 Estimation of the albedo

In the previous section, we estimate the illumination conditions and the surface from which we can obtain the surface normals $\mathbf{n}(\mathbf{p})$. The only remaining unknown in Equation 9.2 is the albedo $\boldsymbol{\rho}$. To calculate the albedo, we can solve Equation 9.2 for every pixel value, but we observe that there are sometimes erroneous effects in the albedo caused by the user independent shadow model $e_l(\mathbf{p})$, because the shadow mapping is not precise. To overcome this problem, we use Bayes theorem which allows us to calculate a MAP estimate for the albedo. The MAP estimate of the albedo at a certain location is defined as follows:

$$P(\rho(\mathbf{p})|b(\mathbf{p})) = \frac{P(b(\mathbf{p})|\rho(\mathbf{p}))P(\rho(\mathbf{p}))}{P(b(\mathbf{p}))} \quad (9.13)$$

In order to obtain the albedo term, we can maximize the following Equation:

$$\rho_{noshadow}(\mathbf{p}) = \arg\max_{\rho(\mathbf{p})} P(b(\mathbf{p})|\rho(\mathbf{p}))P(\rho(\mathbf{p})) \quad (9.14)$$

$$= \arg\max_{\rho(\mathbf{p})} \mathcal{N}\Big(\rho(\mathbf{p}) - \frac{b(\mathbf{p})}{\mu_r(\mathbf{p})}, \sigma_r(\mathbf{p})\Big) \times \mathcal{N}\Big(\mu_\rho(\mathbf{p}), \sigma_\rho(\mathbf{p})\Big) \quad (9.15)$$

In Equation 9.15, we define a mean reflection term $\mu_r(\mathbf{p}) = \sum_l \mathbf{n}(\mathbf{p})^T \mathbf{s}_l i_l e_l(\mathbf{p})$ and the varations of the reflection term $\sigma_r^2(\mathbf{p}) = \sum_l \mathbf{n}(\mathbf{p})^T \mathbf{s}_l i_l \sigma_l^2(\mathbf{p})$ (see Section 9.2.1 for σ_l^2) and assume that the shadow maps are Gaussian distributed. Using the training set from which we calculate the PCA models, we can also easily determine the mean albedo $\mu_\rho(\mathbf{p})$ and standard deviation of the albedo $\sigma_\rho(\mathbf{p})$ at certain locations necessary in Equation 9.15. By taking the derivative of the log probabilities, we find a shadow free albedo term $\rho_{noshadow}(\mathbf{p})$. From this shadow free albedo $\boldsymbol{\rho}_{noshadow}$, we computed the variations $\mathbf{u}_\rho$ using a linear least square solver giving us all the subspace model parameters, see Section 9.2.2.

9.2.8 Evaluation of the obtained illumination conditions, surface and albedo

In Sections 9.2.6 and 9.2.7, we compute both the variations of the surface and albedo. With these variations, we can determine the variation of the combined models, see Equation 9.5 and give the estimates for the surface and albedo using Equations 9.6 and 9.7. Given the estimated albedo and surface, together with the illumination conditions found in Section 9.2.4, we can reconstruct an image $\hat{\mathbf{b}}$ which should be similar to the original image. This can be measured using the sum of the square differences between the pixel values. It is also interesting to

monitor the variations from the PCA model, which shows if overfitting occurs at certain light directions. For this reason, we use an evaluation measure, which is similar to the measure used in [20]:

$$E = \frac{1}{\sigma_b}\|\mathbf{b} - \hat{\mathbf{b}}\|^2 + \sum_{k=1}^{K} \frac{u_{\mathbf{t}}^2(k)}{\lambda_{\mathbf{t}}(k)} \tag{9.16}$$

In this case, σ_b controls the relative weight of the distance between the original and reconstructed image, which is the most important factor to minimize, $\lambda_{\mathbf{t}}$ are the eigenvalues of the depth map and albedo model, see Equation 9.5. This evaluation measure allows us to determine when the iterative estimation procedure convergence.

9.2.9 Refinement of the albedo

The albedo $\rho_{noshadow}(\mathbf{p})$, calculated from the MAP estimate and subspace model, misses details (Section 9.2.7). To recover these details, we perform two steps. In the first step, we determine the albedo using the original facial image to recover the details removed by the subspace models. In the second step, we filter the recovered albedo by using the correlation between the different positions in the images to remove spikes.
The first step is to recover the details in albedo based on the image, where we use the following equation:

$$\rho_{details}(\mathbf{p}) = \frac{b(\mathbf{p})}{\mu_r(\mathbf{p})} \tag{9.17}$$

In Equation 9.17, we assume that both the surface normals and illumination conditions are correctly estimated.
In the second step, we remove erroneous effects in the albedo caused by using a user-independent shadow model, which usually results in spike in areas containing the transition between shadow and light. To suppress the spikes, we learn the relationship between albedo at neighboring locations using correlation. Based on the correlation between neighboring locations, we filter the albedo removing spikes if the correlation between locations expects a different albedo value. If the albedo is similar to the expected albedo value it will hardly change. The correlation between locations is learned using a training set. To explain the filter, we use a different notation, where the location $\mathbf{p}$ will be replaced by the subscript x and y.

$$\hat{\rho}_{x,y}^{x,y+1} = \overline{\rho}_{x,y} + r\frac{\sigma_{x,y}}{\sigma_{x,y+1}}(\rho_{x,y+1} - \overline{\rho}_{x,y+1}) \tag{9.18}$$

Using statistics, we can predict the value $\hat{\rho}_{x,y}^{x,y+1}$ at location (x, y) using the value at location $(x, y+1)$ given the correlation r between both positions and the means $\overline{\rho}$ and standard deviations σ of the albedo determined from a training set. We perform a similar prediction from all the locations $((x, y+1)\ (x, y-1)$,

$(x+1, y)$ and $(x-1, y)$), which surround (x, y). In order to compute the final albedo map, we use all surrounding locations in the following equation:

$$\rho_{x,y}^{final} + \lambda[\quad (\rho_{x,y}^{final} - \hat{\rho}_{x,y}^{x,y+1}) + (\rho_{x,y}^{final} - \hat{\rho}_{x,y}^{x,y-1}) \\ (\rho_{x,y}^{final} - \hat{\rho}_{x,y}^{x+1,y}) + (\rho_{x,y}^{final} - \hat{\rho}_{x,y}^{x-1,y}) \quad] = \rho_{x,y}^{details} \tag{9.19}$$

The final albedo $\rho_{x,y}^{final}$ is estimated by taking into account the correlation between surrounding location with a weight factor of $\lambda = 0.2$ and as an initial estimate we use $\rho_{x,y}^{details}$. To solve Equation 9.19 for an entire grid of albedos, we use a multigrid method for boundary value problems (Simultaneous Over-Relaxation) described in [93]. This allows us to determine the final albedo $\rho_{x,y}^{final}$. Using this method, we are able to remove the spikes, but at the same time, we preserve the details in the final albedo $\rho_{x,y}^{final}$.

9.3 Experiments

9.3.1 Training VIG

In order to create the PCA models, it is necessary to obtain a dataset from which we can calculate both the depth maps and the albedo. One of the important limitations of this method can be the 3D database used for training. This database has to be sufficient in order to obtain a model which can be used to reconstruct a probe image. We have experimented with two publicly available databases, namely the 3D FRGC training set (Spring 2003 range images) [90] and the 3D Bosphorus Face Database [99]. Both databases contain facial images together with their range images, which gives us for each pixel a 3D coordinate. We register the facial images to a common coordinate system, using the landmarks provided by the databases. From the range maps provided with the images, we calculate both depth maps and the surface normals, where we use some simple spike removal and hole filling methods to obtain smooth depth maps. Because the illumination in the images is controlled, it is possible to estimate the illumination conditions from both the surface normals and the appearance in the image. This also allows us to compute the albedo. A disadvantage of the 3D FRGC training set is that the images are overexposed, this makes the albedo estimation less accurate.
In our earlier work, we used the 3D FRGC training set to create our face model. The 3D FRGC training set contains individuals that are also present in Experiment 4 of the FRGCv2 database. A disadvantage of this set, however, is that all faces have a neutral expression. The 3D Bosphorus Face Database contains all kinds of expressions and other variations and this database has no overlapping individuals with Experiment 4. Still the results that are achieved by training using 3D Bosphorus Face Database are better, mostly because of the presence of expressions. For this reason, we use the 3D Bosphorus Face Database to obtain the PCA models of the depth map and albedo. We have observed that a better

training set has a large effect on the face recognition results. Important is that this database contains all kinds of variations (expression, race differences), but at the same time is taken under controlled conditions, like illumination, registration and pose.

