Face Recognition in a border control environment:

Non-zero Effort Attacks' Effect on False Acceptance Rate

Tom Fladsrud

Master's Thesis
Master of Science in Information Security
30 ECTS
Department of Computer Science and Media Technology
Gjøvik University College, 2005

The MSc programme in Information Security
is run in cooperation with the Royal Institute
of Technology (KTH) in Stockholm.

Institutt for
informatikk og medieteknikk
Høgskolen i Gjøvik
Postboks 191
2802 Gjøvik

Department of Computer Science
and Media Technology
Gjøvik University College
Box 191
N-2802 Gjøvik
Norway

Abstract

Face recognition is a biometric authentication method that has become more and more relevant in the recent years. From being too inaccurate, it is becoming a more mature technology deployed in large scale systems like the new Visa Information System. From the earlier FERET evaluations to the more recent Face Recognition Vendor Test 2000 and 2002 evaluations we have observed significant improvements in face recognition. Systems based on 3D face recognition even claims to distinguish between identical twins. During our research we have seen that even though face recognition have greatly matured since the earliest forms, there still exists several possible attacks against this technology. Some of the attacks reviewed in this report are specific to face recognition, while others apply for all authentication methods. During the deployment process of face recognition, these attacks should be taken in consideration. As Kosmerlj stated in her thesis; there is still work to be done to improve face recognition before it can be applied in high security settings or applied in large scale applications. One method to reduce the number of people being falsely accepted is by combining the face recognition system with human supervision.

To survey the additional value of a human supervisor, we conducted an experiment where we investigated whether a human would detect false acceptances made by a computerized system, and the role of hair in human recognition of faces. The study showed that, on average, humans were able to detect almost 80 % of the errors made by the computerized system. More over, the study shows that the ability of an individual to recognize a human face is a function of hair: the false acceptance rate was significantly higher for the image-pairs where the hair was removed compared to where it was present. This indicates that there is in fact a substantial opportunity for an impostor to circumvent the human guard using simple and cheap methods. Hair is a feature that may be easily manipulated, and this is perhaps the easiest and cheapest form of non-zero-effort attack on a face recognition system.

Keywords: Face recognition, False Acceptance Rate, False Rejection Rate, attacks on face recognition systems, biometrics, Visa Information System, human supervision, the effect of hair on human recognition of faces

Sammendrag

Ansiktsgjenkjenning er en biometrisk autentiseringsmetode som har blitt mer og mer relevant de siste årene. Fra å være for unøyaktig, har det blitt en mer moden teknologi som skal brukes i det nye Visa Information System. Fra de tidligere FERET evalueringene til de mer nylige Face Recognition Vendor Test 2000 og 2002 evalueringene har vi sett en betydelig forbedring innen ansiktsgjenkjenning. Det er til og med påstått at man ved bruk av 3D ansiktsgjenkjenning kan skille mellom identiske tvillinger. Gjennom forskningsprosessen har vi sett at ansiktsgjenkjennings teknologien helt klart har modnet siden den første formen for ansiktsgjenkjenning, men det er fortsatt flere angrep som er mulige mot bruk av ansiktsgjenkjenning. Noen av angrepene som er gjennomgått i denne rapporten er spesifikke for ansiktsgjenkjenning, mens andre vil gjelde også for andre autentiseringsmetoder. Det bør taes hensyn til disse angrepene når man skal benytte ansiktsgjenkjenning. Som Kosmerlj poengterer i sin rapport så gjenstår det fortsatt en del arbeid for å bedre ansiktsgjenkjenningsmetodene før de kan benyttes i høysikkerhetsinstallasjoner eller i stor-skala applikasjoner. En metode for å redusere antall mennesker som blir falskt akseptert er ved å kombinere et ansiktsgjenkjenningssystem med menneskelig overvåking.

For å evaluere den ekstra verdien av å benytte en menneskelig vakt, utførte vi et eksperiment hvor vi undersøkte om et menneske vil kunne detektere de falske aksepter som et system har gjort, og rollen hår har for menneskelig gjenkjenning av ansikter. Studien viste at et menneske i gjennomsnitt detekterer nesten 80 % av feilene et datamaskinbasert system gjør. Videre viser studien at menneskers evne til å gjenkjenne ansikter er en funksjon av hår; falsk aksept raten var signifikant høyere for bildepar hvor hår var fjernet i motsetning til når det ikke var fjernet. Dette indikerer at en bedrager faktisk har en betydelig mulighet til å omgå en menneskelig vakt ved å bruke enkle og billige metoder. Hår er et ansiktstrekk som lett kan manipuleres, og dette er kanskje den enkleste og billigste form for såkalt *non-zero-effort attacks* på et ansiktgjenkjenningssystem.

Nøkkelord: Ansiktsgjenkjenning, Falsk Aksept Ratio, Falsk Avvisnings Ratio, angrep på ansiktsgjenkjennings systemer, biometri, Visa Information System, menneskelig overvåking, effekten hår har for menneskelig gjenkjenning av ansikter

Contents

Abstract . . . iii
Sammendrag . . . v
Contents . . . vii
List of Figures . . . xi
List of Tables . . . xiii
Acknowledgments . . . xv
1 Introduction . . . 1
1.1 Statement of the problem . . . 1
1.2 Need for the study . . . 2
1.3 Purpose of the study . . . 2
1.4 Research questions . . . 2
1.5 Research methods . . . 3
1.6 Delimitations . . . 4
1.7 Reading guide . . . 4
2 Biometric overview . . . 5
2.1 Authentication . . . 5
2.2 False Acceptance and False Rejection . . . 5
2.3 Multimodal Biometric systems . . . 6
3 Face Recognition . . . 9
3.1 Face recognition methods . . . 9
3.2 Evaluation of face recognition products . . . 10
3.2.1 Face Recognition Grand Challenge . . . 11
3.2.2 The Face Recognition Vendor Test 2005 . . . 12
3.2.3 Other aspects to an evaluation process . . . 12
3.3 Differences between human and computer based recognition of faces . . . 13
4 Visa Information System and border control environments . . . 17
4.1 Biometrics in Visa Information System . . . 17
4.2 The process when applying for a visa . . . 17
4.3 Threats to the security of visas . . . 19
5 Circumvention of face recognition products . . . 21
5.1 Methods to gain false acceptance . . . 22
5.1.1 Photographs in front of camera . . . 22
5.1.2 Identical twins . . . 23
5.1.3 Replay and alteration of templates . . . 23
5.1.4 Swamping attack . . . 24
5.1.5 Piggy-back attack . . . 24
5.1.6 Illegitimate enrollment . . . 24
5.1.7 Coercive attack . . . 25
5.1.8 Trojan horse . . . 25
5.2 Security threats to facial recognition and countermeasures . . . 26

5.2.1 Impersonation attack . . . 26
5.2.2 Attacks between the sensor and the biometric system . . . 28
5.2.3 Back doors . . . 29
5.2.4 Hill climbing attack . . . 29
5.2.5 Liveliness detection in face recognition systems . . . 30
5.3 A successful attack – definition . . . 31
6 Experiment description . . . 33
6.1 Introduction . . . 33
6.2 Procedure . . . 34
6.3 Purpose of the experiment . . . 36
6.4 What data is possible to obtain from such an experiment? . . . 37
6.5 Face image databases and algorithms . . . 38
6.6 Restrictions . . . 41
7 Experiment results . . . 43
7.1 Hair . . . 43
7.2 Other aspects . . . 43
7.2.1 Gender . . . 43
7.2.2 Age . . . 44
7.2.3 Educational degree . . . 44
7.2.4 Time . . . 45
7.2.5 Experience . . . 48
8 Discussion . . . 49
8.1 The theory . . . 49
8.2 The role of hair . . . 50
8.3 Other aspects . . . 50
8.3.1 The role of the age of the participants . . . 50
8.3.2 The role of the educational degree of the participants . . . 51
8.3.3 The role of the time spent on evaluating the image-pairs . . . 51
8.3.4 The role of the gender of the participant . . . 52
8.3.5 The role of experience with face recognition . . . 52
8.3.6 Various considerations . . . 52
8.4 The added value of the work . . . 53
9 Conclusions . . . 55
9.1 The research questions . . . 55
10 Further work . . . 59
Bibliography . . . 61
A Appendix – Definitions . . . 69
B Appendix – Applications developed . . . 71
B.1 IC_Client 1.0 . . . 71
B.2 IC_Administrator 1.0 . . . 71
B.3 RAW Image Converter 1.0 . . . 71
B.4 ImageConverter 1.0 . . . 72
B.5 SFI Analyzer 1.0 . . . 74
C Appendix – Database used in the experiment . . . 77
D Appendix – Results from the experiment . . . 79
D.1 Gender . . . 79

D.2 Age 80
D.3 Educational degree 81
D.4 Time 82

List of Figures

1 Watchlist Reciever Operating Characteristic 6
2 False Acceptance Rate vs. False Rejection Rate 7
3 Multimodal biometrics 8
4 Laptop screen in front of a web-camera 22
5 The importance of eyebrows 28
6 Attack between the sensor and the biometric system 29
7 Registration of participants to the experiment 36
8 Presentation of face-image-pairs in the experiment without hair 37
9 Evaluation time exceeded 38
10 Presentation of face-image-pairs in the experiment with hair 39
11 Screen shot of the administration application for the experiment 40
12 Example of result files generated from the administration module 41
13 Differences in false acceptances between the participant groups 45
14 Histogram of false acceptances in each participant groups 46
15 Distribution of errors when hair was present 46
16 Distribution of errors when hair was removed 47
17 False acceptances vs. age interval 47
18 Application for converting from RAW to JPEG image format 72
19 The bat file generated from the RAW Image Converter 1.0 73
20 The ImageConverter 1.0 74
21 SFI Analyzer 1.0 75
22 Cross tabular – the composition of educational degree between the groups 81

List of Tables

1 T-test False Acceptance and False Rejections vs. hair 44
2 False Acceptance and False Rejections with and without hair 44
3 T-test False Acceptance and False Rejections vs. age 79
4 False Acceptance and False Rejections vs. age 80
5 Anova test Age vs. False Acceptances showing the significants of age . . . 80
6 T-test showing the influence of hair and age 80
7 Levene's test for equality and variance on age 80
8 The significants of educational degree 81
9 Overview of the differences in performance due to educational degree . . 82
10 Frequency distribution – False acceptances 82
11 Frequency distribution – Comparison time 82
12 Correlation False acceptances vs. Total time of comparisons 83

Acknowledgments

Several people have contributed one way or another to the result of this thesis, and I would like to use this opportunity to thank them all for their guidance and support through this process. I would thank my supervisor Erik Hjelmås for guidance and helpful feedback during the writing of this report and experiment. Asbjørn Hovstø, the leader of the Norwegian committee for biometrics K188 [1], who directed me into this subject, and who have provided constructive feedback throughout the process. Marijana Kosmerlj, who provided helpful feedback on the report and necessary data to conduct the experiment. Tom Halvorsen, Senior adviser in UDI and project co-ordinator for the Visa Information System, who have supplied me with information and guidance on the process of applying for a visa and about the Visa Information System in general. He has also given me helpful feedback on my report. I also want to thank Frode Volden for the help he provided in the analysis process of the results from the experiment. The library at Gjøvik University College have also contributed with good help and service.

I also owe the providers of the AR Face Database [2] and the CVL Face Database [3] used in the experiment acknowledgement for access to their face databases. The experiment would not have been possible without their help. The databases were provided by the Computer Vision Laboratory, University of Ljubljana, Slovenia [3] and Computer Vision Center (CVC) at the U.A.B [2].

I would like to thank all these people for the help I have received during the masters' thesis. Finally I would like to thank my fiancée Monica S. Engebakken for patience and support during the process.

1 Introduction

A growing security issue today is the increased occurrences of identity fraud [4] used in terror-related crimes to gain access to resources and locations [5, 6], and illegal immigration with false passport and visa [7, 8]. These are issues that the new Visa Information System (VIS) will try to defeat. Applicants trying to get a visa might not give the correct information about their name or place of living, and they might also try to get a visa under several different names. If the authorities checking the information receive applications containing only written data, they have no way of checking if the applicant has tried to apply under a different name. This is a problem VIS will try to defeat using biometric authentication such as face recognition (mandatory) and fingerprint (optional) [9, 10] as a supplement to manual control. When the applications in addition contain a photograph of the applicants face, and this is registered in a central database, the authorities can check the information by searching with given criteria over registered faces. In this thesis we will examine the possibility of circumventing face recognition products available today using methods of low cost, which with the necessary knowledge is possible to conduct for the average person. A survey of found methods with higher cost will also be provided. The reviewed methods for circumventing face recognition systems will then be evaluated towards a border control setting for visa applicants, which will be supervised by a human guard.

1.1 Statement of the problem

When the authorities decide which face recognition product to use, two important criteria are the False Acceptance Rate (FAR) (See chapter 2.2) and False Rejection Rate (FRR) (See chapter 2.2) of the products. Traditional estimation of FAR of face recognition products is usually based on zero effort impostors' [11]. In a real border control environment traditional estimation of FAR with *zero effort* impostors are not necessarily representative for the real amount of false acceptances. Potential attackers with or without plenty of resources could use several technological and physical techniques to circumvent the system. This could involve physiological alteration of their appearances using masks, facial make-up, different facial hair or plastic surgery, or technological techniques to alter information about an applicant for a visa. Also, identical twins is traditionally problematic when using face recognition, although a supplier of 3D face recognition claims to have countered this problem [12]. To obtain a more realistic evaluation of FAR it is therefore important to examine possible attacks, their influence on the FAR, and resources needed to perform them. Such research will enable the authority and other users of face recognition products to perform more enlighten evaluation of face recognition products, and make them aware of the problems so that they can execute necessary countermeasures.

1.2 Need for the study

There exists little or no publicly available data on face recognition products response to attackers that perform an effort other than simply supplying their own biometric data hoping that they will circumvent the system. How can those employing such systems know which system to use when the evaluation is based surely on zero effort circumvention? This could very well result in the choice of the lesser product. Also, those who employ such system should be aware of the different approaches that exist for circumvention, so that they can make measures to thwart this. This thesis will provide an overview of these attacks and how they are done.

Institutions that are employing face recognition products will undoubtedly benefit from a survey that has demonstrated the effect of non-zero effort impostors, since this would make them more aware of the potential differences between traditional estimation of FAR, and when it is based on non-zero effort attacks. Hopefully this will make for a demand for more realistic evaluations of FAR, more in accordance with the environments in which it will be employed. Users may then avoid potentially costly pitfalls. In Norway such stakeholders could be UDI, which are heavily involved in the introduction of face recognition in the new NORVIS (the Norwegian version of VIS) system, and other institutions that decide to use face recognition products.

1.3 Purpose of the study

The intention of the thesis is to see if the face recognition products available today are adequate in a setting such as the new Visa Information System. To evaluate this, the authority performing the evaluation should have information of the potential threats that could arise. This thesis will provide such information by giving an overview of some of the threats that exists towards face recognition software, and an evaluation of the probability for such threats occurring in a border control environment, in a setting like the new Visa Information System.

1.4 Research questions

The new visa system involves biometrics such as fingerprint and face recognition. This thesis will examine the different methods for circumventing face recognition products, involving the resources and skills needed and the potential cost. The focus of this thesis will be on the use of face recognition in a border control environment with non-zero effort attackers and the effect these will have on face recognition products reliability and performance. In order to find out how impostors will affect the face recognition systems, the following issues will be examined:

1. What efforts does an impostor need to make to deceive a face recognition product in a border control environment?
2. How will the resources of an attacker influence the security of a face recognition product in a border control environment?
3. Could today's procedures for calculating FAR result in a positive evaluation of insecure products?

4. What effect will non-zero effort attacks have on the FAR of a face recognition system in a border control environment?

1.5 Research methods

When deciding which research methods to use, we used J.W. Creswell's book *Research design* [13] as a basis. In this thesis we look at different methods for circumvention of face recognition products that affect the false acceptance rate. The primary method used for analyzing this problem is literature survey. To establish the impact non-zero effort attacks have on the FAR of a face recognition product in a border control environment, we have used a mixed methods approach. The intention was to contact individuals in the face recognition community, to see if they had literature or knowledge of literature about circumvention of face recognition systems that could be used in the thesis work. This however did not result in any usable material on circumvention, but we did receive information on face recognition in general and more specific information about the Visa Information System. A thorough examination of available literature through the Internet and different libraries provided the material necessary to conduct my thesis.

The methods were used for gathering information to this thesis:

- A literature study to find general information on face recognition and issues regarding circumvention of face recognition products and the human perception of faces. The literature study was used to gain an increased knowledge within these areas, and to obtain ideas on a useful experiment. To be able to use a literature study, there has to be relevant literature available, and we should have access to the necessary databases. Access to the Gjøvik University College library and the databases available through this library, IEEE and Citeseer, combined with web searches provided most of the necessary literature to perform the study. Contacts within the face recognition community and the VIS provided additional literature that were necessary to perform this study.
- Email correspondence with contacts within the face recognition community and the government to obtain knowledge beyond what is possible from searching the web and using the library. To be able to correspond with such contacts, information about such contacts should be available. This was achieved through a former employer and the teaching supervisor we where able to come in contact with such contacts. We did receive vital information about face recognition and the Visa Information System using this method.
- An experiment on human comparison of face image-pairs with and without hair, to evaluate the effect hair has on human ability to recognize faces. The image-pairs used in the experiment had already been accepted as the same individuals of a computer-based face recognition system. The success of this experiment depended on enough people being willing to participate in the experiment. By using this method we were able to measure the effect hair has on human ability to recognize faces. This way we were provided an indication of how easy or difficult it is to circumvent both a computer-based face recognition system and a human supervisor.

1.6 Delimitations

This thesis focuses on the non-zero-effort attempts (See Appendix A) effect on false acceptance rate in general, and in the discussion the angle taken is that of a border control environment. Non-zero-effort attempts that affect the false rejection rate is not part of this thesis, and are only briefly mentioned.

1.7 Reading guide

We will first review the basic terminology within biometrics in chapter 2 and theory of face recognition in chapter 3, before we are introduced to the Visa Information System and the process of applying for a visa in chapter 4. These chapters will provide a thorough introduction to readers unfamiliar with authentication in general and face recognition and the Visa Information System in particular, to make them more able to grasp the remaining content of the report. One of the main contributions of this thesis is chapter 5, which provides a thorough review of methods for circumventing face recognition products. It also provides some methods for preventing or reducing such circumvention. This is part of our contribution to research question 1, 2, 3 and 4. Chapter 6 provides a description of the experiment on human ability to detect false acceptances made by a computerized system, and the role of hair in human recognition of faces. In chapter 7 we present the results obtained from the experiment, while we in chapter 8 discuss the results obtained from the experiments and other findings throughout the work on the masters' thesis. In the conclusion in chapter 9 we summarize our findings and work, before we suggest further work within our topic in chapter 10. The appendix A provides definitions on words that may be unfamiliar to the reader, while appendixes C and B provides an overview of the database and the applications developed for this thesis. Appendix D provides further details from the experiment than those presented earlier in chapter 7.

2 Biometric overview

2.1 Authentication

When a user is authenticated the person concerned lets the system know his identity. There are two modes for authentication; verification and identification [14, 15]. In addition D.M. Blackburn [16] adds another task to a biometric system; the watchlist.

- Identification: (Who am I?) this mode is used when the identity of the individual is not known in advance. The entire template database is then search for a match to the individual concerned, in a one-to-many search. If a match is made the individual is identified [14]. It is important to note that a match does not mean a sample that is identical to the template, but rather is within a given threshold [17, 15].
- Verification: (Am I whom I claim I am?) this mode is used when the person provides an alleged identity. The system then performs a one-to-one search, comparing the captured biometric characteristics with the biometric template stored in the database. If a match is made the identity of the person is verified [14].
- The watchlist task: in the watchlist task the person does not claim any identity. The biometric sample of the individual is compared with the stored samples in a watchlist to see if the individual concerned is present in the watchlist [16, 18]. Examples of watchlist tasks could be comparing a flight passenger towards a database of known terrorists, or comparing a John Doe patient with a list of missing persons. When a person is found that have a resemblance to one or more samples in the watchlist that is higher than the given threshold, the system should give an alarm and return the samples that triggered the alarm. When this alarm goes for an individual that is actually present in the watchlist and this person has the highest similarity score, it is called a *correct detect and identify*. An alarm that goes of even though the person is not present in the watchlist is called a *false alarm*, while the frequency which false alarms encounters is called the *false alarm rate* [16]. In an ideal system we want the false alarm rate to be 0% and the correct detect and identify rate to be 100%. However this is not possible, so we must compromise. To better see this give-and-take relationship, we can plot the detect and identify rates and their associated false alarm rate in a *Watchlist Receiver Operating Characteristic* (See figure 1). The decision on whether to choose a system with a low false alarm rate and a medium correct detect and identify rate, or if we want a medium false alarm rate and a high correct detect and identify rate, depends on the usage of the system.

2.2 False Acceptance and False Rejection

There are two types of error a biometric system cam make [15]: False rejection which is when a legitimate user is rejected, and false acceptance which is when an illegitimate user is accepted as someone else. The probability that a genuine person is rejected is called false rejection rate (FRR), while false acceptance rate (FAR) is the probability that

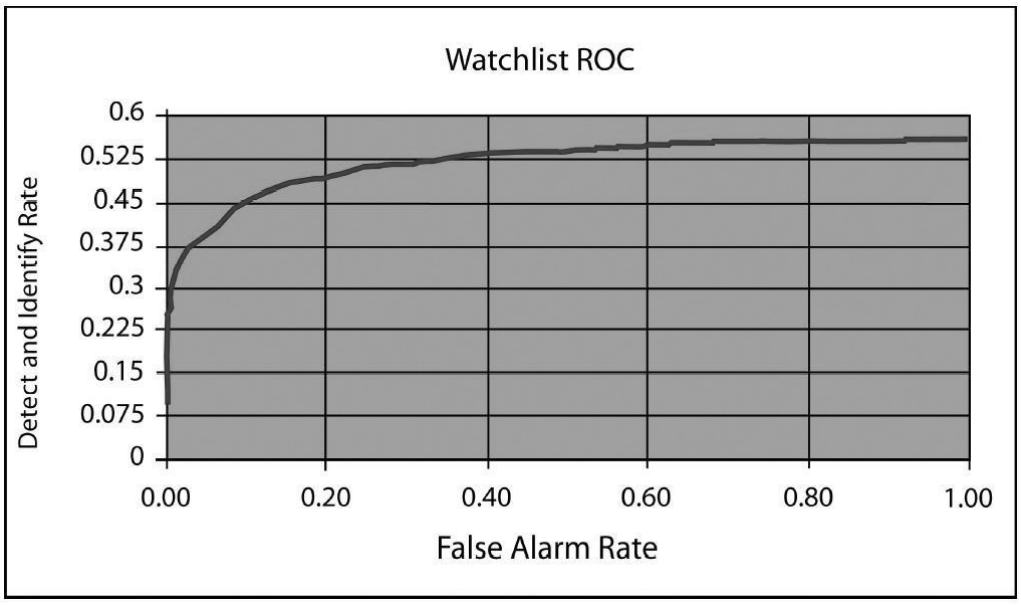

Figure 1: Watchlist Receiver Operating Characteristic – The figure is taken from *Biometrics 101, Duane M. Blackburn* [16]. The detect and identify rates and their associated false alarm rate is plotted into the diagram. A WROC helps to better see the give-and-take relationship between false alarm rate and the correct detect and identify rate.

an impostor is accepted as a legitimate person. The point were FRR and FAR are equal, is called equal error rate (EER) (See figure 2). In addition there are some individuals that do not have the biometric feature from which there can be produced repeatable templates. The expected proportion of the population for whom the system is unable to obtain repeatable templates is called the failure to enroll rate. A system may also be unable to capture or locate an image of sufficient quality [19]. This could be because their finger is plastered or the quality of the image inadequate [20]. The expected proportion of transactions for which this is the case is called the failure to acquire rate.

