KUNGL
TEKNISKA
HÖGSKOLAN

HØGSKOLEN
I GJØVIK

NISlab
Norwegian Information
Security Laboratory

Passport of the Future:
Biometrics against Identity Theft?

Marijana Kosmerlj

Institutionen för
Data- och Systemvetenskap

Examensarbete 20 poäng
i data- och systemvetenskap
inom magisterprogrammet i informations- och kommunikations säkerhet,
Kungl Tekniska Högskolan

Examensarbete
Nr 2004-x-164
2002

Abstract

The purpose of biometric passports is to prevent the illegal entry of travellers into a specific country and limit the use of counterfeit documents by more accurate identification of an individual. After September 11, 2001, the interest in use of biometric-based passports in border control has increased rapidly. Biometric-based passports will include a high-capacity contactless smart card chip containing raw image files of the holder's face in addition to other identity information. In border control, a photo of the passport holder will be taken and compared to the image stored in the passport. The biometric system's threshold will decide whether the two images are sufficiently similar in order to verify the identity of the traveller.

There are many problems associated with biometric technologies such as error rates, spoofing attacks, non-universality and interoperability problems. Recent papers have proven how easy it is to spoof biometric systems using methods such as static images. This report goes one step further and takes a closer look at the adversaries and their capabilities in the border control environment. The traditional way of calculating the false acceptance rate of biometric systems might not give the true false acceptance rate in such an adversary environment. For example, the percentage of the adversaries who have at least twenty "look-alikes" in the target population from a biometric system's perspective might be a better indicator for the true false acceptance rate.

An overall security process that involves people, technology and procedures can overcome limitations of the biometric technologies. The biometric based passports will provide a new "speed bump" that will reduce identity theft by "zero-effort" and "small-effort" impostors. Smart adversaries with a large international network and many resources will be stopped by this "speed bump" only for a limited time - until they have discovered new ways of forging and counterfeiting passports.

Sammendrag (Abstract in Norwegian)

Formålet med biometriske pass er å forhindre ulovlig adgang av reisende inn i et land og å begrense bruken av forfalskede reisedokumenter ved en mer nøyaktig autentisering av reisende. Etter den 11. september 2001 har interessen for bruk av slike biometriske pass økt kraftig. Biometriske pass vil bestå av et høykapasitets smartkort som vil inneholde et bilde av passinnehaveren i tillegg til annen type identitetsinformasjon. I passkontrollen vil ett foto bli tatt av passeieren og sammenlignet med fotoet lagret i passet. Et systemdefinert parameter vil avgjøre om disse to fotoene er like nok til å fastslå om personen i de to fotoene er en og samme person.

Mange utfordringer er knyttet til biometriske systemer slik som feilrater, “spoofing”-angrep, ikke-universalitet og interoperabilitetsproblemer. Forskningen har vist hvor lett det er å lure biometriske systemer ved bruk av for eksempel statiske foto. Denne rapporten går ett skritt videre og tar en nærmere titt på fiender og deres ressurser i et grensekontrollmiljø. Den tradisjonelle måten å beregne feilakseptraten til biometriske systemer på, vil ikke gjenspeile den virkelige feilakseptraten i dette miljøet. For eksempel, vil andelen av fiendene som har minst tyve ”look-alikes” i målpopulasjonen sett fra et biometrisk systems perspektiv være en bedre indikator for den virkelige feilakseptraten.

Et sikkerhetsprogram som involverer mennesker, teknologi og prosedyrer kan kompensere for begrensninger i biometriske systemer. De nye biometriske passene vil representere en ny “fartsdemper” for små svindlere. Smarte bedragere med et stort internasjonalt nettverk og masse ressurser vil bli stoppet av denne “fartsdemperen” bare for en begrenset tid - til de har funnet nye måter å forfalske pass på.

Table of Contents

Acknowledgments

This study would not be possible without the help and the support from many people.
Professor Einar Snekkenes was very active in the guidance of this study. Erik Hjelmås, the Research Scholar at the Gjøvik University College, helped me to conduct the experiments. Asbjørn Hovstø, the leader of the Norwegian committee for biometrics K188, has shared with me some of his experiences with the biometric technologies and he supported and encouraged this study from the beginning.
Kjetil Einar Godø, the Section Leader and Eamonn J. Casey, the Technical Lead at the Norwegian Police Data and Procurement Service, have supported this study by providing both the practical help and valuable information.
Louise Yngström, the Associate Professor at the Department of Computer and Systems Sciences at the KTH Royal Institute of Technology in Sweden has given me valuable feedbacks on this study in the very beginning.
Kostas Papadopoulos, the Product Manager at Epsys in Norway, introduced me to the smart cards and biometrics.

The face images used in this work have been provided, among others, by the Computer Vision Laboratory, University of Ljubljana, Slovenia [39], Computer Vision Center (CVC) at the U.A.B. [38], Centre for Vision, Speech and Signal Processing at the University of Surrey [21] and the Gjøvik University College [42].

I want to thank all of these people for their contributions.
Finally, I want to thank my family and my friends for the patience and understanding they have shown for my work.

Preface

It will be most productive to think of security not as a way to provide ironclad protection, but the equivalent of speed bumps, decreasing the velocity and impact of electronic attacks to a level where other protection mechanisms can operate.

––Andrew M. Odlyzko

Since the preface chapter is the only chapter in a scientific report where the author can be personal, I could not resist including it in this report. This study is as much personal as it is scientific.

I want to compare this study with a journey which goal is to climb to the top of a very high mountain. I have been climbing for several months now and I can see the top of it, but I will not manage to reach it by the end of the journey. Nevertheless, I am not sad for this; I am happy that I took this challenge and experienced many exciting views.

Oslo, June 2004
Marijana Kosmerlj

1 Introduction

The International Civil Aviation Organization has adopted a global, harmonized blueprint for the integration of biometric identification information into machine readable passports. The purpose of the new biometric passports is to prevent the illegal entry of travellers into a specific country and to limit the use of fraudulent documents by more accurate authentication of individuals. This study aims to find out to what extent the integration of biometric identification information into passports will improve their robustness against identity theft.

1.1 Passport of the future

After September 11, 2001, the interest in use of physiological and behavioural characteristics to identify and verify identity of an individual has increased rapidly worldwide. These physiological and behavioural characteristics are believed to be distinct to each individual and can therefore be used to increase the binding between the travel document and the person who holds it.

The International Civil Aviation Organization (ICAO), which plays a major role in setting global travel standards, has adopted a global, harmonized blueprint for the integration of biometric identification information into passports and other machine readable travel documents [1], [2]. The blueprint requires that a high-capacity contact-less integrated circuit containing a raw image file of the holder's face in addition to other identity information such as name and date of birth be included in the machine readable passports and other travel documents. Inclusion of the additional biometric technologies, fingerprint and iris, are recommended, but not mandatory.

The purpose of biometric passports is to prevent the illegal entry of travellers into a specific country and limit the use of fraudulent documents, including counterfeit and modified documents and the impostor's use of legitimate documents [2].

Since the facial image is the mandatory biometric identifier to be included in the future passports, we will in this study focus on the use of the facial image for the identity verification of passport holders.

1.2 Need for the study

Security does not come for free, and therefore it is important to find out what are the benefits of the new technology compared to the costs associated with it.
A new technology assessment report by the USA's General Accounting Office [2] estimates the initial cost associated with using biometric based passports for border security in the USA in excess of US$8 billion. The consequences can be too big if such an expensive solution is implemented before further research is done to find out if and to what extent the integration of biometric identification information into passports will improve their robustness against the identity theft.

1.3 Purpose of the study

The primary objective of the study is to produce new knowledge with respect to security of biometric techniques in a national border control setting. The results of the work should be useful for those making design decisions with respect to biometric technologies in a border control setting or other large-scale applications.

1.4 Statement of the problem

The purpose of biometric passports is to prevent the illegal entry of travellers into a specific country and to limit the use of fraudulent documents by more accurate identification of individuals.
It is interesting to find out to what extent the integration of biometric identification information into passports will improve their robustness against identity theft.

1.5 Research questions

In order to find out to what extent the integration of biometric identification information into passports will improve their robustness against identity theft, the following questions need to be answered:

- How robust are today's machine readable passports against identity theft?
- How different/similar are people in terms of their biometric identifiers from a biometric system's perspective?
- How should the false acceptance rate be measured in a border control setting?

1.6 Delimitations

This study focuses on the ability of the biometric authentication and the face technology to prevent identity theft in a border control setting with an assumed adversary environment. The adversaries and their capabilities are described in Chapter 6.

1.7 Outline of chapters

Chapter 2 introduces biometric recognition and challenges in the world of biometric systems. Chapter 3 presents the answers to the research questions found in the related literature. In Chapter 4, we present the type of research design used in the study.
Chapter 5 is our contribution to the second research question: How different/similar are people in terms of their biometric identifiers from a biometric system's perspective? This chapter describes the experimental design and the experimental results.
Chapter 6 is our contribution to the third research question: How should the false acceptance rate be measured in a border control setting? In this chapter, we describe

the border control application, its adversaries and their capabilities, and then propose another measure for the false acceptance rate.
The answers to the research questions found in the related literature and by our own contributions are discussed in Chapter 7. Chapter 8 summarizes the conclusions of the study. Some suggestions for future work are given in Chapter 9. Directions on how to obtain a copy of the data sets and tools used in the experiment can be found in Appendix A. Appendix B contains a simple list of issues that should be considered when choosing a biometric system. The most important standards related to the travel documents and the biometric technologies are given in Appendix C.

2 Theory

Biometric technologies are automated methods of recognizing an individual based on their physiological or behavioural characteristics such as face, iris and fingerprints. Biometric systems are applications of biometric technologies and can be used to verify the claimed identity of a person and to establish a person's identity.

In an ideal biometric system, every person possess the characteristic, no two persons have the same characteristic, the characteristic remain permanent over time and does not vary under the conditions in which it is collected and the biometric system resists countermeasures. Evaluation of biometric systems quantifies how well biometric systems accommodate the properties of an ideal biometric system. All of existing biometric systems suffer from the same problems: false acceptance and false rejection caused by the variability of conditions at the human-machine interface. A common feature of any system that uses biometric is a trade-off between high security and a more usable system.

2.1 Biometric technologies and systems

Biometric technologies are automated methods of recognizing an individual based on their physiological or behavioural characteristics which are thought to be distinctive to each person [1], [23]. Some examples of common biometric technologies are face, fingerprint, iris, dynamic signature and voice.

Every biometric identifier (characteristic) can be characterized by the following properties [34]:

- Universability – every person should have the characteristic
- Uniqueness – any two persons should be sufficiently different in terms of the characteristics
- Permanence – the characteristic should be sufficiently permanent over a period of time
- Collectability – the characteristic can be measured quantitatively

In addition to above properties, the following properties should be considered in a practical biometric system [34]:

- Performance – an achievable identification accuracy, the resource requirements to achieve an acceptable identification accuracy, and the working factors that affect the identification accuracy
- Acceptability – refers to what extent people are willing to accept the biometric system
- Circumvention – refers to how easily is or how much effort is required to fool the biometric system by fraudulent techniques

No biometric identifier is perfect [3]. Each biometric identifier has its strengths and limitations. Table 1 gives an overview of the strength and limitations of a selection of biometric technologies.

Table 1 – Comparison of biometric technologies (source: [3], Table 1.1 on page 16)

Biometric technology	Universality	Uniqueness	Permanence	Collectability	Performance	Acceptability	Circumvention
Face	High	Low	Medium	High	Low	High	Low
Fingerprint	Medium	High	High	Medium	High	Medium	High
Hand Geometry	Medium	Medium	Medium	High	Medium	Medium	Medium
Iris	High	High	High	Medium	High	Low	High
Retinal scan	High	High	Medium	Low	High	Low	High
Signature	Low	Low	Low	High	Low	High	Low
Voice Print	Medium	Low	Low	Medium	Low	High	Low
Facial thermograms	High	High	Low	High	Medium	High	High

Biometric systems are applications of biometric technologies. Figure 1 shows a generic biometric system that consists of five subsystems: data collection, transmission, signal processing, decision and data storage [11].