9.3.2 Experimental Setup

In order to test the performance of illumination correction methods, we use both the CMU-PIE [109] and the FRGCv2 [90] database. Most methods for illumination correction are evaluated using databases recorded in a laboratory, with images illuminated using a single light source on a predetermined grid. The CMU-PIE database is such a database, which allows us to compare VIG with other methods under different predetermined illumination conditions. A problem of these kinds of databases is that bootstrap data from the same predetermined light sources is often used, which positively biases the results. In this research, we chose to also evaluate the methods on images taken under uncontrolled conditions. For this reason, we have performed FRGCv2 experiment 4, where a single gallery image taken under controlled illumination conditions is compared with a probe image taken under uncontrolled illumination conditions. This matches the real-life problem that all illumination correction methods aim to solve. Although one of the biggest challenges is to remove the illumination variations from these images, there are more challenges in face recognition, see [78], that are not addressed in this paper. Examples of problems other than illumination are (small) pose variations, expressions, occlusions due to caps and glasses, which all have negative effects on the final recognition results.
In order to compare our illumination correction method, we have used the PCA-LDA Likelihood ratio classifier described in [126] and the Kernel Correlation Filter (KCF) together with the normalized cosine distance [101] for face recognition. The latter method already achieves good performance on the FRGCv2 database, see [101], together with the illumination correction method of Gross et al [50]. In the literature, several illumination correction methods that use local regions show promising results on the FRGCv2 database. For this reason, we use Gross et al [50] and Tan et al [119] for comparison. One of the best methods that uses the entire image for illumination corrections is the method of Zhang et al in [139], which uses the Spherical Harmonics representation. For comparison, we have developed our own implementation of this method. This method is trained on the same database as is used to train our illumination correction methods.

9.3.3 Face Recognition Results on CMU-PIE database

We performed a simple experiment on the CMU-PIE [109] database in order to compare our illumination correction method with other methods. In this experiment, the difficult images of CMU-PIE illuminated without ambient light are used. In Figure 9.2, we show that our method is able to render the face images with different illumination conditions, although this becomes difficult, if almost

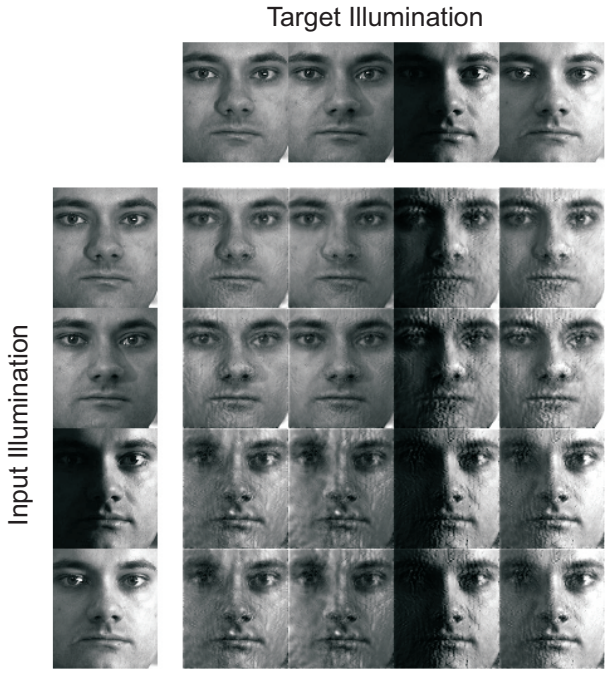


Figure 9.2: Rendering faces under different illumination conditions: the first row contains the recorded images which VIG has to render given the input images on the first columns. The rest of the rows contain the rendered images given the input image

Uncorrected	Zhang et al	Gross et al	Tan et al	VIG
72.2%	74.5%	86.8%	88.1%	90.8%

Table 9.1: The face recognition results of the PCA-LDA likelihood ratio on the CMU-PIE database in recognition rates

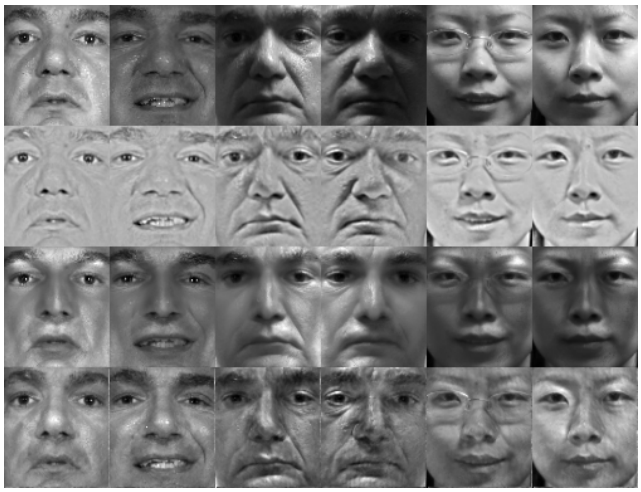


Figure 9.3: The output of the illumination correction methods: The first row contains the original images, the second row shows the output of local preprocessing method of Gross et al, in the third row are the images corrected the global illumination correction method of Zhang et al, the last row shows the images obtained using VIG

half of the face contains shadow (third row of Figure 9.2). For the face recognition experiments, the images are corrected to frontal illumination in order to make them comparable. We used 20% of the 68 subjects for training and used the frontal illuminated images as gallery images. We repeat this experiment 20 times, randomly assigning subjects to the training and test set. In Table 9.1, we show the results in face recognition. The illumination correction methods easily improve the recognition results because it mostly contains illumination variations. We observe that VIG performs better than the other illumination correction methods on this face database.

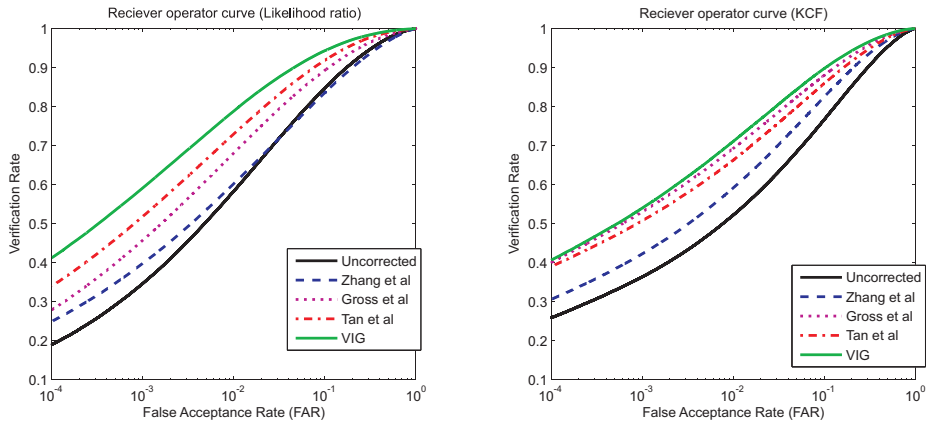
9.3.4 Face Recognition Results on FRGCv2 database

In Experiment 4 of the FRGCv2 database, 3 sets of images are defined: a training set, a target set and a query set. All the images of these sets are corrected using all

illumination correction methods. The output of the different illumination correction methods is shown in Figure 9.3, where the uncorrected images are in the first row and the corrected images with the methods of respectively Gross et al, Zhang et al and VIG are in the second, third and fourth row. Figure 9.3 shows that the illumination correction methods are also used on images taken under controlled conditions (the first two colums), where VIG seems to only change the overall light intensity of the image. The other four images are taken under uncontrolled conditions, which include sometimes difficult illumination conditions as can be seen in the fourth image. We observe that VIG is able to remove most of the cast shadows caused by the nose, especially visible in the fourth image. Furthermore, it is able to correct for the dark areas (right side) in the fourth image. We can, however, not correct for the cast shadows on the cheeks caused by the glasses in the fifth image, because our model does not include glasses nor the reflections they cause. In order to evaluate the effects of the illumination corrections methods on the face recognition, we setup the following experiment. After the illumination correction on all images, the face recognition methods are trained using the corrected images in the training set. We perform a one-to-one comparison described by FRGCv2 database, where we compare one image from the target set (controlled illumination) with one image from the query set (uncontrolled illumination). The face recognition methods used for this comparison are the PCA-LDA Likelihood ratio and the Kernel Correlation Filter (KCF). The Receiver Operating Characteristics (ROC) of these methods are presented in Figures 9.4(a) and 9.4(b). We observe from Figure 9.4(a) that the Virtual Illumination Grid method clearly performs best in combination with the Likelihood ratio, followed by the method of Gross et al. The other global illumination correction method of Zhang et al performs better than uncorrected images at FAR ≤ 1 %. Using the KCF, we observe that the difference between VIG and Gross et al is small. The method of Gross et al reaching similar results as reported in Figure 9.4(b) of the paper on KCF [101], where they performed the same experiment. In the same paper, the KCF achieves even better recognition results, by using the gallery images to train a SVM. We did not perform this experiment, because it deviates for the FRGC protocol. The results of Zhang et al are in this experiment better than uncorrected images. By comparing Figures 9.4(a) and 9.4(b) with each other, the likelihood ratio reaches best results in combination with VIG.

9.3.5 Fusion

In [28], we have fused local and global illumination method in order to improve the performance. In this paper, we fuse VIG with the method of Gross et al in order to investigate if the combination of both methods improves the performance. To fuse both methods, we use simple z-score normalization on the similarity scores. The ROC curves of this experiment are shown in Figure 9.5. The combinations improve the results of the Likelihood ratio slightly and the KCF much more. We also combine all scores achieving a much better performance than each of the separate classifiers. Of course, other face recognition methods can be used to even further improve the face recognition. However, here the goal is to show that the



(a) (b)

Figure 9.4: ROCs for the FRGCv2 database Experiment 4 of the PCA-LDA likelihood ratio (left) and Kernel Correlation Filters (right) for all the illumination correction methods

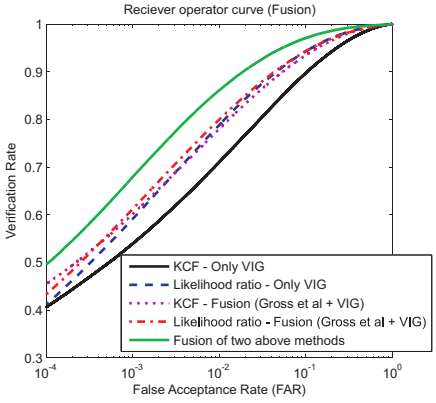


Figure 9.5: ROCs for the FRGCv2 database Experiment 4 obtained by fusion of VIG and the method of Gross et al, for each face recognition method separately and also for the combination of all methods

combination of VIG and Gross et al improves the performance of well-known face recognition methods.