Face recognition products that shall be used by VIS in a border control environment, where the intention is that as many previously registered candidates as possibly are recognized, shall operate on a small FRR when registering a new visa applicant to prevent multiple registrations of visa applicants. Further, the face recognition products must have a smallest possible FAR at the border control when the applicant is checked before they are granted access to the country in order block as many illegal attempts as possible.

2.3 Multimodal Biometric systems

There seems to be no single biometric feature that is able to be as accurate and reliable as some systems require. Fingerprints can be copied [21] and altered by cuts and bruises [15], face recognition has too many false acceptances [11] and has not yet been proven to distinguish between identical twins – although Aurora claims that they can [12]. There are also several other drawbacks in other biometric authentication schemes [15]. To cope with this we can use a multimodal biometric system [15, 22, 23]. That is a system that

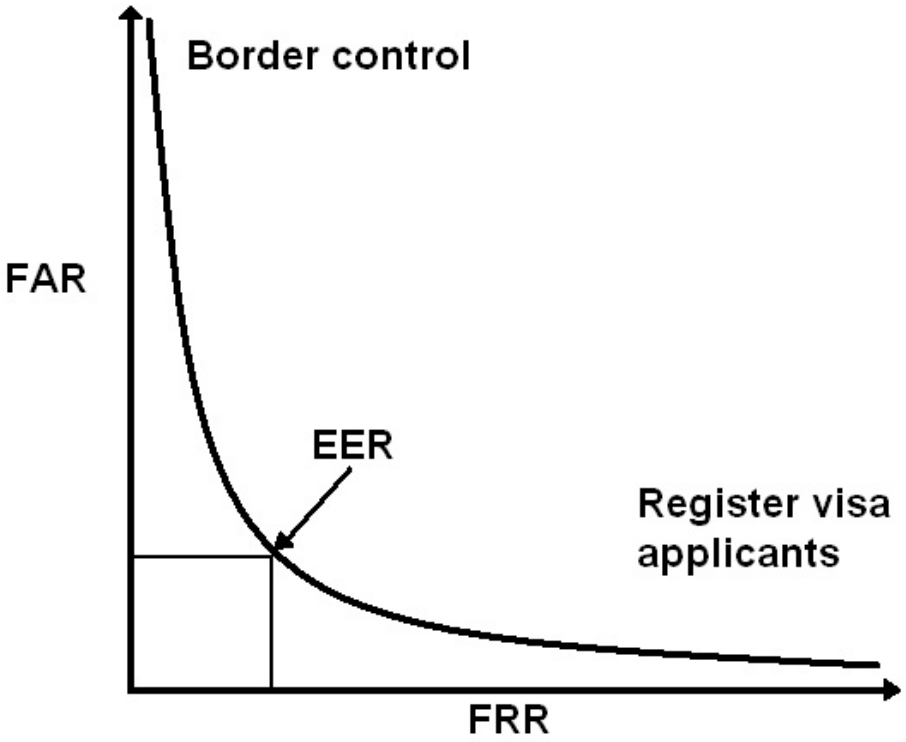

Figure 2: The figure provides a graphical illustration of the relation between false acceptance and false rejection in a border control environment. In a border control we want a small False Acceptance Rate to prevent impostors using stolen visa's to pass themselves off as someone with a legal visa. At the registration of the visa, we want a small False Rejection Rate when we search to find whether or not the applicant previously have been issued a visa with another identity or whether or not the applicant is registered in a watchlist.

combines the use of more than one biometric feature. For example, the system could use fingerprint recognition combined with face recognition as shown in figure 3. This way the system accommodate for the problem with distinguishing between people with similar faces like for instance identical twins by using fingerprints, while at the same time the problem with worn fingerprints and people without hands are handled by using face recognition. L. Hong et al. [24] demonstrated in their paper a multimodal biometric system that combined face recognition with fingerprint recognition. The system showed significantly improvements in recognition performance. Face recognition was first applied to limit the search to the top five matches, followed by fingerprint recognition to make the final decision. Not only the recognition accuracy was improved, but also the CPU time was improved compared to sheer fingerprint recognition because only the top five had to be computed with fingerprint recognition. Also a combination of different methods of recognition within the same biometric feature could be used to accommodate for drawbacks in one scheme by combining it with another scheme that does not have the same drawbacks [15]. For example combining a facial recognition system that has good performance on faces exposed to illumination changes, with a system that has greater performance on face images taken in a controlled environment. This way we

can accommodate for both situations. Biometric systems match scores that are generated by noisy input has large variance. By installing multiple sensors that capture different biometric traits, much of this variance could be accommodated. This is also a kind of multimodal biometric system [23]. A multimodal system also provide anti-spoofing measures [15, 25, 23] by making it more difficult for an adversary to simultaneously provide several different features of a legitimate user.

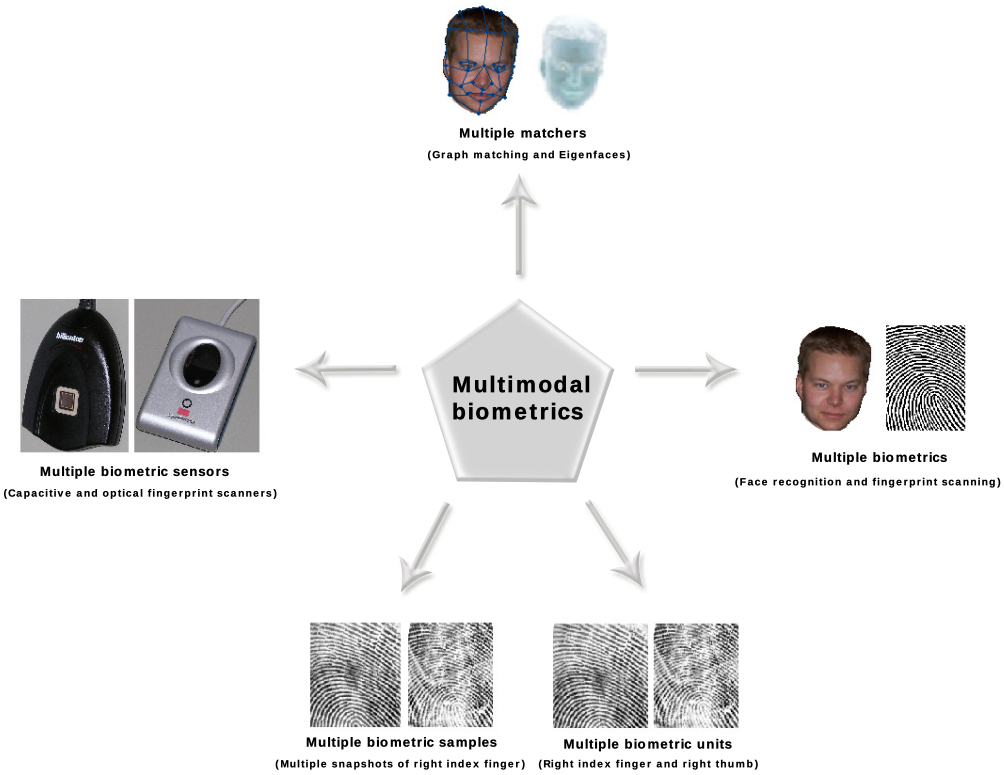

Figure 3: Presentation of different variations of multimodal biometrics. Multimodal biometric authentication could be employed using different biometric features (two different fingers, or combining iris-scan and fingerprints or face recognition and fingerprints), different scanning technology, different matching technology, duplicate snapshots etc. The illustration is made by inspiration of *An Introduction to Biometric Recognition, by Anil K. Jain, Arun Ross and Salil Prabhakar, published in IEEE Transactions on circuits and systems for video technology* [15].

3 Face Recognition

3.1 Face recognition methods

Face recognition, like other biometrics, has various methods for recognizing people. Most of which are resistant against moderate changes in hairstyle [26], as these techniques do not use the areas near the hairline. The process flow in face recognition consists of four phases: capture of samples, feature extraction, template comparison and matching. There are several methods used in face recognition, some more suited than others for specific applications. Recent surveys and reviews on face recognition or specific face recognition technologies are provided in Zhao et al. [27], Kong et al. [28], Li and Lu [29], and Li and Jain [30].

The most used techniques for face recognition are Eigenfaces, Local feature analysis and Elastic Graph Matching as described below.

Eigenfaces [31, 32, 27] was developed by Massachusetts Institute of Technology (MIT), and was motivated by a technique developed by Sirovich and Kirby in 1987 for efficiently representing pictures of faces using principal component analysis (PCA). Variations of eigenfaces are often used as the basis of other face recognition methods. It has been argued that this technique does not bear any resemblance to the way humans recognize and measure similarity between faces. However, according to Woodward et al. [33] the mathematical properties of the eigenface representation and matching process have been demonstrated to achieve reasonable results in certain minimally controlled environments. Like all facial recognition technology, the eigenface recognition method is best utilized in well-lit, frontal image capture situations [26].

Local feature analysis (LFA) [32, 26] is one of the most widely used facial biometric technology today, and can accommodate for some changes in facial expression and aging. Local feature analysis refers to a class of algorithms that extract a set of geometrical metrics and distances from facial images and uses those features as the basis for representation and comparison. The features used are typically the eyes, mouth, nose, jaw line, eyebrows and cheeks. These features are represented with their position, size and general outline. The good performance compared with some other techniques, are among the factors that has made this technique popular. One drawback for this method is that it is dependent on a relatively constant environment and the quality off the image.

Elastic Graph Matching [34, 35, 36] is another method used in face recognition. Its main advantage is that it can provide face recognition that is invariant to affine transformations and localized changes in facial expressions [37]. In Elastic Graph Matching, local features are extracted at specified locations of the face. Also the distances between these nodes are recorded. Some features are more reliable and important for recognition than others, and because of this an approach with the use of weights have been intro-

duced [38]. An extension of the Elastic Graph Matching approach has been introduced that uses several images of the same individual, typically from different angels. This is called Elastic Bunch Graph Matching [34, 35, 36]. Each node on the graph then contains several values. This improves the recognition because it will be more robust to differences in posture and facial expressions. Elastic Graph Matching is called elastic because the match is preserved approximate instead of rigid [34].

Previous face recognition data sets has been restricted to still images, but now the Face Recognition Grand Challenge invite vendors of three-dimensional face scans [39]. This is an element in achieving FRGC's goal of the development of algorithms that have substantial improvement in performance compared to the FRVT 2002. As mentioned earlier, Aurora [40] claims according to [12] that their 3D software is able to distinguish between identical twins. This is a significant improvement compared with 2D face recognition. However, these claims are not substantiated by independent tests that also examine the systems ability to recognize faces over time. A problem with the current two-dimensional method is that although it works well under conditions similar to that of training, there still remains much work to accommodate for changes in illumination and pose [41]. The 3D face recognition method used by Gang et al. [41] showed an increase in performance under different pose and lighting conditions, a result substantiated by [37] and [42]. Medioni et al. [42] performed a test with 3D face recognition and compared their results with 2D face recognition on images of 100 subjects, each acquired in seven different poses within $\pm$ 20°. The comparison showed a considerable improvement in FAR and corresponding FRR using 3D compared to the 2D systems; 2 % equal error rate for the 3D face recognition method, versus 6 % for 2D face recognition method. Although the advantages of 3D face recognition is apparent, there are also, as pointed out by Bowyer et al. [43], several disadvantages with current 3D face recognition. Among others, the methods do not handle variations in facial expression very well, and the tests performed are often biased and not based on large and challenging datasets. Further, illumination do affect the quality of the sensed data, and create *holes* (an area of missing as a result of the sensor being unable to acquire data) and *spikes* (an outlier error in the data resulting from disturbance, for example reflection of light).

3.2 Evaluation of face recognition products

Warren Court provides in his paper [44] an introduction to biometric evaluation for organizations that want do perform their own biometric studies, including established methodologies and criteria from which to develop a test plan. For personnel not familiar with testing of biometrics, this paper provides a basic knowledge to the subject, and is a good place to start the first time one is involved in the evaluating process of biometric authentication systems.

One of the evaluating procedures that are the most referred to is the Face Recognition Vendor Tests (FRVT). It provides independent government evaluations of commercially available face recognition products, and are designed to provide U.S. Government and law enforcement agencies with information to assist them in determining where and how facial recognition technology best can be deployed [18]. The earlier FERET evaluating

methodology (1994, 1995 and 1996) for face recognition algorithms [45] and the FERET database is perhaps the most referred to in the literature, and has helped advancing face recognition to the prototype stage. By the year 2000, face recognition technology had matured to commercial systems. The improvements of the technology from the FERET test were measured in the Face Recognition Vendor Test 2000 (FRVT 2000). And the performance progress from 2000 to 2002 were evaluated in the FRVT 2002 on large real-life databases [18]. The FRVT 2002 report [18] showed a considerable improvement in error rates. From 2000 to 2002 there was an error rate reduction of 50%.

3.2.1 Face Recognition Grand Challenge

Since FRVT 2002, a number of new face recognition technologies have been developed that have the promise of improving performance by an order of magnitude. Among others Aurora [40] claim that their 3D face recognition product can distinguish between identical twins [12]. Previous face recognition data sets have been restricted to still images. To develop face technologies that include high resolution still images, multi-images of a person and three-dimensional face scans, the Face Recognition Grand Challenge (FRGC) is being conducted from May 2004 to July 2005 [39]. The goal with FRGC is to develop algorithms that have substantial improvement in performance compared to the FRVT 2002.

The FRGC is divided in two challenges; version 1 and version 2. Version 1 is designed to introduce the participant to the FRGC challenge problem format and its supporting infrastructure provided by the Biometric Experimentation Environment (BEE). This is an XML based framework for describing and documenting computational experiments. The BEE provides a framework that makes it possible to describe the experiment, record the raw results and provide the analysis, presentation and documentation of the experiment in a common format.

Version 2 is designed to challenge researchers to meet the FRGC performance goal. Participation in the FRGC is free and open to all interested researchers. The FRGC version 2 consists of six experiments:

1. **Experiment 1:** Experiment 1 is a controlled experiment where the gallery consists of a single controlled still image of a person, and each probe consists of a single controlled still image.
2. **Experiment 2:** Experiment 2 studies the effect using multiple still images of a person has on performance. Each biometric sample consists of the four controlled images of a person taken in a subject session. For example, the gallery is composed of four images of each person where all the images are taken in the same subject session. Likewise, a probe now consists of four images of a person.
3. **Experiment 3:** Experiment 3 measures the performance of 3D face recognition. The gallery and probe set consist here of 3D images of a person.
4. **Experiment 4:** Experiment 4 measures recognition performance from uncontrolled images. The gallery consists of a single controlled still image, and the probe consists of a single uncontrolled still image.

5. **Experiment 5:** Experiment 5 examine and compare 3D and 2D images. The gallery consists of 3D images and the probe consists of a single controlled still image.
6. **Experiment 6:** Experiment 6 also examine and compare 3D and 2D images. The gallery consists of 3D images, but in contrast to experiment 5 the probe here consists of a single uncontrolled still image.

3.2.2 The Face Recognition Vendor Test 2005

The Face Recognition Vendor Test (FRVT 2005) [46] will be conducted by the National Institute of Standards and Technology (NIST) in the time frame of August and September 2005. It follows five previous face recognition technology evaluations – three FERET evaluations (1994, 1995 and 1996) and FRVT 2000 and 2002. FRVT 2005 will determine if the goal of the FRGC are meet by measuring the progress of face recognition systems since FRVT 2002 and the effectiveness of new face recognition technologies. The FRVT 2005 is planning to evaluate performance on high resolution still imagery (5 to 6 mega-pixels), three dimensional facial scans, multi-sample still facial imagery and preprocessing algorithms that compensate for pose and illumination.

The accuracy of the evaluation will be guaranteed by using images not previously available to researchers or developers. The test environment, called Biometric Experimentation Environment (BEE), and the test data will be provided by the government.

3.2.3 Other aspects to an evaluation process

When evaluating face recognition products it would perhaps also be wise to use the Common Criteria (CC) and its Strength Of Function (SOF). SOF investigates the strength of the underlying security mechanism of what is evaluated. In this context, that is the ability to correctly identify or verify a user. According to M. Krechel et al. [47], it has been proposed that all sets of security requirements and specifications, which are used as the basis for evaluation of a biometric product, should include a claim for SOF and a rationale to explain the claim.

The CSU Face Identification Evaluation System [48] evaluates the performance of face identification systems, and may also be considered as a guidance in the evaluating process for such systems.

Mansfield and Wayman have produced a paper where they demonstrate a best practice in testing and reporting performance of biometric devices [19] using technical performance tests. This report is based in an earlier report with the same subject [49] and feedback from that report. They acknowledge that this is not the only form of biometric testing, and mention other types of testing like reliability, availability, vulnerability and security. These are all tests that will be highly relevant when evaluating face recognition products to be used in systems like VIS and NORVIS. In connection with the new visa system, there is a need for further study of the other areas of evaluating face recognition products suitability for the system, such as security, reliability, availability and vulnerability. According to Mansfield and Wayman [19], other groups are also considering methods and philosophies for these other types of tests. Issue 1 of the best practice report [49]

has been used when conducting a performance evaluation of seven biometric systems, including a face recognition system [20]. These systems were tested for a scenario of positive identification in normal office environment, with cooperative users.

All of these papers could work as a good basis when conducting an evaluation of face recognition products. However, the security threats must also be taken in consideration during the evaluation process.

When evaluating face recognition products there are several requirements that should be fulfilled [50]. For starters, the details of the procedures used for the evaluation must be published along with the test results and representative examples of the dataset used for the test. The details of the information of the evaluation process should be such that others can repeat the evaluation process. Further, on must take considerations regarding how hard or easy the tests shall be. It is important that the test is not too easy, because that would lead to a score of about 100% of most of the products, and it would be hard to differentiate the products adequacy. With that in mind, the test cannot be too hard either, because then it would be beyond the capability of the existing face recognition techniques. The conclusion is that evaluating procedures should be based on a middle way of the two [50], often referred to as the *three bears* problem [18].

A problem with evaluation of biometric technologies like face recognition is that the performance depends much on the environment and method used in the evaluation. And the result of such test has concluded with a better performance than the products has performed in real life. As the FRVT 2002 points out, there is no face recognition system that is *right* for all applications [18]. Some systems may be favourable in one specific setting, while it performs inadequately in another. Keep in mind that face recognition systems is application dependent, and evaluate the face recognition system according to the conditions given in your usage of the system. Considerations should be made on whether the system is going to be used in verification, identification or watchlist mode, if the images will be exposed to changes in illumination and so on. These are all aspects that should influence the choice of face recognition technology.

3.3 Differences between human and computer based recognition of faces

From right after birth humans are able to recognize faces. Recognizing faces is a natural talent, and we are better at recognizing faces than other objects. There are parts of our brain that are more involved in recognizing faces than others, and the recognition of faces is a process that is done by other parts of the brain than dose involved in deciding facial expressions and state of mind [51].

We will here provide some differences and similarities between human and computer-based face recognition. According to Bruce et al. [52],

> *there is no necessary link between techniques developed by engineers to automate face recognition, and natural mechanisms used by the human visual system to achieve the same end.*

Their article presents the result of an experiment that compared human face recognition with two computer-based recognition systems. The observers were told to sort the images they found similar. 40 observers were presented with images were the hair was visible and 40 observers were presented images were the hair were removed to acquire data from face recognition were the hair was not dominant. Each system produced significant – but numerically small – correlations with human similarity data. The graph face recognition system provided similar correlations to the humans' ratings of faces both when the hair was present and when it was removed. PCA, however, provided much higher correlations to the ratings obtained from the observers presented with faces with hair. Graph matching is more similar to humans' ability to recognize faces where the image varies.

We have seen that computer based face recognition systems have performed ratings of similarity that correlates to human perception of face similarity. However human and computer-based perception of similarity differed somewhat. In the master thesis of Kosmerlj [11], the face recognition products used found several different identities similar. According to Kosmerlj, she did not find these faces to be similar from her perspective. It should be noted that no test panel were used to verify the human perspective, and that these images presented faces with hair, which might have influenced the human evaluation of similarity. Kalocsai et al. [53] however, performed an experiment that correlated the performance of a global feature based system and a global template matching based system with human face recognition performance on the same data set. They argued that the best artificial system would be one that performed as well as humans. An oval area around the face blocked out everything outside the face, eliminating the effect of hair and background. A test-panel of 64 observers was told to decide if two images were of the same person. Some of the algorithms used in the Kosmerlj's thesis is also represented in the article of Kalocsai et al.. The results from Kalocsai et al. shows that the Gabor-filter based system correlated very high with the performance of human error on different trials (different trials refer to measuring similarity between two different individuals). It also correlates high to human error on same-trial (same-trial refers to two images of the same individual, but with different expressions), however the humans make less mistakes on highly similar same-pairs. The PCA-DLA method received similar results, however with somewhat lower correlation to human performance. While both methods corresponded somewhat to human performance, the Gabor-filter correlation coefficients were higher, indicating that local features are necessary when seeking face recognition that resembles to human performance.