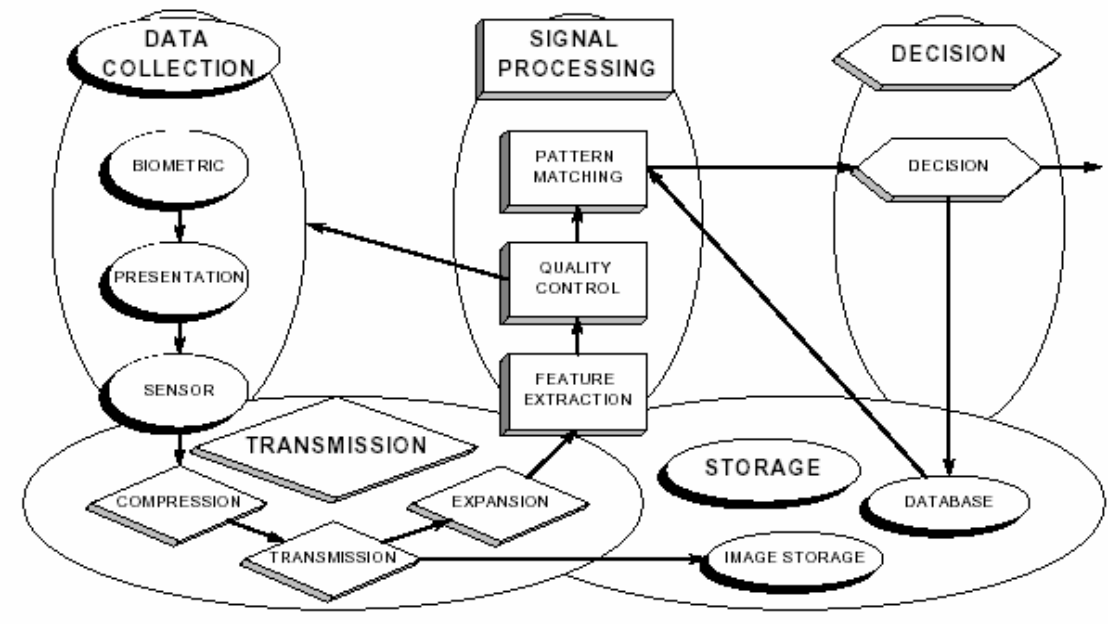


Figure 1 – Generic biometric system (source: [11], Figure 17.1 on page 348)

In the data collection subsystem, the biometric characteristic of a person is scanned by a biometric sensor device to acquire a digital representation of the characteristic.

The digital representation is then transmitted to the signal processing subsystem, where the digital representation is further processed to generate a compact but expressive representation, called a template (also called biometric identifier or biometric signature).
The generated template is the input to a pattern matching algorithm that compares the template with previously stored templates to establish similarity score between the templates. The resulting similarity score and the system threshold are then sent to the decision module that will decide whether the templates are sufficiently similar.
The templates can be stored in a database or in a smart card issued to the individual. When a person's template has been stored in the database or in the smart card, we say that the person has been enrolled into the biometric system.

Biometric systems can operate in three modes [23]:
- Verification is used to verify a person's claimed identity
- Identification is used to establish a person's identity from a list of known individuals
- Watchlist is used to find out if a person is present on the list of known individuals and, if so, to establish the person's identity.

2.1.1 Verification

In the verification task, the user claims an identity and the biometric system determines whether the user's claim is correct or not [23]. From the biometric system's perspective, the user's claim is correct if the similarity score between the claimed and the biometric identifier of the user is above the system's verification threshold. The user's claim is not correct if the similarity score is below the system's verification threshold.

Biometric system's decision can be correct or false. The decision is correct in the two cases:
- the biometric system has decided that the user's claim is correct and the claimed and the biometric identifier of the user belong to the same person. In this case we say that a genuine individual has been correctly accepted or verified by the biometric system. The probability of making the correct verification is called the probability of correct verification or probability of verification (PV)
- the biometric system has decided that the user's claim is not correct and the claimed and the biometric identifier of the user do not belong to the same person. We say that an impostor has been correctly rejected by the biometric system.

The decision is not correct in the two cases:

- the biometric system has decided that the user's claim is correct and the claimed and the biometric identifier of the user do not belong to same person. We say that an impostor has been falsely accepted as a genuine user. This error is called false acceptance, and the probability of making the false acceptance is called the false acceptance rate (FAR)
- the biometric system has decided that the user's claim is not correct and the claimed and the biometric identifier of the user belong to the same person. We say that a genuine individual has been falsely rejected. This error is called false rejection, and the probability of making the false rejection is called the false rejection rate (FRR). The probability of verification is equal to one minus the false reject rate.

The FAR and FRR come in pairs: a smaller FRR implies a larger FAR and vice versa. The trade off between FAR and FRR depends on the security and throughput requirements of the system and are highly dependent upon the specifics of the application environment [7] . Equal error rate (EER) is the point where FAR and FRR are equal.

Receiver Operating Characteristic (ROC) and Detection Error Trade-off (DET) curves are used to show the performance of biometric systems [8].
The ROC curve shows the relationship between the probability of verification and the false acceptance rate. The DET curve is a modified ROC curve that shows the relationship between the false rejection rate and false acceptance rate. Figure 2 below is an example of a DET curve.

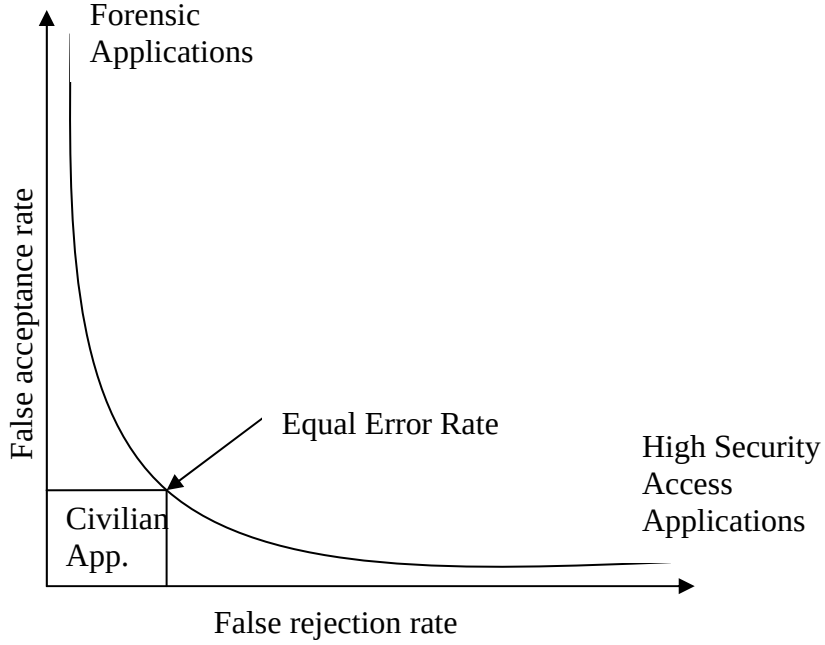


Figure 2 – An example DET curve (source [3], Figure 1.25 on page 33)

High-security access applications require a small FAR because they want only genuine persons to be granted access. In forensic applications where the desire to

catch a criminal outweighs the inconvenience of examining a large number of falsely accused individuals operate at a small FRR. Civilian applications usually operate at EER (Equal Error Rate) [3].

2.1.2 Watchlist

In the watchlist task, the biometric system determines if the individual's biometric identifier matches a biometric identifier of someone on the watch list [23].
The biometric system's output is the list of the individuals that matches the unknown individual's biometric identifier sorted by the similarity score in ascending order.
If the individual to be found is present on the watch list, and the individual with the highest similarity score returned by the biometric system is the correct individual, then we say that the biometric system has correctly detected and identified the person.
The probability of making correct detect and identify decision is called correct detect and identify rate.

If the individual to be found is present on the watch list and the resulting list does not include the correct person, or the individual with the highest similarity score returned by the biometric system is not the correct individual, then we say that the biometric system has made a false alarm error. The probability of making false alarm is called false alarm rate.

There is a trade off between the probability of correct detect and identify rate and the false alarm rate. If we increase the probability of correct detect and identify rate, the false alarm rate will increase. A Watchlist Receiver Operating Characteristic curve is used to show the relationship between the probability of correct detect and identify rate and the false alarm rate. Figure 3 below is an example of the Watchlist ROC curve.

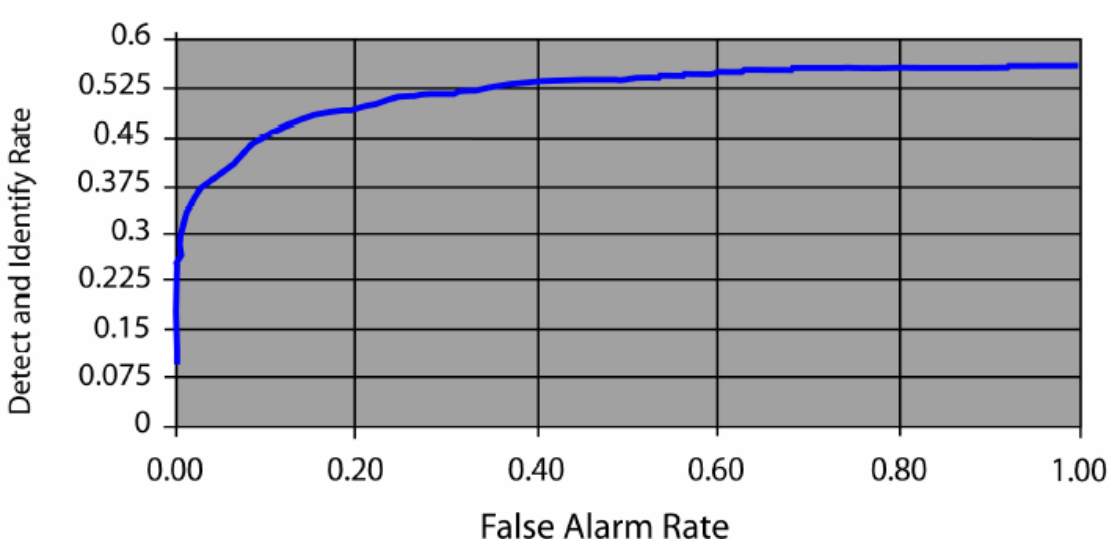


Figure 3 - An example Watchlist ROC curve (source [23], Figure 8 on page 10)

The size of the watch list affects the performance of watchlist tasks. Face Recognition Vendor Test 2002 shows that the watchlist performance (detect and identity rate) decreases linearly with the logarithm of the number of the users on the watch list [20].

2.1.3 Identification

In the identification task, a biometric signature of an individual is presented to the biometric system, and the task of the biometric system is to establish the identity of the user [23]. The identification task is a special case of the watchlist task. In the identification task, we know that the user to be identified is present in the user database.

The similarity scores between the user's biometric identifier and the biometric identifiers in the database are calculated. If the correct user has the highest similarity score, then we say that the biometric system has made correct identification at rank one. If the correct user has the next highest similarity score, we say that the biometric system has made correct identification at rank two and so on.

The probability of identification at rank k is defined as the probability that the user's true identity is within the top r matches returned. The cumulative match characteristics (CMC) curve plots the probability of identification versus the top r matches. Figure 4 is an example of the CMC curve.

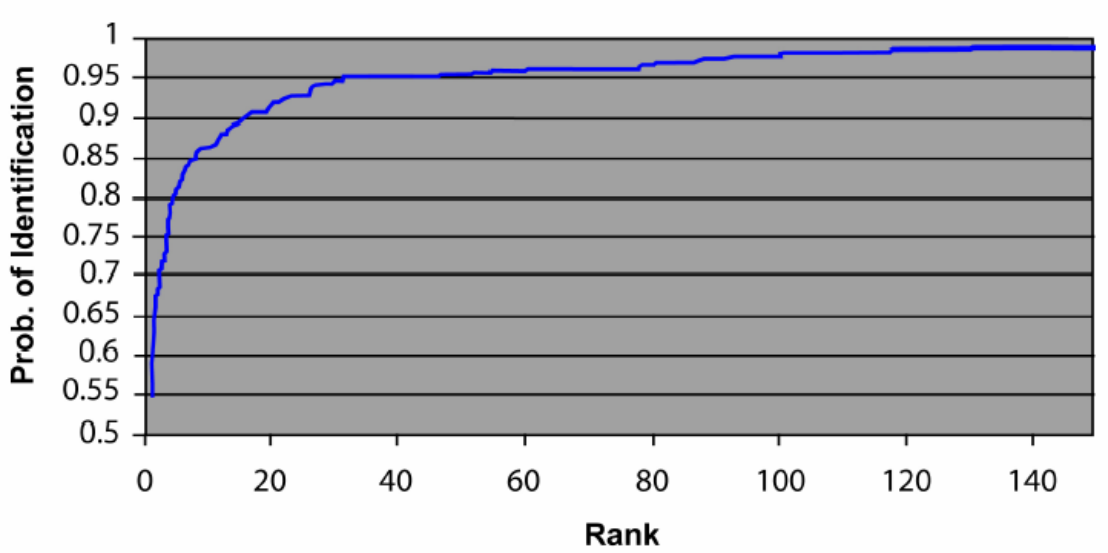


Figure 4 – An example CMC curve (source [23], Figure 10 on page 13)

The number of the users in the database affects the identification performance. Face Recognition Vendor Test 2002 shows that the identification performance decreases linearly with the logarithm of the number of registered users [20].