9.4 Discussion

In the previous section, we have shown that VIG improves the face recognition results. In order to provide more inside information about VIG, we discuss the following two issues: The first issue concerns the limitations of our method and we discuss the influence of these limitations on the face recognition results. The second issue is the accuracy of the depth maps in relationship with the reflectance models.

9.4.1 Limitations

We have shown that VIG performs well for the uncontrolled conditions in Experiment 4, for instance in Figure 9.4(a), although fusion with the method of Gross et all still shows room for improvement. We observed that the main reason is the use of reflectance and face models, which are not able to include all exceptions. Baseball caps, for instance, cause a cast shadow in the face, making it impossible to estimate the correct illumination conditions using the entire image, as can be observed in Figure 9.6. Another well-known problem in face recognition are glasses, which reflect the light in other directions than anticipated by the reflection model, making the estimation of the face shape difficult. In order to solve these problems, a detection method for glasses and cap can be developed, allowing us to ignore parts of the image in these cases. Although VIG might not always work in case of the mentioned exceptions, the results show a clear contribution of VIG in most other cases. By comparing VIG with the method of Gross et al, which gives us an illumination invariant representation of the face, we observe from the fusion results that both methods make different errors. The illumination invariant representation of Gross et al is able to deal with for instance caps and glasses, because it uses local assumptions. On the other hand, VIG creates a reconstruction based on more global assumptions which cannot be modelled by Gross et al. The advantage is that the reconstruction will give us more information, like the estimated depth map of the face. This can be used to incorporate 3D and 2D face recognition, which is impossible with illumination invariant representation of Gross et al.

9.4.2 Accuracy of the Depth Maps

We estimate the depth maps based on the Lambertian reflectance model and a PCA model of depth maps of faces. From literature, we know that the Lambertian reflectance model is a reasonably good estimate of the reflectance of the skin, but more accurate reflectance models are known. Bidirectional Reflectance Distribution Functions (BRDF) like the Phong and Torrance-Sparrow model probably

Figure 9.6: Examples of face images containing baseball caps - Because of the cast shadow in the facial images caused by the baseball caps, VIG is unable to determine the correct illumination conditions. The first column shows the entire face, the second column is the region of interest used by VIG, the last column contains the corrected image

provide a better explanation. The disadvantage of these models is that they are non-linear, which has large consequences on the computation time. Another possibility is to measure a BRDF for human skin, using 3D face acquisition equipment, although we are not sure that this BRDF holds for different types of light. Because we use the Lambertian reflectance model throughout this paper, we expect that the depth maps are not always accurate in certain regions. We also observe that PCA models have limitations in modelling details. On the other hand, PCA models are able to correct certain mistakes by enforcing domain specific knowledge. From Figure 9.7, we observe that the depth maps estimated using VIG fail to recover the sharp contours. Instead, we obtain a smooth version of the surface. Similar results can be observed in the research presented in [8], where shape from shading is compared with measured profiles. However, Figure 9.7 also shows that VIG is to a certain extend able to model expressions, like the round cheeks in the second row. Although, we show that the accuracy of the depth maps is limited, this does not have to affect the face recognition results. The reason that VIG might calculate an incorrect depth maps is that our model can be biased to certain explanations. As long as for individuals the same mistakes are made in calculating the depth maps, these mistakes do not have to affect the face recognition results.

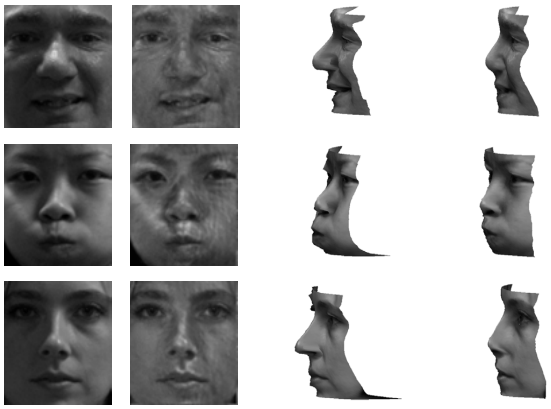


Figure 9.7: Comparison of depth maps using the 3D FRGC database: First column contains the original image, second column shows the correction of VIG, third column is the 3D range map acquired with a laser-scanner and in last column our estimate of the depth map is given.

9.5 Conclusions

We present a new method to correct for illumination effects in facial images. Although multiple methods for illumination correction in facial images are described in the literature, this method has some advantages with respect to most other methods. First of all, we assume and model multiple light sources using a Virtual Illumination Grid (VIG). This allows us to achieve good recognition results, especially for facial images taken under uncontrolled conditions. Secondly, we use two PCA models, both for the depth maps and the albedo and we couple them, taking into account the correlation between depth map and albedo. Thirdly, we are able to estimate a depth map of the surface, which can be useful for improving face recognition under pose variations or for comparison with 3D face recognition. We test multiple methods on the FRGCv2 database using Experiment 4, where faces are recorded under uncontrolled conditions. Our experiment is different from most other illumination correction methods, which are tested on databases recorded in laboratories. In these cases, the illumination directions are predetermined and they include usually only a single light source, while in truly uncontrolled conditions, both the illumination directions and the number of light sources are usually unknown. We show that VIG is able to improve the results of different face recognition methods significantly under these uncontrolled conditions. Furthermore, we fuse VIG, which is a global illumination correction method, with a local illumination correction by Gross et al [50], which further improves the recognition results.

Conclusion Part III

In this part, we have first answered the specific research question: Which measures can be taken to improve the face recognition system for facial images captured under uncontrolled illumination conditions? In camera surveillance, face images are recorded in uncontrolled conditions, which causes a serious decrease in the performance of the face recognition methods in comparison with images recorded under controlled conditions. The final research question is: How much improvement of the face recognition performance is obtained with the measures mentioned before? We have shown in Chapter 6 that model-based illumination correction methods are able to improve the face recognition results. We improved this method in Chapter 7 and Chapter 9, using more light sources to model the uncontrolled conditions. This also resulted in a performance gain in face recognition. We combined our model-based method, which uses a global physical model of the illumination, with a correction method, which performs correction based on the local region. By combining these methods, we achieved significant better results in face recognition.

The illumination correction method determines the illumination conditions using an iterative approach in which illumination, the 3D surface and corresponding albedo are estimated to find the best explanation for the appearance in an image. In the case of video recording, the illumination conditions change only slowly over time. In this situation, the illumination correction can be performed more efficient by using information determined for the previous frames.

In Chapter 7 and 9, we have shown that our illumination correction method is also able to estimate the surface. This surface estimate can be useful for 3D face recognition and pose correction. We know that the accuracy of this estimation is limited, because of the reflectance and surface models that are used in our illumination correction method. Our opinion is that further research on the reflectance properties of the face can improve the estimation of the 3D surface of the face. This can be exploited in problems like pose correction and 3D versus 2D face recognition.

10

SUMMARY & CONCLUSIONS

10.1 Summary

We have presented a detailed investigation of face recognition in camera surveillance. In this investigation, research has been performed on the various components of the face recognition system: face detection, face registration, face intensity normalization and face comparison. We have studied the different characteristics of faces recorded by CCTV systems that might cause problems for face recognition. In this research, the focus has been on the problems that occur due to the low resolution of the faces in recordings and the uncontrolled illumination effects that change the appearance of the faces. The effects of these problems have been investigated and we have developed methods to improve the performance of the face recognition system under these specific problems.
In chapter 2, we reviewed the components of a face recognition system. The first component is face detection, which localizes the face in an image or video recording. Background substraction, skin color detection and appearance based learning methods can be applied for this purpose. The second component is face registration, which aligns the images to a common reference coordinate system. Face registration normally relies on landmarks and can be based on both the appearance of the landmarks and the relations between landmark positions. The third component is the face intensity normalization, which changes the intensity values

in order to make the face better comparable. This methods aims to compensate, for instance, the illumination effects in a face image. We make a distinction between methods that use local regions to normalize the images and methods that estimate a global physical model. The last component is the face comparison (often called face recognition), which compares a probe image with gallery images in order to determine or verify the identity of the user. There are two categories of face comparison methods, the first category is the holistic face comparison and the second category is face comparision based on local features.
In camera surveillance, the resolution of the faces in the recordings is usually low. In part I, we have performed research on the effects of the resolution on the face recognition system. In this research, we have focussed on the effect that the face resolution has on the face registration and comparison component, because these components seems to be most sensitive to resolution changes. We have shown that our face recognition system achieves the best results at a face resolution of 32×32 or higher. At lower resolutions, the results in face recognition decrease rapidly. We have also shown that correct registration using manually labelled landmarks achieve much better results in face recognition than face registration based on automatic landmarking.
In part II, we have improved the face registration, focussing on low resolution facial images. Accurate face registration is important for face recognition. Face registration is usually performed by landmark based face registration methods. These methods perform poorly if the resolution of the faces is low, because the resolution of the individual landmarks becomes even lower. For this reason, we have developed holistic face registration methods. These methods find the optimal registration parameters by maximizing the similarity which is determined by holistic face recognition methods. Firstly, we have shown that these registration methods find the registration parameters as accurately as registration based on manually labelled landmarks. Secondly, these registration methods are able to register face images under low resolutions while maintaining a high accuracy.
In camera surveillance, faces are recorded under uncontrolled illumination conditions. In part III, we have corrected for the illumination in faces by developing face intensity normalization methods that model the reflectance on the face surface. In these methods, we have used a model of the face surface and albedo, which is determined from a grey-value image together with a 3D range image. In these methods, we have focussed on uncontrolled illumination conditions, using models that explain the reflectance even if difficult shadow regions are present. We have shown that our face intensity normalization methods improve the face recognition results. A byproduct of our normalization is an estimate of the face surface. This estimation could be further used for 2D vs 3D face recognition and pose correction. We have combined methods that use the local region in the images to correct for illumination with our face intensity normalization methods. In order to combine these methods, we have used both score-level and decision-level fusion to use the advantages of both categories. By combining two methods from different categories, the final results in face recognition have improved significantly.