Another aspect influencing computer-based face recognition performance is illumination. Experiments show that changes in illumination have greater effect on similarity than changing the identity. Humans however are less affected by such changes [54]. There are several approaches in face recognition to overcome the problem of illumination; edge map, filtering the image with 2D Gabor-like functions, derivations of the gray-level distribution and logarithmic transformation. With a 34 degree change in horizontal illumination angle these methods perform poorly for changes in illumination angle from left to right, however with a smaller change in angle (17 degrees) from left to center, the results improved [54]. However, for a large database the results will be unsatisfactory,

emphasizing the importance of controlling lighting sources in face recognition. Experiments performed by Adini et al. [54], showed that for changes in expression, a simple gray-level comparison was sufficient to recognize all the faces that was represented with the whole face (except hair). Other methods, and recognition of only parts of the face, performed poorer. Cross et al. [55] have however since [54] conducted an experiment to see how changes in illumination affects face recognition. Their experiment showed that available algorithms was able to handle illumination quite well, however as they point out, illumination could still cause major problems when combined with other changes like expression or pose.

Experiments have shown that the expectation of the observer recognizing faces influences the identifying process. People looking for individuals expected to have broad shoulders and round faces, will choose the person presented with these characteristics – and some resemblance – as the right person, although the resemblance may not be great [51]. In a border control environment this could be exploited. If the person presented, according to the computer system, has similarity to the alleged identity, and the weight, height and hair is similar, this could be enough to pass, although the resemblance might be insignificant. This is specially the case when the person in question is of a race different of the observers'. According to Brigham [56] people have a tendency to be less accurate when recognizing faces of a different race, leading to higher false acceptance rate (22% higher false acceptance rate for cross-race faces than for own-races). This is also supported by Chiroro et al. [57] who found the false positives to be significantly smaller on own race than for different races. People in general, have more problems with separating faces of human races different from their own [58, 59]. It is shown that one main reason for this is that these people are unfamiliar with recognizing faces of different races. On the other hand, people accustomed to associating with other human races recognize faces of these races with the same accuracy as people from their own race. A study performed on African-American students showed that those with considerable contact with people of Caucasian race, recognized faces of people of Caucasian race as good as they recognized faces of their own race. On the other hand, students that had seen few faces with other skin-color than their own emphasized the cross-race effect [57, 51]. Levin argues that the cross-race effect is due to people collecting race-specifying information at the cost of recognition accuracy [58]. Results from the Face Recognition Vendor Test 2002 [18] suggest that the race also has an impact on the performance of face recognition algorithms. Among others it suggested that people of Chinese origin are easier to recognize than people born in Mexico. This is substantiated by a study of Furl et al. [59], that showed that experience-based algorithms recognized minority-race faces more accurately than majority-race faces. A learning process favorable to own-race faces was established as the reason for this effect. Experiments reviewed in [56] show that training in recognizing cross-race faces have an significant improvement in recognition of cross-race faces, suggesting that the cross-race effect might correlate somewhat to experience in recognizing faces of a particular race.

The eyebrows are traditionally believed to be less significant compared to eyes in computer-based face recognition. Human perception of faces however is shown to rely heavily on the presence of eyebrows, it is even more important than the eyes [60]. J.

Sadr et al. performed an experiment demonstrated in [60] that indicate that the absence of the eyebrows has even greater negative effect on human identification of faces than the absence of the eyes. Even at distance the eyebrows make an important role in recognition of the face because it separates the forehead and orbit, and because the eyebrows are less affected by shadow and illumination changes. Removal or manipulation of the eyebrows could then have great effect on a human observer's ability to recognize a person.

4 Visa Information System and border control environments

4.1 Biometrics in Visa Information System

At the end of 2006, the Visa Information System (VIS) will be launched. An important part of this European co-operation will be the use of face recognition. Although other biometric authentication methods like for example iris provides better performance and accuracy [20, 15], the VIS will use face recognition and fingerprints. According to [10], face recognition will be mandatory and fingerprint will be optional, even though fingerprints is a more mature biometric authentication method [61] with international standards and a higher recognition rate [20, 17]. There are several arguments for these choices. Among others fingerprints are considered sensitive and there are legal considerations towards using fingerprint in some countries [1]. Necessary alterations of current laws in these countries may not be in place for several years. As pointed out earlier iris-scan is a method that has performed well in tests concerning accuracy, and is a biometric that should be considered in such settings. However, this is a method that is patented by the company Iridian Technologies [62], and EU would prefer to own the rights to the algorithms themselves. This is also the case for algorithms for face recognition. Today the only thing that is agreed on is that a JPEG image shall be stored in the VIS [2]. In what way this shall be used in identification of individuals in 2007 is still not settled. Recent development in the process of choosing a authentication method for the new VIS is that fingerprint shall be the primary method for automatic recognition [3].

4.2 The process when applying for a visa

When an applicant is applying for a visa, the applicant first delivers his visa application [4]. The controlling authority then checks the identity of the applicant. If there already exist a visa sticker in the passport, this is used to search in the VIS. First the visa number will be used to find if the visa is already registered. When this is the case, the data returned by the VIS will be compared with the applicant by a human controlling authority. If the information is correct the applicant is identified and the data and image (and/or fingerprint) of the applicant could be reused, or if necessary corrected. In cases were the visa is not already registered, the procedure will be to search by name and birth of the applicant in the VIS database. If the applicant is registered in the database, this data will be compared with the applicant. In cases where the information is the same, the applicant is identified and the registered data is reused or corrected.

[1] According to Tom Halvorsen in UDI – January 2005, face recognition will be mandatory and fingerprints will be optional, partially due to legal issues in some countries.

[2] The compression of the images will be such that no vital information is lost from the face images', preserving the facial features. See appendix A for further information about the JPEG and JPEG 2000 format.

[3] According to Tom Halvorsen in UDI – June 2005, fingerprint will be the primary source for automatic authentication. Face images is still to be captured for visual evaluation. This is still not formally decided, because the use of biometrics is to be defined in August/September.

[4] This chapter is mainly based on information provided by email-correspondence with Tom Halvorsen, UDI

Situations can occur where the search procedures presented earlier do not return any data of the applicant from the database, or where the authority is in doubt. In such situations, a one-to-many search with the fingerprint (possible also with face image in the future) of the applicant will be performed towards the VIS database to see if the applicant is registered under a different identity. If this search results in a match with a record of the person under a different name, the case must be solved.

When none of the above results in a match with the VIS database and adequate documentation is delivered, a new case is created in the VIS. The applicant is then interviewed and the application data is registered in the computer system NORVIS (in Norway), and stored in the VIS. This stored data contains one image in JPEG format and eventually 10 fingerprint-scans, which is controlled before it is stored in the VIS database. First an image that is received with the application is scanned. From the time NORVIS is operative (from the fall of 2005) to the end of 2006, images are only obtained by scanning. This image is then only written onto the sticker. When VIS is activated (at the end of 2006), an image will be taken with a digital camera during the application process.

For each application one image only is registered per applicant. If an image from a previous application is available this could be reused. The image is controlled according to a manual from the ICAO document 9303 [63], which guides all biometric data used towards travel documents and visa. The automatic face recognition process will always be supplemented by a human evaluation of the images. The lighting, distance between applicant and camera, resolution, height and width on the image, and background color will be gathered into one standard. A frontal picture without expressions will be taken where all of the head of the applicant is visible. Further guidance on capture of face images are provided in [64], *Annex A* in particular, which provides a *Best practices for Face Images*. Face recognition algorithms available today are able to handle regular glasses [65]. Because of this, applicants do not need to take their glasses off, given that the spectacle frame does not cover too much of the face and do not reflects to much light. However sunglasses must be taken of [63]. By the end of 2006 it is planned that VIS will consist of complete systems handling image and fingerprints. These will operate with rigidly mounted lighting or flash. The image is stored as JPEG or JPEG 2000 in the VIS database. At a later time the image might be processed for template generation for use in automatic face recognition. After this procedure the NORVIS performs a search towards the Schengen Information System (SIS) art. 96 database, to check if the applicant is reported as unwanted within the Schengen. If the applicant is registered in the SIS database, a visa is normally denied.

NORVIS performs a check towards lists in accordance with the consular instructions (CCI) of the Schengen to see if the applicant's native country is listed on such a list (CCI annex 5b). If this is the case, the country that has reported the country to this list will be consulted in forehand of an emission of a visa. A consultation process will then be started, where data will be provided to the country in the Schengen that brought the country of interest on the list. This country is then given seven days to respond (this deadline may be postponed by up to 90 days). Should the country not respond or asks

for a postponement within the deadline of seven days, this will be regarded as consent to the emission of the visa. Countries responding within the deadline may either give their consent or refuse emission of visa. If the visa is denied, Norway may not issue a Schengen visa. In distinct circumstances the applicant can apply for a visa that applies to Norway only. In such cases, the country refusing the visa shall be notified about this. Several checks will also be performed on lists of restrictions to see if the applicant for some reason should not be granted entry to Norway. If these and the previous evaluations are in order and the visa fee is received, the applicant may be granted a visa. A positive decision will be registered in the VIS database, and a visa with an image on the label is issued. This visa applies for a maximum period of three months. All alphanumeric data regarding visas handled in NORVIS are stored both in the NORVIS database DUF and in the VIS. Whether or not the applicants' biometric data will be stored in both databases is yet not decided.

4.3 Threats to the security of visas

A visa have many threats that must be accommodated for [66] at the process of manufacturing the material used in the visa, at the process of applying the portrait, signature and biometric data to the document, and at the security printing. There are several threats that could arise, among others:

- Substitution of the photograph
- Alteration of text both in the machine readable zone and in the visual zone
- Removal or substitution of entire pages or visas
- Theft of a blank document
- Counterfeiting a complete travel document
- Impostors

The latter will be further studied in the experiment where we will see whether or not an impostor approved by the face recognition software is revealed by a human supervisor. To prevent or reduce the chance of the above threats to occur, the *Doc 9303* [66] states several basic and additional features to follow from the process of manufacturing the visas to the inspection process at the border control. A further review of these countermeasures is provided in the *ICAO Doc 9303, Part 2 October 30, 2002* [66].

5 Circumvention of face recognition products

> *A practical biometric system should meet the specified recognition accuracy, speed, and resource requirements, be harmless to the users, be accepted by the intended population, and be sufficiently robust to various fraudulent methods and attacks to the system.*
>
> — A. K. Jain et al. 2004 [15].

In the book *Biometrics: Personal Identification in Network Society* [67] there is stated a claim that any human physiological or behavioral characteristic could be a biometric provided it has the characteristics; universability, uniqueness, performanence, collectabillity, performance, acceptability and circumvention. In the case of information security the uniqueness and how easy it is to fool the system may perhaps be the most significant characteristic of a biometric. These characteristics are defined as follows:

- Uniqueness – no two persons should have the same characteristics
- Circumvention – how easy is it to fool the system by fraudulent techniques

The uniqueness of the biometric feature is crucial for the authentication system to work properly. To identify a person, the biometric feature used must be unique for the person to be identified. If the uniqueness of the feature is low, the system would return the identity of too many individuals when used in a larger system. The supervisor will then encounter problems trying to find the right individual. This would result in a high false acceptance rate, and time-consuming problems of finding the right person from the returned result in a supervised authentication system. An automatic system based on a biometric feature with low uniqueness would result in an insecure system that would allow several unauthorized personnel to enter. Kosmerlj [11] provides in her masters' thesis a good indication of the uniqueness of face recognition products. Her thesis shows that the uniqueness between people's facial characteristics from computer-based face recognition system at that time was too low for use in an application involving hundreds of thousands or millions of face templates.

This section will give a contribution to the second characteristic listed above; circumvention. How easy a system is to circumvent, says much about the value of the system. An authentication system that can easily be fooled has little value, even if all other characteristics of the system are good. Methods to see how robust a face recognition system is towards circumvention, can be divided into two main categories; *methods to gain false rejections* – how one can fool the system so that you will not be recognized – and *methods to gain false acceptance* – how one can fool the system to believe one is someone that one is not. This thesis will survey the latter method: *methods to gain false acceptance*.

5.1 Methods to gain false acceptance

5.1.1 Photographs in front of camera

The early face recognition systems were fooled simply placing an image of a legitimate user in front of the camera. However, improvements and inclusion of liveliness detection has defeated this weakness [68]. A LCD screen could instead be placed in front of the camera with a video clip of the legitimate user, as illustrated in figure 4. This method has fooled some face recognition systems. A disturbing fact is that 80% of all cyber-crimes is committed by insiders [17]. These are people with great opportunities to take video clip of colleagues, clips that may be used to gain illegitimate access.

Figure 4: This illustrates how the PC may be placed in front of the web-camera, attempting to fool the face recognition based authentication mechanism.

One approach to gain false acceptance that has been tested, is the use of life-sized photograph using a flat color photograph taken with a 35mm film camera and blown up to life size. In [69], this approach did not work. However, when they used a 3D mask made from photographs, they were able to circumvent the systems using default levels. But when the systems where set at a sensitivity level of 90%, they were not able to fool the system. The system FaceGuardian from Keyware's Biometric NT Logon, gave however somewhat better results at a lower sensitivity level (80%). High sensitivity levels did however result in a false rejection of enrolled users that had a bad hair day, and of people with swollen faces due to extensive dental work. This last point rises a new

potential way of disable the face recognition systems ability to recognize people that do not want to be recognized, reviewed later.

The scheme above could also be made using a digital camera to take pictures of the individual and place these on the notebook (PC). The notebook could then be placed in an appropriate distance from the camera, showing the image of the legitimate user [70]. However, Cognitec has in their FaceVACS included Live-Check, making every attempt with still images useless. On the downside the user-friendliness drops considerably, since the users seldom is recognized right away [70]. An important note in this context is that schemes of this kind would not work in environments that combine face recognition systems with human supervision.

5.1.2 Identical twins

One weakness, that is easy to exploit and difficult to counter when only using face recognition, is distinction between identical twins. In [69] identical twins where used to circumvent the face recognition systems. This resulted in a successful identification of the wrong twin every time, even when the applications where set to maximum sensitivity.

Aurora [40] claims according to [12] that their 3D software is able to distinguish between identical twins. However, these claims are not substantiated by independent tests that also examine the systems ability to at the same time have an acceptably low false rejection rate. Also this systems ability to recognize faces over time should be tested. The Aurora systems camera uses a near infrared light to put a virtual mesh on the face. This is done 16 times and these images are merged into one template and the measurements of the features are then calculated.

Problems with distinction between identical twins being falsely identified may be reduced or eliminated using a second security measure like fingerprints – a biometric that differs even for identical twins [71] – or thermal facial scanning. Fingerprint is however not totally robust towards attacks such as production of artificial fingers. Artificial fingers that circumvent modern fingerprint technology has been made [21], and further development on fingerprint systems must be performed.

5.1.3 Replay and alteration of templates

Some face recognition systems are such that it is possible to play back data collected with the aid of for example a sniffer-program listening to the USB port. This enables the attacker to bypass the face recognition system. One such sniffer-program is the *USB snoop for windows* [72] which inserts itself between the driver of the USB adapter and the device driver. The captured data is written to a separate log file, which then may be analyzed. These filter drivers are quite easy to detect, and need administrator rights to be installed under Windows 2000 and Windows XP. However other tools like the USB Agent by Hitex [73], eavesdrop the USB cable directly and are virtually invisible. A USB agent attached on to the cable records all transmitted data and transfer it to a foreign PC. An attacker can then, with the aid off the software that goes with device, analyze the protocols used by the target PC, and filter out the data packages that is relevant. After

analysis it is then possible to generate the data required to log in to the system.

In [70] the ID Mouse by Siemens was compromised with the aid of USB data packets and a few lines of Perl script. The data obtained were then used to reconstruct the image of a fingerprint. All that is required to replay the data gathered by eavesdropping, is a micro controller with USB support and some storage capacity. Together these actions constitute a device capable of impersonating the previously removed biometric scanner on the target PC. The firmware required to do so is fairly easy to program. Upon configuration requests, the device needs to respond with answers identical to those of the actual scanner and then at the right moment play back the stored biometric data. The use of challenge-response procedures [74, 75], in the course of which the biometric scanner and the application mutually authenticate one another and thereafter communicate with one another exclusively in an encrypted fashion, foil this kind of attacks. Sensors used for scanning is assumed to have enough intelligence to respond to such challenges [74].

Another method to circumvention is bribery or blackmail of administrators that has control of the databases containing the facial template. These individuals have the opportunity and necessary rights to alter templates to match an attacker so that he may bypass the system as a legitimate subject.

In some cases Denial of Service attacks on biometric systems is also a goal of an adversary. Either to be falsely rejected a user during screening or for some other reason. The system is then overloaded with requests, making it unable to handle legitimate operations. Alteration of the database, making some users template differ from the original, is also regarded as a Denial of Service against the person associated with the corrupted template.

5.1.4 Swamping attack

A swamping attack [68] is similar to brute force attacks. You exploit weakness in the algorithm to obtain a match for the wrong data. For a fingerprint the attacker present a print with hundreds of minutiae in hope that at least the threshold number of them will match the stored template.

5.1.5 Piggy-back attack

Piggy-back attacks [68] are performed in that an unauthorized user gain access by slipping past the security check at the same time as an authorized user passes. For example the attacker may go through a door as an authorized user opens the door with his biometric data. This kind of attack could also involve threats, or following close to a car in front of you through the toll.

5.1.6 Illegitimate enrollment

Illegitimate enrollment [68] is a security threat that concerns all security systems. If an individual present his biometric data, and is able to enroll these with a false identity, this individual will in the future be able to have full access although he actually should

not have. This would also be a problem in a border control environment. If the person concerned is not identified as someone else, and there are no other way to check the information the subject have provided, he or she will be able to pass themselves off as someone they are not. Fraud of this kind is hard to detect, and when the attacker first is enrolled the chance of discovering this at a later time is limited.

According to Schneier, Hollywood can make people's faces look like others, but this requires special skills and is expensive [76]. Imagine a remote system using face recognition for authentication. The users take his picture and mail it in for authentication. This is easy to fool. An adversary can take the users picture and send it in with his own email, or using a email program were you can choose the addresser (such programs are easy to make – I have made one such program in Java using only a few hours on implementing it). The adversary will then be falsely authenticated.

5.1.7 Coercive attack

A coercive attack [68], where the right biometric data are represented by a illegitimate user, can be used to fool a sensor. The most obvious way of doing this is by forcing the legitimate user to authenticate to the system. There are several ways of doing this; bribery, blackmail, use of physical force and cutting off the biometric body-part to mention some [25]. The last of these have been, and still are, a concern for users of biometric authentication. Because of this, there has been done much to prevent these kinds of attacks by using liveliness detection. In face recognition one such method is to detect changes in facial expression. For example, the person could be challenged to make a particular expression to distinguish a real person from a prerecorded image [77]. Also, a measure of the stress level, a human guard and video surveillance are methods to prevent these kinds of attacks. A control should also in some cases be made towards the operators of such systems to prevent corruption and similar threats. If the person controlling the system is bribed, the attacker does not even have to resemble a legitimate user from a human perspective to pass the system. There could also be that the administrator replaces the face of a legitimate user with the highest bidder, or insert an extra user in the system. Actions to prevent this should be taken.

5.1.8 Trojan horse

Also a Trojan horse [68] could be placed in the system, making the owner able to do approximately what she wants. This could include everything from accepting everyone attempting to get authenticated within a time frame, to denying all users. However, a smart Trojan horse would limit the activity to activities that does not make the system behave in a manner that is susceptible. Rahta et al. call attention to usage of a Trojan horse against biometrics. If the feature extractor is attacked by a Trojan horse, this could be used to replaces feature sets [74] to be stored in the system, with feature sets preselected by the intruder.

5.2 Security threats to facial recognition and countermeasures

In [32] several countermeasures are suggested to reduce the weaknesses of facial recognition. We have in this section briefly discussed those of these suggestions that we find most important.

Issues like lightning and other environmental conditions should be taken in consideration. If these issues cannot be controlled, technology that can accommodate this should be chosen. Facial expression is another issue in this context that in some cases may be controlled, but in other range of application it may not. If there is no way of controlling the facial expressions of the users, a technology that can handle different expressions in a satisfactory manor should be preferred. Local feature analysis is such a technology. Most facial expression cause facial deformation in the lower part of the face [55], leaving sufficient invariant information in the upper part of the face for recognition. However a facial expression like a scream is still a problem to face recognition, because it affects both the upper and lower part of the face.

FRVT 2002 [18] showed that aging and sex had an impact on the performance of face recognition systems in that aging lead to an increased false rejection rate. The differences in performance for the sexes decreased as the age increased. Possible solutions to this are frequently re-enrolling the users, and the use of technology that can accommodate this (for example, a neural network).

Facial recognition is not considered to be as secure as other biometric technologies [20]. To accommodate this, technologies that are robust, that is known to distinguish between legitimate users and a fake, and that involve liveliness detection should be preferred. Combining facial recognition with other biometrics, passwords or PIN's, or with human supervision provides better security.

5.2.1 Impersonation attack

Impersonation attack [32], is when an unauthorized user alters his face in such a way that it resembles an authorized user, or alter his face in such ways that he will not be identified in a screening situation. The last is simpler by using plastic surgery, masks or makeup and disguises.

Occlusion

Studies show that occlusion of the eyes affects the performance of face recognition systems more than occlusion of the mouth [78, 79]. It is believed that the eyes carry the most discriminant information of a persons face, making this result less surprising. Covering the eyes with sunglasses then would be an efficient method to avoid or at least reduce the chance of being recognized by the face recognition systems. This would work in automatic systems like surveillance cameras in airports and stadiums like the one used at the Super Bowls [80], and to fool street surveillance like the one used by the Tampa Police Department in Florida [81]. If in addition the subject wear a scarf the chance of being recognized is insignificant [78]. Regular glasses however do not lead to the same problem in recognition. As mention earlier face recognition algorithms available today

are able to handle regular glasses [65]. However, if the spectacle frame cover a large portion of the face, recognition will be harder to accomplish. Other form of occlusions is the use of nylon stockings over the head, foam pad or cotton in the mouth, or growing a mustache or beard.

Orthodontic treatment

Swollen faces due to extensive dental work, could make it hard for a face recognition system to recognize a subject [69]. Taister et al. [82] also states that orthodontic treatment in some case may alter the face. Among others, removal of the wisdom teeth will cause the face to appear narrower. Removing premolars from mandible and pulling the remaining lower teeth together by means of orthodontics can markedly alter the profile of the face. Some treatments like this may alter the face to a degree that makes it more difficult to recognize a face. Also lack of necessary treatment may alter the face if this leads to swelling.