2.2 Evaluating biometric systems

2.2.1 Ideal biometric system

In order to understand the strengths and weaknesses of the biometric systems, it is important to understand the properties of an ideal biometric system [4].
An ideal biometric system has the following properties:

- every person possess the characteristic that the biometric identifies
- no two persons have the same characteristic
- the characteristic remain permanent over time and does not vary under the conditions in which it is collected
- the biometric system resists countermeasures

Evaluation of biometric systems quantifies how well biometric systems accommodate the properties of an ideal biometric system [4].

2.2.2 Best practices in testing and reporting performance of biometric devices

The major player in developing standardized evaluation methods for biometric systems is the UK Biometrics Working Group (BWG). BWG is a UK governmental organization that has developed "best practices" standards for testing and reporting on biometric system performance [8]. This standard is based on "An introduction to Testing Biometric Systems" [4].

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) have also been working on development of testing and reporting methodologies and metrics that cover biometric technologies, systems and components. So far, a new work item proposal for a multi-part standard (ISO/IEC 19795) was approved and a call for contributions was issued. The draft has been developed from the BWG's "best practices" document.

According to the best practices, the evaluations of biometric systems should satisfy the following requirements [4]:

- the test should be administered by independent groups and tested on biometric data not previously seen by a system
- the details of the evaluation protocol, performance results and representative examples of the data set should be published so that others could repeat the evaluations
- an evaluation protocol must also not be too difficult or too easy in which case it would not be possible to make difference in performance between different vendors
- a three-step evaluation plan: a technology evaluation, a scenario evaluation and an operational evaluation.

2.2.3 Technology evaluation

The technology test evaluates the technology itself: it measures the performance of the matching algorithms under controlled conditions in a laboratory. The purpose of a technology evaluation is to measure the state of the art and to determine the most promising approaches [4]. In a technology evaluation, the algorithms to be tested are given a database of biometric identifiers (e.g. facial images). One part of the database is given to the participants so that they can be familiar with the biometric identifiers in the database and the other part is used for testing. The results from a technology test are repeatable since the technology tests are done under controlled conditions. The products of the technology evaluation are the verification, identification and watchlist performance metrics.

2.2.4 Scenario and operational evaluations

Scenario evaluations test complete biometric systems under conditions that model a class of applications such as access control [4]. The purpose of the scenario tests is to determine if the algorithms that performed well in the technology evaluation perform well under the field conditions too. Scenario evaluations cannot be repeated because they test systems under the field conditions.

The operational test measures the overall performance for a specific algorithm for a specific application [4]. The purpose of an operational test is to find out if a biometric system meets the requirements of a specific application. One example of such test is a test of a face recognition algorithm for use in the border control.

The scenario and operational evaluation of biometric systems measure in addition to the metrics used in the technology evaluation, several other metrics:

- failure to enrol: the percentage of the subjects who could not be enrolled under the pre-determined enrolment policy [8] [40],
- failure to acquire: the percentage of the subjects for which the biometric system was unable to acquire the biometric sample of a satisfactory quality [8] [40],
- throughput rate/user throughput: the number of users processed per unit time based on both computational speed and human–machine interaction [40].

2.2.5 Factors affecting performance of biometric systems

The performance of the biometric systems in technology, scenario and operational evaluation depends on the three main factors:

- the biometric technology,
- the environment in which the biometric data is collected and tested,
- the characteristics of the persons in the data set.

Each biometric technology has its strengths and limitations that will affect the performance of biometric systems. Some of them have a high degree of distinctiveness (e.g. fingerprint, iris), others have a lower degree of distinctiveness

(e.g. face). Some will change over a period of time, and others will not. Table 2 gives an overview of factors that affect performance of biometric systems.

Table 2 – Factors affecting performance of biometric systems (source [40], Table 2-1 on page 37)

Biometric	Factors causing errors
All	Template aging as a result of age, gender, other factors
Fingerprint	Degradation of fingerprints caused by occupation, age, trauma
Face	Environmental factors such as lighting, background, contrast; pose, movement of subject, glasses
Iris	Positioning, eye angle, glasses
Voice	Illness, age, quality of communication system
Hand	Injury, trauma to hand

The users of the application and the environment in which the biometric system will be used will affect the performance [11]. The habituated users will usually produce lower error rates than the less habituated users. The biometric system will produce higher error rates if it is placed in a stressful location in comparison to a location where the conditions that can influence the performance are controlled. The number of users affects the performance of watchlist and identification tasks [20].

The characteristics of the persons in the data set will also affect the performance of the system. Recent studies have shown that there are differences in recognizability of different persons with respect to their biometric identifiers such as fingerprints and speech [10], [29]. Some persons will produce higher error rates than other persons.

The performance of a biometric system in a real-world application will in addition be dependent on the fraud rate [11]. The fraud rate is the percentage of the population that is attempting to defraud the system.

2.3 Problems with biometric technologies

All of existing biometric systems suffer from the same problems: false acceptance and false rejection caused by the variability of conditions at the human-machine interface [3], [2]. A common feature of any system that uses biometric is a trade-off between high security and a more usable system.

Performance of biometric technologies depends greatly on how and where they are deployed and tested [3]. The test results of biometric systems have proven to be more impressive than real life performance data. Biometric companies have been testing the accuracy and their technologies in highly controlled environments, using static or artificially generated templates.
Most live checks in biometric systems do not work efficiently. Recent publications [18], [19] show that with little effort many leading biometric technologies are

susceptible to spoofing attacks in which fake fingerprints, static facial images, and static iris images are used to fool the biometric systems.

More research needs to be done in the field of biometric identification [3]: metrics used for assessment of quality of a biometric measurement are subjective and inconsistent; extraction of invariant and discriminatory information from a given measurement is still an extremely difficult problem.

Another problem of biometrics is non-universality: for each biometric there are always persons who do not show the characteristic to a degree sufficient for capturing with a biometric device or who do not posses the characteristic at all (e.g., the people who miss one or several fingers will not be able to use a fingerprint sensor) [5].

In today's biometric industry, the majority of the biometric devices and their software are proprietary and as such may hide possible flaws and may be incompatible with other biometric devices or software [2].

Some of the limitations of biometric technologies can be overcome by an overall security process that involves people, procedures and technology [47]. For example, procedures for handling people who are not able to enrol into a biometric system should be developed.

2.4 Biometrics in passports

There are several scenarios for use of biometric technologies in the travel documents [1]:

- when applying for a new passport the biometric identifier of the applicant can be compared to the existing identifiers in diverse databases in order to determine whether the applicant holds a passport under a different identity
- the identity of the applicant can be verified against the one stored in the passport when she comes to collect her passport
- at the border, the captured biometric identifier can be used to verify the passport holder’s identity and/or to determine whether the passport holder is a known criminal or terrorist.

More information on the use of biometric technology in the future travel documents can be found in the standards listed in Appendix C.

3 Literature Review

Today's machine readable passports are exposed to many threats such as reproduction and alteration of passports. In order to improve the security of the machine readable passports, the International Civil Aviation Organization has published two sets with security features that the issuing states are recommended to implement. The security of the passports will depend on the capabilities of potential adversaries will vary between different countries since each country is free to decide which of the security features to implement.

Iris is the most distinctive biometric characteristic, followed by fingerprint and face. Multiple biometrics can provide higher distinctiveness in addition to higher reliability and robustness against spoofing attacks.

The traditional way of calculating the false acceptance rate is based on ”zero effort” impostors. There is a need for a more realistic false acceptance rate in an adversary environment.

3.1 Today's machine readable passports

3.1.1 Threats to the security of today's machine readable passports

The machine readable passports are exposed to the following threats [22]:

- reproduction of a complete passport or construction of a fraudulent document using materials from legitimate documents
- alteration of a passport:
  - photo substitution
  - alteration of the text on the biographical data page
- theft of genuine document blank
- impostors with assumed identity and altered appearance

3.1.2 Security of today's machine readable passports

ICAO has published two sets with the security features that the issuing states can use in order to improve the security of their machine readable passports [22]. The purpose of the security features is to prevent against unauthorized reproduction and alteration of travel documents and other forms of tampering, to detect altered documents and to enable authentication of the document.

The first set contains the most essential security features. The other set contains additional security features that will provide an enhanced level of security. ICAO recommends the issuing states to implement all of the essential security features and to select one or more additional security features after they have completed a full risk assessment of their travel documents.

The security of travel documents will vary between different countries since each country is free to decide which of the security features to implement.
According to Anita Alden (the National Criminal Investigation Service in Norway), the Norwegian machine readable passports are robust against the alteration: it is not possible to alter the photo page of a Norwegian machine readable passport without destroying the passport [30]. No forged or counterfeit Norwegian machine readable passport has been discovered so far.[1]

The security of the machine readable passports will depend on the effect of the implemented security controls on the potential adversary. One measure of the effect of a security control is the effort needed for an adversary to defeat the security control. This effort is an indirect measure for the capabilities of the potential adversary [46].

3.1.3 Identity theft

According to a report on identity fraud [14], written by United States General Accounting Office, the prevalence of identity theft appears to be growing. In years from 1999 through 2001, the border control inspectors in the US have intercepted over 100,000 fraudulent documents annually.

The number of fraudulent travel documents has also been growing in Norway in the last years. The National Criminal Investigation Service in Norway has registered 81 fraudulent passports in 1998 compared to 190 in 2003 [31]. By the end of 2003 the Norwegian police authorities have registered ca. 37 000 stolen Norwegian passport in the Schengen-database SIS [30]. Only one of these passports was a machine readable passport.

The real number of fraudulent documents is difficult to measure and it is probably higher than the number of detected fraudulent documents.

One of the threats to today's machine readable passports is the impostors with assumed identity and altered appearance. One interesting question in this context is how good the border control inspectors are to detect impostors with assumed identity and altered appearance. It is difficult to answer this question since the number of falsely accepted impostors is difficult to estimate.

[1] Information provided by the Norwegian Police Data and Procurement Service

3.2 Distinctiveness of biometric identifiers

The false acceptance rate is an indicator of how distinctive we humans are from a biometric system's perspective [32]. This chapter gives an overview of false acceptance rates in the face, fingerprint and iris biometric technologies.

3.2.1 Fingerprints

FVC2002 [44], the Second International Competition for Fingerprint Verification Algorithms, shows that the current state of the art in fingerprint recognition is 99.8% verification at 0.2% false accept rate.

S. Pankanti, S. Prabhakar, and A. K. Jain have developed a theoretical model for individuality of fingerprints (the probability of correspondence between fingerprints which belong to different persons) and tested this model against the empirical results obtained from an Automatic Fingerprint Matching System. The results show that the individuality of fingerprints is much lower in practice than in theory [16].

Several experiments on similarity of different biometric identifiers have been conducted recently. K. Jain, S. Prabhakar, and S. Pankanti demonstrate that the state-of-the-art fingerprint verification systems have lower accuracy when tested with identical twin fingerprints in comparison to non-twin fingerprints [15]. Table 3 shows the results from their tests.

Table 3 – False acceptance and false rejection rates with different threshold values for the twin-twin and twin-nontwin matchings in an identical twin database (source: [15], Table 1 on page 2660)

Threshold	FRR (%)	FAR (twin–twin) (%)	FAR (twin–nontwin) (%)
16	1.05	8.51	2.20
20	2.20	4.79	1.02
24	3.00	2.13	0.48
26	3.49	1.06	0.29

K. Jain, S. Prabhakar, and S. Pankanti also demonstrate that the fingerprint verification systems have lower accuracy when tested with the same class of fingerprints in comparison to different classes of fingerprints [15]. Table 4 below shows the results from their tests.

Table 4 – False acceptance and false rejection rates with different threshold values for within-class and between-class matchings in the NIST4 database (source: [15], Table 3 on Page 2661)

Threshold	FRR (%)	FAR (within-class) (%)	FAR (between-class) (%)
16	1.05	10.65	6.11
20	2.20	6.30	3.24
24	3.00	3.83	1.82
26	3.49	3.01	1.39

3.2.2 Face

Several independent technology evaluations in face recognition community has been conducted in the period from 1996 to 2002 [6], [20], [43] in order to measure the effect of numerous variables such as pose, lighting and temporal changes on the performance of face recognition algorithms.
These evaluations show that the face has low discrimination capability: variations in the template of a single face (within-class variations) under different conditions such as pose, illumination, expression and temporal changes are too big in comparison to variations in the templates of different faces (between-class variations) under constant conditions [17]. The relation between within-class and between class variations sets the limits to the performance of the face recognition systems.
The minimum false acceptance rate of these systems will be equal to the birth rate of identical twins (0.82 %) since the face recognition systems are not able to reliably distinguish between identical twins [17].