10.2 Conclusions

The goal of this research was to improve the face recognition performance for camera surveillance. Face recognition for camera surveillance is difficult, because of the low resolution of the face in the recording, the uncontrolled illumination conditions, pose variations, and occlusions of the face due to, for instance, caps and sun glasses. In this research, our focus has been on solving the problems that occur due to the low resolution and illumination variations. In order to solve these problems, we investigate their effects on the various components of the face recognition system. We have focussed on answering the following questions:

What is the effect of both low resolution and illumination on the different components (Face Detection, Face Registration, Face Intensity Normalization and Face Recognition) of the face recognition system?
Resolution: In part I, we have performed a study on the effect that face resolution has on the performance of the different components in the face recognition system. This study has focussed on the face registration and recognition components, because they seem to be most sensitive to the decrease in facial resolution. The face recognition system with which we have experimented, gives good results for face resolution of 32×32 and higher. Below a face resolution of 32×32 the performance of face recognition decreases rapidly. A face resolution of 32×32 is normal for the camera surveillance scenarios we have focussed on. Although the automatic face registration performs also best at 32×32 or higher, we have discovered a large difference in face recognition results between registration using automatic and manually found landmarks.
Illumination: During this study, we have performed experiments with images taken under controlled and uncontrolled illumination conditions. These experiments show that both the face registration and recognition components suffer from uncontrolled illumination conditions. During these experiments, we have not used an illumination correction method in the face recognition system. This has shown the importance of a proper face intensity normalization and it has given an indication of the improvements that can be achieved by using such methods.

Which measures can be taken to improve the face recognition system for low resolution facial images and images captured under uncontrolled illumination conditions?
Registration: For the low resolution facial images, we have observed that the accuracy of registration could be improved. For this reason, we have developed a holistic based registration method. The advantage of this method is that is uses the entire facial image to find the registration parameters. In case of a face recorded under low resolution, landmark based face registration methods fail to find the landmarks accurately. Our holistic based registration method can find the registration parameters accurately by maximizing the similarity score produced by holistic face recognition methods. In chapter 5, we improved this registration method further by using edge features and an evaluation criterion for registration

instead of maximizing the scores for holistic face recognition methods.
Illumination: In order to correct for the illumination in the face images, we have developed three illumination correction methods. In these illumination correction methods, we model the behavior of the reflectance. We also use models of the face surface and albedo, obtained from facial images together with 3D range images. Using both reflectance and face models, we can reconstruct the illumination conditions in the face image. This also allows us to render a face image under frontal illumination. The face rendered under frontal illumination are easier to compare for face recognition methods. An advantage of this method is that we also obtain an estimate of the facial surface, which can be used in pose correction or 2D vs 3D face recognition. We have also combined our illumination correction methods with methods that use the local region around a pixel to correct for illumination. To combine these two methods, we have used score or decision level fusion. The results in face recognition have become better than the results of the separate methods.

How much improvement of the face recognition performance is obtained with the before mentioned measures?
Registration: For the face registration, manually labelled landmarks are often used as a groundtruth. Especially on the FRGCv2 database, there are no registration methods published in literature that outperform the manually labelled landmarks. The holistic registration methods developed by us have a similar performance as registration using manually labelled landmarks. The best landmark based methods give of 87% verification rate at 0.1% FAR in face recognition, while the holistic registration method achieves a performance in face recognition of around 91% verification rate at 0.1% FAR on FRGCv2 database experiment 1. This registration methods is also able to maintain accurate registration results at lower resolutions down to 32×32. This makes our registration method suitable for camera surveillance applications, as well as other face recognition applications.
Illumination: By using our illumination correction methods, we have achieved a substantial improvement from a 30% verification rate at 0.1% FAR with uncorrected images to a 53% verification rate at 0.1% FAR . Our illumination correction method outperform the other illumination correction methods which we used for comparison. The methods are tested on the FRGCv2 database, where experiment 4 is used to compare facial images recorded in controlled conditions with facial images recorded in uncontrolled conditions. Our illumination correction method is one of the first model-based methods with reported results on facial images taken under uncontrolled conditions. Most other model-based methods only report results on databases recorded in a laboratory, where a predefined light grid is used. We also combined our model-based method with local preprocessing methods, achieving significant improvements in comparison with the best separate illumination correction method.

10.3 Recommendations

One of our recommendations is that a standard experiment needs to be designed for camera surveillance scenarios. This will highlight new problems and make comparison with other methods easier. We have performed some experiments with CCTV recording, but this data could not be made publicly available and it also did not contain cross-session data. Databases with cross-session data contain face images recorded over different time sessions, which is important for realistic results. The FRGC database contains cross-session data making this database realistic for multiple applications. This means that even if some of the methods might not always be suitable for camera surveillance, these methods can also be applied in other application, like access control. Recently, the creators of the FRGC database have developed new challenges in the Multiple Biometric Grand Challenge [84]. These challenges contain video recordings of persons, which might give more realistic camera surveillance conditions. But still, these created scenarios are not similar to real camera surveillance scenarios. In the other recommendations, we will focus more on the technical questions that are left unanswered. These recommendation are categorized for the three major parts of our thesis:

- **Resolution:** In part I, we investigated the effects of the resolution of the faces on the different components of the face recognition system. In this investigation, we focussed especially on the face registration and recognition, because they were the most sensitive. In part III, we experimented with more advantage face intensity normalization methods. One of the questions that remains is: what are the effects of resolution on these face intensity normalization methods? Another issue is that in this research, we have experimented with several well-known holistic methods for face recognition. An interesting question is how state of the art face recognition methods perform at low resolutions? Although, we have limited the research in part I to the face resolution using images, new research can be done on video recordings. In case of video recording, super-resolution methods to enhance the resolution can be further investigated.
- **Registration:** In part II, we developed new holistic registration methods to perform accurate face registration on low resolution images. Although this already seems to work well on face images recorded under uncontrolled conditions, further improvements might be obtained by using the illumination correction methods.. In this case, illumination correction methods based on the local regions have to be used because they do not rely on a face model that requires registration. Another recommendation is to look at video sequences, the face registration can be performed in multiple frames, which improves the computation time of the method if the difference between the registration in the frames changes only slightly. Finally, the face registration can be extended to correct for pose variation. In this case, registration parameters to model the poses and a 3D model of the face to render images under poses are required to maximize a similarity functions.
- **Illumination:** In part III, a model-based illumination correction method

is presented which is able to correct for uncontrolled illumination in facial images. The models that are used in this method to describe the behavior of the illumination are simplifications of reality. More research has to be performed on the errors caused by these models, which can be done using face images together with accurate 3D depth maps. Another important issue in this research is the models of the surface and albedo. Acquisition in more controlled environments may improve the estimation of the surface and albedo models, which can improve the estimate of the surface and albedo. The estimate of the surface can also be used for pose correction or comparison under pose. The last recommendation is that it can be interesting to perform the illumination correction method in video sequences, where the illumination conditions usually change only slowly in time. This probably narrows the search for the illumination conditions, which can make the method faster and more accurate.

REFERENCES

[1] M. Abdel-Mottaleb and A. Elgammal. Face detection in complex environments from color images. In *International Conference on Image Processing, ICIP 99.*, volume 3, pages 622–626 vol.3, 1999.

[2] E. Acosta, L. Torres, and A. Albiol. An automatic face detection and recognition system for video indexing applications. *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP02)*, 4:3644–3647, 2002.

[3] T. Ahonen, A. Hadid, and M. Pietikinen. Face recognition with local binary patterns. In *Computer Vision - ECCV 2004*, pages 469–481, 2004.

[4] A. Aizerman, E. M. Braverman, and L. I. Rozoner. Theoretical foundations of the potential function method in pattern recognition learning. *Automation and Remote Control*, 25:821–837, 1964.

[5] F. R. Al-Osaimi, M. Bennamoun, and A. Mian. Illumination normalization for color face images. In *Advances in Visual Computing*, pages 90 – 101, 2006.