Facial make-up and eyebrows

An experiment performed by Sadr et al. [60], using face-images of celebrities where the eyes or eyebrows were removed as reproduced in figure 5, showed that the eyebrows had great impact on the human ability to recognize faces. In fact the eyebrows were more important in the recognition process than the eyes. In automatic face recognition systems, the eyebrows may sometimes be less important in the recognition process than other features such as the eyes and mouth. If an attacker cooperates with a legitimate user that resemble from a computers perspective, he or she could alter their eyebrows in such ways that their eyebrows look alike. Making the eyebrows prominent could affect a human's ability to recognize the two persons as the same person. Combining this with other makeup and alterations of the hair could be enough to fool the guard. How face recognition systems react to facial makeup and the role of the eyebrows is a subject that should be further studied. Sadr et al. [60] claims that a study in that direction would contribute to surveying which parts of the face has more importance in facial recognition than others, and would better show the degree of difficulty of fooling face recognition systems. Such studies will also better survey the impact of facial surgery, since face lifts specially targets the appearance of the eyebrows.

Countermeasures

By using multiple biometrics the chances for successful impersonation attacks will be reduced. For example by combining fingerprints and face recognition, the system could even distinguish between most identical twins and people with similar fingerprints, since both their prints and face must be similar to fool the system [68]. When the biometric feature, in this case your face, first is compromised it is no way of altering your face so it again can be used for authentication purposes, except from facial surgery, which is a non-option. Ratha et al. propose a scheme they call *cancelable biometrics* to counter this problem. This consists of an intentional, repeatable distortion of the biometric signal based on a chosen transform. The face is distorted in the same fashion for enrollment and each authentication. This way when one variant of the transformed face templates is compromised, the transform function can be altered to produce a different template for the given user. In general this transforms are made to be non invertible, so that even

Figure 5: This illustration shows the face of Demi Moore with all features intact, the eyebrows removed, and the last image is with her eyes removed. A similar image-pair was used by Sadr et al. [60], and showed that the eyebrows had great impact on the human ability to recognize faces.

if the transformed template is captured, the original face can not be recovered. Although this scheme works fine when the image is captured between the sensor and the system, it does not prevent attacks performed on the sensor with the original image such as using a LCD screen with images of an authorized user in front of the sensor. Other security measures should be taken to prevent these kinds of attacks, such as combining the face recognition with a security guard or liveliness detection.

5.2.2 Attacks between the sensor and the biometric system

A face recognition system may also be attacked between the sensor and the biometric system [68, 25, 77]. If a physical access to the connection between the sensor and system is available, it is, in theory, easier to attack the connection between the sensor and the system than attacking the sensor. A template could be replayed with different user credentials, making the user able to authenticate to the system as two different individuals. This is an attack that could be realistically used related to social security fraud. Also, a spoofed signal can be acquired and later replayed to gain access. For example by listening to the connection between the camera and the evaluation system [77] as illustrated in figure 6. This will however result in an identical biometric sample as the one received earlier, something that is rare working with biometrics. If the system tests for such situations, the attack will be detected. Also the use of digital encryption combined with time-stamping would prevent such attacks. It is not only the input to the system that could be overwritten or replaced, but also the output returned to the system. If a system is combined with a human guard that let the user pass if the system returns a match, it could be altered to return a high match score for one or several particular users.

Ratha et al. [74, 83] proposes a method for transactions of fingerprints that are secure against replay attacks. Although their method is made for fingerprint recognition, they state that it also could be extended to other biometrics like face recognition. To achieve a secure transaction, they propose that the service provider issues a different verification string for each transaction, and mix this with the fingerprint image. The string is placed in different places based on the structure of the image itself, making the scheme less vulnerable to attacks than if it was placed at a fixed location for all images. At the service

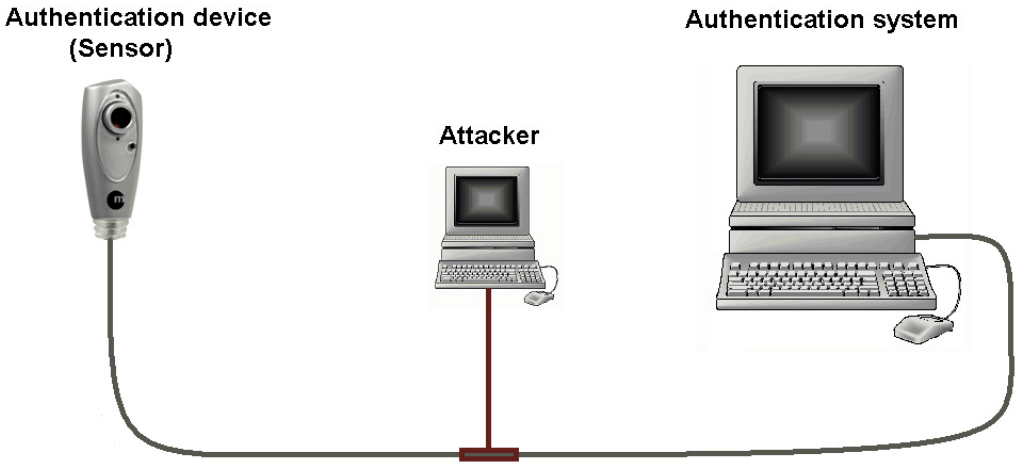

Figure 6: An attack can be executed between the sensor and the biometric system. The attacker – illustrated with a red connection – may then catch a packed sent by the biometric capture device (the sensor), and resend this packed at a later time to falsely authenticate her self to the authentication system. The attacker must then have access to the connection between the sensor and the authentication system.

provider, this image is decompressed and the checked for the right one-time verification string. The messages hidden in the images have minimal impact on the appearance of the decompressed image. In [83], they also propose a challenge/response method for countering replay attacks. After the initiation of the transaction, a transaction server generates a pseudo-random challenge (i.e. 3, 10, 50) for the current transaction that this send to the sensor. The sensor then computes a response to the challenge, i.e. to select the 3rd, 10th and 50th pixel and compute the response from them, and return a response such as '113, 25, 192'. The sensor can then check that the client in addition to knowing the correct response function, also was using the image received by the server.

5.2.3 Back doors

Recognition software could also contain back doors made for easy update and easy access of a super-user, who can override the system in special cases like when the users does not contain the necessary biometric feature such as fingerprint. Such loopholes should be avoided. If information of such loopholes leaks, the whole system would be compromised. The problem is that you never know whether an adversary have this knowledge or not. This emphasizes the importance of avoiding back doors. The face recognition decision could be overridden if the the system involves back doors, or otherwise is controlled by an attacker. In such cases the system becomes useless even if the recognition algorithm has excellent performance [74].

5.2.4 Hill climbing attack

A brute force attack in face recognition is when you try producing every possible variation of templates, and try to authenticate to the systems with these until you find one template that gets above the threshold. The possibility for a successful brute force at-

tack on face recognition systems is relatively limited, and requires a tremendous computation capacity. A more realistic approach is instead the use of a Hill climbing attack [68, 84, 85, 86]. In a hill climbing attack, you repeatedly send biometric data to a system with small changes and record the score returned by the system. Then small changes is made to the data that is sent to the system, and the data that returns the best scores is kept and further altered and sent towards the system. Eventually you get high enough score to get above the threshold, and you are in. This method is suitable when you have no knowledge of a legitimate user in advance. By not allowing multiple attempts, having a maximum limit, or by quantifying the returned score in such ways that small changes goes undetected [87], such attacks can be prevented. However the latter have been circumvented [88]. The system could also return *approved* or *denied*, also preventing such attacks. A similar approach is described by A. Adler [89]. In his approach an initial estimate of the targets face is chosen. This image does not need to resemble the target image. At each iteration a small alteration is made and the modification that has the greatest improvement on the match score is kept. The iterations are performed until no significant improvement in match score is made. His experiment showed a maximum match score that corresponded to a confidence above 0.999, indicating that the calculated estimate would be accepted as an image of the genuine target person. The results from the test also indicate differences in how face recognition systems recognizes faces. Among others did alterations in the hair affect the performance, while the lower face shape and beard was not substantially altered, likely because that information is not encoded in the template. Adler also point out the security concern this may have related to the biometric visas'. This approach is immune to template encryption, indicating that biometric templates and biometric match scores should be made unavailable to un-trusted parties. Adler states in his article that traditional claims that templates can not be recreated is false, and he points out the importance of securing stored biometric data and not returning more information from the system than necessary.

5.2.5 Liveliness detection in face recognition systems

As we have seen, the use of liveliness checks could be used to prevent several of the reviewed attacks. Liveliness detection in biometric systems is still a concern for the users of biometrics systems, and is therefore an issue of great importance. Hopefully in a near future this technology will mature further, so that users no longer will have to worry about being separated from their biometric feature by an attacker. In this chapter we will provide a brief introduction to this aspect of face recognition security. Patrick G. McLean proposes in [90] a solution where the face recognition system is combined with a weight sensor, that measures the individual that is authenticating himself. In applications where the users use the system frequently, this solution might be adequate. However in systems where the users are authenticated less frequently, perhaps months apart, these users weight might vary considerably. This makes this kind of liveliness check insufficient. Another concern with this approach is that an attacker could use a weight belt to accommodate for any potential variance in weight from the individual of whom the face really belongs. People could also be threatened to come with an attacker to the sight so that he or she gains access to the facility. A combination of face recognition and a weight sensor could at least prevent the attacker from being in the immediate proximity of the

victim, making him able to escape, but it would probably not prevent this kind of attack. A method that could prevent this kind of attack however is combining the face recognition system with a human guard.

Another method for liveliness detection, proposed by Klosterman et. al. [77], is to challenge the user trying to authenticate himself to make a particular expression to distinguish a real person from a prerecorded image [77]. Also, a measure of the stress level, or video surveillance are methods that could be used for liveliness detection.

5.3 A successful attack – definition

We will in this section provide a definition on what may be considered a successful attack on a face recognition product in a border control environment.

> *A successful attack on a face recognition product in a border control environment, is when a user with criminal intent, different from the individual with a legitimate visa, is accepted by both the recognition system and the authorities in the border control as the legitimate visa applicant, or when an applicant, stored in a watchlist or as an existing and rejected visa applicant, successfully circumvent the system to believe that he or she is not present in either.*

6 Experiment description

6.1 Introduction

This chapter provides a description of the experiment that will survey the differences between human and machine-based methods ability to recognize faces and how hair and the presence of hair affect this process.

Kalocsai et al. [53] have in their experiment examined similarities between human and system based face recognition performance. They used two images of each person, one with an angry face expression and one with a neutral expression. An oval area was used around the face to eliminate influence of background and hair on recognition performance. A test-panel of 64 candidates were told to ignore differences in facial expression and to press a button deciding if the images were of the same person or not. The response time and failure rate from the observations were recorded, and the stimuli-frequency provided to the test-panel was as follows:

- An image (not face image) was presented for 500 msec
- The first face image was presented for 150 msec
- A mask was presented for 500 msec
- The second face image was presented for 150 msec. This was the image that the test-panel should decide if was of the same person as the first face image
- Then a second mask was presented for 500 msec

In a real world border control environment, a security guard would probably look at the image for a longer period than in this experiment. Another point that should be noted is that the same images were used more than once, something that could affect the decisions made by the participants, as they may have remembered that they had seen the image before, making the mistake that it was the same as the first face image. Further, this experiment was conducted in 1998, and the technology has evolved and matured since then. This experiment did not compare the results from the experiments with similar experiments where hair was part of the recognition elements. In addition this experiment examined the human ability to remember whether or not the face image were of the same person. In our experiment on the other hand the participants will be provided two images that they can see at the same time. In [52] however, Bruce et al. compared face similarity with and without hair, performed by both humans and machine. They found that there were correlations between human and machine recognition, but that the correlations differed from the Graph matching algorithm and the PCA algorithm. PCA gave a much higher correlation with the samples containing hair, while Graph matching had correlations to human ratings both with and without hair. Their experiments were however on whether or not the faces had similarities, and the participants were told to sort the images in piles of similar images. They were however not instructed to evaluate whether or not these images were of the same individual. Kos-

merlj [11] found in her MS'c that using face recognition for identification resulted in a high false acceptances rate. By looking at the images that resulted in false acceptances, she was unable to find any similarities between these persons and the subject. Although the hair on these image-pairs was removed when evaluated by the computer-based face recognition system, her human evaluation of the image-pairs was based on images that included the complete face including hair. The evaluation would subsequently also include hair as one of the elements affecting the recognition.

Hair is a feature that may be manipulated without affecting computer based face recognition, because this parameter is excluded from the recognition process. This makes the hair a favourable feature to manipulate when performing a non-zero-effort attack to gain false acceptance.

Our experiment is based on the results obtained by Kosmerlj, and we use the face images that resulted in a high false acceptance in her experiment and the corresponding images that gave an incorrect match when authenticated by an automatic system. This way we can see whether the humans face recognition process of these individuals correlate to the computer-based recognition process or not. Alteration of hair is publicly accepted, and will probably be a more discreet way of altering the appearance than facial surgery, mask and so on. Altering your hair is also an easy and cheap method to alter your appearance, and will not affect the system based face recognition. If an attacker is able to find someone he or she resembles, this person may alter his hair style, color etc. to amplify the similarities between her and the target person. If hair is an important part of the recognition process, this non-zero-effort attack might be enough to circumvent the human guard. This calls for a study to establish whether hair influences the human face recognition process or not, and if it does to what degree.

6.2 Procedure

The image set containing the faces used in this experiment is the same set of images that resulted in high false acceptance rates in Kosmerlj's MS'c thesis [11], consisting of a set of persons with a corresponding set of pictures of persons regarded as similar.

A control group of 61 persons were divided into two groups. The division was made simple by having every other participant evaluate images of faces with hair, while the other evaluated faces where the hair was removed. Half of each group was presented the images in reverse order to eliminate variance due to difficult images instead of variance due to mental weariness:

- Group 1 consisted of 31 participants that were presented with image-pairs where an oval was used to remove the hair and background from the pictures.
- Group 2 consisted of 30 participants that were presented image-pairs where the depicted persons' hair was visible.

The groups were informed about the fact that:

- The image-pair are by the computer evaluated to be of the same person. But that this

does not necessary mean that they actually are the same person.

- The images taken of the same individual was taken at different times, and subsequently their hair may differ in length, shape and color. Also the background, glasses and clothes may differ.
- The image-pairs are randomly selected and composed, and the number of same-pairs and different-pairs are not known at the time of the experiment.
- The participant was then informed about the time-limit of 10 seconds, and that a new dialog box would appear should this time be exceeded.

In the experiment the participants were presented with 60 image-pairs, where 15 was composed by two images of the same individual taken at different times and 45 image-pairs was composed by one image of one individual and the other image consisting of an image of his or her look-alike. An image could appear more than once, but the same composed image-pair will only appear once. Although if a face image should appear more than once, this should not affect the performance because the participant were given the two images to compare at the same time, and subsequently memory of an earlier appearance of the image should not affect the decision.

The same face-pairs were given to each participant in all groups, but in different order. This way the results may be compared to see if there are a correlation between the results, and whether or not the similarities found are due to pure coincidence. To prevent influence from other participants, each participant was given the same images, but in a different order regarding both image-pairs. Each participant had to mark the image-pairs as either being of the same individual or of someone else. The time used on each comparison, the result, the age, gender and educational background of the participant were recorded. Also whether or not the participant have been involved in identification of faces from images earlier was recorded to see whether or not this influences the recognition.

To present these images to the participants, store the data obtained and use this data, two applications have been developed in Java. The first application is a client application that presents the images to the participant (figure 7, 8, 10), and stores the information provided. In the application each participant is given a maximum time of 10 seconds to evaluate the images, after which the images is made invisible. A new dialog will then appear, which guides the participant to make a decision of whether they were of the same individual or not (figure 9).

The second is an administration application (figure 11) that presents results from the experiment in a semicolon separated file that could be directly imported in analysis tools (figure 12). This application has several choices for generation of semicolon separates files that can be used for several purposes.

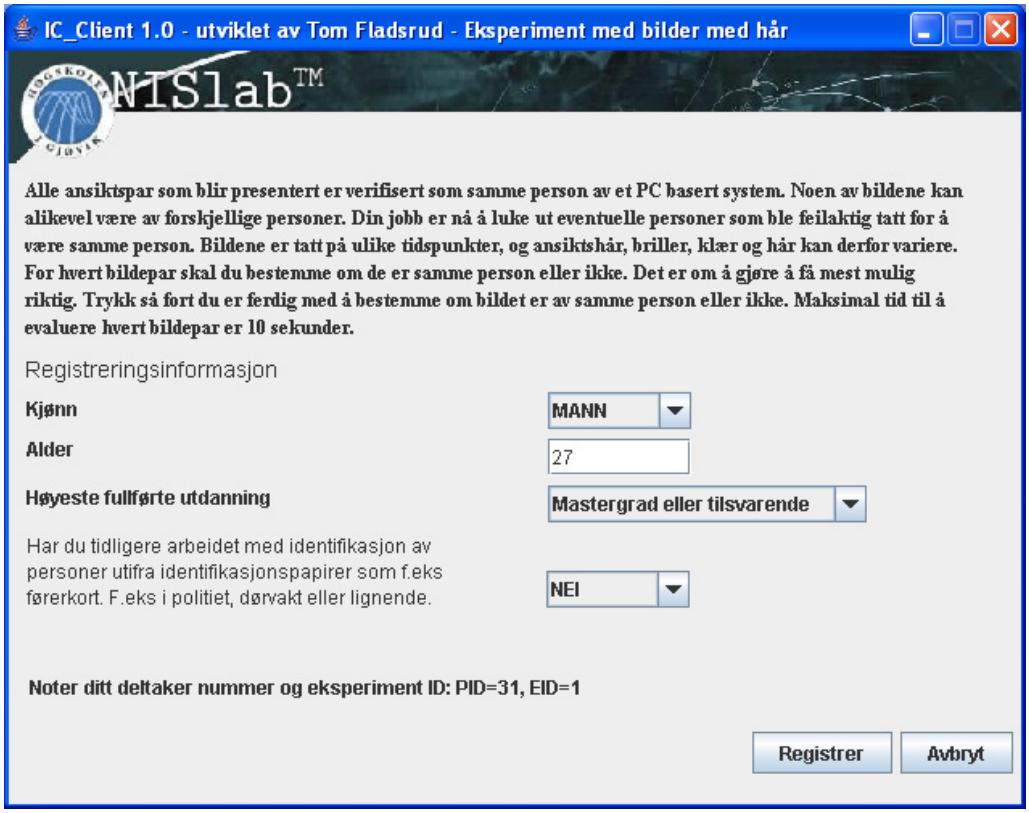

Figure 7: Registration of information about the participants in the experiment. The information stored includes gender, age, highest completed educational degree and whether or not the participants have had work related to check of ID towards faces.

6.3 Purpose of the experiment

If a substantial amount of impostors is able to circumvent both the face recognition system and the human border control guard by manipulating appearance, thereby being falsely identified as someone they are not, the use of face recognition systems in such settings should be revised. One the other hand, if the impostors that is able to manipulate the system into believing that they are whom they claim to be, and this in most cases is revealed by the human supervisor, the practical implications of the problem would be much smaller.

The purpose of this experiment is:

- To find out to what extent individuals that are able to bypass computer based recognition systems as legitimate users although they are not, also are able to pass a human guard. This may produce valuable contributions to research questions 1 and 2.
- To study the impact of alterations of a person's hair and the effect this has on a person's ability to bypass a face recognition system. As above, this may also contribute to answering research questions 1 and 2.
- To study whether human face recognition results in false acceptance of the same individuals as the computer based face recognition systems or not. If this is not the

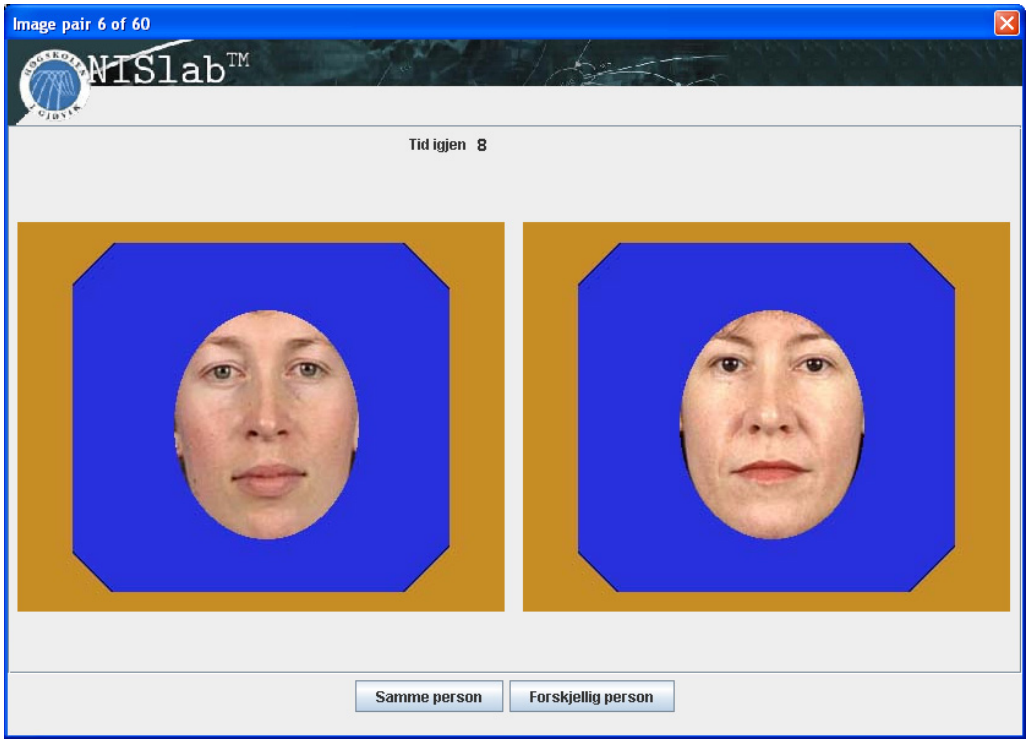

Figure 8: Presentation of face images during the experiment where the hair and background are excluded. Two face-images are presented at the time. The participant shall evaluate whether are of the same person or not when presented with face images where hair and background are excluded. The participants are instructed to press the button at the left if they think the two images are of the same individuals, while they are told to press the button to the right if they think the face images are of different persons. In the upper part of the window the participants can see how much time they have left on each image-pair.

case, then what differences are there?

- The experiment will contribute to an increased understanding of differences in human and computer-based face recognition. And will be a basis for evaluating to what extent a human guard will increase the security when combined with a face recognition system in a general application.