The Face Recognition Vendor Test 2002 [20], an independently administered technology evaluation of mature face recognition systems, shows that:

- the current state of the art in face recognition is 90% verification at 1% false accept rate under the assumption of the controlled indoor lighting
- the elapsed between the enrolled and new images of a person affects the performance: for the top systems, performance degraded at approximately 5% per year
- identification and watchlist performance decreases linearly with the logarithm of the database size
- males are easier to recognize than females and older people are easier to recognize than younger people.

3.2.3 Iris

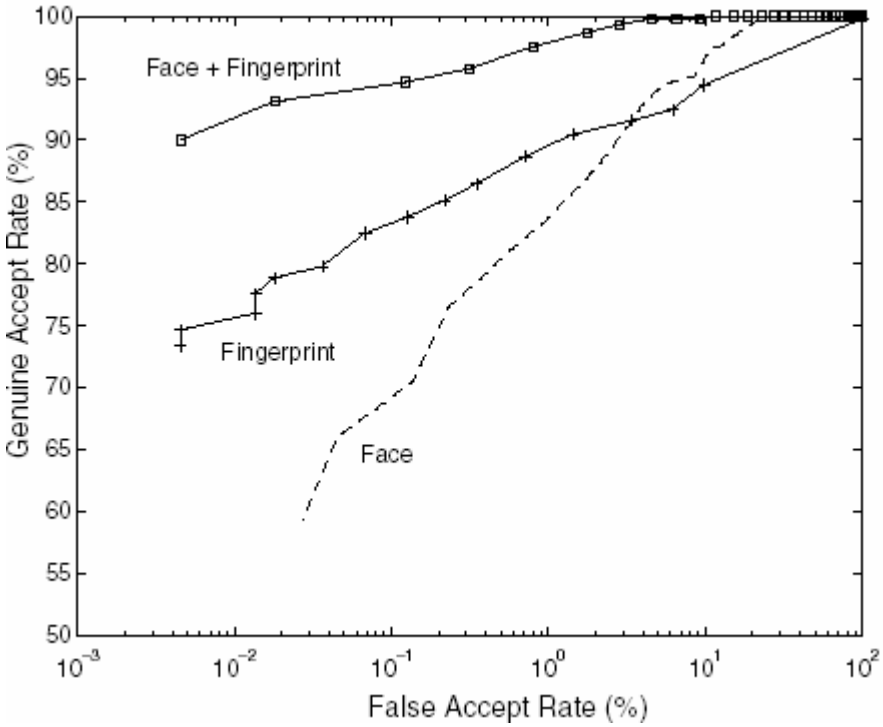
Iris is a biometric identifier with a high level of distinctiveness. Iris patterns of genetically identical twins appear to be as uncorrelated as they are among unrelated eyes [35].

Algorithms developed by John Daugman [36][37] for encoding and recognizing iris patterns have been tested by several different organizations (British Telecom, US Sandia Labs, UK National Physical Lab, NBTC, Panasonic, LG, Oki, EyeTicket, IBM SchipholGroup, Joh.Enschede, IriScan, Iridian, and Sensar). All testing organizations have reported a false match rate of 0 in their tests which is in accordance with the theoretically calculated equal error rate of 1 in 1.2 million [17].

3.2.4 Multimodal biometric systems

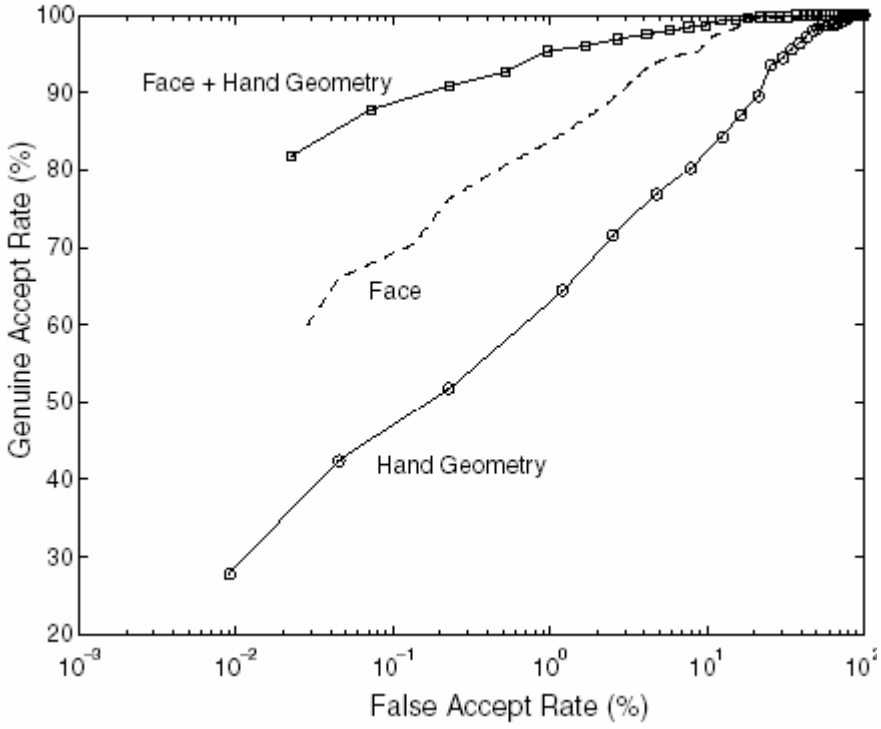
Experimental results demonstrate that the integration of multiple biometrics can provide better verification performance than the individual biometrics [24], [25], [26], [27], [28]. Multiple biometrics will also increase robustness of the biometric systems against the spoofing attacks and solve the problem of non-universality [28].

It is shown that the integration of face and fingerprint or face and hand geometry provides better verification performance than the individual biometrics [24] and [28]. The ROC curves below show an improvement in verification performance when matching scores from multiple biometric identifiers are combined using the sum rule.



Threshold	Genuine Accept Rate (%)	False Accept Rate (%)
41	100.00	49.19
43	99.77	33.28
46	99.55	6.77
47	98.66	2.20
48	97.55	0.32
51	84.00	0.00

Figure 5 - A combination of fingerprint and face matching scores provides better verification performance (source: [24], Fig. 9 on page 2122)



Threshold	Genuine Accept Rate (%)	False Accept Rate (%)
57	100.00	1.59
58	99.77	0.65
59	99.33	0.19
60	98.22	0.03
62	95.55	0.00

Figure 6 – A combination of face and hand geometry matching scores provides better verification performance (source: [24], Fig. 9 on page 2122)

The integration of face and speech [26], [27] and the integration of face and iris [33] show similar results.

According to Ross Anderson, a combination of a very good biometric and a very imprecise one will in some cases result in the worse overall performance [45].

3.3 False acceptance rate and "zero-effort" impostors

The way in which the false acceptance rate and the false rejection rate are calculated is defined in a testing protocol. One example of such a protocol is given in [11]. In this protocol, a matching algorithm is given two data sets: a gallery and a probe set. The gallery set consist of known subjects (persons who are registered in the biometric system) while the probe set consist of subjects to be recognized. The probe set can contain both the subjects present in the gallery set and the subjects not present in the gallery set. The matching algorithm calculates similarity scores between images in these two sets. The result of the computation is a similarity matrix that is used to compute the algorithm error rates.

In "best practices" standard for testing and reporting on biometric system performance [8], the calculation of the false acceptance rate is based on the "zero effort" impostors. These impostors submit their biometric identifier as if they were attempting successful verification against their own template.
The authors of the "best practices" document realise the need for another measure for the false acceptance rate in case of "non-zero effort" impostors.

In Face Recognition Vendor Test 2002, the true false acceptance rate is introduced based on the comparisons between the gallery images and the images of the persons not present in the gallery set [13]. This way of computing the false acceptance rate models the situation in which a person who does not already have access to the system attempts verification. Even though this model might represent a more realistic situation, the calculation of the true false acceptance rate is still based on the "zero effort" impostors.

Thus, in the environments where it is realistic to assume that impostors will actively try to fool a biometric system, the false acceptance rate computed in the traditional way will not be representative for the actual percentage of impostors falsely accepted by the biometric system. James L. Wayman claims that the actual percentage of users falsely accepted will be the product of the false acceptance rate and the fraud rate [32]. The problem with this formula is that it is difficult to estimate the actual fraud rate (see Chapter 3.1.3).

4 Method

One of the criteria for selecting the research approach was the research problem itself. We found out that we needed to use a mixed research approach in order to solve the research problem [48]. The mixed research approach combines the quantitative and qualitative methods of collecting and analysing data.

We used the following methods:
- literature study (a qualitative method)
- experiment (a quantitative method)
- use of known theories for constructing new theories (a qualitative method).

Literature study gave answer to the first research question and partial answers to the last two questions. The experiment was conducted to find out more information on similarity of human faces from biometric systems' perspective. Since little information was found on the intelligence and capability of the adversaries in a border control setting, we have used the existing theories and models[2] to model an adversary environment and to propose a new measure for false acceptance rate based on the modelled adversary.

[2] The course in Security Metrics at the Gjøvik University College

5 Experimental Design and Results

One measure of similarity of human faces from biometric systems' perspective is the percentage of people having one or more look-alikes in a generic population.
The experiment described in this section estimates the percentage of Norwegian people having one or more look-alikes in the Norwegian population.

5.1 Experimental design

5.1.1 Data sets

Subjects in the experiment are selected from several face databases: the Ljubljani CVL Face Database, the XM2VTS database, the AR database, a collection of photos of students and employees at Gjøvik University College (we will refer to this collection as the HIG photos) and a collection of the Norwegian passport photos (we will refer to this collection as the passport photos):

- Ljubljani CVL Face Database [39] - 114 persons, 7 images per person; resolution 640x480; colour images; uniform background and lighting; frontal and side views; varying facial expressions
- XM2VTS Database [21]: 295 subjects, 4 recordings per subject over 4 months; colour images; resolution 720x576; varying posture; varying illumination (controlled)
- AR Face Database [38]: 126 subjects, 2 recordings per subject separated by two weeks time (13 images per recording); colour images; resolution: 768x576; images feature frontal view faces with different facial expressions, illumination conditions, and occlusions
- HIG photos [42]: 2762 subjects, 1 image per subject, colour images; resolution 194x234; varying facial expressions; varying posture; varying illumination. 45 % of the subjects are males; approximately 83 % of the subjects are between 18 and 35 years old. Figure 7 below shows a couple of example photos.
- Passport photos: several thousands of the Norwegian passport photos, 1 image per subject; varying facial expressions; varying posture; varying illumination.

Figure 7 – Example photos from the HIG database

The CVL, AR and XM2VTS databases have been used before in various experiments whereas the HIG and passport photos have never been used before.

In order to limit the effect of side views, lighting conditions and occlusions (such as sunglasses and scarfs) on the verification performance, frontal and approximately frontal facial images without occlusions and with varying but controlled lighting conditions were selected for the experiment. The expression and pose are varying among the different subjects, which are also expected to vary in the border control setting.

5.1.2 Instrumentation

To study similarity of facial images we needed to calculate the similarity score (distance) between every facial image in our database and all the other facial images. We found out that we could use the CSU Face Identification Evaluation System 5.0 [12] to generate similarity scores.

The CSU Face Identification Evaluation System is developed at the Colorado State University and provides standard face recognition algorithms and software to support statistical analysis techniques that aid in evaluating the performance of face identification systems.

CSU Face Identification Evaluation System can be split into four basic phases:

- image data pre-processing: pre-processing reduces unwanted image variation by aligning the face imagery, equalizing the pixel values, and normalizing the contrast and brightness
- algorithm training: the face recognition algorithms are trained to recognize human faces with a set of example (training) images
- algorithm testing: the face recognition algorithms are given a set of images for which they will create the distance matrix
- analysis of results: the identification performance statistics are calculated based on the resulting distance matrices.

The software includes four different facial recognition algorithms:

- A standard Principal Component Analysis (PCA) algorithm
- A combined Principal Component Analysis and Linear Discriminant Analysis (LDA) algorithm
- A Bayesian Intrapersonal/Extrapersonal Image Difference Classifier (BIC)
- An Elastic Bunch Graph Matching Algorithm (EBGM)

More information about these algorithms can be found in the CSU Face Identification Evaluation System User's Guide: Version 5.0 [12].

We used only the three algorithms from the CSU software in the experiment. The training phase for these algorithms is very similar and they can use the same training images. Face recognition algorithms included in the CSU software require the eye coordinates to be given. Two matlab scripts written by Erik Hjelmås, the research scholars at the Gjøvik University College (http://w3.hig.no/~erikh/) were used for automatic and manual detection of eye coordinates.

5.1.3 Experimental procedures

The purpose of the experiment is to find out the percentage of Norwegian people who have one or more "look-alikes" in the Norwegian population.
Figure 8 shows the main steps in the experiment.