[6] S. Arca, P. Campadelli, and R. Lanzarotti. A face recognition system based on automatically determined facial fiducial points. *Pattern Recognition*, 39(3):432 – 443, 2006.

[7] C. G. Atkeson, A. W. Moore, and S. Schaal. Locally weighted learning. *Artificial Intelligence Review*, 11:11–73, 1997.

[8] G. A. Atkinson, M. L. Smith, L. N. Smith, and A. R. Farooq. Facial geometry estimation using photometric stereo and profile views. In *The 3rd IAPR/IEEE International Conference on Biometrics, Alghero, Italy*, pages 1–11, 2009.

[9] S. Baker and T. Kanade. Hallucinating faces. In *4th International Conference on Automatic Face and Gesture Recognition*, pages 83–89, March 2000.

[10] S. Baker and T. Kanade. Limits on super-resolution and how to break them. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(9):1167 – 1183, September 2002.

[11] S. Baker and I. Matthews. Lucas-kanade 20 years on: A unifying framework. *International Journal of Computer Vision*, 56:221–255, 2004.

[12] M. Balcan, A. Blum, P. P. Choi, J. Lafferty, B. Pantano, M. R. Rwebangira, and X. Zhu. Person identification in webcam images: An application of semi-supervised learning. In *International Conference on Machine Learning Workshop on Learning from Partially Classified Training Data*, pages 1–9, 2005.

[13] M. S. Bartlett, M. H. Lades, and T. J. Sejnowski. Independent component representations for face recognition. volume 3299, pages 528–539. SPIE, 1998.

[14] R. Basri and D. Jacobs. Lambertian reflectance and linear subspaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(2):218–233, Feb 2003.

[15] A. Bazen, R. Veldhuis, and G. Croonen. Likelihood ratio-based detection of facial features. In *14th Annual Workshop on Circuits, Systems and Signal Processing (ProRISC 2003)*, pages 323–329, Veldhoven, The Netherlands, nov 2003.

[16] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):711–720, 1997.

[17] G. Beumer, A. Bazen, and R. Veldhuis. On the accuracy of EERs in face recognition and the importance of reliable registration. In *SPS 2005*. IEEE Benelux/DSP Valley, April 2005.

[18] G. Beumer, Q. Tao, A.M.Bazen, and R. Veldhuis. A landmark paper in face recognition. *7th International Conference on Automatic Face and Gesture Recognition (FGR06)*, pages 73–78, 2006.

[19] V. Blanz. Face recognition based on a 3d morphable model. In *7th International Conference on Automatic Face and Gesture Recognition. FGR '06.*, pages 617–624, Washington, DC, USA, 2006. IEEE Computer Society.

[20] V. Blanz and T. Vetter. A morphable model for the synthesis of 3d faces. *SIGGRAPH '99: Proceedings of the 26th annual conference on Computer graphics and interactive techniques*, pages 187–194, 1999.

[21] B. J. Boom, G. M. Beumer, L. J. Spreeuwers, and R. N. J. Veldhuis. The effect of image resolution on the performance of a face recognition system. In *ICARCV'06*, December 2006.

[22] B. J. Boom, G. M. Beumer, L. J. Spreeuwers, and R. N. J. Veldhuis. Matching score based face registration. In *17th Annual Workshop on Circuits, Systems and Signal Processing (ProRISC 2006)*, Veldhoven, The Netherlands, 2006. STW.

[23] B. J. Boom, L. J. Speeuwers, and R. N. J. Veldhuis. Model-based reconstruction for illumination variation in face images. In *8th International Conference on Automatic Face and Gesture Recognition (FGR08)*, 2008.

[24] B. J. Boom, L. J. Speeuwers, and R. N. J. Veldhuis. Subspace-based holistic registration for low-resolution facial images. *EURASIP Journal of Advance in Signal Processing*, 2010.

[25] B. J. Boom, L. J. Speeuwers, and R. N. J. Veldhuis. Virtual illumination grid for correction of uncontrolled illumination in facial images. *Accepted in: Pattern Recognition*, 2010.

[26] B. J. Boom, L. J. Spreeuwers, and R. N. J. Veldhuis. Automatic face alignment by maximizing similarity score. *7th International Workshop on Pattern Recognition in Information Systems (PRIS 2007)*, pages 221–231, 2007.

[27] B. J. Boom, L. J. Spreeuwers, and R. N. J. Veldhuis. Model-based illumination correction for face images in uncontrolled scenarios. In *Computer Analysis of Images and Patterns 2009, Munster*, 2009.

[28] B. J. Boom, Q. Tao, L. J. Spreeuwers, and R. N. J. Veldhuis. Combining illumination normalization methods for better face recognition. In *The 3rd IAPR/IEEE International Conference on Biometrics, Alghero, Italy*, 2009.

[29] B. J. Boom, R. T. A. van Rootseler, and R. N. J. Veldhuis. Investigating the boosting framework for face recognition. In *Proceedings of the 28th Symposium on Information Theory in the Benelux, Enschede, The Netherlands*, pages 189–196, Eindhoven, May 2007.

[30] R. Brent. *Algorithms for Minimization without Derivatives*. Prentice-Hall, Englewood Cliffs, N.J., 1973.

[31] J. Cai, A. Goshtasby, and C. Yu. Detecting human faces in color images. In *International Workshop on Multi-Media Database Management Systems, 1998.*, pages 124–131, Aug 1998.

[32] M. Castrilln-Santana, O. Dniz-Surez, L. Antn-Canals, and J. Lorenzo-Navarro. Face and facial feature detection evaluation. In *3rd International Conference on Computer Vision Theory and Applications (VISAPP)*, 2008.

[33] C.-P. Chen and C.-S. Chen. Lighting normalization with generic intrinsic illumination subspace for face recognition. volume 2, pages 1089–1096 Vol. 2, Oct. 2005.

[34] L. Chen, L. Zhang, L. Zhu, M. Li, and H. Zhang. A novel facial feature localization method using probabilistic-like output. In *Asian Conference on Computer Vision*, pages 1–10, 2004.

[35] U. K. D. P. Commission. *CCTV Code of Practice*. July 2000.

[36] P. Comon. Independent component analysis, a new concept? *Signal Process.*, 36(3):287–314, 1994.

[37] T. Cootes, D. Cooper, C. Taylor, and J. Graham. Active shape models - their training and application. In *Computer Vision and Image Understanding*, volume 61, page 3859, 1995.

[38] T. Cootes, G. Edwards, and C. Taylor. Active appearance models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(6):681–685, jun 2001.

[39] T. Cootes and C. Taylor. On representing edge structure for model matching. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2001*, volume 1, pages I–1114–I–1119 vol.1, 2001.

[40] T. F. Cootes, A. Hill, C. J. Taylor, and J. Haslam. The use of active shape models for locating structures in medical images. In *13th International Conference on Information Processing in Medical Imaging (IPMI '93)*, pages 33–47, London, UK, 1993. Springer-Verlag.

[41] D. Cristinacce, T. Cootes, and I. Scott. A multi-stage approach to facial feature detection. In *15th British Machine Vision Conference, London, England*, pages 277–286, 2004.

[42] J. Czyz and L. Vandendorpe. Evaluation of lda-based face verification with respect to available computational resources. *PRIS 2002*, pages 59–66, 2002.

[43] Y. Dai and Y. Nakano. Face-texture model based on sgld and its application in face detection in a color scene. *Pattern Recognition*, 29(6):1007 – 1017, 1996.

[44] S. C. Dass, K. Nandakumar, and A. K. Jain. A principled approach to score level fusion in multimodal. *Audio- and Video-Based Biometric Person*

Authentication, pages 1049–1058, 2005.

[45] H. K. Ekenel and B. Sankur. Multiresolution face recognition. *Image and Vision Computing*, 23:469–477, 2005.

[46] M. Everingham and A. Zisserman. Regression and classification approaches to eye localization in face images. In *7th International Conference on Automatic Face and Gesture Recognition(FGR06)*, pages 441–446, 2006.

[47] S. Flight, Y. van Heerwaarden, and M. van Aalst. Evaluatie cameratoezicht amsterdam centrum. Technical report, O+S, het Amsterdamse bureau voor Onderzoek en Statistiek, October 2004.

[48] A. Georghiades, P. Belhumeur, and D. Kriegman. From few to many: Illumination cone models for face recognition under variable lighting and pose. *IPMI '93*, 23(6):643–660, 2001.

[49] H. Graf, E. Cosatto, D. Gibbon, M. Kocheisen, and E. Petajan. Multimodal system for locating heads and faces. In *International Conference on Automatic Face and Gesture Recognition, 1996*, pages 88–93, Oct 1996.

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

[51] G.-D. Guo, H.-J. Zhang, and S. Z. Li. Pairwise face recognition. *IEEE International Conference on Computer Vision (ICCV 2001)*, 02:282, 2001.

[52] G. Hager and P. Belhumeur. Efficient region tracking with parametric models of geometry and illumination. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(10):1025–1039, Oct 1998.

[53] L. Hempel and E. Topfer. Urbaneye: Cctv in europe. final report. Technical report, Centre for Technology and Society, Technical University Berlin, August 2004.

[54] Z. Henderson, V. Bruce, and A. M. Burton. Matching the faces of robbers captured on video. *Applied Cognitive Psychology*, 15:445–464, 2001.