6.4 What data is possible to obtain from such an experiment?

- The difference between False Acceptance Rate obtained from the test of human recognition with hair as a parameter, compared to the test conducted on human recognition with images where hair was not a parameter.
- The difference between False Acceptance Rate obtained from computer-based face recognition systems and human face recognition.
- To find out whether people using very limited time to evaluate each image-pair have increased or lower failure rate than those spending more time on evaluation.
- To find out whether or not the evaluators' educational degree affects the recognition performance.

Figure 9: When the participant have not decided whether the two face-images are of the same person within the given time space, the a new dialog appears where the participant is instructed to make his or her decision.

- To find out whether or not the evaluators age affects the recognition performance.
- To find out to what extent, if any, the recognition rate for a human guard drops or increases after a given amount of comparisons and time.
- Whether or not stress due to a limited available time affects the false recognition rates. The limited time frame for evaluating the similarity will provide a stress factor that makes the experiment more similar to that of a border control environment.
- A contribution in whether one gender gives higher false acceptance rates from human face recognition by recording the gender of the faces in the images.
- The data obtained should give an indication of how much the presence of a human guard helps reducing FAR when used in addition to a automatic face recognition system in i.e. a border control.

6.5 Face image databases and algorithms

The human face recognition experiment with and without hair is based on the data obtained from the best performing face recognition system used in the masters thesis of Kosmerlj [11]. She used in her thesis the CSU Face Identification Evaluation System 5.0 [91] to generate similarity scores. These similarities were stored in SFI files that had to be

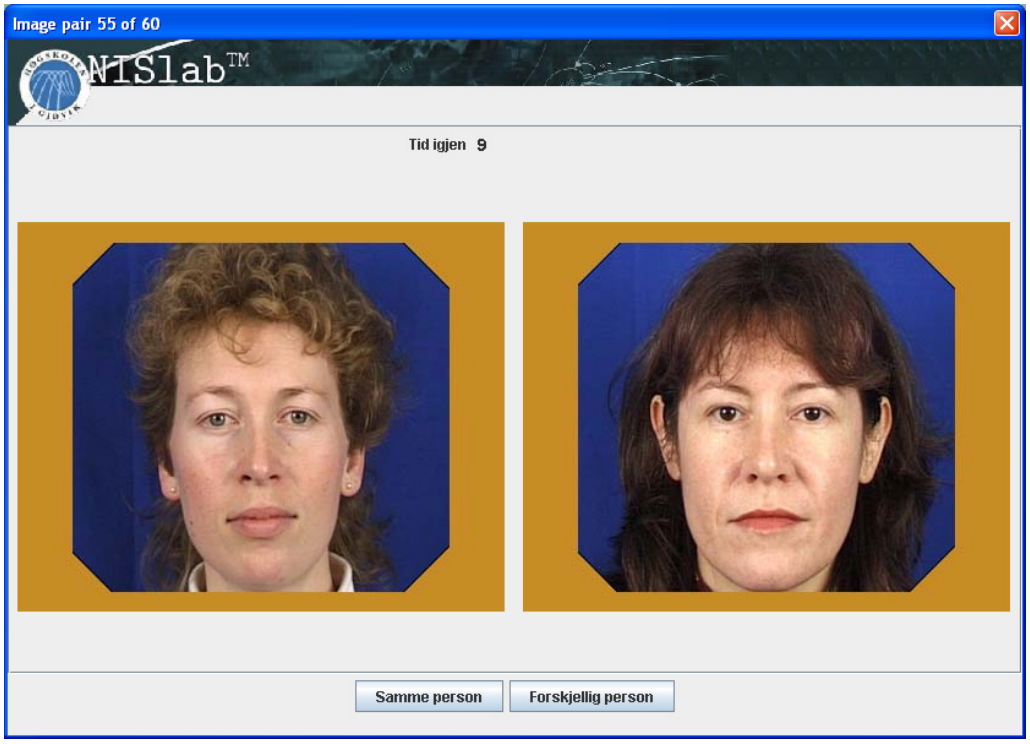

Figure 10: Presentation of face images during the experiment where the hair and background are included. Two face-images are presented at the time. The participant shall evaluate whether are of the same person or not when presented with face images where hair and background are included. The participants are instructed to press the button at the left if they think the two images are of the same individuals, while they are told to press the button to the right if they think the face images are of different persons. In the upper part of the window the participants can see how much time they have left on each image-pair.

evaluated in order to identify the face images that provided a similarity within the given threshold. The data obtained from Kosmerlj's thesis involved searching and comparing hundreds of SFI files, something that would be too time-consuming if done manually. Because of this I developed an application I have called SFI Analyzer 1.0 to analyze and return the files that is within the threshold. This application is reviewed in appendix B.

The images used in this experiment are a selection of the face images that resulted in false acceptances, using a false acceptance rate of 0.1% (threshold -0.42), and a corresponding false rejection rate of 33%, with the face recognition algorithm that gave the best results in Kosmerlj's thesis, the PCA_MahCosine. The selected images was part of the following face databases:

- Ljubljani CVL Face Database [3]: This database contains face images of 114 subjects, 7 images per subject, resolution 640x480, color images, uniform background and lighting, frontal and side views and varying facial expressions.
- XM2VTS Database [92]: This database contains face images of 295 subjects, 4 recordings per subject over 4 months, color images, resolution 720x576, varying posture,

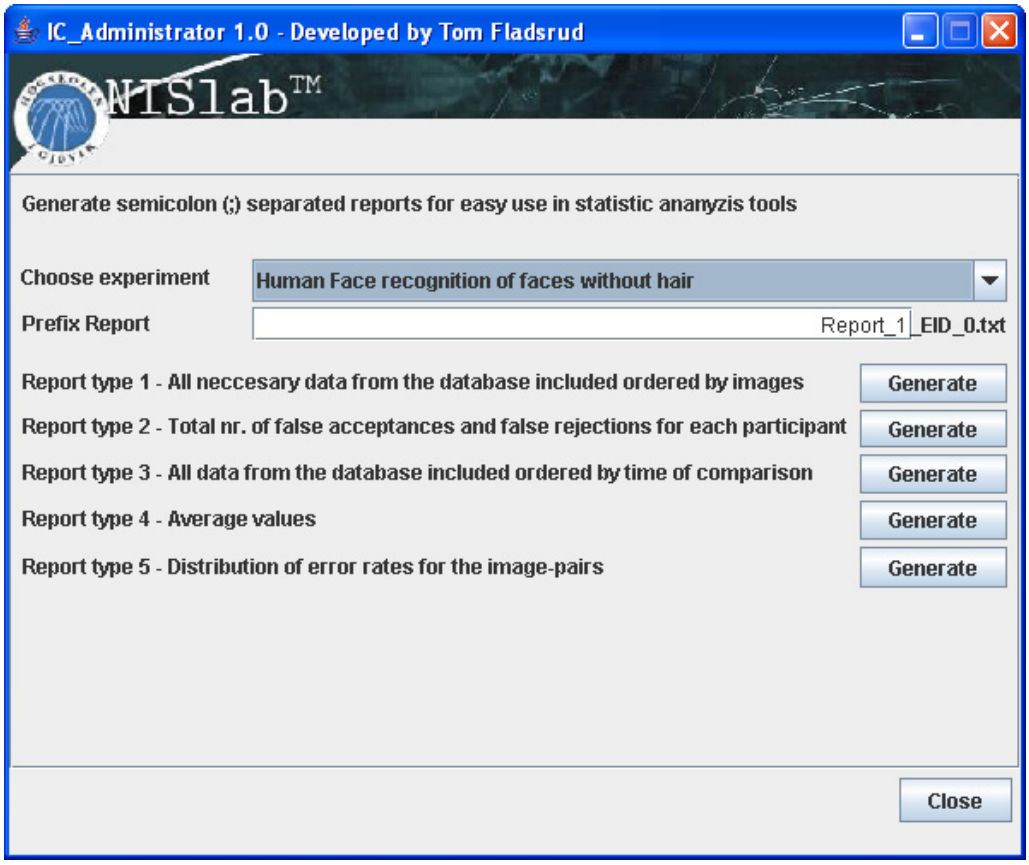

Figure 11: The administrator application provides reports with semicolon separated data to import in a Excel document or other analytical software like the SPSS.

varying illumination (controlled).

- AR Face Database [2]: This database contains face images of 126 subjects, 2 recordings per subject separated by two weeks time, 13 images per recording, color images, resolution 768x576, frontal view with different facial expressions, illumination and occlusions.

The images in AR Face database [2] are stored in RAW format. To use these images in the test application, referred to as IC_Client 1.0, they had to be converted into JPEG format. Because of the huge amount of images, and because I had no conversion tools available, I developed an application I have called RAW Image Converter 1.0 to do this automatically. This application is described in appendix B.

Also the face images in the XM2VTS Database were less than straight-forward to use. Each image here was zipped and stored in PPM format. As with the AR Face database I had to convert these into JPEG format. Further, each subject was represented with his or her own folder where images of them were stored. Because of this I developed an application that unzipped the images in all sub-folders of a given start folder, and convert all images found with PPM format and store this in a specified output folder. This

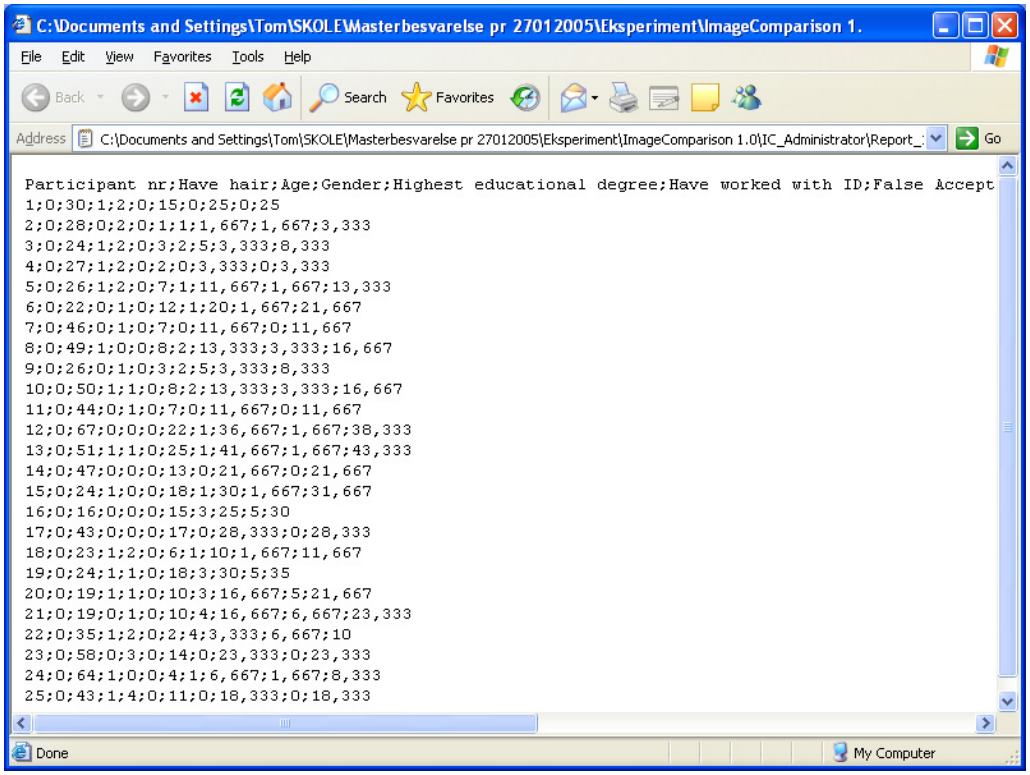

Figure 12: This figure illustrates one of the files that are generated by the IC_Administrator 1.0 module. This results can be directly imported in an statistical analysis software.

application, which I have called ImageConverter 1.0, is described in appendix B.

All images used in Kosmerlj's and subsequently our experiment consisted only of frontal and approximately frontal facial images without occlusions and with varying but controlled lighting conditions. These correlates to the use of facial images in VIS, where the lighting conditions will be controlled, and were the subjects will be instructed to face forward when the picture is captured. All face-images used in the experiment were encircled with a frame, and the images without hair were encircled with an oval that removed the hair and background (See figure 8 and 10).

6.6 Restrictions

In a border control environment there can be a tremendous amount of traffic passing through the checkpoint, and therefore a limited amount of time available for the control guard to evaluate the picture of each visa applicant. Because of this, to make the experiment more similar to a real situation, each participant will be given a stress-factor in form of a limited time to evaluate each image. Each participant will be given 10 seconds for evaluation of the image.

The thesis of Kosmerlj [11] also included a face database of passport applicants and the photograph database of students Gjøvik University College. However these face data-

bases did not include more than one image of each subject and our experiment is therefore based on the PCA_MahCosine evaluation of the faces from the face databases CVL-, XM2VTS- and the AR Face database.

7 Experiment results

The IC_Client 1.0 was used to present and store the data processed from the experiment into a database. These were then fetched by IC_Administrator 1.0 that produced semicolon separated files with a structured collection of the data. These files were then imported in the analysis tool SPSS (Statistical Package for the Social Sciences) for analysis. To establish whether or not hair have an impact on humans ability to recognize faces, we have tried to find whether or not there are a significant difference in the error rate when evaluating faces where the hair is visible and those where the hair is removed. Additional tables with results from the experiment are provided in appendix D.

7.1 Hair

The analysis reveals, as shown in table 1 and 2, that there are in fact a significant difference between the number of false acceptances when the hair is removed and when it is not. The results show an increased false acceptance rate for the images where the hair is removed compared to where it is not. As we can see in figure 13 the participants presented with image-pairs where the hair was removed provided an mean number of false acceptance of 9.61 which is significantly higher than for those who were presented with face images where the hair was present, which resulted in a mean false acceptance of 3.90. The variation in false acceptances between the two groups are further illustrated in figure 14, where we in addition can see the variations in number of false acceptances within each group. When looking at false rejections however there seem to be no significant difference in this error rate.

In figure 15 and 16 we can see the distribution of the false acceptances and false rejections in the two experiments. As we can observe from these figures there are only 3 image-pairs that have not been evaluated wrongly from one or more participants in the experiment where the hair was removed, while there were 18 image-pairs that was never evaluated wrongly from the image-pairs where the hair was present. We can also observe that most of the image-pairs have been falsely evaluated more than once when the hair was removed.

7.2 Other aspects

7.2.1 Gender

Another interesting factor that we wanted to look into regarding human face recognition performance, was whether or not the there are differences in the face recognition performance between the genders of the participants. As we can see from table 4 and 3, the difference in performance between the genders were however not significant.

Error type	Have hair	Number of Participants	Mean Number of Errors	Std. Deviation	Std. Error Mean
False Acceptances	No	31	9.61	6.168	1.108
	Yes	30	3.90	4.139	0.756
False Rejections	No	31	1.23	1.283	0.231
	Yes	30	0.80	1.031	0.188

Table 1: The table shows the mean false acceptances and false rejections with and without hair present on the image-pairs in the experiment. As we can observe from the table there are a significant difference between the number of false acceptances when evaluating image-pairs where the hair is present (3.90) and when it is not (9.61). Regarding the mean number of false rejections the differences between the evaluation results from the two participant groups are smaller.

Error type		F	Sig.
False Acceptances	Equal variances assumed	5.860	0.019
	Equal variances not assumed		
False Rejections	Equal variances assumed	2.457	0.122
	Equal variances not assumed		

Table 2: The Levene's test for equality and variance on False Acceptance and False Rejections on the two groups of image-pairs shows that there are a significant difference between the false acceptances for human evaluation of face with hair and those where the hair is removed. There are, as can be seen in table 1, about 3 times higher mean value for false acceptances when faces where the hair is removed compared to the mean value for false acceptances where the hair is present. Regarding false rejections however there seems to be no significant difference between the two groups.

7.2.2 Age

If we look at the importance of the age however, we could, by performing a Anova test as shown in table 5, see that the age affected the face recognition performance. However, this could possibly be due to imbalance in the age of the participants in the two experiments. Because of this we performed a T-test to see whether or not the difference was due to the composition of the two participant groups. The T-test, shown in table 6 and 7, shows that the influence of age is significant. It should be noted however that, as we can see in table 6, the mean age in the two groups vary from 38.58 to 31.87. Figure 17 shows how the number of false acceptances increase and decrease as the age interval changes.

7.2.3 Educational degree

Further, the effect educational degree have on human ability to recognize faces was also considered in this experiment. A cross tabular, as shown in figure 22, showed that there were a clear predominance of people with a PhD degree in the participant group tested with images including hair (11), compared to in the group with participants tested with images where the hair was removed (1). The results, when comparing educational degree with false rejection rate, shows, as we can see in table 9, a decrease in false acceptances as the educational degree increased with the exception of the one individual that had a PhD degree or higher. The Anova test provided in table 8 show that there were a signifi-

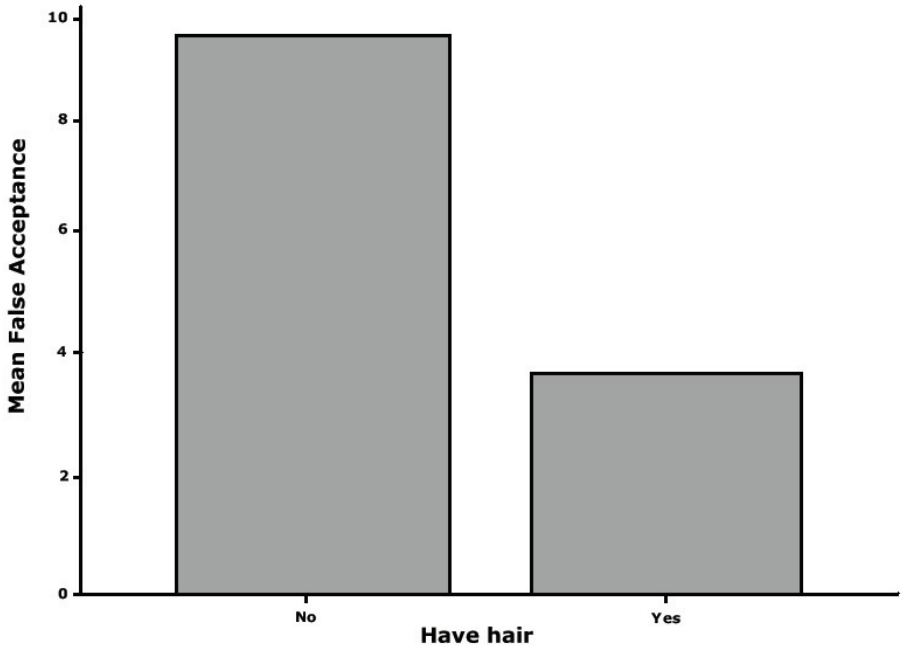

Figure 13: The histogram shows a graphical overview of the mean number of false acceptances of the two groups with and without hair. As we can see there is a clear difference between the false acceptances provided by the two groups. The group where the hair is removed has a number of false acceptances that are almost three times the amount of false acceptances in the group where hair is present.

cant difference in false acceptance between the groups of educational degree, while the difference between the false rejection rate were not significant.

7.2.4 Time

We also wanted to see whether or not the error rate increased, decreased or remained constant over time. To evaluate this every other participants were provided the image-pairs in reverse order to be able do distinguish between differences over time due to more difficult image-pairs at the beginning or the end. As we can see from table 10 the number of false acceptances remains approximately constant over time between the four time-frames. In table 11 however, we observe that the evaluation time decreases over time.

Another interesting factor regarding the time when evaluating faces, is whether or not the time spent on each comparisons affect the error rates, and in that case how the error rates are affected. By performing a correlation on the number of false acceptances and the total time spent on comparison on the data-sample that were sorted in by time of comparison, as shown in table 12, we could see that there were no correlation between number of false acceptances and total time used on comparison. By comparing the total time on comparison and number of total errors however we can see that high total time on comparison provides a high total error rate.

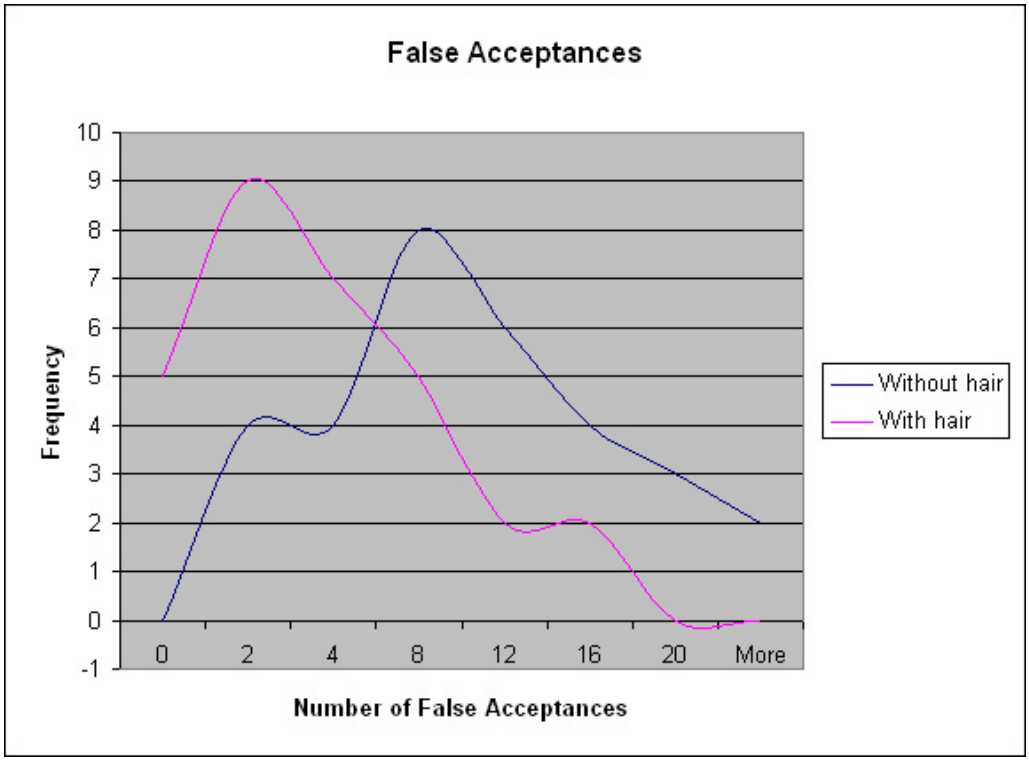

Figure 14: The histogram shows a graphical overview of the false acceptances of the two groups and the frequency of each number of false acceptances.