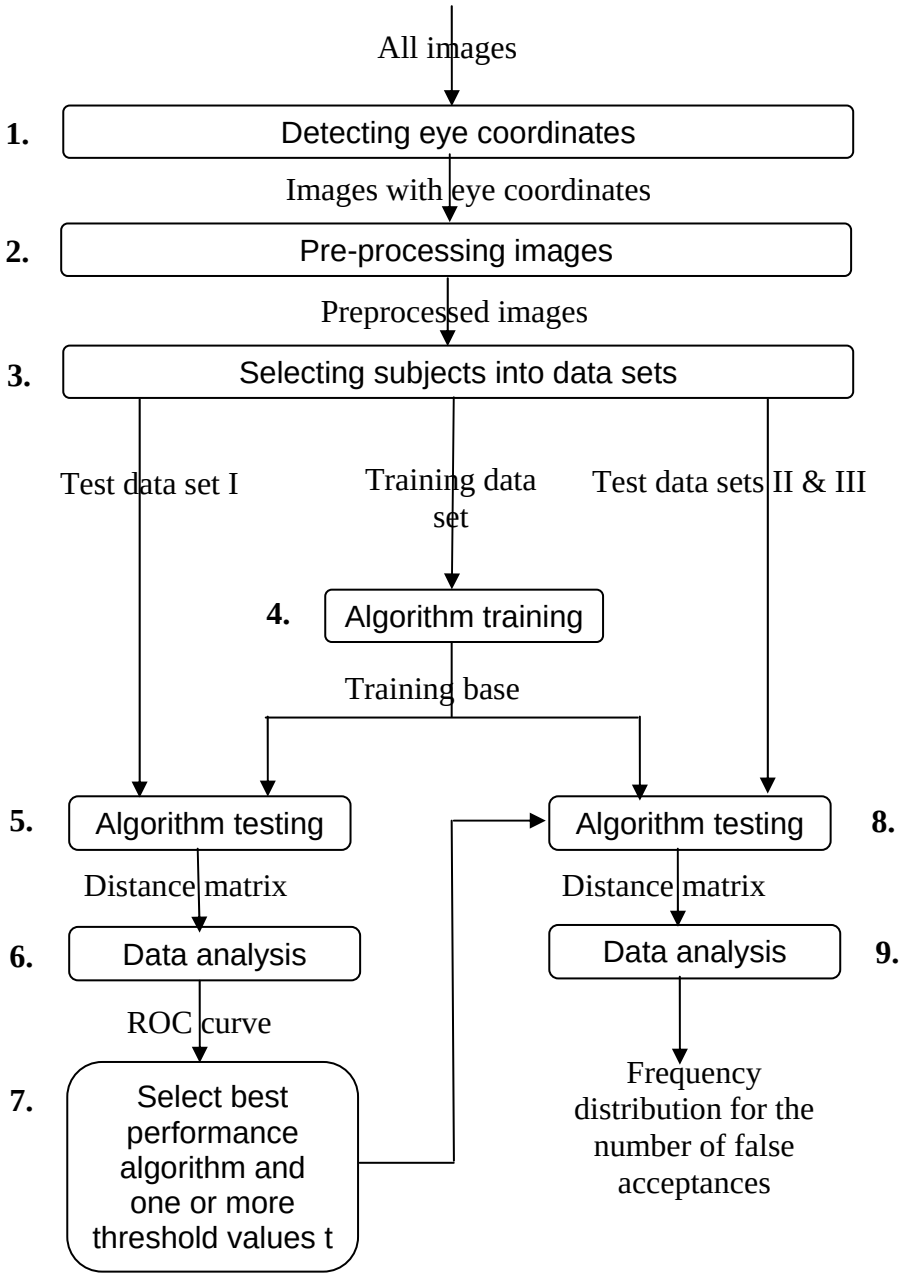


Figure 8 – Experimental procedure

Here are the main steps of the experiment:

1. The eye coordinates of the images were determined using the two matlab scripts: one of the scripts determines the eye coordinates automatically whereas the other allows for manual selection of eye coordinates by clicking on the images. The eye

coordinates for the HIG and passport photos were automatically determined. The HIG photos that got the wrong eye coordinates (approximately 16%) were manually processed and assigned the correct coordinates. The correctness of the eye coordinates for the passport photos was not checked due to the large number of photos. The eye coordinates for the images in the XM2VTS database were extracted from the eye coordinate files included in the database.

2. The images with the eye coordinates were input to the preprocessing script of the CSU software that generated normalized images.
3. The images were randomly assigned to four disjoint data sets: one training data set and three test data sets. The training data set was created by random selection of 1336 subjects from the HIG photo database, 50 subjects from the CVL database, 100 subjects from the XM2VTS database and 50 subjects from the AR database. The images from the AR database with screaming and angry facial expressions were excluded from the training set. The CVL, AR, and XM2VTS databases were included in the experiment because the HIG and passport photo databases contained only one image per subject. We needed at least two images per subject in the training set and two images per subject for calculation of false rejection rate (FRR). The passport photos and the rest of the images from the XM2VTS database, the AR database, the CVL database and the HIG database were used for testing. The first test data set was created by random selection of two images of each subject from the XM2VTS database, the CVL database and the AR database. We will refer to this data set as the data set *I*. The data set *I* were used to determine the verification performance of the selected face recognition algorithms. The second test data set contained the rest of the HIG photos whereas the third set was created by random selection of 10 000 images from several thousands of passport photos. We will refer to these two data sets as the data set *II* and the data set *III*, respectively. These two sets are used to find out the percentage of people having one or more look-alikes in the Norwegian population.
4. The face recognition algorithms were trained using the training data set.
5. Distance matrices for the data set *I* are calculated.
6. Based on the distance matrices, FAR and FRR for different threshold values *t* (unique distance values found in the distance matrix) are calculated.
7. The results from the step 6 were used to identify the face recognition algorithm with best performance and to select one or more acceptable FARs and their matching threshold values *t* for to this algorithm
8. Distance matrices for the data sets *II* and *III* are calculated.
9. Frequency distributions for the number of false acceptances in the data set *II* and *III* are calculated for each selected threshold value *t.*

Appendix A contains the eye coordinates files, the scripts used for random assignment of subject to data sets, the resulting data set files, the scripts used for training and testing of the face recognition algorithms, the scripts used for calculation of verification performance statistics, and the scripts used for calculation of frequency distributions.

The testing protocol used in the calculation of verification performance statistics is taken from the FERET verification testing protocol [6]. This protocol has been used in technology evaluations of face recognition algorithms since 1994.

5.1.4 Data analysis

We did not find a way to estimate the sample size prior to conducting the experiment. Methods for establishing "confidence intervals" on the ROC are not well-understood [9].

Since the CSU software does not contain statistical analysis tools for evaluating the performance of the face verification systems, we developed our own tools for this task. We used the Matlab program to write procedures for calculation of FAR, FRR, plotting the ROC curves and calculation of frequency distribution for the number of false acceptances in a data set (see Appendix A). The formulas used for calculation of FAR and FRR are taken from the FERET verification testing protocol [6].

We used the Microsoft Excel to present frequency distributions for the number of false acceptances in graphical form (histogram).

5.1.5 Validity and reliability issues

Validity is defined as "the extent to which any measuring instrument measures what it is intended to measure" [41]. Many factors can affect the validity of the experimental results and their generalizability. The software used in the experiment can influence the experiment's internal validity. The logical errors might be present in the statistical scripts we have written for the Matlab program. The CSU face recognition software may also contain errors, but this is not very likely since the software is open source and has been used in several research studies.

The characteristics of the subjects chosen for the experiment present a threat to the external validity of the experiment. The data sets used in the experiment are convenience samples that might limit the generalization of the results to the whole Norwegian population. We have tried to compensate for this by random assignment of subjects to different data sets.

Reliability is defined as "the extent to which an experiment, test or any measuring procedure yields the same results on repeated trials" [41]. Our experiment should produce the same results on repeated trials as long as the same software and the same data sets together with the determined eye coordinates are used. The HIG and the Passport databases are not available to the public and this represents a threat to the reliability of the experimental results. Researchers at the Gjøvik University College can obtain a copy of the HIG photo database and repeat the experiment. The researchers at the company that provided the passport photos will be able to repeat the experiment with the passport photos.

5.2 Results

The CSU Face software was used to generate the distance matrix for each face recognition algorithm based on the facial images in the test set I. The distance matrices were input to a matlab script that calculated the false acceptance rate and the false rejection rate for all the unique distance values.

Figure 9 is the ROC curve that shows the relationship between the calculated FAR and FRR pairs.

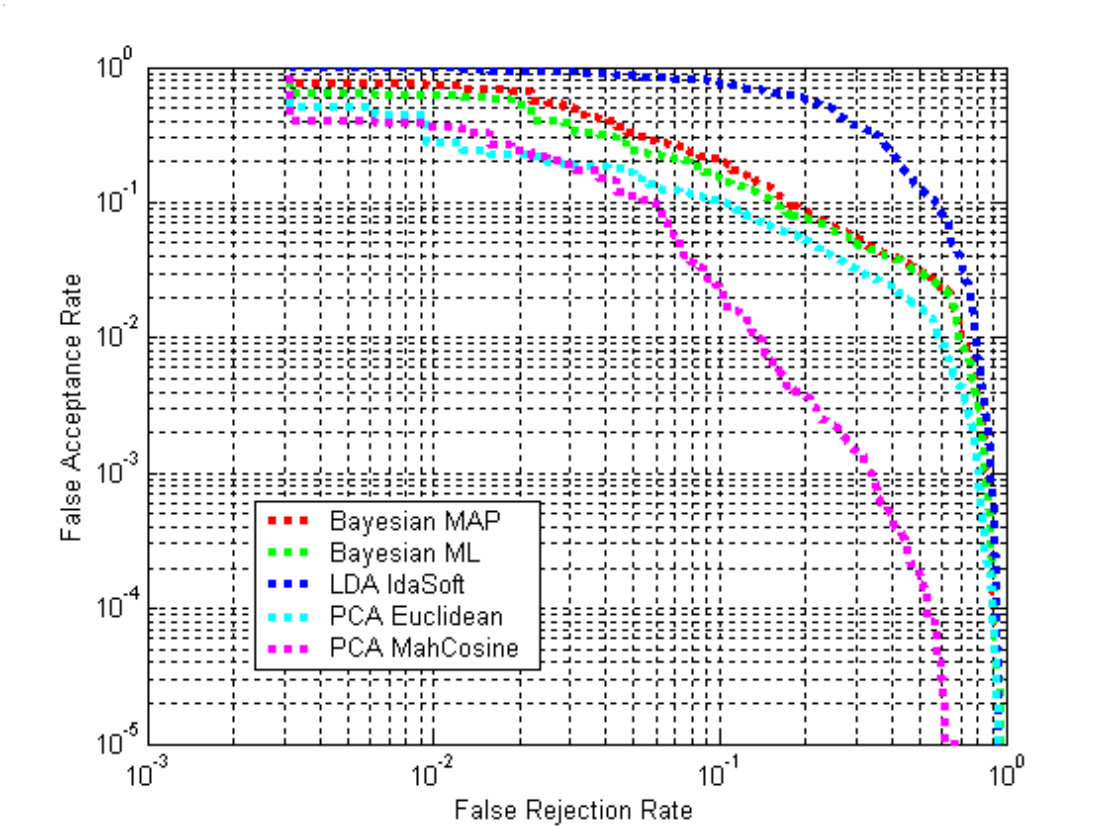


Figure 9 – The ROC curve for the test set *I* (note that the logarithmic axis are used in order to better visualize the trade-off between FAR and FRR)

The ROC curve shows the FAR/FRR pairs for the three algorithms (the two of the three algorithms calculated two distance matrices each based on the different distance measures). We can see that the PCA_MahCosine algorithm has best performance that is in accordance with the previous experience of the authors of these algorithms.

Table 5 shows the false acceptance rate, the false rejection rate and the probability of verification for different threshold values for the PCA_MahCosine algorithm.

Table 5 – Performance statistics: false acceptance rate, false rejection rate and probability of verification taken from the ROC curve based on the test set *I* (the numbers are rounded off to the nearest integer)

Threshold t	FAR (%)	FRR (%)	PV (%)
-0.27	1.0	14	86
-0.32	0.5	17	83
-0.42	0.1	33	67
-0.46	0.05	40	60

The best performance algorithm, PCA_MahCosine together with two threshold values that corresponds to 1% and 0.1% FAR were selected for the second part of the experiment. In this part of the experiment, the CSU software is used to calculate the distance matrix for the HIG images. Then the selected threshold values were input to a script that computed the number of false acceptance for each subject in the set.

Table 6 and Figure 10 below show the relative frequency distribution for the number of false acceptances in the test set *II* for the threshold value that corresponds to 1% FAR.