[55] G. Heusch, Y. Rodriguez, and S. Marcel. Local binary patterns as an image preprocessing for face authentication. *7th International Conference on Automatic Face and Gesture Recognition (FGR06)*, pages 6 pp.–, 10-12 April 2006.

[56] S. Hissel and S. Dekkers. Evaluatie cameratoezicht op openbare plaatsen (tweemeting). Technical report, Regio Beleidonderzoek, May 2008.

[57] X. Hou, S. Li, H. Zhang, and Q. Cheng. Direct appearance models. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, volume 1, pages I–828–I–833 vol.1, 2001.

[58] R.-L. Hsu, M. Abdel-Mottaleb, and A. Jain. Face detection in color images. *IEEE Transactions On Pattern Analysis and Machine intelligence*, 24(5):696–706, May 2002.

[59] M. hsuan Yang and N. Ahuja. Gaussian mixture model for human skin color and its applications in image and video databases. In *Its Application in Image and Video Databases*, pages 458–466, 1999.

[60] HumanScan. Bioid face db. `http://www.humanscan.de/`.

[61] Intel. Open computer vision library. `http://sourceforge.net/projects/opencvlibrary/`.

[62] L. D. Introna and D. Wood. Picturing algorithmic surveillance: The politics of facial recognition systems. *Surveillance & Society*, 2:177–198, 2004.
[63] A. K. Jain and S. Z. Li. *Handbook of Face Recognition*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2005.
[64] K. Jia and S. Gong. Generalized face super-resolution. *IEEE Transactions on Image Processing*, 17(6):873–886, June 2008.
[65] K. Jia, S. Gong, and A. Leung. Coupling face registration and super-resolution. In *British Machine Vision Conference*, volume 2, pages 449–458, September 2006.
[66] M. J. Jones and J. M. Rehg. Statistical color models with application to skin detection. *Int. J. Comput. Vision*, 46(1):81–96, 2002.
[67] M. J. Jones and P. Viola. Face recognition using boosted local features. In *International Conference on Computer Vision*, 2003.
[68] K. Jonsson, J. Matas, J. Kittler, and S. Haberl. Saliency-based robust correlation for real-time face registration and verification. In *Proc British Machine Vision Conference BMVC98*, pages 44–53, 1998.
[69] M. Kirby and L. Sirovich. Application of the karhunen-loeve procedure for the characterization of human faces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(1):103–108, Jan 1990.
[70] G. Kukharev and P. Forczmanski. Data dimensionality reduction for face recognition. *Machine Graphics & Vision*, 13:99–121, 2004.
[71] J. Lee, B. Moghaddam, H. Pfister, and R. Machiraju. A bilinear illumination model for robust face recognition. pages 1177–1184, 2005.
[72] K. Lee, J. Ho, and D. Kriegman. Acquiring linear subspaces for face recognition under variable lighting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(5):684–698, 2005.
[73] Y. Li, S. Gong, and H. Liddell. Support vector regression and classification based multi-view face detection and recognition. In *4th IEEE International Conference on Automatic Face and Gesture Recognition*, pages 300–305, 2000.
[74] R. Lienhart, A. Kuranov, and V. Pisarevsky. Empirical analysis of detection cascades of boosted classifiers for rapid object detection. *Pattern Recognition*, pages 297–304, 2003.
[75] S.-H. Lin, S.-Y. Kung, and L.-J. Lin. Face recognition/detection by probabilistic decision-based neural network. *IEEE Transactions on Neural Networks*, 8(1):114–132, Jan 1997.
[76] C. Liu, H.-Y. Shum, and W. T. Freeman. Face hallucination: Theory and practice. *International Journal of Computer Vision*, 75:115–134, 2007.
[77] A. Mahalanobis, B. V. K. V. Kumar, and D. Casasent. Minimum average correlation energy filters. *Applied Optics*, 26:3633–3640, 1987.
[78] A. M. Martínez. Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(6):748–763, 2002.
[79] J. Matas, K. Jonsson, and J. Kittler. Fast face localisation and verification. In *Image and Vision Computing*, volume 17, pages 575–581, 1999.
[80] B. Moghaddam and A. Pentland. Probabilistic visual learning for object representation. *IEEE Transactions on Pattern Analysis and Machine Intel-*

ligence, 19(7):696–710, 1997.

[81] J. Nelder and R. Mead. A simplex method for function minimization. *The Computer Journal*, 7:308–313, 1965.

[82] J. Neyman and E. S. Pearson. On the problem of the most efficient tests of statistical hypotheses. *Royal Society of London Philosophical Transactions Series A*, 231:289–337, 1933.

[83] NIST. Frgc face database. `http://www.frvt.org/FRGC/`.

[84] NIST. Mbgc face database. `http://www.frvt.org/FRGC/`.

[85] C. Norris and G. Armstrong. Cctv and the social structuring of surveillance. *Crime Prevention Studies*, 10:157–178, 1999.

[86] T. Ojala, M. Pietikainen, and D. Harwood. Comparative study of texture measures with classification based on feature distributions. *Pattern Recognition*, 29:51–59, 1996.

[87] E. Osuna, R. Freund, and F. Girosit. Training support vector machines: an application to face detection. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1997*, pages 130–136, Jun 1997.

[88] V. Perlibakas. Distance measures for pca-based face recognition. *Pattern Recognition Letters*, 25(6):711 – 724, 2004.

[89] J. P. Phillips, T. W. Scruggs, A. J. Otoole, P. J. Flynn, K. W. Bowyer, C. L. Schott, and M. Sharpe. Frvt 2006 and ice 2006 large-scale results. Technical report, National Institute of Standards and Technology, March 2007.

[90] P. Phillips, P. Flynn, T. Scruggs, K. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek. Overview of the face recognition grand challenge. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '05)*, 1:947–954 vol. 1, June 2005.

[91] P. Phillips, P. Flynn, T. Scruggs, K. Bowyer, and W. Worek. Preliminary face recognition grand challenge results. In *7th International Conference on Automatic Face and Gesture Recognition (FGR06)*, pages 15–24, April 2006.

[92] M. Powell. An efficient method for finding the minimum of a function of several variables without calculating derivatives. *Computer Journal*, 7:155–162, 1964.

[93] W. Press, S. Teukolsky, W. Vetterling, and B. Flannery. *Numerical Recipes in C*. Cambridge University Press, Cambridge, UK, 2nd edition, 1992.

[94] R. Ramamoorthi and P. Hanrahan. On the relationship between radiance and irradiance: determining the illumination from images of a convex lambertian object. *J. Opt. Soc. Am. A*, 18(10):2448–2459, 2001.

[95] T. Riopka and T. Boult. The eyes have it. In *Proceedings of ACM SIGMM Multimedia Biometrics Methods and Applications Workshop*, pages 9–16, Berkeley, CA, 2003.

[96] D. Roth, M. hsuan Yang, and N. Ahuja. A snow-based face detector. In *Advances in Neural Information Processing Systems 12*, pages 855–861. MIT Press, 2000.

[97] H. A. Rowley, S. Baluja, and T. Kanade. Neural network-based face detection. *IEEE Transactions On Pattern Analysis and Machine intelligence*, 20:23–38, 1996.

[98] A. A. Salah, H. Cinar, L. Akarun, and B. Sankur. Robust Facial Landmarking For Registration. *Annals of Telecommunications*, 62(1-2):1608 – 1633,

2007.

[99] A. Savran, N. Alyüz, H. Dibeklioğlu, O. Çeliktutan, B. Gökberk, B. Sankur, and L. Akarun. Bosphorus database for 3d face analysis. In *First European Workshop on Biometrics and Identity Management: BIOID 2008*, pages 47–56, 2008.

[100] M. Savvides, R. Abiantun, J. Heo, S. Park, C. Xie, and B. Vijayakumar. Partial & holistic face recognition on frgc-ii data using support vector machine. *Conference on Computer Vision and Pattern Recognition Workshop, 2006. CVPRW '06.*, pages 48–48, June 2006.

[101] M. Savvides, R. Abiantun, J. Heo, S. Park, C. Xie, and B. Vijayakumar. Partial & holistic face recognition on FRGC-II data using support vector machine. In *Conference on Computer Vision and Pattern Recognition Workshop, 2006. CVPRW '06.*, pages 48–48, June 2006.

[102] M. Savvides and B. Vijaya Kumar. Efficient design of advanced correlation filters for robust distortion-tolerant face recognition. *IEEE Conference on Advanced Video and Signal Based Surveillance, 2003.*, pages 45–52, July 2003.

[103] B. Schlkopf, A. Smola, and K.-R. Mller. Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation*, 10(5):1299–1319, 1998.

[104] H. Schneiderman and T. Kanade. A statistical method for 3d object detection applied to faces and cars. In *Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on*, volume 1, pages 746–751 vol.1, 2000.

[105] S. Shan, W. Gao, B. Cao, and D. Zhao. Illumination normalization for robust face recognition against varying lighting conditions. *IEEE International Workshop on Analysis and Modeling of Faces and Gestures (AMFG 2003)*, pages 157–164, 17 Oct. 2003.

[106] X. Shang. *Grip-pattern recognition: Applied to a smart gun.* PhD thesis, University of Twente, December 2008.

[107] A. Shashua and T. Riklin-Raviv. The quotient image: class-based rerendering and recognition with varying illuminations. *IEEE Transactions On Pattern Analysis and Machine intelligence*, 23(2):129–139, Feb 2001.