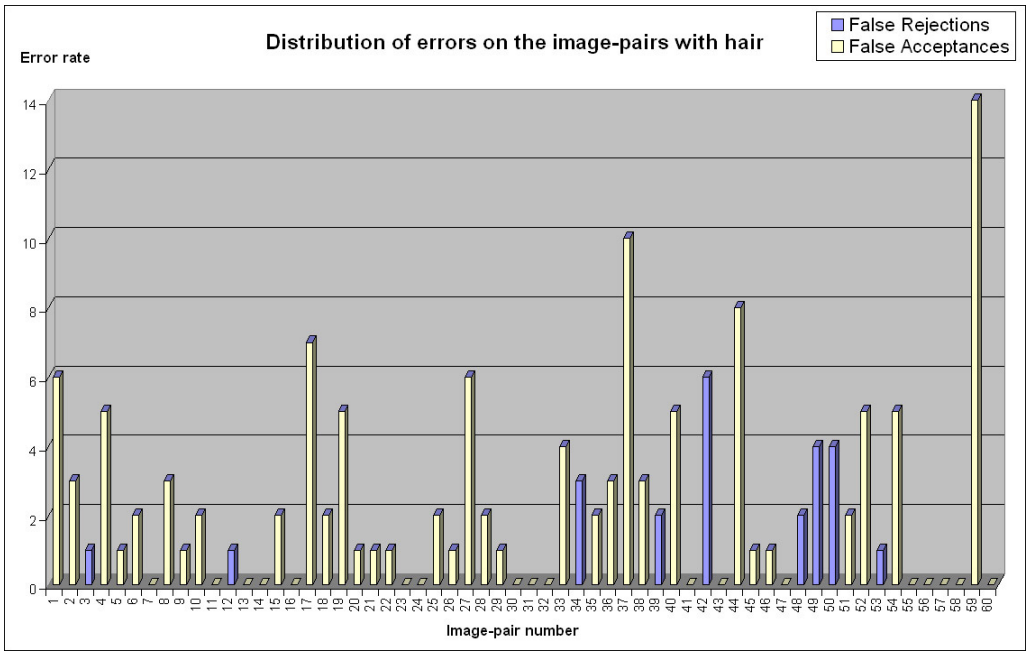

Figure 15: The summarized distribution of false acceptances and false rejections for the image-pairs used in the experiment where the hair was present.

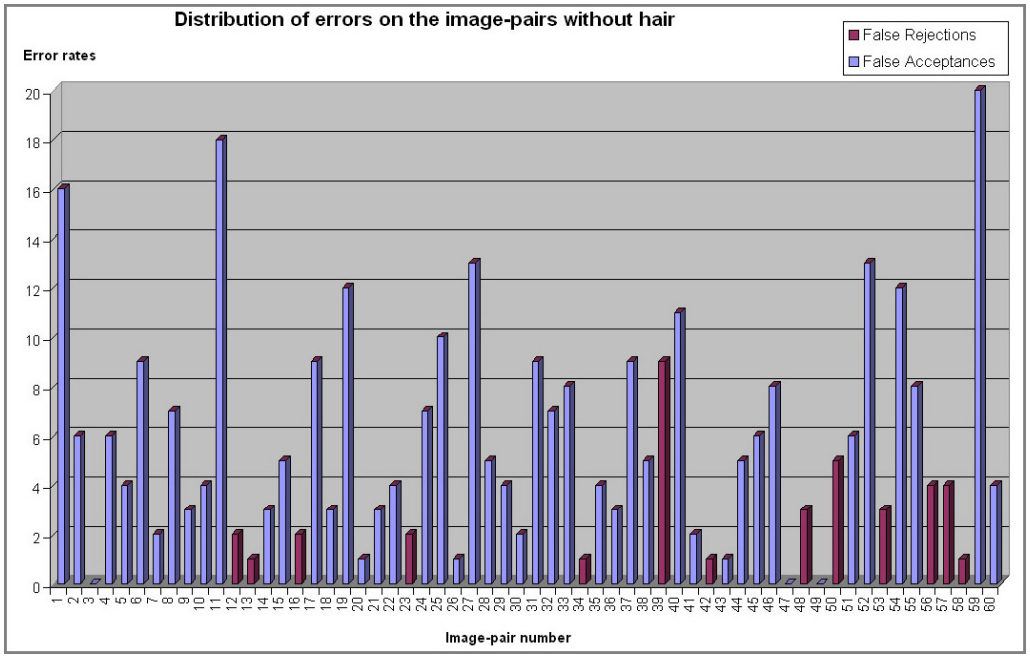

Figure 16: The summarized distribution of false acceptances and false rejections for the image-pairs used in the experiment where the hair was removed.

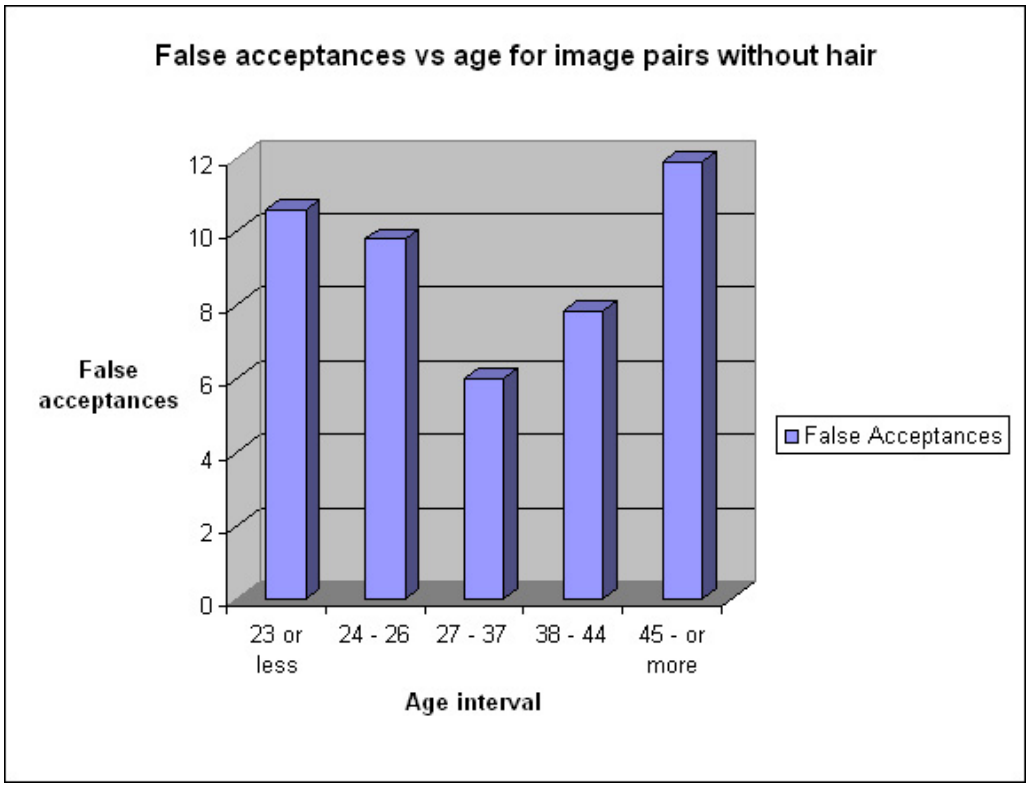

Figure 17: The histogram shows how the number of false acceptances decrease as the age interval approaches 27-37, while it increases as it get further from this age interval.

7.2.5 Experience

Before attending the experiment we also wanted to know whether or not the experience with identification of faces affected the error rates. However only one of the participants had this kind of experience, and although this person scored very well in the test, the differences were not significant.

8 Discussion

During the experiment we have looked at different methods for circumventing face recognition products, and in the experiment we have examined the influence hair has on humans face recognition performance. We will now discuss the results of the experiment, the findings from the literature study and how these results will affect the security of face recognition products.

8.1 The theory

As we have seen from this study it is not enough to physically secure the connection between the central system and the clients. If an attacker manages to get access to the connection, he could with the right equipment manipulate the electronic data between the two units as he wants. To prevent this, the system should use signatures and cryptology. But still this is inadequate, and there should also be taken action to prevent replay attacks, Trojan horses and viruses. If a Trojan is inserted to the system, this could be used to produce a high score for a given candidate [83]. If this person is able to find a look-alike from a human perspective, this could be enough to bypass the security check.

Further, it should be noted that an important aspect to the use of biometric authentication systems is the human factor. It does not matter if system is secure in it self if you can not trust the humans operating it. Also, humans do fail, and considerations must be made to prevent or at least reduce the possibility for occurrences of this. To prevent the operators from failing, one important aspect of such system is how user friendly it is. The easier things are to operate, the harder it is to fail. Also, a thorough training should be performed.

Studies show that the expectation of the observer influences his decision [51]. A human guard that will evaluate faces that are identified as the right person will then be somewhat biased in his evaluation. Because the person is already accepted as the right person by the computer system, he will perhaps be more likely to also be evaluated as the same individual by the human guard, although this may be incorrect. This could be because the human guard is biased in his evaluation or because he relies on the system and therefore perform a less thorough evaluation of the person.

In our study we have reviewed literature that have dealt with the role of the eyebrows, the cross race effect, the expectation of the observer and so on, and in our experiment we have examined the role of the hair. These are all factors that influence the human ability to correctly identify another individual based on the face. An attacker that are aware of this factors, and who are able to exploit this to his advantage will have a great chance of succeeding in fooling a human observer.

8.2 The role of hair

As we can see from the results of our experiment, the false acceptance rate on the image-pairs where the hair is removed are significantly higher than for those where the hair is present. The hair is a feature that can be easily manipulated, indicating that there is in fact a great opportunity for an impostor to circumvent both the system and the human guard using simple and cheap methods. When combining this with facial makeup and the influence eyebrows, the color of the eyes and beard have on human face recognition performance we see that using a human supervisor to increase the security may be insufficient. An impostor would have an even greater chance of succeeding if the owner of the visa corporate with the impostor and applies for a visa when he has facial hair, glasses, manipulated his eyebrows in such a way that the impostor is able to get similar eyebrows and so on. A better solution to achieve higher security would then be to employ multimodal biometric systems [14, 15].

As we observed from figure 15 and 16 there were only 3 image-pairs that have not been guessed wrong from one or more participant in the experiment where the hair was removed, while when hair was present there were 18. This may indicate that the hair is a feature that plays a major role in distinguishing several of the faces. It may also indicate that the face-images are very much alike. This makes it even more likely that they may be falsely considered as the same person also in a border control environment. In such environment the human supervisor may also relay more on the decision of the computer based system and this could affect his decision.

It should be noted that although there were an average number of false acceptances of 9.61 for the experiment where the hair and background was removed, only 45 of the 60 image-pairs were actually composed of face images of different persons. This produces a false acceptance rate of 21.36% on average. From this we can observe that the human supervisor is able to eliminate some of the errors that the face recognition systems performed. Still over 20% of the errors that the system has made goes undetected. Combining this with the observation that most of the face image-pairs where evaluated wrong by more than one individual, we have an indication that the human supervision does not provide sufficient additional security in a high security setting. It would in this context also be interesting to, in a further study, see how alteration of hair combined with manipulation of other features such as eyebrows, teeth and beard would affect the human recognition performance.

8.3 Other aspects

8.3.1 The role of the age of the participants

As we observed earlier there was some invariance in the mean age between the groups where hair was present and the group where the hair was removed. However, the Levene's test in table 7 provide a significant-score of 0.005, indicating that the chance that the difference is caused by another factor is very small. The mean age of the women in the experiment was, although insignificant, somewhat higher (38.62) of the 26 women that participated than for the 35 men (32.80).

From figure 17 we could see that the number of false acceptances is lower the nearer the age is the age of 27 - 37. This could indicate that the experience in recognition faces has an effect, because those over 27 would have more experience in the field, and better knowledge on what to look for in a face. The number of false acceptances increases for those over 37. This could be due to poorer sight of those over the age of 37. Also some in this latter age-group had little experience with computers, and they might then focus more than others on the operation on the mouse, making somewhat less attention to the images. This would surely affect the results.

8.3.2 The role of the educational degree of the participants

From the results of the experiment we observe a significant difference in recognition performance based on the educational degree. To conclude that education alone affects the performance would be a hasty and perhaps wrong assumption. This result may very well be because people with a higher degree also would be closer to the age in which the best face recognition performance is observed. We can also observe that the one participant with a PhD degree, actually have a worse performance than the people with a masters degree. There is actually only one participant with a PhD degree, resulting in a score that may be of pure chance. The participant with a PhD was actually in upper part of the group of participants between 38-44 years of age, and his high false acceptance rate could also be a outcome of this.

8.3.3 The role of the time spent on evaluating the image-pairs

The results from the experiment showed that there was a correlation between the time spent on comparison of the images where the hair was present, but that there were no correlation on the other images. This could be a result of misleading distribution because some of the participants have spent considerable time on some of the images, while others have been evaluated much faster. From table 10 we can observe that the false acceptance rate remains constant over time, while we in table 11 observed that the participants overall time-usage on comparison decreased linearly for each quartile. This indicate that experience have an effect on face recognition performance. The participants evaluate the images faster, but their error rate remains constant. This could indicate an increased focus on the features to evaluate and that the participants use less time on finding these, but that they still make some errors. The number of image-pair used in this experiment may not be enough to examine whether or not the participant get tired and less aware to details over a longer time period. However on the 60 images used in this experiment there seem to be no such effect.

The fact that they evaluate faster as they have evaluated for some amount of time may also indicate that they get restless, and because of this spend less time on evaluating, and that how long time they spend on evaluating has little effect on their error rate.

8.3.4 The role of the gender of the participant

From the experiment-data we can observe that there were a few more men attending the experiment than women, however the differences were small. The women (6.73) provided a smaller mean value for false acceptances than men (6.86), and a higher mean value for false rejections (women: 1.04, men: 1.00). These differences is too small to state that one gender is more strict in their evaluation than the other.

Another factor to observe in this context is that there were 15 women and 16 men that participated in the experiment without hair, while there were 11 women and 19 men that participated in the experiment with hair. Although this is slightly different from the expected distribution between the groups, it is so close that there should be little variations if the distribution had been 50/50.

Whether or not the gender of the persons on the images affected the performance, was not a part of this study because there were too few images of women in the sample, and because this has been established before [18].

8.3.5 The role of experience with face recognition

In the experiment on human face recognition performance with and without hair the participants consisted mainly of people with no experience with face recognition. It should be kept in mind that, as reviewed earlier in this report, experience with identification of faces has an effect on the recognition performance. Even thought the only person in this test with such experience did have only one false acceptance, and he knew which one he missed when he was confronted with the results, this is not enough to make a conclusion that experience matter. It should be noted that this person participated in the experiment where the hair was visible and that there also were a few others that performed as well as this person. Because of his enthusiasm, he was also allowed to try evaluating image-pairs where the hair where removed, but then on a database that was identical but not part of the real experiment. His score here were however, although better than the average for this group, much higher than his score on face images where the hair was present.

As we have reviewed earlier in this report, experience with recognition of faces influences the face recognition performance of humans. In our experiment however none of the participants had substantial experience in this field, but one had some experience. This could be a factor that would influence the outcome of our experiment in both groups, but the difference between the groups would probably be similar. However further studies in this directions are recommended.

8.3.6 Various considerations

One of the participant commented that some of the face images looked similar, but that he still could see that they were different. He could however not pinpoint what made him aware that they were not the same person. In that context it would have been interesting to see whether or not manipulation of these individuals' facial features could increase the similarity.

Some of the participants may have taken the experiment less serious and not given the task enough focus, so that the results have been affected. However, the entire process was supervised, and none were found to not take the task seriously. Either way this would apply for both groups and potential variations will be leveled out in the total score of the group.

In our experiment we used the face images that were already accepted from the face recognition system. When looking at the false acceptance rate obtained from the experiment on should keep in mind that this is the false acceptance rate of the images that have already been accepted, and not the entire population. With that in mind we can observe that the false acceptance rate when combining the face recognition system with a human guard will decrease considerable. However, this combination is far from perfect, and other solutions should be regarded in high security settings where no one should be falsely accepted. The use of a human guard could however have a psychological preventive effect on impostors.

8.4 The added value of the work

This work contributes to an increased understanding in human and computer-based face recognition. The difference between human face recognition on face images where the hair is removed vs. where the hair is present have not, as far as we know, been examined earlier. This experiment provides an additional insight in human face recognition performance, and will serve as a good basis when evaluating the additional security a human security guard will provide.

9 Conclusions

Aging is a problem with face recognition, but in VIS, the visas will be valid for a limited timeframe of three months, and the face images will be updated frequently. Because of this, aging will not affect the error rate significantly. If applicants registered in a watchlist do not apply for a visa for several years, their appearance may however change so much during this time that the system is unable to identify them in the watchlist. They are then falsely rejected. Should the applicants alter their appearance by manipulating their eyebrows, beard or performing plastic surgery this will as we have seen affect the error rate.

A major problem with existing face recognition is that it produces too many false acceptances to be used in large scale applications like the Visa Information System. This could be compensated by a human supervisor. Our experiment however, shows that a human guard does not have sufficient effect on reducing the false acceptances in large scale and high security application. This was especially the case when the hair was not included in the images. This implies that the hair plays a considerable role in human recognition of human faces. Alteration of the hair is also a cheap method that any attacker can afford, and it is therefore likely that this method will be used. As our study shows there are also other factors, like alteration of the eyebrows, alteration of the teeth and colored contact lenses to resemble another person that could affect the false acceptance rate. When combining these methods, attackers that are able to circumvent a face recognition product have a considerable chance of also getting passed the security guard. Because of this, other authentication methods combined with face recognition should be considered.

Until there have been substantial performance improvements in face recognition products, the security of large scale authentication applications should not be based only on face recognition. A suggestion, to improve the recognition performance, is using multimodal biometric systems, where you combine face recognition with for instance fingerprint scanning. The face recognition could then be used to reduce the search, while fingerprint is used to narrow the sample further or find the identity. A possibility is to combine several face recognition products that use different methods for authentication. This method however, have a greater possibility to let an illegitimate individual pass than the combination of different biometric features, because that would require more biometric features to correspond with the stored templates. This is also the intention in the new VIS, but restrictions in some countries prevent this. Our experiment emphasizes the need for changes in these laws.

9.1 The research questions

Our contribution to the research questions on which this thesis is based are discussed below. Regarding research question 1, there seems to be little effort an impostor has to

make if he or she has a look-alike from the face recognition systems point of view. For an impostor to succeed he or she should have their own copy of the face recognition program or a similar system to be able to see whether or not they have a look alike with a legitimate visa. They could also use the identity of an individual suitable for a visa and register with this identity. If the impostor is able to make him or herself look similar to the person on the identity document, the impostor have a great chance of fooling the security guard and get a visa on false premises. By altering the hair, eyebrows, beard and so on the impostor will have a significant chance of succeeding even with limited resources. Hopefully this thesis will contribute to increase the understanding of the limitations and threats to face recognition systems, making vendors and employers more aware of the challenges, so that countermeasures may be initiated.

Our answer to research question 2 is that the resources of an attacker will affect the security, because he will then have more options to circumvent the system. But our experiment shows that even without considerable resources it possible to perform a successful attack. Keep in mind though that also knowledge is an important resource for circumventing such systems.

Further, in research question 3 we wanted to examine whether or not today's procedures for calculating FAR could result in a positive evaluation of insecure products. We have reviewed the existing methods for circumventing face recognition products. Should there be other methods to bypass face recognition than providing your own face to pass as someone else, this could very well affect the false acceptance rate. We found several methods to circumvent face recognition products, but most of these can with careful planning be avoided. However 100% security can not be achieved and we can assume that although face recognition have low false acceptance rate, it could be circumvented by other means than manipulating the face. We may also assume that today's methods for calculating FAR could result in a misleading false acceptance rate because manipulation of the eyebrows, beard and plastic surgery will affect the similarity scale. How much face recognition systems similarity scale is affected by these alterations is a case for further study, but when employing such systems you should keep in mind not to only focus on the false acceptance rate but also the overall security of the system. For instance a system that returns the exact similarity score to the client is vulnerable to a hill climbing attack.

As we can observe from the experiment, covering human ability to recognize faces where the hair is removed and where hair is present, an impostor that alters his hair to look more similar to the person he simulates will have a increased chance of fooling the human guard if he is able to circumvent the face recognition system. Combined with facial makeup, colored eye-lenses and manipulation of the beard and the eyebrows we observe that the answer to research question 4 is that non-zero effort attacks may have a great effect on the FAR of a face recognition systems in a border control environment.

To summarize, human supervision may not be the best answer to reduce the false acceptance rate of face recognition products in large scale and high security applications, and one should consider combining face recognition with other biometric authentication technologies rather than combining it merely with human supervision.

10 Further work

After working with this thesis there have surfaced aspects related to face recognition that need further study. Suggestions on areas for further research on face recognition include:

- A survey of what impact non-zero effort attacks will have on the False Rejection Rate in a border control environment.
- A survey of the impact non-zero effort attacks will have on the False Rejection Rate on face recognition systems in general or in a specific environment or application.
- Further research regarding how people of different ethnic races affect the False Acceptance and False Rejection in state of the art face recognition systems.
- Research regarding 3D face recognition systems and their performance on siblings and twins (both identical and not).
- A study similar to ours, but that examine the role of hair when evaluated on participants with substantial experience with identification of faces.
- Research on the influence the information provided to persons evaluating face images have on the false acceptance rate and false rejection rate. The intention of such research will be to examine the effect knowledge about the outcome of a face recognition systems evaluation of the image-pairs has on a human guards ability to correctly evaluate whether or not the individual correspond to the claimed identity.
- A study of what impact facial makeup and eyebrows have on current face recognition systems.

There have been a great focus on face recognition for the past 30 years, and the interest is increasing as the technology matures. However, when working with this thesis we have seen that there are still aspects of face recognition that needs further study. We have in this thesis surveyed different methods for circumventing the current face recognition system with the goal of gaining false acceptance. Even so, there are yet to be carried out a similar study that thoroughly surveys methods that affect false rejection rate in different face recognition systems. This would be interesting in order to see whether face recognition is suitable technology for identification of individuals already recorded in a border control environment.

In our review of the cross-race effect on human's ability to identify people with another race, another field of face recognition surfaced; research on cross-race effect on face recognition systems performance. This is a subject that should be further studied.

Further, the importance of eyebrows, and how alteration of these, and the use of facial makeup affects the performance of face recognition systems should be further studied. According to Sadr et al. [60] such studies could contribute to surveying which parts of the face that is more important for facial recognition than others. Such studies would further contribute to a better understanding of the level of resources and skills needed

to fool a face recognition system.