Table 6 – The frequency table for the number of the false acceptances in the test set *II* (1% FAR, 14% FRR)

The number of false acceptances	The number of subjects having the specified number of the false acceptances	The percentage of subjects having the specified number of the false acceptances
0	44	3.29 %
1-3	506	37.87 %
4-6	490	36.68 %
7-9	209	15.64 %
10-12	64	4.79 %
13-15	14	1.05 %
15-18	6	0.45 %
19-21	2	0.15 %
22-24	1	0.07 %

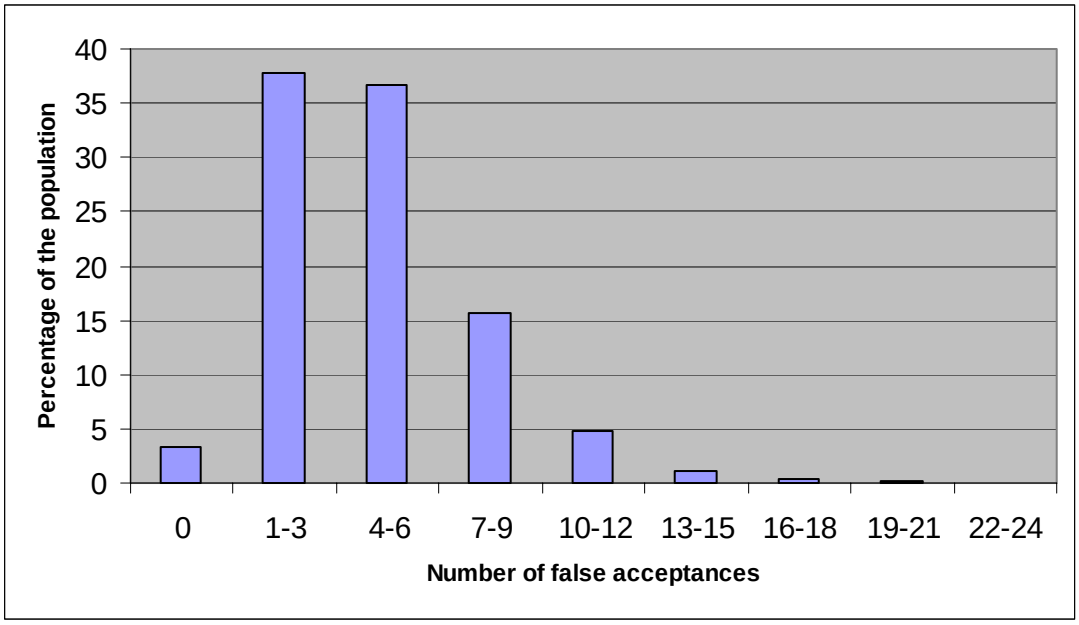


Figure 10 – The frequency distribution for the number of false acceptances in the test set *II* (1% FAR, 14% FRR)

The results presented in Table 6 and Figure 10 indicate that there exist different degrees of distinctiveness in the data set: number of false acceptances varies from 0 to 22. Approximately 97 % of the subjects generated one or more false acceptances.

We looked at the images of the persons who have generated many false acceptances and the images of their look-alikes. We could not find any similarities between any of these persons and their look-alikes.

Table 7 and Figure 11 below show the relative frequency distribution for the number of false acceptances in the test set *II* for the threshold value that corresponds to 0.1% FAR.

Table 7 - The frequency table for the number of the false acceptances in the test set *II* (0.1% FAR, 33% FRR)

The number of false acceptances	The number of subjects having the specified number of the false acceptances	The percentage of subjects having the specified number of the false acceptances
0	1144	85.63 %
1	142	10.63 %
2	42	3.14 %
3	8	0.60 %

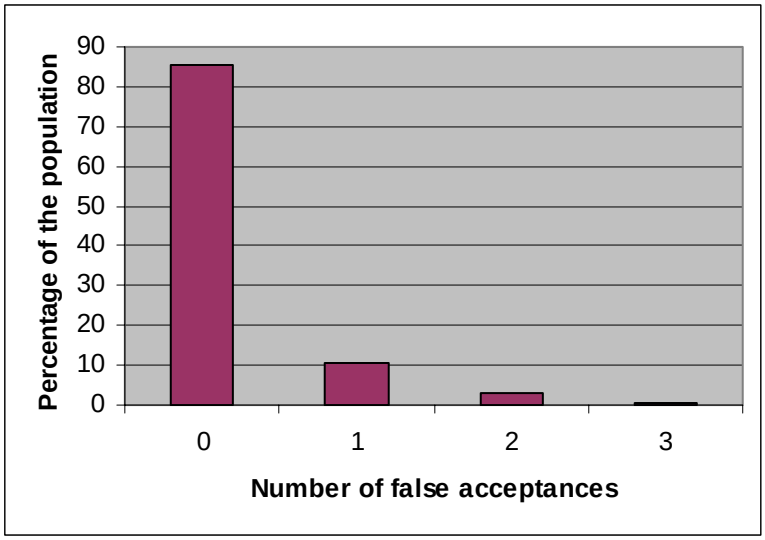


Figure 11 - The frequency distribution for the number of false acceptances in the test set *II* (0.1% FAR, 33% FRR)

The results presented in Table 7 and Figure 11 show that majority of the subjects did not generate any false acceptances.

We repeated the last part of the experiment with the 10 000 passport photos. Table 8 and Figure 12 show the relative frequency distribution for the number of false acceptances in the test set *III* for the threshold value that corresponds to 1% FAR.
The passport photos generated much more false acceptances than the HIG photos: 99.99 % of the subjects generated more than one false acceptance.

Table 8 - The frequency table for the number of the false acceptances in the test set III (1% FAR, 14% FRR)

The number of false acceptances	The number of subjects having the specified number of the false acceptances	The percentage of subjects having the specified number of the false acceptances
0	1	0.01 %
1-200	3421	34.21 %
201-400	2806	28.06 %
401-600	1558	15.58 %
601-800	1017	10.17 %
801-1000	609	6.09 %
1001-1200	328	3.28 %
1201-1400	178	1.78 %
1401-1600	64	0.64 %
1601-1800	17	0.17 %
1801-2000	1	0.01 %

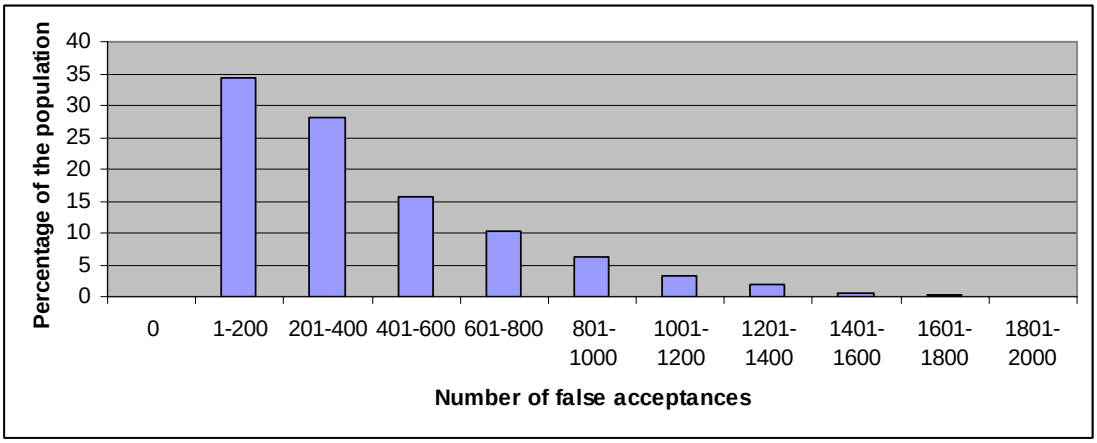


Figure 12 - The frequency distribution for the number of false acceptances in the test set III (1% FAR, 14% FRR)

Table 9 and Figure 13 show the relative frequency distribution for the number of false acceptances in the test set *III* for the threshold value that corresponds to 0.1% FAR. The passport photos generated much more false acceptances than the HIG photos at the same security level: 92 % of the subjects generated more than one false acceptance.

Table 9 - The frequency table for the number of the false acceptances in the test set III (0.1% FAR, 33% FRR)

The number of false acceptances	The number of subjects having the specified number of the false acceptances	The percentage of subjects having the specified number of the false acceptances
0	800	8.00 %
1-50	6987	69.87 %
51-100	1196	11.96 %
101-150	500	5.00 %
151-200	251	2.51 %
201-250	133	1.33 %
251-300	79	0.79 %
301-350	35	0.35 %
351-400	16	0.16 %
401-450	1	0.01 %
451-500	1	0.01 %
501-550	1	0.01 %

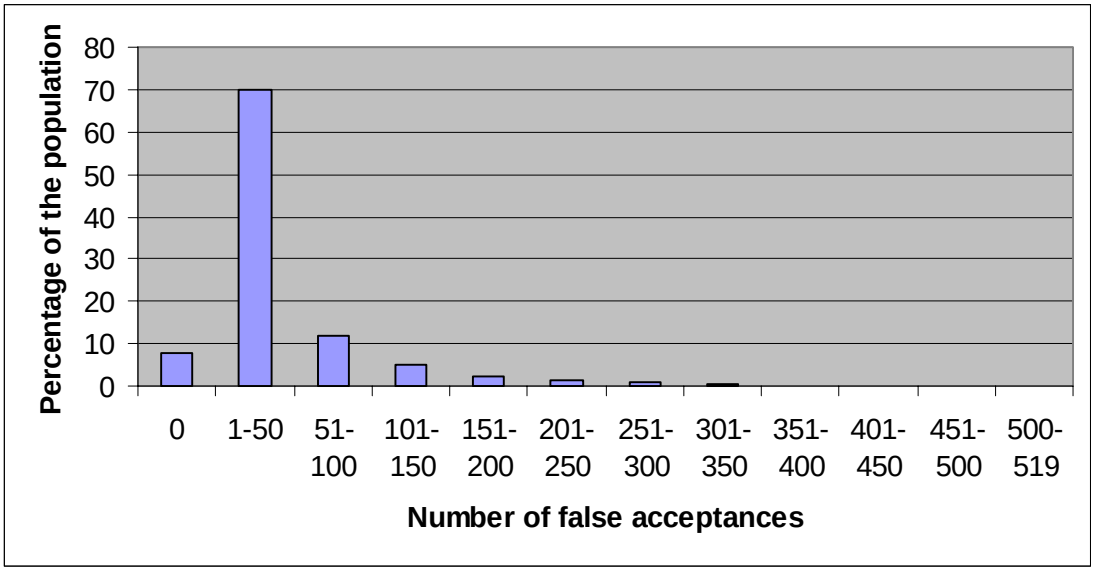


Figure 13 - The frequency distribution for the number of false acceptances in the test set III (0.1% FAR, 33% FRR)

6 Characteristics of Border Control Application

The security of the future passport will be dependent not only on the implemented security features, but also on the capability of the adversaries in a border control setting. We have not found any study in the reviewed literature that discusses the adversaries in a border control setting and their capabilities. This chapter is our contribution to the third research question.
Based on our assumptions about the future border control application, its adversaries and their capabilities, we propose an indicator for the true false acceptance rate.

6.1 Future border control

Border control application will have a large number of users. A user of a border control application is a passport holder who wants to enter a country.
The face recognition system used in the border control application will operate in the environment where some of the factors that can affect performance will be controlled such as lighting conditions. Factors such as facial expression, pose and changes caused by aging or esthetic surgery might not be possible to control.

When a user comes to border control, he would need to pass several control checks as shown in the Figure 14. All control checks except the manual control of identity, will be fully automatic processes and they will be supervised by the border control inspectors.
First, a face verification system will be used to verify the claimed identity and to check if the person is present on the watchlist. If the claimed identity and the true identity of the user do not match or the user is present on the watchlist, the user will have to go to the manual check. If neither of this is true, the user will be allowed to proceed to the next control point. At this control point, the passport will be examined for tampering attempts, then the authenticity of the passport will be checked and finally, the passport will be looked up in the database of stolen passports.