[108] J. Shi, A. Samal, and D. Marx. How effective are landmarks and their geometry for face recognition? *Computer Vision and Image Understanding*, 102(2):117 – 133, 2006.

[109] T. Sim, S. Baker, and M. Bsat. The cmu pose, illumination, and expression (pie) database of human faces. Technical Report CMU-RI-TR-01-02, Robotics Institute, Pittsburgh, PA, January 2001.

[110] T. Sim and T. Kanade. Combining models and exemplars for face recognition: An illuminating example. In *CVPR Workshop on Models versus Exemplars in Computer Vision*, december 2001.

[111] W. Smith and E. Hancock. Recovering facial shape using a statistical model of surface normal direction. *IEEE Transactions On Pattern Analysis and Machine intelligence*, 28(12):1914–1930, Dec. 2006.

[112] W. Smith and E. Hancock. Recovering face shape and reflectance properties from single images. In *8th IEEE International Conference on Automatic Face & Gesture Recognition, 2008. FG '08.*, pages 1–8, Sept. 2008.

[113] K. Sobottka and I. Pitas. Extraction of facial regions and features using color and shape information. In *Proceedings of the 13th International Conference on Pattern Recognition, 1996*, volume 3, pages 421–425 vol.3, Aug 1996.

[114] K. Sobottka and I. Pitas. Face localization and facial feature extraction based on shape and color information. In *International Conference on Image Processing, 1996*, volume 3, pages 483–486 vol.3, Sep 1996.

[115] L. Spreeuwers, B. Boom, and R. Veldhuis. Better than best: matching score based face registration. *Proceedings of the 28th Symposium on Information Theory in the Benelux, Enschede, The Netherlands*, pages 125–132, 2007.

[116] S. Srisuk and W. Kurutach. A new robust face detection in color images. volume 0, page 0306, Los Alamitos, CA, USA, 2002. IEEE Computer Society.

[117] C. Stauffer and W. Grimson. Adaptive background mixture models for real-time tracking. In *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on.*, volume 2, pages 246–252 Vol. 2, 1999.

[118] K.-K. Sung and T. Poggio. Example-based learning for view-based human face detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(1):39–51, Jan 1998.

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

[120] Q. Tao and R. Veldhuis. Hybrid fusion for biometrics: Combining score-level and decision-level fusion. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2008.*, pages 1–6, June 2008.

[121] Q. Tao and R. N. J. Veldhuis. Illumination normalization based on simplified local binary patterns for a face verification system. In *Biometrics Symposium 2007 at The Biometrics Consortium Conference, Baltimore, Maryland*, pages 1–7, USA, September 2007. IEEE Computational Intelligence Society.

[122] Q. Tao and R. N. J. Veldhuis. Optimal decision fusion for a face verification system. In *2nd International Conference on Biometrics, Seoul, Korea*, Image Processing, Computer Vision, Pattern Recognition, pages 958–967, Berlin Heidelberg, August 2007. Springer Verlag.

[123] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of cognative neuroscience*, pages 71–86, 1991.

[124] V. N. Vapnik. *The nature of statistical learning theory.* Springer-Verlag New York, Inc., New York, NY, USA, 1995.

[125] M. A. O. Vasilescu and D. Terzopoulos. Multilinear analysis of image ensembles: Tensorfaces. In *Computer Vision ECCV 2002*, pages 447–460. Springer-Verlag, 2002.

[126] R. Veldhuis, A. Bazen, W. Booij, and A. Hendrikse. Hand-geometry recognition based on contour parameters. In *Proceedings of SPIE Biometric Technology for Human Identification II*, pages 344–353, Orlando, FL, USA, March 2005.

[127] P. A. Viola and M. J. Jones. Rapid object detection using a boosted cascade of simple features. In *CVPR (1)*, pages 511–518, 2001.

[128] H. Wang, S. Li, Y. Wang, and J. Zhang. Self quotient image for face recognition. In *International Conference on Image Processing*, volume 2, pages

1397–1400, 2004.

[129] J. Wang, C. Zhang, and H. Shum. Face image resolution versus face recognition preformance based on two global methods. *Proceedings of Asia Conference on Computer Vision*, 2004.

[130] P. Wang, M. Green, Q. Ji, and J. Wayman. Automatic eye detection and its validation. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops, 2005*, pages 164–164, June 2005.

[131] P. Wang, L. C. Tran, and Q. Ji. Improving face recognition by online image alignment. In *18th International Conference on Pattern Recognition (ICPR 2006)*, volume 1, pages 311–314, 2006.

[132] R. Willing. Airport anti-terror systems flub tests. *USA Today*, September 2, 2003.

[133] L. Wiskott, J.-M. Fellous, N. Krüger, and C. von der Malsburg. Face recognition by elastic bunch graph matching. In L. C. Jain, U. Halici, I. Hayashi, and S. B. Lee, editors, *Intelligent Biometric Techniques in Fingerprint and Face Recognition*, chapter 11, pages 355–396. CRC Press, 1999.

[134] C. R. Wren, A. Azarbayejani, T. Darrell, and A. P. Pentland. Pfinder: Real-time tracking of the human body. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):780–785, 1997.

[135] M.-H. Yang and N. Ahuja. Detecting human faces in color images. In *International Conference on Image Processing (ICIP '98)*, volume 1, pages 127–130 vol.1, Oct 1998.

[136] P. Yang, S. Shan, W. Gao, S. Z. Li, and D. Zhang. Face recognition using ada-boosted gabor features. *International Conference on Automatic Face and Gesture Recognition*, page 356, 2004.

[137] G. Zhang, X. Huang, S. Z. Li, Y. Wang, and X. Wu. Boosting local binary pattern (lbp)-based face recognition. In *Chinese Conference on Biometric Recognition*, volume SINOBIOMETRICS 2004, pages 179–186, 2004.

[138] L. Zhang, S. Z. Li, Z. Y. Qu, and X. Huang. Boosting local feature based classifiers for face recognition. In *IEEE International Conference Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'04)*, page 87, Washington, DC, USA, 2004. IEEE Computer Society.

[139] L. Zhang and D. Samaras. Face recognition under variable lighting using harmonic image exemplars. In *IEEE International Conference on Computer Vision and Pattern Recognition*, pages 19–25, 2003.

[140] L. Zhang and D. Samaras. Face recognition from a single training image under arbitrary unknown lighting using spherical harmonics. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(3):351–363, 2006. Student Member-Zhang,, Lei and Member-Samaras,, Dimitris.

[141] W. Zhao, R. Chellappa, and P. Phillips. Subspace linear discriminant analysis for face recognition. Technical Report CAR-TR-914, Center for Automation Research, University of Maryland, College Park, 1999.

[142] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *ACM Comput. Surv.*, 35(4):399–458, 2003.

[143] S. Zhou, G. Aggarwal, R. Chellappa, and D. Jacobs. Appearance Characterization of Linear Lambertian Objects, Generalized Photometric Stereo, and Illumination-Invariant Face Recognition. *IEEE Transactions on Pattern*

Analysis and Machine Intelligence, 29(2):230–245, 2007.

[144] S. K. Zhou, R. Chellappa, and D. W. Jacobs. Characterization of human faces under illumination variations using rank, integrability, and symmetry constraints. *Computer Vision - ECCV 2004*, pages 588–601, 2004.

[145] Z. Zivkovic and F. van der Heijden. Efficient adaptive density estimation per image pixel for the task of background subtraction. *Pattern recognition letters*, 27:773–780, January 2006.

Samenvatting

Dit proefschrift beschrijft een gedetailleerd onderzoek naar gezichtsherkenningssystemen voor cameratoezicht (camera surveillance). In dit onderzoek wordt er gekeken naar de technische kant van de gezichtsherkenningssystemen. Een normaal gezichtsherkenningssysteem is op te delen in verschillende componenten: gezichtslokalisatie, gezichtsregistratie, gezichtsnormalisatie en gezichtsvergelijking. Er is onderzocht welke problemen er zijn met gezichten die zijn opgenomen door camera surveillance systemen. In dit onderzoek hebben we ons beperkt tot twee problemen, namelijk de lage resoluties van de gezichten in afbeeldingen en de veranderingen die ongecontroleerde belichting op het gezicht veroorzaakt in een afbeelding. We hebben eerst onderzocht welke effecten deze twee problemen hadden op een gezichtsherkenningssysteem, waarna we nieuwe methodes hebben ontwikkeld om de prestaties van gezichtsherkenningssystemen in cameratoezicht te verbeteren.
Om gezichtsherkenningssystemen te verbeteren, hebben we eerst gekeken uit welke componenten een gezichtsherkenningssysteem bestaat. Er wordt eerst een overzicht gegeven worden van deze componenten: De eerste component is gezichtslokalisatie, dat de positie van een gezicht in een afbeelding vindt. Monitoren van de achtergrond in video, onderscheiden van huidskleur en de vorm van het gezicht kunnen gebruikt worden voor de lokalisatie. De tweede component is de gezichtsregistratie, waarbij het gezicht in een standaard positie wordt gebracht met een standaard grotte, zodat afbeeldingen kunnen worden vergeleken. Gezichtsregistratie gebeurt normaal op basis van "landmarks" (orintatiepunten in het gezicht), waarbij zowel de intensiteitswaarden in de afbeelding als the relatie tussen de "landmarks" worden gebruikt. De derde component is de intensiteitnormalisatie, dat de intensiteitswaarden in een afbeelding kan standaardiseren om gezichten beter te vergelijken. Dit heeft voornamelijk het doel om de belichting in afbeeldingen te compenseren. Hierbij hebben wij onderscheid tussen lokale en globale methodes. De laatste component is de gezichtsvergelijking (ook vaak gezichtsherkenning genoemd), die een inkomende ("probe") afbeelding van een gezicht vergelijkt met gezichten die al in het systeem zitten, de "gallery" wordt gebruikt om de identiteit van de gebruiker te achterhalen. Er zijn twee categorien van gezichtsvergelijking methodes, de eerst categorie gebruik het gehele plaatje, terwijl in de tweede categorie gebruik wordt gemaakt van lokale kenmerken.
In camera surveillance is de resolutie van de gezichten in de opnames meestal laag. In Deel I van dit proefschrift is onderzoek gedaan naar het effect dat deze lage resolutie heeft op het hele gezichtsherkenningssysteem. Er was al wel onderzoek gedaan naar het effect van lage resolutie op de gezichtsvergelijking component, maar er was nog niet gekeken naar het complete systeem. In ons onderzoek hebben we ons vooral gericht op de effecten van de resolutie op de gezichtsregistratie en de gezichtsvergelijking, omdat deze componenten het meest gevoelig waren voor de lagere resoluties. We hebben laten zien dat onze gezichtsherkenningssysteem bij een resolutie van 32×32 of hoger het beste werkt. Op lagere resoluties nemen de resultaten in gezichtherkenning snel af. We hebben ook laten zien dat correcte gezichtregistratie met handmatig verkregen "landmarks" veel betere resultaten