As mentioned, Aurora [40] claims according to [12] that their 3D software is able to distinguish between identical twins. These claims however are not substantiated by independent tests. An independent study of 3D face recognition systems would therefore be useful. An important drawback with existing 2D face recognition systems is their inability to correctly recognize identical twins. As a result, it would be useful to study 3D face recognition systems performance regarding siblings and twins, both identical and fraternal.

Bibliography

[1] The Norwegian committee for biometrics K188, K 188 - Person-ID. `http://www.standard.no/imaker.exe?id=3925`.

[2] A. M. Martinez and R. Benavente. June 1998. The AR Face Database. *CVC Technical Report #24*.

[3] CVL Face Database. `http://www.lrv.fri.uni-lj.si/facedb.html`.

[4] J. Madubuko. September 18, 2001. Identity theft is fastest-growing threat. *Silicon.com*.

[5] A. Norman, J. R. Willox, and T. M. Regan. March 2002. Identity fraud: Providing a solution. *LexisNexis*.

[6] G. R. Newman. June 2004. Identity Theft, Problem-Oriented Guides for Police. *Problem-Specific Guides Series No. 25, U.S. Department of Justice, Office of Community Oriented Policing Services*.

[7] J. Hopkins. December 29, 2003. Increase in Arrests for Visa and Passport Fraud. *U.S. Visa News*.

[8] C. Groening. November 17, 2003. INS Vet Warns Public About Illegal Immigration-Identity Theft Link. *Religion News*.

[9] T. A. Aass and B. Rygh. March 30, 2004. Nyhetsbrev om norsk flyktning- og innvandringspolitikk Nr. 13. *Kommunal- og regionaldepartementet og Utlendingsdirektoratet*, volume 11.

[10] February 24, 2004. EU Visa Information System gets go-ahead. *eGovernment News*.

[11] M. Kosmerlj. Passport of the Future: Biometrics against Identity Theft? MSc thesis, Gjøvik University College, NISlab, June 30, 2004.

[12] T. Geoghegan. November 25, 2004. How your face could open doors. *BBC News Magazine*.

[13] J. W. Creswell. 2003. *Research Design Qualitative, Quantitative, and Mixed Methods Approaches, Second Edition*. SAGE Publications.

[14] D. Maltoni, D. Maio, A.K. Jain, and S. Prabhakar. June 2003. *Handbook of Fingerprint recognition*. Springer Verlag, New York, USA.

[15] A. K. Jain, A. Ross, and S. Prabhakar. January 2004. An Introduction to Biometric Recognition. *IEEE Transactions on circuits and systems for video technology*, 14(1).

[16] D. M. Blackburn. March 2004. Biometrics 101, version 3.1. *Federal Bureau of Investigation*.

[17] A. K. Jain, S. Pankanti, and S. Prabhakar. 2003. Biometric Recognition: Security and Privacy Concerns. *IEEE Security & Privacy, The IEEE Computer Society*.

[18] P. J. Phillips, P. Grother, R. J. Michaels, D. Blackburn, E. Tabassi, and M. Bone. March 2003. Face Recognition Vendor Test 2002: Evaluation Report. `http://www.frvt.org`.

[19] A. J. Mansfield and J. L. Wayman. August 2002. Best Practices in Testing and Reporting Performance of Biometric Devices, Version 2.01. *Centre for Mathematics and Scientific Computing, National Physical Laboratory, Queens Road, Teddington, Middlesex, TW11 OLW*.

[20] T. Mansfield, G. Kelly, D. Chandler, and J. Kane. March 19, 2001. Biometric Product Testing Final Report Issue 1.0. *Centre for Mathematics and Scientific Computing, National Physical Laboratory, Queens Road, Teddington, Middlesex, TW11 OLW*.

[21] M. Sandström. Liveness Detection in Fingerprint Recognition Systems. MSc thesis, Linköpings tekniska högskola, 2004.

[22] R. de Luis-Garcia, C. Alberola-López, O. Aghzout, and J. Ruiz-Alzola. 2003. Biometric identification systems. *Signal Process.*, 83(12), pages 2539–2557.

[23] A. Ross and A. Jain. 2003. Information fusion in biometrics. *Pattern Recognition Letters*, 24, pages 2115–2125.

[24] L. Hong and A. K. Jain. December 1998. Integrating Faces and Fingerprints for Personal Identification. *IEEE transactions on pattern analysis and machine intelligence*, 20(12), pages 1295–1307.

[25] S. A. C. Schuckers. December 2002. Spoofing and Anti-Spoofing Measures. *Information Security Technical Report*, 7(4), pages 56–62.

[26] 2002. Facial Scan Technology: How it works. Technical Report, International Biometrics Group.

[27] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld. December 2003. Face Recognition: A Literature Survey. *ACM Computing Surveys*, 35(4).

[28] S. G. Kong, J. Heo, B. R. Abidi, J. Paik, and M. A. Abidi. 2005. Recent advances in visual and infrared face recognition: a review. *Computer Vision and Image Understanding*, volume 97, page 1.

[29] S.Z. Li and J. Lu. 2005. Face detection, alignment, and recognition. *In G. Medioni and S.B Kang, editors, Emerging Topics in Computer Vision, chapter 9. Prentice Hall PTR*.

[30] S.Z. Li and A. Jain. Handbook of Face Recognition. *Springer, New York, 2005*.

[31] M. A. Turk and A. P. Pentland. 1991. Face recognition Using Eigenfaces. *Proc. IEEE Conference on Computer Vision and Pattern Recognition, Maui, Hawaii*.

[32] J. Chirillo and S. Blaul. June 2003. *Implementing biometric Security*. Wiley Publishing, Inc.

[33] J. D. Woodward Jr., N. M. Orlans, and P. T. Higgins. 2003. *Biometrics*. McGraw-Hill/Osborne, New York.

[34] J. Zhang, Y. Yan, and M. Lades. September 1997. Face Recognition: Eigenface, Elastic Matching, and Neural Nets. *Proceedings of the IEEE*, 85(9), pages 1423–1435.

[35] L. Wiskott, J. M. Fellous, N. Kruger, and C. von der Malsburg. July 1997. Face Recognition by Elastic Bunch Graph Matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7), pages 775–779.

[36] D. S. Bolme. Elastic Bunch Graph Matching. MSc thesis, Master of Science, Colorado State University, Fort Collins, Colorado, May 2003.

[37] J. Wayman, A. Jain, D. Maltoni, and D. Maio. 2005. *Biometric Systems: Technology, Design and Performance Evaluation*. Springer-Verlag London Limited.

[38] B. Duc, S. Fischer, and J. Bigun. April 1999. Face Authentication with Gabor Information on Deformable Graphs. *IEEE TRANSACTIONS ON IMAGE PROCESSING*, 8(4).

[39] J. Phillips. The Face Recognition Grand Challenge (FRGC). `http://www.frvt.com/FRGC/`.

[40] Aurora Computer Services Ltd. `http://www.facerec.com/`.

[41] G. Pan, Y. Wu, Z. Wu, and W. Liu. July 2003. 3D face recognition by profile and surface matching. *Neural Networks, 2003, IEEE*, volume 3, pages 2169–2174.

[42] G. Medioni and R. Waupotitsch. October 2003. Face modeling and recognition in 3-D. *Proceedings of the IEEE International Workshop on Analysis and Modeling of Faces and Gestures (AMFG'03)*, pages 232– 233.

[43] K. W. Bowyer, K. Chang, and P. Flynn. August 2004. A Survey Of 3D and Multi-Modal 3D+2D Face Recognition. *International Conference on Pattern Recognition*.

[44] W. Court. June 30, 2003. Biometrics: Evaluation Criteria and Scenario Based Performance Testing. *GSEC*.

[45] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss. October 2000. The FERET evaluating methodology for face recognition algorithm. *IEEE Transactions on pattern analysis and machine intelligence*, 22(10).

[46] J. Phillips. The Face Recognition Vendor Test (FRVT 2005). `http://www.frvt.com/FRVT2005/default.aspx`.

[47] M. Krechel and N. Tekampe. Common Criteria Protection Profile For Biometric Verification Mechanisms, Version 1.0. Technical report, TUV Informationstechnik GmbH, Germany, October 2004.

[48] R. Beveridge, D. Bolme, M. Teixeira, and B. Draper. May 1, 2003. The CSU Face Identification Evaluation System User's Guide: Version 5.0. *Computer Science Department Colorado State University*.

[49] February 2000. Best practices in testing and reporting performance of biometric devices, Issue 1. Report for CESG and Biometrics Working Group.

[50] P. J. Phillips, A. Martin, C.L. Wilson, and M. Przybocki. February 2000. An Introduction to Evaluating Biometric Systems. *National Institute of Standards and Technology*.

[51] V. Bruce and A. Young. 1998. *In the eye of the beholder*. The science of face perception. Oxford University Press Inc., New York.

[52] V. Bruce, P. J. B. Hancock, and A. M. Burton. 1998. Comparison between Human and Computer Recognition of Faces. *ICAFGR, Nara, Japan*, (408-413).

[53] P. Kalocsai, W. Zhao, and E. Elagin. April 1998. Face similarity space as perceived by humans and artificial systems. *Automatic Face and Gesture Recognition, IEEE*, pages 177–180. .

[54] Y. Adini, Y. Moses, and S. Ullman. July 1997. Face Recognition: The Problem of Compensating for Changes in Illumination Direction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7), pages 721–732.

[55] R. Cross, J. Shi, and J. F. Cohn. December 2001. Quo vadis Face Recognition? *In Third Workshop on Empirical Evaluation Methods in Computer Vision, Kauai, Hawaii*, pages 119–132.

[56] J. C. Brigham. 1986. The influence of race on face recognition. *Aspects of face processing, Nijhoff*, pages 170–177.

[57] P. Chiroro and T. Valentine. 1995. An Investigation of the Contact Hypothesis of the Own-race Bias in Face Recognition. *The Quarterly Journal of Experimental Psychology*, 48A(4), pages 879–894.

[58] D. T. Levin. 2000. Race as a Visual Feature: Using Visual Search and Perceptual Discrimination Tasks to Understand Face Categories and the Cross-Race Recognition Deficit. *Journal of Experimental Psychology: General*, 129(4), pages 559–574.

[59] N. Furl, P. J. Phillips, and A. J. O'Toole. 2002. Face recognition algorithms and the other-race effect: computational mechanisms for a developmental contact hypothesis. *Cognitive Science*, volume 26, pages 797–815.

[60] J. Sadr, I. Jarudi, and P. Sinha. 2003. The role of eyebrows in face recognition. *Perception*, volume 32, pages 285–293.

[61] A. K. Jain, S. Prabhakar, L. Hong, and S. Pankanti. May 2000. Filterbank-Based Fingerprint Matching. *IEEE TRANSACTIONS ON IMAGE PROCESSING*, 9(5), pages 846–859.

[62] Iridian Technologies. `http://www.iridiantech.com`.

[63] ICAO Doc 9303, Machine Readable Travel Documents, Sixth Edition – 2005. *ICAO*.

[64] ICO/IEC. 2005. International Standard, ISO/IEC FDIS 19794-5, Information technology – Biometric data interchange formats, Part 5: Face image data.

[65] X. Jiang, M. Binkert, B. Achermann, and H. Bunke. 1998. Towards Detection of Glasses in Facial Images. *Proceedings of 13th ICPR*.

[66] ICAO. ICAO Doc 9303, Machine Readable Travel Documents, Part 2 Machine Readable Visas – October 30, 2002.

[67] A. Jain, R. Bolle, and S. Pankanti. 1999. *Biometrics: Personal Identification in Network Society*. The Kluwer international series engineering and computer science. Boston: Kluwer.

[68] R. M. Bolle, J. H. Connell, S. Pankanti, N. K. Ratha, and A. W. Senior. 2003. *Guide to Biometrics*. Springer-Verlag New York, Inc.

[69] T. Bracco. October 5, 2000. Our 'not so impossible' mission. *Network World*.

[70] L. Thalheim, J. Krissler, and P.-M. Ziegler. Nowember 2002. Body Check Biometric Access Protection Devices and their Programs Put to the Test. *c't 11/2002, page 114 - Biometrie*.

[71] A. K. Jain, S. Prabhakar, and S. Pankanti. Can Identical Twins be Discriminated Based on Fingerprints? Technical Report MSU-CSE-00-23, Department of Computer Science, Michigan State University, East Lansing, Michigan, October 2000.

[72] USB Snoop for Windows, SourceFORGE.net. `http://sourceforge.net/projects/usbsnoop/`.

[73] USB Agent by Hitex, Hitex Solutions. `http://www.hitex.com/products.html?con_usb_agent.html~content`.

[74] N. K. Ratha, J. H. Connell, and R. M. Bolle. 2001. Enhancing security and privacy in biometric-based authentication systems. *IBM systems journal*, 40(3).

[75] L. O'Gorman. December 2003. Comparing Passwords, Tokens, and Biometrics for User Authentication. *Proceeedings of the IEEE*, 91(12), pages 2019–2040.

[76] B. Schneier. August 1999. The uses and abuses of biometrics. *Communications of the ACM*, 42(8), page 136.

[77] A. J. Klosterman and G. R. Ganger. May 2000. Secure Continuous Biometric-Enhanced Authentication. *CMU-CS-00-134, School of Computer Science, Carnegie Mellon Univeristy, Pittsburg*.

[78] A. M. Martinez. 2002. 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), pages 748–763.

[79] A. M. Martinez. 2000. Recognition of Partially Occluded and/or Imprecisely Localized Faces Using a Probalistic Approach. *IEEE Computer Vision and Pattern Recognition, CVPR*.

[80] P. E. Agre. 2001. Your Face Is Not a Bar Code: Arguments Against Aoutomatic Face Recognition in Public Places. *Department of Information Studies University of California, Los Angeles*.

[81] J. Stanley and B. Steinhardt. January 3, 2002. Drawing a Blank: The failure of facial recognition technology in Tampa, Florida. *ACLU*.

[82] M. A. Taister, S. D. Holliday, and H. I. M. Borrman. April 2000. Comments on Facial Aging in Law Enforcement Investigation. *U.S. Department of Justice Federal Bureau of Investigation, Forensic science communications*, 2(2).

[83] N. K. Ratha, J. H. Connell, and R. M. Bolle. 2003. Biometrics break-ins and band-aids. *Pattern Recognition Letters, Elsevier Science*, volume 24, pages 2105–2113.

[84] U. Uludag and A. K. Jain. 2004. Attacks on Biometric Systems: A Case Study in Fingerprints. *In Proc. SPIE-EI 2004, Security, Seganography and Watermarking of Multimedia Contents VI*.

[85] C. Soutar, R. Gilroy, and A. Stoianov. 1999. Biometric System Performance and Security. *Conf. IEEE Auto. Identification Advanced Technol*.

[86] A. Adler. July 20-22, 2005. Vulnerabilities in biometric encryption systems. *Audio- and Video-based Biometric Person Authentication - AVBPA 2005, Tarrytown, New York, USA*.

[87] The BioAPI Consortium. March 2001. BioAPI Specification (Version 1.1).

[88] A. Adler. May 2004. Images can be regenerated from quantized biometric match score data. *IEEE, Niagara Falls*.

[89] A. Adler. May 2003. Sample images can be independently restored from face recognition templates. *IEEE, Montreal*.

[90] P. G. McLean. A Secure Pervasive Environment. *Trusted Computer System Group, Information Networks Division, Defence Science and Technology Organisation, PO Box 1500, Edinburgh 5001, South Australia*.

[91] R. Beveridge, D. Bolme, M. Teixeira, and B. Draper. May 1, 2003. The CSU Face Identification Evaluation System User's Guide: Version 5.0. *Computer Science Department Colorado State University*.

[92] K. Messer, J. Matas, J. Kittler, J. Luettin, and G. Maitre. March 1999. XM2VTSDB: The Extended M2VTS Database. *In Proceedings of the Second International Conference on Audio- and Video-based Biometric Person Authentication (AVBPA 1999)*, pages 72–77.

[93] U.S. Government Biometric Verification Mode Protection Profile for Medium Robustness Environments, Version 1.0. Technical report, Information Assurance Directorate, November 15 2003.

[94] ICO/IEC. 2004. International Standard, ISO/IEC 15444-1, Information technology – JPEG 2000 image coding system: Core coding system.

[95] J. Niessen. May 2004. Five years of EU migration and asylum policy-making under the Amsterdam and Tampere mandates. *Paper prepared for the German Council of Experts for Immigration and Integration*.

[96] P. Chaston. December 27, 2003. Schengen Information System and Biometrics. *White Rose, London*.

A Appendix – Definitions

Impostor: *A person who submits a biometric sample in either an intentional or inadvertent attempt to pass him/herself off as another person who is a legitimate enrollee.* [93]

JPEG: Joint Photographic Experts Group, JPEG is a standardized image compression mechanism (www.tufts.edu/orgs/edmedia/gloss.shtml)

JPEG 2000: Joint Photographic Experts Group 2000, is a wavelet-based image compression standard. It was created by the Joint Photographic Experts Group committee with the intention of superseding their original discrete cosine transform-based JPEG standard. The usual file extension is .jp2. JPEG 2000 can operate at higher compression ratios without generating the characteristic 'blocky and blurry' artifacts of the original DCT-based JPEG standard. It also allows more sophisticated progressive downloads. Part of JPEG 2000 has been published as an ISO standard [94]. Further evaluations on the differences between JPEG and JPEG 2000 are provided in [64].

Non-zero-effort attack: An attack using means other than submitting own original and unaltered biometric template to gain unauthorized access. This expression include attacks from where the attacker alter his or her appearance to look similar to a legitimate user, to attacks performed against the system and its surroundings like a man in the middle attack, Hill climbing attack, bribery, and the use of threats and force to mention some.

PPM: Portable Pixel Map, The PPM format is a lowest common denominator color image file format (http://netpbm.sourceforge.net/doc/ppm.html)

Probe: *A probe is one signature in a probe set* [18].

Probe set: *A probe set contains the images (biometric signature) of unknown individuals presented to the system for recognition* [18].

SIS art. 96 database: Schengen Information System is a database of individual's names and details for the purpose of *by means of an automated search procedure, to have access to reports on persons and objects for the purposes of border checks and controls and other police and customs checks carried out within the country in accordance with national law and, in the case of the single category of report referred to in Article 96, for the purposes of issuing visas, the issue of residence permits and the administration of aliens in the context of the application of the provisions of this Convention relating to the movement of persons.* [95]

SIS II is an expansion of the second generation of the SIS. This system is a multi-access international database on cross border crime, used by police and border authorities of Schengen Member States. It lists information on wanted persons and stolen goods [96]. This database will store biometric data and digital photographs, and will be integrated

with the Visa Information System [95].

SOF: *Strength Of Function (SOF) – A qualification of a TOE security function expressing the minimum efforts assumed necessary to defeat its expected security behaviour by directly attacking its underlying security mechanisms.* [47]

Template: *A template is a mathematical representation of a biometric sample (i.e. a facial image). A template enables algorithms to work more quickly than would otherwise be possible, by encoding the relevant information in a series of bits and bytes* – Identix, 2005 (http://www.identix.com/trends/faqs/template.html, 2005).

TOE: *Target of Evaluation – An IT product or system (and its associated documentation) that is the subject of a Common Criteria evaluation.* [47]

Zero-effort attack: *An arbitrary attack on a specific enrollee identity in which the impostor masquerades as the claimed enrollee using his or her own biometric sample.* [93]

Zero-effort attempt: *An impostor attempt is classified as 'zero effort' if the individual submits their own biometric feature as if they were attempting successful verification against their own template* [19].

B Appendix – Applications developed

To be able to execute the experiment, I have developed several applications in Java. These are IC_Client 1.0, IC_Administrator 1.0, RAW Image Converter 1.0, ImageConverter 1.0 and SFI Analyzer 1.0. To make the database used in the experiment I have also used an application for administration of database that I have developed, and that are available at my home site http://master.fladsrud.com. To demonstrate how analysis of the SFI files, and unpacking and converting of the images could be done in fast and easy fashion I have described the applications I have developed for this thesis below. This chapter also provides an overview of how the applications handling the experiment are composed. Source code for the applications IC_Client 1.0 and IC_Administrator 1.0 developed for this masters' thesis are available at http://master.fladsrud.com.

B.1 IC_Client 1.0

The IC_Client 1.0 application is used to present the face images to the participants in the experiment. This application uses the face_experiment database, which is an hsqldb database. The face_experiment database and its creation are described in appendix C. The participants are here first presented with a window where they register their age, highest educational degree, their gender and whether they previously have worked with identification of faces. Second the participants are presented the different image-pairs and are guided to press either a button saying they are of the same individual or if they are of different individuals. Should the participant use more than the specified time in deciding, a new window will appear, hiding the images and guiding the participant to make a decision. IC stands for Image Comparison.

B.2 IC_Administrator 1.0

The IC_Administrator 1.0 application is used to get useful information stored about the experiment from the face_experiment database, described in appendix C. The IC_Administrator application displays all available experiment types in a dropdown list, and the user can then generate a text file by selecting an experiment type, and press the generate button. The application then fetches information from the database, and generates a semicolon separated text file that can be directly imported into an statistical analysis tool.

B.3 RAW Image Converter 1.0

Images from the AR Face database [2] are in .raw format. To use these images I had to convert them into JPEG images. To convert the images into JPEG images I used the ImageMagick from http://www.imagemagick.org. However to avoid manually converting each image, something that would be too time consuming, I developed an application I have called RAW Image Converter 1.0. This program takes all images in a given folder,

and its sub folders and converts all .raw images found and place them in a given output folder. This program uses the ImageMagick program in the following way: The RAW Image Converter 1.0 generates an executable .bat file as shown in figure 19, which contains commands to use the ImageMagick program to convert all .raw images into JPEG images. This application could with small alterations be extended to perform the same task for all image conversion supported by ImageMagick. The bat file is the open, and the user may choose to run it. If he does, the conversion of the images will start. The bat file is stored with the time of creation as part of the file name, and will therefore not be overwritten if the program is executed more than once. A screen-shot of the application is provided in figure 18.