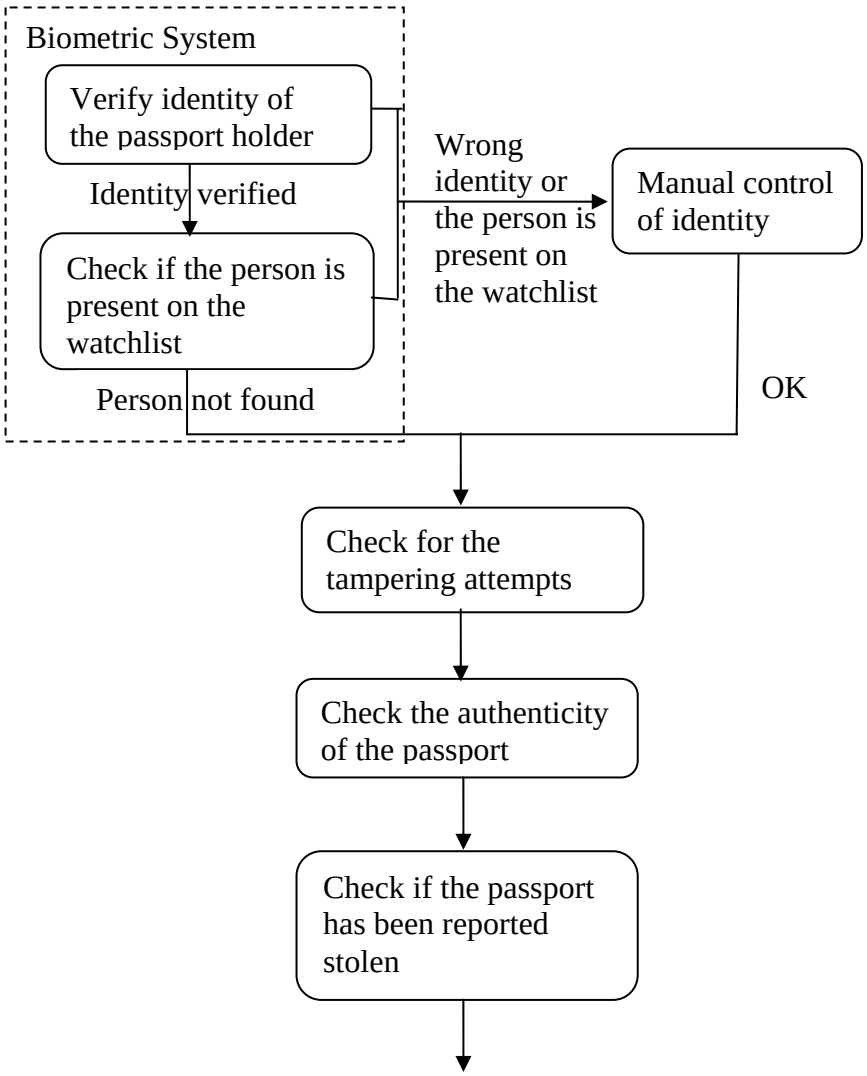


Figure 14 – Main steps in the future border control application

6.2 Adversary model and true false acceptance rate

The adversary in our model is an organization that sells travel documents to people who for some reason need a new travel document. The organization consists of several cells that operate around the world. The employees of the organization are people who have the knowledge and the skills about the reproduction and alteration techniques for travel documents. They cooperate with people who are willing to sell or lend their own travel documents.

Since the ICAO has recommended use of face as mandatory biometric identifier, they have been preparing for these new biometric based machine readable passports. They have obtained access to several face databases of people i different countries and they have purchased their own face recognition system which is used to found ”look-alikes” for their customers (impostors).

The traditional way of computing the false acceptance rate assumes that impostors are ”zero effort” impostors. These impostors will try to represent themselves to be other persons regardless of whether they look alike these persons or not. It is more realistic to assume that the impostors will try to claim identity of a person they know they look alike from a biometric system’s perspective. In such adversary environment, a more adequate measure for the true false acceptance rate would be the proportion of the impostors who will be falsely accepted as their look-alikes in the target population.

7 Discussion

7.1 Robustness of today's machine readable passports against the identity theft

According to the National Criminal Investigation Service in Norway, the Norwegian machine readable passports are difficult to alter. Only one of the 37 000 stolen Norwegian passports was a machine readable passport, and no forged or counterfeit machine readable passport has been detected so far. These facts may indicate that the adversaries steal the old kind of the passports because it costs less to alter them and because they already know how to do it, or that they have already found a smart way of altering and reproducing the Norwegian machine readable passports. Norway began issuing the machine readable passports a couple of years ago. The most of the issued passports are non readable passports and this can be one explanation for the fact that only one machine readable passport has been stolen so far.
The reported number of stolen and counterfeit passports is not a valid measure for their robustness against identity theft since the true number of stolen and counterfeit passports is not known.

Since it is up to the issuing state to decide whether to implement the recommended security features, it does not help much if, for example, Norway has implemented a lot of security features while some other country has implemented few of them. Thus, a system is only as secure as its weakest link. The adversaries may discover this weakest link and exploit only the weakest passports. In order to compensate for this, the passports with low security level could be examined more carefully in border control in order to find out if they have been counterfeit or tampered with.
Ideally, there should be a standard set of security features that all issuing countries had to implement.

The robustness of the today's machine readable passport against identity theft will depend both on the security measures implemented in the passports and the capabilities of the adversaries.

7.2 Similarity of people from biometric systems' perspective

The related literature demonstrates that the face has much lower distinctiveness in comparison to fingerprint and iris. A combination of several biometric identifiers can increase the distinctiveness between people.

The conducted experiment with the HIG photos showed that 97% of the subjects had at least one look-alike when the face recognition algorithm was set to operate at the security level of 1% FAR and 14% FRR. This result indicates that the face has low distinctiveness that is in accordance with the results from the recent evaluation tests in the face recognition community.
The experiment with the HIG photos showed that 14% of the subjects had at least one look-alike when the face recognition algorithm was set to operate at the security level

of 0.1% FAR and 33% FRR. Thus, the distinctiveness of the human faces was much higher at this security level. According to the results of the FRVT 2002 evaluation test, the verification performance will degrade at approximately 5 % per year. When the probability of verification decreases, the false rejection rate increases accordingly. In a period of five years, the false rejection rate of the face recognition algorithm used in the experiment will increase to approximately 58% (33% + 5*5%). In the border control setting, this means that the face recognition system will fail to verify the identity of 58% of travellers. This percentage may even get higher due to the changes in face appearance caused by esthetic surgery, make-up, diet, scar etc.

The experiment with the passport photos showed that 99.99 % of the subjects had at least one look-alike when the face recognition algorithm was set to operate at the security level of 1% FAR. The maximum number of false acceptances was 519. The security level of 0.1% FAR resulted in 92% of the subjects having more than on look-alike. The maximum number of false acceptances was 1831. Thus, the passport photos generated much more false acceptances than the HIG photos. There might be several reasons for such a high number of false acceptances. One reason might be that the subjects included in the training data set are not representative for the Norwegian population. For a border control application it would be essential that the face recognition algorithms be trained with a representative set. This raises a new research question: is it possible to create a training set that will be representative for the whole world. If not, than the face recognition system used in border control might be population dependent: the people who do not belong to the population, with which the face recognition algorithm is trained, will probably generate higher number of false acceptances.
The eye coordinates of the passport photos were generated automatically, which means that the 16 % of the eye coordinates were not correct. This is more obvious reason for the high number of false acceptances in the passport data set.

A commercially available face recognition system would probably generate fewer false acceptances than the baseline algorithms used in our experiment. Thus, the real number of people having one or more ”look-alikes” is probably much lower than the experiment demonstrated. On the other hand, the photos used in the experiment were mainly frontal facial images without any occlusions. The false acceptance rate would be higher if we included side view images or facial images with sunglasses, scarfs and other kind of occlusions.

7.3 True false acceptance rate in the border control application

The false acceptance rate calculated in the traditional way is not an adequate indicator for the true false acceptance rate in a real world application such as border control.
A better indicator for the true false acceptance rate would be the percentage of the impostors who have at least one ”look-alike” in the target population.

In order to justify this new false acceptance rate, we have created an adversary model in the future border control application. In this model, the adversary is a large international organization that sells travel documents to people who for some reason need a new travel document. The organization has obtained access to several face

databases of people in different countries, and has purchased a face recognition system. They use this face recognition system to find ”look-alikes” for their customers (impostors). The organization cooperates with people who are willing to sell or lend their own travel documents.

Our adversary model may contain some unrealistic assumptions. For example, we assumed that the adversary is able to obtain access to several face databases of people i different countries. This may not be a realistic assumption. However, the adversary can probably buy photos and personal information from companies that sell personal data. Alternatively, they may collect personal data on the Internet.
Stolen passports will probably be detected in border control assumed that the revocation routines are effective and in place.

The validity of the proposed indicator for the true false acceptance rate is questionable. A more realistic indicator would be the percentage of the impostors who has at least twenty ”look-alikes” in the target population since the probability of obtaining the passport would be greater in the case of a higher number of ”look-alikes”.

The true false acceptance rate can be estimated from similarity scores between a target sample data set and an ”impostor” sample data set. If the impostors belong to the target population, then the true false acceptance rate is equal the percentage of the people in the target population who have at least one ”look-alike”. In this case, our experimental design can be used to estimate the true false acceptance rate. The true false acceptance rate estimated in this way may be lower than the real world false acceptance rate since the impostors who do not have a look-alike in the target population will probably alter their appearance in order to look alike the person in the stolen passport.

7.4 Robustness of biometric based passports

There are many problems with biometric technologies such as error rates, spoofing attacks, non-universality and interoperability problems. More research needs to be done in the area of biometrics.

An overall security process that involves people, technology and procedures can overcome the limitations of biometric technologies. For example, the verification performance of face recognition algorithms could be improved by using the user-specific thresholds: the users that generate many false acceptances could have the higher verification threshold than the users with fewer false acceptances.

Our experiment showed that the people and their look-alikes were not similar to each other from the human perspective. This may indicate that the people are better at recognizing human faces than the face recognition algorithms are. People and technology could be combined to decrease the identity theft. For example, border control inspectors could verify the identity of pass holders by visually comparing the photo stored in the passport and the face of the pass holder while a face recognition system could be used to check whether the pass holder is on the watch list.

Recent studies show that the biometric systems that use multiple biometrics are much more reliable, have better performance and are more robust against spoofing attacks. In the border control application, the face should be used together with some other biometric identifier in order to increase the ability of the biometric system to correctly verify an individual's identity.

The inclusion of biometric identification information into machine readable passports will improve their robustness against identity theft if additional security measures are implemented in order to compensate for the limitations of the biometric technologies. The robustness of the biometric based passports will depend on the capabilities of the adversaries in a border control setting. As they their skills and knowledge increase, the robustness of the passport will decrease.

8 Conclusions

There are many problems associated with biometric technologies such as error rates, spoofing attacks, non-universality and interoperability problems. An overall security process that involves people, technology and procedures can overcome limitations of face recognition and other biometric technologies. Some of the issues that are relevant for designing the security process are given in Appendix B.

In the border control that will use face for identity verification, several security measures can be used to improve the overall performance of the system:

- User-specific thresholds: the users that generate many false acceptances could have the higher verification threshold than the users with fewer false acceptances
- Multiple biometrics: the face could be used together with some other biometric identifier in order to increase the ability of the biometric system to correctly verify an individual's identity
- Since people are better at recognizing human faces than the face recognition systems, border control inspectors could verify the identity of passport holders by visually comparing the photo stored in the passport and the face of the passport holder. A face recognition system could be used to check whether the passport holder is present on the watch list.

The false acceptance rate as measured in the face recognition community does not give the correct picture of the true false acceptance rate that can be expected in a real border control application with non-"zero-effort" impostors. A more representative measure for the true false acceptance rate might be the percentage of the impostors who have at least twenty (or some other number) "look-alikes" in the target population.

The robustness of today's machine readable passports will vary between the issuing states. The adversaries might exploit the passports with the lowest level of security. The inclusion of biometric identification information into machine readable passports will improve their robustness against identity theft if additional security measures are implemented in order to compensate for the limitations of the biometric technologies. As the skills and the knowledge of the adversaries increases, the robustness of the biometric passports will decrease.

Thus, the biometric based passports will provide a new "speed bump" that will reduce identity theft by "zero-effort" and "small-effort" impostors. Smart adversaries with a large international network and many resources will be stopped by this "speed bump" only for a limited time - until they have discovered new ways of forging and counterfeiting passports.

9 Suggestions for future research

- Issues that are relevant to practical biometric systems such as usability, privacy and acceptability issues
- Collection of large database of facial images for use in various experiments
- Large scale technology test whose goal is to estimate the proportion of people having one or more look-alikes in a population
- Large scale technology test whose goal is to estimate the proportion of impostors having one or more look-alikes in a target population
- Scenario test whose goal is to find out how easy it is to fool a biometric system by different fraudulent techniques.