oplevert dan de gezichtregistratie met automatisch gevonden "landmarks".
Door het grote verschil tussen handmatig en automatische gezicht registratie, hebben we in Deel II methodes ontwikkeld die zelfs op lage resolutie een goede registratie kunnen uitvoeren. Nauwkeurig gezichtsregistratie is belangrijk voor de opvolgende gezichtsvergelijking component. Zoals al opgemerkt, gezichtsregistratie wordt normaal gesproken gedaan door methodes die gebruik maken van "landmarks". Deze methodes werken slecht als de resolutie van de gezichten erg laag is, omdat dan ook de resolutie van de regio waar de individuele "landmarks" zich bevinden nog lager wordt. Wij hebben een holistische gezichtsregistratie methode ontwikkeld. Deze methode vindt de optimale registratie parameters door de overeenkomst tussen gezichten bepaald door een holistische gezichtsherkenning methodes te maximaliseren. We hebben hierbij laten zien dat deze methodes in staat zijn de registratieparameters net zo nauwkeurig te vinden als registratie gebaseerd op handmatig geselecteerde "landmarks". Daarnaast is deze registratiemethode in staat afbeelding van gezichten op lage resolutie met hoge nauwkeurigheid te registreren.
In camera surveillance worden gezichten vaak opgenomen onder ongecontroleerde belichtingscondities. In Deel III, hebben we gezichtsnormalisatie methodes ontwikkeld voor het corrigeren van belichting in gezichten. Deze methodes maken gebruik van de reflectie van het licht op de 3D vorm van het gezicht. Er zijn modellen gebruikt van de 3D vorm van het gezicht en de "albedo" (reflectie factor) op het gezicht. Onze intensiteitnormalisatie richt zich vooral op het corrigeren van ongecontroleerde belichtingsomstandigheden, waarbij er modellen zijn gebruikt de reflectie zelf met ingewikkelde schaduw regio's kunnen verklaren. Door gebruik te maken van onze intensiteitnormalisatie verbeteren de resultaten van de gezichtsherkenningssysteem. Onze intensiteitnormalisatie geeft ons ook nog een schatting van de 3D vorm van het gezicht. Dit kan worden gebruikt voor 2D/3D gezichtsherkenning en het corrigeren van gezichten waarbij mensen niet in de camera kijken. We hebben ook methodes onderzocht die locale regio's gebruiken voor de belichtingcorrectie en deze methodes hebben we gecombineerd met onze gezichtsnormalisatie methode. Om deze methode te combineren is er gebruik gemaakt van "decision-level" en "score-level" fusion, zodat de voordelen van beide methodes worden versterkt en significant betere resultaten worden gehaald in gezichtsherkenning.

Dankwoord

In September 2005 verhuisde ik van het westen van Nederland naar Enschede om te promoveren. Dit was het begin van een nieuwe periode in mijn leven. Als een informaticus, kwam ik terecht bij de vakgroep Signalen en Systemen van elektrotechniek. Vooral in het begin realiseerde ik mij dat ze daar wel een beetje een ander dialect hebben, niet dat Twents zo moeilijk is, maar vooral mijn achtergrond in informatica was anders. Door de jaren heen ben ik dit gelukkig meer en meer gaan begrijpen, dit heb ik hoofdzakelijk te danken aan al mijn collega's bij Signalen en Systemen.

Ik wil vooral mijn twee begeleiders Raymond Veldhuis en Luuk Spreeuwers bedanken. Ik heb het heel erg gewaardeerd dat ze mij de vrijheid en het vertrouwen hebben gegeven om mijn eigen ideen te onderzoeken, waardoor ze een heleboel ideen hebben moeten aan horen. Door hun begeleiding zijn sommige ideen verbeterd en verwezenlijkt, veelal omdat ze vaak ook met hun achtergrond er op een andere manier naar konden kijken. Hun advies op het gebied van mijn werk stelde mij instaat dit proefschrift afronden, maar hun advies en hulp buiten dit werk stelde mij in staat ook veel andere dingen te kunnen bereiken. Daarnaast heb ik de samenwerking met mijn promotor Kees als zeer prettig ervaren.

Mijn tijd bij zijn vakgroep Signalen en Systemen heb ik als heel aangenaam ervaren. Naast gezellig koffiedrinken, lunchen en borrelen met mijn collega, heb ik ook genoten van andere groepsactiviteiten. Vooral de wintersport vakanties waar ik altijd samen met Anne, Dirk-Jan, Gerbert en Rene heen ben geweest waren erg leuk. Ondanks dat ik heb laten zien dat zulke extreme sporten misschien toch niet helemaal geschikt zijn voor mij.

Tijdens mijn periode bij Signalen en Systemen was ik niet de enige die op gezichtsherkenning heeft gewerkt. In mijn eerste jaren, heb ik veel geleerd over gezichtsherkenning methodes van Gert, met wie ik ook samen op conferentie en vakantie in Singapore ben geweest. Een andere collega en kamergenoot was Qian, die haar kennis en ideen op het gebied van gezichtsherkenning heeft gedeeld. Deze samenwerking heeft geleid tot een mooie publicatie voor ons beide op een conferentie in Sardini. Robin was de eerste "master" student die ik heb begeleid en de samenwerking met hem tijdens zijn tijd als student en later ook collega heb ik als heel fijn ervaren. De meest wiskundige persoon in de biometrie groep was Anne. Hij heeft mij met wiskundige problemen meer dan eens geholpen. Naast gezichtsherkenning, delen we ook dezelfde passie voor volleybal. We zijn samen op vakantie naar Sardinie geweest, wat ontzettend leuk was. Naast de personen in de biometrie groep heb ik ook veel geleerd van mijn collega's in de biomedical groep, de problemen en technieken die we gebruiken lijken vaak op elkaar. Dit leidde tot veel interestante discussies met Dirk-Jan, Gerbert, Rene en vooral niet te vergeten mijn kamergenoot Almar die mijn kennis en interesses op het gebied van beeldverwerking hebben verbreed.

Ik ben dankbaar voor de hulp die Geert-Jan heeft geboden op het gebied van het regelen van computers die al mijn berekeningen konden uitvoeren, meer dan tienduizend afbeeldingen werd toch wat veel voor mijn eigen computer. Daarnaast wil

ik Anneke en Sandra bedanken voor het helpen bij mijn administratie.
Tijdens de jaren dat ik bij de Universiteit Twente werkte, heb ik altijd volleybal gespeeld bij Harambee. Dit stelde mij in de gelegenheid even mijn werk te vergeten, zodat ik mij na die tijd weer met een frisse blik op problemen kon storten. Ik heb in mix 4 (2005-2006), heren 10 (2006-2007), heren 10 (2007-2008), heren 9 (2008-2009), heren 9 (2009-2010) en heren 9 (2010-2011) gespeeld en over de jaren veel andere leden van Harambee leren kennen. Veel van mijn teamgenoten beschouw ik nu als goede vrienden en ik wil hun bedanken voor de heel leuke tijd in Enschede.
Als laatste wil ik mijn vrienden en familie uit Nieuw-Vennep bedanken voor hun steun. Ik heb af en toe proberen uit te leggen waar ik allemaal mee bezig was op het gebied van gezichtsherkenning met computers en ik hoop dat dit proefschrift dit misschien nog wat duidelijker maakt. Ik wil vooral mijn broer Pieter bedanken voor het controleren van de engelse spelling en grammatica in dit proefschrift. Verder wil ik mijn moeder bedanken die het mogelijk heeft gemaakt dat ik dit niveau kan halen, na alle moeite die je heb gestopt in mijn opleiding. De laatste persoon die ik wil bedanken is mijn vriendin Jolanda, voor haar liefde en zorg. Zij heeft mijn gestimuleerd mijn promotie af te ronden en ze inspireert mij om steeds weer nieuwe doelen te behalen.