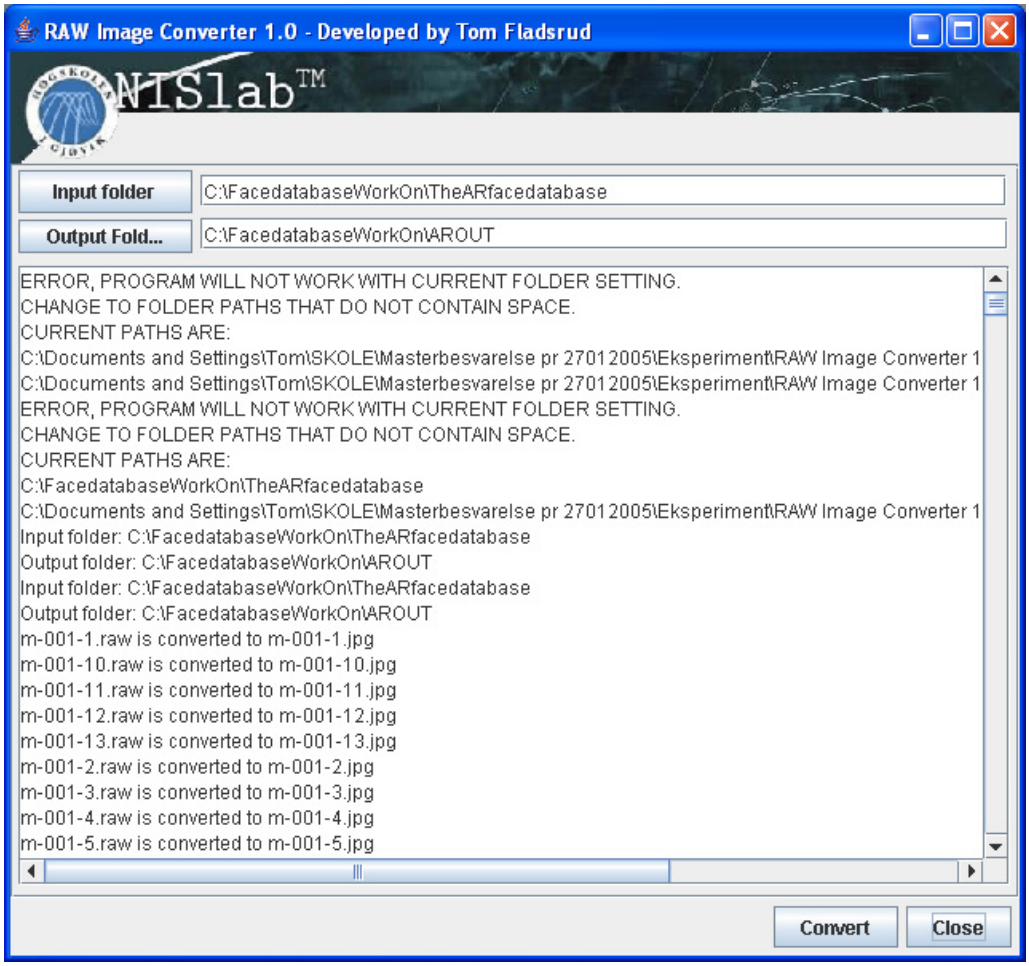

Figure 18: The RAW Image Converter 1.0 – This application convert images of .raw format, from a selected folder and its sub folders, into JPEG format, an places them in the selected output folder.

B.4 ImageConverter 1.0

The XM2VTS face database [92] contains images in PPM format. These are also placed in a sub folder for each individual. To manually unzip and convert these images would be too time consuming. Because of this, and because I did not have a converter for

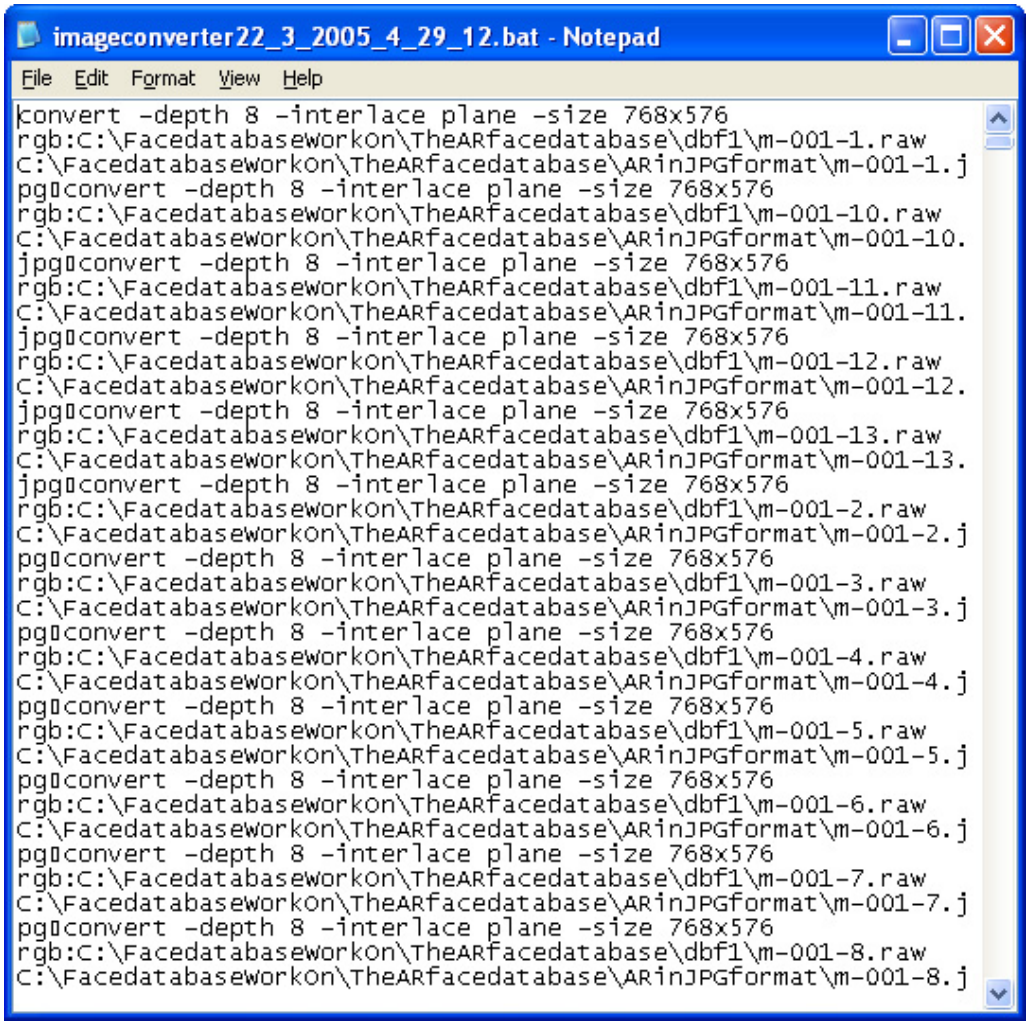

Figure 19: The RAW Image Converter 1.0 generates a bat file that do the actual conversion of the image files. This figure illustrates a part of one such bat file.

PPM too JPEG format, I developed an application for automatically unzipping all files and subsequent zip-files in a folder and its sub folder. The application also converts the images from PPM into JPEG format. I have also included conversion between JPEG, BMP and PPM format, and to convert all formats into gray scale images. I have been able to make this application by using the open source algorithms provided by Sieuwert van Otterloo and Ernst van Rheenen (blueRing Software Development), available at http://www.bluering.nl/imagetoolz/index.html , for conversion between different image formats. By using these algorithms for conversion in an application I have developed, that uses the algorithms to convert all images in each folder and also unzip the files found there, I have developed an effective tool for conversion of the images in the AR face database. Figure 20 provides a screen-shot of this application.

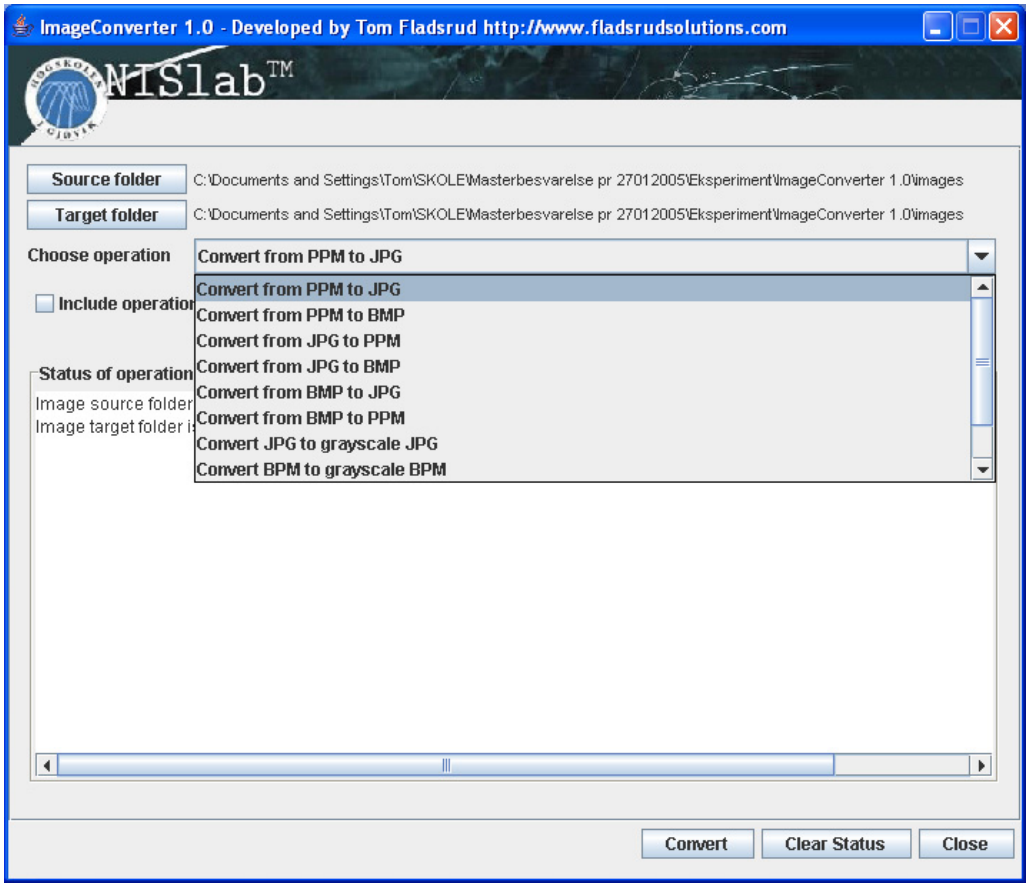

Figure 20: The ImageConverter 1.0 – This is an application for conversion between different image formats. You choose the root folder where the images are stores, and the folder where the converted images shall be stored. Then you choose what action to perform. Either you could unzip images in the folder, and subsequent folders (if the option is checked), or you could convert between different image formats. If the action should be done also on all subsequent folders, the *Include operation to involve operation on all subsequent folders* option must be selected. Then press the *Convert* button to execute the task.

B.5 SFI Analyzer 1.0

In the masters thesis of Kosmerlj [11] the CSU application used there produced SFI files that contains similarity scores for each images compared. These SFI files are several hundreds in number, and each image is compared to all other images, producing SFI files with several hundreds of lines with data. As we can see, this results in so much data to analyze that it would be to time consuming for humans to do manually. Because of this, I developed the application SFI Analyzer 1.0, as shown in figure 21. This application goes through all SFI files found in a specified folder and its sub folders, and analyze the data there. In the application the user can specify a threshold to search for. Each line with a threshold within the given threshold is written to a separate file. This way all images that is more similar than the given threshold is separated to a separate file, making it easier to find these files.

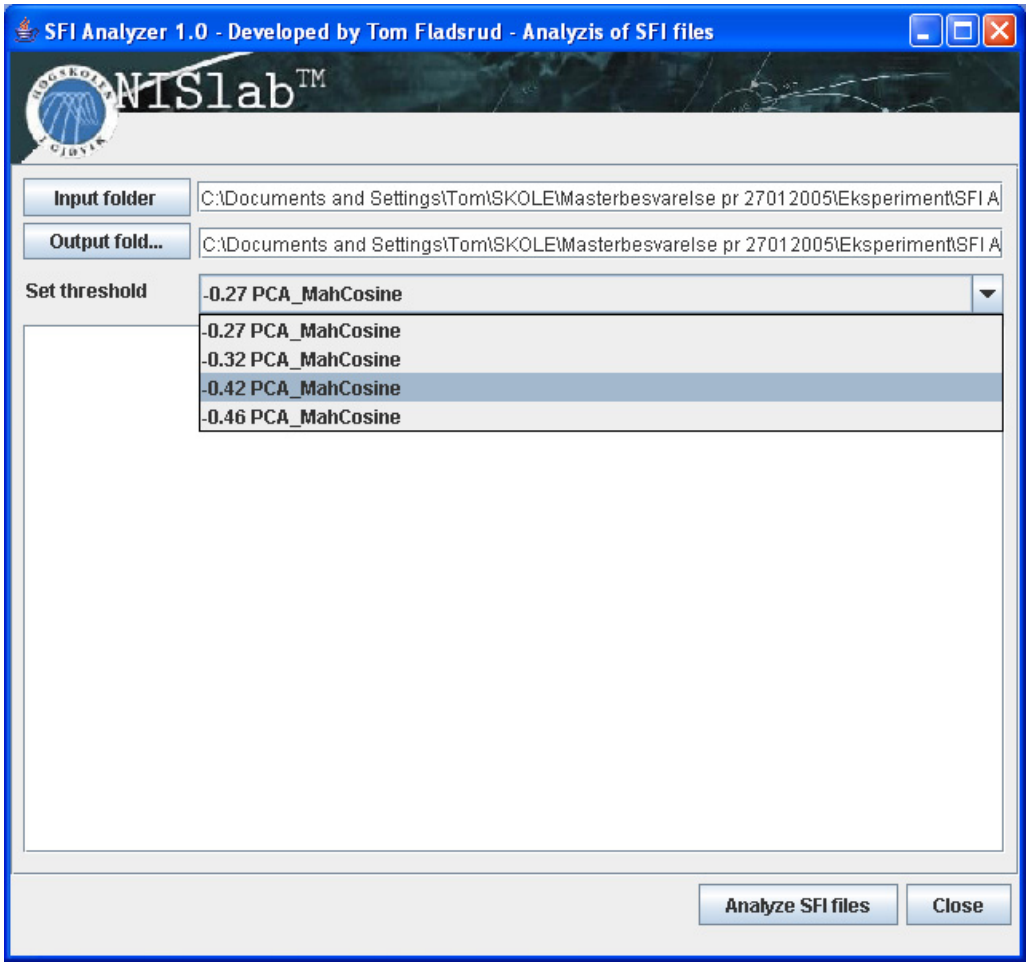

Figure 21: SFI Analyzer 1.0 – analysis of SFI files to find which images to use in the experiment. The SFI Analyzer analyzes the .sfi files that were generated from Kosmerlj's experiment in her masters thesis [11]. Each such .sfi file represent one face image, and contains one line of data for each face image they have been compared with. Each line contains the name of the face image, and a similarity score represented with scientific notation. The SFI Analyzer runs through all the files in the root folder that is selected, and its subsequent folders to find .sfi files. Each such file is the analyzed to see whether or not it contains images with a similarity below the given threshold. For each .sfi file that contains more than one image that have a similarity below the threshold the name of the .sfi file and the faces that were below the threshold are written to the result file.

C Appendix – Database used in the experiment

For readers wanting to reconstruct the database used in this thesis, the SQL commands used to generate the hsqldb database for use in the experiment, in the applications IC_Administrator and IC_Client, is reproduced below. The database *face_experiment* was build using the DatabaseCreator 1.0 available at http://master.fladsrud.com. This is an application developed for easy creation and alteration of hsqldb databases. The SQL commands used for creating this database, and the SQL commands for inserting initial values to the database are available at http://master.fladsrud.com.

CREATE CACHED TABLE Imagepair (Imagepair_ID INTEGER, Image_ID1 INTEGER, Image_ID2 INTEGER, Experiment_ID INTEGER, Same_Individual INTEGER, PRIMARY KEY (Imagepair_ID));
CREATE CACHED TABLE Image (Image_ID INTEGER, Image_path VARCHAR(200), Experiment_ID INTEGER, PRIMARY KEY(Image_ID));
CREATE CACHED TABLE Degrees (Degree_ID INTEGER, Degree_name VARCHAR(50), PRIMARY KEY (Degree_ID));
CREATE CACHED TABLE Experiment_type (Experiment_ID INTEGER, Experiment_name VARCHAR (30), PRIMARY KEY (Experiment_ID));
CREATE CACHED TABLE Participant (Experiment_ID INTEGER, Participant_ID INTEGER, Age INTEGER,Gender INTEGER, Highestdegree_ID INTEGER, Job_with_ID_check INTEGER, PRIMARY KEY (Experiment_ID, Participant_ID));
CREATE CACHED TABLE Comparison (Experiment_ID INTEGER, Participant_ID INTEGER, Imagepair_ID INTEGER, Alike INTEGER, Time_spent BIGINT, PRIMARY KEY (Experiment_ID, Participant_ID, Imagepair_ID));

D Appendix – Results from the experiment

This chapter provides the results, in form of tables, from the experiment conducted in connection with this thesis. The tables are divided in one section for each feature it represents.

D.1 Gender

Error type	**Gender of the participants**	**Number of Participants**	**Mean Number of Errors**	**Std. Deviation**	**Std. Error Mean**
False Acceptances	Woman Man	26 35	6.73 6.86	5.814 6.160	1.140 1.041
False Rejections	Woman Man	26 35	1.04 1.00	1.113 1.237	0.218 0.209

Table 3: The table shows the difference between false acceptances and false rejections between the genders. As we can see from the table the differences in number of false acceptances between the genders of participants are minimal.

D.2 Age

Error type		**F**	**Sig.**
False Acceptances	Equal variances assumed Equal variances not assumed	0.084	0.773
False Rejections	Equal variances assumed Equal variances not assumed	0.582	0.448

Table 4: The Levene's test for equality and variance on False Acceptance and False Rejections vs. age shows that there are no significant difference between the number of false acceptances and false rejections between the two genders.

		F	**Sig.**
Age * False Acceptances	Between Groups (Combined) Within Groups Total	2.593	0.040

Table 5: The Anova test shows that the impact age has on false acceptances is significant.

	Have hair	**Number of Participants**	**Mean Number of Errors**	**Std. Deviation**	**Std. Error Mean**
Age	No Yes	31 30	38.58 31.87	14.953 9.673	2.686 1.766

Table 6: The T-test on the age vs. the groups with and without hair shows that age have a statistically significant influence in the human evaluation of faces. Note however that the mean age in the two groups vary from 38.58 to 31.87.

		F	**Sig.**
Age	Equal variances assumed Equal variances not assumed	8.627	0.005

Table 7: The Levene's test on the age vs. the groups with and without hair shows that age have a statistically significant influence in the human evaluation of faces.

D.3 Educational degree

			Highesteducationaldegree					
			0	1	2	3	4	Total
Havehair	No	Count	7	10	12	1	1	31
		Expected Count	5,6	8,6	10,2	6,1	,5	31,0
		% within Havehair	22,6%	32,3%	38,7%	3,2%	3,2%	100,0%
		% within Highesteducationaldegree	63,6%	58,8%	60,0%	8,3%	100,0%	50,8%
		% of Total	11,5%	16,4%	19,7%	1,6%	1,6%	50,8%
	Yes	Count	4	7	8	11	0	30
		Expected Count	5,4	8,4	9,8	5,9	,5	30,0
		% within Havehair	13,3%	23,3%	26,7%	36,7%	,0%	100,0%
		% within Highesteducationaldegree	36,4%	41,2%	40,0%	91,7%	,0%	49,2%
		% of Total	6,6%	11,5%	13,1%	18,0%	,0%	49,2%
Total		Count	11	17	20	12	1	61
		Expected Count	11,0	17,0	20,0	12,0	1,0	61,0
		% within Havehair	18,0%	27,9%	32,8%	19,7%	1,6%	100,0%
		% within Highesteducationaldegree	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%
		% of Total	18,0%	27,9%	32,8%	19,7%	1,6%	100,0%

Figure 22: The cross tabular on highest finished educational degree and whether or not hair was present showed that there a clear predominance of people with a PhD degree in the group with images with hair (11) compared to the group where the hair was removed (1).

		F	Sig.
False Acceptances* Highest educational degree	Between Groups (Combined) Within Groups Total	4.780	0.002
False Rejections* Highest educational degree	Between Groups (Combined) Within Groups Total	1.488	0.218

Table 8: Anova test on the highest educational degree vs. False Acceptances shows that the difference between the number of false acceptances in the two groups due to educational degree is significant. The differences between the false rejections in this context is however not significant.

Highest educational degree		False Acceptances	False Rejections
0	Mean	10.91	1.09
	N	11	11
	Std. Deviation	7.341	0.944
1	Mean	8.65	1.53
	N	17	17
	Std. Deviation	6.274	1.419
2	Mean	5.50	0.80
	N	20	20
	Std. Deviation	3.663	1.196
3	Mean	2.25	0.67
	N	12	12
	Std. Deviation	3.911	0.778
4	Mean	11.00	0.00
	N	1	1
	Std. Deviation	-	-
Total	Mean	6.80	1.02
	N	61	61
	Std. Deviation	5.966	1.176

Table 9: The table shows a decrease in false acceptances as the highest finished educational degree increases. The mean value for those with the highest educational degree of junior high school or lower (0) is 10.91, the mean value for those with the highest educational degree of college or lower (1) is 8.65, the mean value for those with the highest educational degree of 1-3 years at a university or lower (2) is 5.50, the mean value for those with the highest educational degree of a masters degree or lower (3) is 2.25, while the mean value for those with the highest educational degree in form of a PhD or higher (4) is 11.00 (this last group consists however of one individual only).

D.4 Time

Measure	1. quartile	2. quartile	3. quartile	4. quartile
Mean	0.30	0.31	0.11	0.30
Std. Deviation	0.615	0.647	0.321	0.691

Table 10: The table shows a frequency distribution of the number of false acceptances within each quartile. The number of false acceptances remains approximately constant over time between the four time-frames.

Measure	1. quartile	2. quartile	3. quartile	4. quartile
Mean	69677.4590	65339.6066	64605.2623	60511.0164
Median	65854.0000	60246.0000	61698.0000	58024.0000
Std. Deviation	23962.57958	23406.41481	23919.48406	21234.38534

Table 11: The table shows a frequency distribution of the evaluation time within each quartile. As we can see from the table the comparison time decreases linearly over time.

		Total time	**False acceptances**
Total time	Pearson Correlation Sig. (2-tailed) N	1 61	0.266(*) 0.038 61
False acceptances	Pearson Correlation Sig. (2-tailed) N	0.266(*) 0.038 61	1 61

Table 12: This Pearson correlation shows whether or not there are a correlation between the number of false acceptances and the total time the participants used on the comparisons. The table shows that there are no correlation between the total time used on comparison and the false acceptance rate.