References

[1] ICAO. *Biometrics Deployment of Machine Readable Travel Documents*. ICAO TAG MRTD/NTWG. Technical Report, Version 1.9. Montreal, May 2003. Available at http://www.icao.int/mrtd/download/documents/Biometrics deployment of Machine Readable Travel Documents.pdf (accessed March 04)

[2] United States General Accounting Office. *Technology Assessment: Using Biometrics for Border Security*. November 14, 2002. Available at http://www.gao.gov/new.items/d03174.pdf (accessed December 03)

[3] Anil Jain, Ruud Bolle and Sharath Pankanti. Introduction to Biometrics. In *Biometrics: Personal Identification in Networked Society*. Kluwer Academic Publishers, Forth Printing 2002, ISBN 0-7923-8345-1

[4] P. J. Phillips, A. Martin, C. L. Wilson, and M. Przybocki. An Introduction to Evaluating Biometric Systems. In *IEEE Computer*. February 2000, pp. 56-63

[5] Dr. Gunter Lassmann. Some results on robustness, security and usability of biometric systems. *Proc. ICME 2002*. Lausanne, Switzerland, Aug. 2002, Vol. 2, Pages 577-580

[6] Syed A. Rizvi, P. Jonathon Phillips and Hyeonjoon Moon. The FERET Verification Testing Protocol for Face Recogntion Algorithms. *Technical Report NIST IR 6281*, October 1998

[7] Mike Bone and James L. Wayman. Evaluating Facial Recognition Technology for Drug Control Applications. Available online at: http://www.frvt.org/DLs/ONDCP2001.pdf.

[8] Mansfield, A. J. Wayman, J. L. *Best Practices in Testing and Reporting Performance of Biometric Devices*. Version 2.01, August 2002

[9] J. L. Wayman. Confidence interval and test size estimation for biometric data. In *Proceedings of IEEE AutoID '99*, pages 177–184, 1999.

[10] G. Doddington, W. Liggett, A. Martin, M. Przybocki, and D. Reynolds, Sheeps, goats, lambs and wolves: a statistical analysis of speaker performance in the NIST 1998 speaker recognition evaluation. In *Proc. of ICSLD '98*, Sydney, Australia, November 1998.

[11] J.L. Wayman. Technical Testing and Evaluation of Biometric Identification Devices. In *Biometrics: Personal Identification in Networked Society*. Kluwer Academic Publishers, Forth Printing 2002, ISBN 0-7923-8345-1

[12] Ross Beveridge, David Bolme, Marcio Teixeira and Bruce Draper. *The CSU Face Identification Evaluation System User's Guide: Version 5.0*. Computer Science Department Colorado State University, May 1, 2003. Available online at: http://www.cs.colostate.edu/evalfacerec/algorithms/version5/faceIdUsersGuide.pdf

[13] Patrick Grother, Ross J. Micheals, and P. Jonathon Philips. Face Recognition Vendor Test 2002 Performance Metrics. In *Proceedings of the Fourth International Conference on Audio-Visual Based Person Authentication*. Guildford, UK, June 2003

[14] U.S. General Accounting Office. *Identity Fraud: Prevalence and Links to Alien Illegal Activities*. GAO-02-830T. Washington, D.C.: June 25, 2002

[15] K. Jain, S. Prabhakar, and S. Pankanti. On the Similarity of Identical Twin Fingerprints. *Pattern Recognition*. Vol. 35, No. 11, pp. 2653-2663, November 2002.

[16] S. Pankanti, S. Prabhakar, and A. K. Jain. On the Individuality of Fingerprints. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. Vol. 24, No. 8, pp. 1010-1025, August 2002.

[17] John Daugman. Phenotypic versus Genotypic Approaches to Face Recognition. *Face Recognition: From Theory to Applications*. NATO ASI Series. Series F: Computer and system sciences, vol. 163. ISBN 3-540-64410-5

[18] T. Matsumoto, H. Matsumoto, K. Yamada, S. Hoshino. Impact of Artificial Gummy Fingers on Fingerprint Systems. In *Proceedings of SPIE*. Vol. 4677. Optical Security and Counterfeit Deterrence Techniques IV.

[19] Lisa Thalheim, Jan Krissler, Peter-Michael Ziegler. Body Check: Biometric Access Protection Devices and their Programs Put to the Test. c't 11/2002, page 114 – Biometrie. Available online at: http://www.heise.de/ct/english/02/11/114/

[20] P.J. Phillips, P. Grother, R.J Micheals, D.M. Blackburn, E Tabassi, and J.M. Bone. *FRVT 2002: Evaluation Report. March 2003*.

[21] K. Messer, J. Matas, J. Kittler, J. Luettin, and G. Maitre. XM2VTSDB: The Extended M2VTS Database. In *Second International Conference on Audio and Video-based Biometric Person Authentication*. March 1999.

[22] ICAO. *Machine Readable Travel Documents (Doc 9303)*. Part 1 Machine Readable Passports. ICAO, Fifth Edition, 2003

[23] Duane M. Blackburn. *Biometrics 101*. Version 3.1, March 2004

[24] Arun Ross, Anil Jain. Information Fusion in Biometrics. *Pattern Recognition Letters*. Vol. 24, No. 13, pp. 2115-2125, 2003

[25] Hong, L., Jain, A.K., Pankanti, S., 1999. Can multibiometrics improve performance? In: *Proc. AutoID_99, Summit, NJ, USA*. pp. 59–64

[26] Souheil Ben-Yacoub and Yousri Abdeljaoued and Eddy Mayoraz. *Fusion of Face and Speech Data for Person Identity Verification*. Research Paper IDIAP-RR 99-03, IDIAP, CP 592, 1920 Martigny, Switzerland.

[27] Roberto Brunelli and Daniele Falavigna. Person Identification Using Multiple Cues. *In IEEE Transactions on Pattern Analysis and Machine Intelligence*. Vol. 17, No. 10, pp. 955-966, 1995

[28] Anil K. Jain and Arun Ross. Multibiometric Systems. In *Communications of the ACM*. Vol. 47, No. 1, January 2004

[29] Ruud M. Bolle, Sharath Pankanti and Nalini K. Ratha. *Evaluation techniques for biometric-based authentication systems (FRR)*. IBM Computer Science Research Report RC 21759, 2000.

[30] Anita Alden. Norske pass blir stadig mer ettertraktet: Betalte 11 000 Euro for norsk pass. Kripos 21.05.04. Available at: http://www.kripos.no/akt_tema/dyrt_pass.html

[31] Årsrapport 2003. Kripos. Available at: http://www.kripos.no/statistikker/bilder/rapport.pdf

[32] James L. Wayman. A scientific approach to evaluating biometric systems using a mathematical methodology. In *Proc. Cardtech/Securetech'97*. pp. 477-492

[33] Y. Wang, T. Tan and A. K. Jain, Combining Face and Iris Biometrics for Identity Verification, *Proc. of 4th Int'l Conf. on Audio- and Video-Based Biometric Person Authentication (AVBPA)*, pp. 805-813, Guildford, UK, June 9-11, 2003.

[34] A. K. Jain, A. Ross and S. Prabhakar, An Introduction to Biometric Recognition, *IEEE Transactions on Circuits and Systems for Video Technology, Special Issue on Image- and Video-Based Biometrics*, Vol. 14, No. 1, pp. 4-20, January 2004

[35] John Daugman. The importance of beeing random: statistical principles of iris recognition. *Pattern Recognition*. Vol. 36, No. 2, pp. 279-291
[36] John Daugman. High confidence visual recognition of persons by a test of statistical independence. *Trans. Pattern Analysis and Machine Intelligence*. 15(11): 1148-1161.
[37] John Daugman. U.S. Patent No.5,291,560: *Biometric Personal Identification System Based on Iris Analysis*. Issue Date: 1. March 1994
[38] A.M. Martinez and R. Benavente, *The AR face database*, CVC Tech. Report #24, 1998
[39] *CVL FACE DATABASE*: http://www.lrv.fri.uni-lj.si/facedb.html
[40] John D. Woodward, Jr., Nicholas M. Orlans, Peter T. Higgins. *Biometrics: Identity Assurance in the Information Age*. McGraw-Hill/Osborne, 2003. ISBN 0-07-222227-1
[41] Edvard G. Carmines and Richard A. Zeller. *Reliability and validity assessment*. Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-017. Newbury Park, CA: Sage, 1979
[42] *Gjøvik University College*: http://www.hig.no
[43] D. Blackburn, M. Bone, and P. J.Phillips. *Facial Recognition Vendor Test 2000 Evaluation Report*. February, 2001. Available online at: http://www.dodcounterdrug.com/facialrecognition/FRVT2000/documents.htm
[44] *FVC2002 – Results(Summary)*. Available at http://bias.csr.unibo.it/fvc2002.
[45] Ross Anderson. *Security Engineering: A Guide to Building Dependable Distributed Systems*. John Willey & Sons, 2001. ISBN: 0-471-38922-6.
[46] Brocklehurst, S. Littlewood, B. Olovsson, T. Jonsson, E. On measurement of operational security. In *Aerospace and Electronic Systems Magazine*, IEEE. Volume: 9, Issue: 10, pp. 7-16, October 1994
[47] Keith A. Rodhes. *Information Security: Challenges in Using Biometrics*. United States General Accounting Office, September 2003.
[48] John W. Creswell. *Research Design. Qualitative, Quantitative and Mixed Methods Approaches*. Second Edition, SAGE Publications, 2003

Appendix A – Scripts and data sets

The scripts and data sets used in the experiment are not included in the report due to the large number of files. If you need a copy of these files, please send an email to the following email address marijana@erdal.biz. The paper version of this report includes a CD-ROM with all the files.

Appendix B - Issues to consider when choosing a biometric system

Here is a simple check list of issues to consider when evaluating biometric systems for use in production of biometric passports and in border control:

- Is the verification and identification performance acceptable?
- Is the failure to enrol acceptable?
- Is the throughput rate acceptable?
- Is the biometric system robust against spoofing attacks such as fake fingers and static images?
- How easy would it be to replace the biometric system with another one?
- How user friendly is the biometric system?
- How reliable is the system (mean time to failure)?
- How good is the support from the supplier of the biometric system? How fast will they be able to give support in case of a system failure?
- Issuance of a new biometric passport should not take longer time than the issuance of today's machine readable passport. The previous experience shows that the number of applications for new passports is usually very high in the spring and summer months, and that during this period the passport producer is usually not able to produce passports within the agreed time requirements[3].
- Procedures that handle exception situations, such as revocation of stolen passports and system failure should be developed.
- Privacy issues?

[3] Source: the Norwegian Police Data and Procurement Service

Appendix C – Standards[4]

The tables below give an overview of the standards related to the travel documents and the biometric technologies.

Table 10 - Travel documents

Ref	Standard / document	Status
[ICAO02]	*ICAO 9303 Second Edition – 2002* *Part 1: Passports* *Part 2: Visas* *Part 3: Official Travel Documents*	Final
[BDTD03]	*ICAO Biometrics Deployment of Machine Readable Travel Documents v. 1.9 – May 2003*	Technical Report
[CLTR03]	*ICAO Use of Contactless Integrated Circuits In Machine Readable Travel Documents – Technical Report – 2003*	Technical Report
[LDTR03]	*ICAO Development of a Logical Data Structure – LDS for Optional Capacity Expansion Technologies – 2003*	Technical Report
[DSTR03]	*ICAO PKI Digital Signatures for Machine Readable Travel Documents – 2003*	Technical Report
[NS9200]	*NS-9200, Elektroniske identitetskort – Visuell ID, type ID-1*	Norwegian standard

Table 11 - ISO/IEC Biometri

Standard / document	Status
ISO/IEC FCD 19784, BioAPI – Biometric Application Programming Interface	FCD
ISO/IEC WD2 19794-1; Biometric Data Interchange Formats — Part 1: Framework	WD2
ISO/IEC FCD 19794-2, Biometric Data Interchange Formats — Part 2: Finger Minutiae Data	FCD
ISO/IEC WD2 19794-3, Biometric Data Interchange Formats - Part 3: Finger Pattern Spectral Data	CD
ISO/IEC FCD 19794-4, Biometric Data Interchange Formats – Part 4: Finger Image Data	FCD
ISO/IEC FCD 19794-5, Biometric Data Interchange Formats – Part 5: Face Image Data	FCD
ISO/IEC FCD 19794-6, Biometric Data Interchange Formats – Part 6: Iris Image Data	FCD
ISO/IEC WD 19794-7, Biometric Data Interchange Formats - Part 7: Signature/Sign Behavioral Data	WD
ISO/IEC FCD 19785-1, Common Biometric Exchange Formats Framework: Part 1: Data Element Specification	FCD

Table 12 - ISO/IEC Other relevant standards

Standard / document	Status
ISO/IEC 14443, Identification cards – Contactless integrated circuit(s) cards – Proximity cards	Final
ISO/IEC 15693, Identification cards – Contactless integrated circuit(s) cards –	Final

[4] Source: Asbjørn Hovstø, the leader of the Norwegian committee for biometrics - K188

Vicinity cards	
ISO/IEC 7810, Identification cards – Physical characteristics	Final
ISO/IEC 7816-3, Information technology - Identification cards - Integrated circuit(s) cards with contacts - Part 3: Electronic signals and transmission protocols	Final
ISO/IEC 7816-4, Information technology - Identification cards - Integrated circuit(s) cards with contacts - Part 4: Interindustry commands for interchange	Final
ISO/IEC 9594-8/ITU-T Recommendation X.509, Information Technology – Open Systems Interconnection: The Directory Authentication Framework, June 1997	Final