# Face Recognition Issues in a Border Control Environment

Marijana Kosmerlj, Tom Fladsrud, Erik Hjelmås, and Einar Snekkenes

NISlab
Department of Computer Science and Media Technology
Gjøvik University College
P. O. Box 191, N-2802 Gjøvik, Norway
erikh@hig.no

**Abstract.** Face recognition has greatly matured since the earliest forms, but still improvements must be made before it can be applied in high security or large scale applications. We conducted an experiment in order to estimate percentage of Norwegian people having one or more look-alikes in Norwegian population. The results indicate that the face recognition technology may not be adequate for identity verification in large scale applications. To survey the additional value of a human supervisor, we conducted an experiment where we investigated whether a human guard would detect false acceptances made by a computerized system, and the role of hair in human recognition of faces. The study showed that, on average, the human guard was able to detect almost 80 % of the errors made by the computerized system. More over, the study showed that the ability of human guard to recognize a human face is a function of hair: false acceptance rate was significantly higher for the images where the hair was removed compared to where it was present.

## 1 Introduction

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.

In May, 2003, the International Civil Aviation Organization (ICAO) adopted a global, harmonized blueprint for the integration of biometric identification information into passports [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 be included in the machine readable passports. Inclusion of the additional biometric technologies, fingerprint and iris, is optional. The new passports will be valid for 5 years.

The purpose of biometric passports is to prevent the illegal entry of travellers into a specific country, limit the use of fraudulent documents and make the border control more efficient [2].

In this paper we focus 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. We claim that the face recognition technology alone is not adequate for identity verification in large scale applications, such as border control, unless it is combined with additional security measures.

The rest of this paper is organised as follows. The next section gives an introduction to distinctiveness of faces, and describes the experiment with Norwegian student and passport photos. Section 3 describes the experiment with human recognition of faces with and without hair. In Sect. 4 we discuss the results from both experiments. Conclusions are made in Sect. 5.

## 2 Face as a Biometric in Border Controls

As a biometric identifier, the face has the advantage that it is socially acceptable and easily collectable. However, the face has large intra-person variability causing face recognition systems to have problems dealing with pose, illumination, facial expression and aging (changes over time). Several independent technology evaluations in face recognition community, conducted in the period from 1996 to 2002 [3–6], show that the face has low discrimination capability. 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 [4].

### 2.1 Adversary Model in a Border Control Context

In "best practices" standard for testing and reporting on biometric system performance [7], 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. In 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. An example of such an environment is a border control.

In order to propose a new way of calculating false acceptance rate in a border control context, we have modelled a possible adversary in this environment. In this model the adversary is a world wide organization that sells travel documents to people who for some reason need a new identity. The organization does not have the knowledge and the skills about the reproduction and alteration techniques for travel documents. Instead it cooperates with people who are willing to sell or lend their own travel documents, and with people who are willing to steal travel documents. Since the ICAO has recommended use of face as mandatory biometric identifier, they have been preparing for these new biometric based passports. They have obtained access to several face databases of people in different countries and they have purchased several face recognition systems which are used to found look-alikes for their customers. In a border control scenario where the identity of passport holders is verified by use of a face recognition

system, there is a high probability that an impostor holding the passport of his "look-alike", will pass the identity verification.

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.

### 2.2 Experimental Results

We conducted an experiment in order to estimate the percentage of Norwegian people having one or more look-alikes in the Norwegian population. Subjects in the experiment were selected from several face databases:

1. Ljubljani CVL Face Database [8]
2. XM2VTS Database [9]
3. AR Face Database [10]
4. photos of Norwegian students at Gjøvik University College (HIG face database) [11]: 2762 subjects, 1 image per subject, colour images; resolution 194x234; varying facial expressions; varying posture; varying illumination
5. Norwegian passport photos: several thousands of passport photos, 1 image per subject; varying facial expressions; varying posture; varying illumination.

In order to limit the effect of side views, lighting conditions and occlusions on the verification performance, frontal and approximately frontal facial images without occlusions and with varying but controlled lighting conditions were selected for the experiment.

We used the CSU Face Identification Evaluation System 5.0 [12] to generate similarity scores between our facial images. The experimental procedure consisted of two steps:

**Preparation** We determined the eye coordinates of the images that did not have eye coordinates. The eye coordinates of the HIG photos were manually determined. The eye coordinates of the passport photos were automatically determined with help of a Matlab script that had an error rate of 16 %. 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 test data set I was created by random selection of two images of each subject from the XM2VTS database, the CVL database and the AR database. The test data set II contained the rest of the HIG photos whereas the data set III was created by random selection of 10 000 images from several thousands of passport photos.

**Experiment** The images with the eye coordinates were processed by the preprocessing script of the CSU software that removed unwanted image variations. In this process the hair is removed from the images such that only the face from forehead to chin and cheek to cheek is visible. After the training of the face recognition algorithms and calculation of distance scores for

data set I, we calculated the verification performance of the face recognition algorithms at several operating points. The face recognition algorithm with best performance was selected for the last part of the experiment where we calculated frequency distributions for the number of false acceptances in the data set II and III at selected operating points.

Figure 1 and Figure 2 show respectively the relative frequency distribution for the number of false acceptances in the test set II and III for the threshold value that corresponds to 1 % FAR.

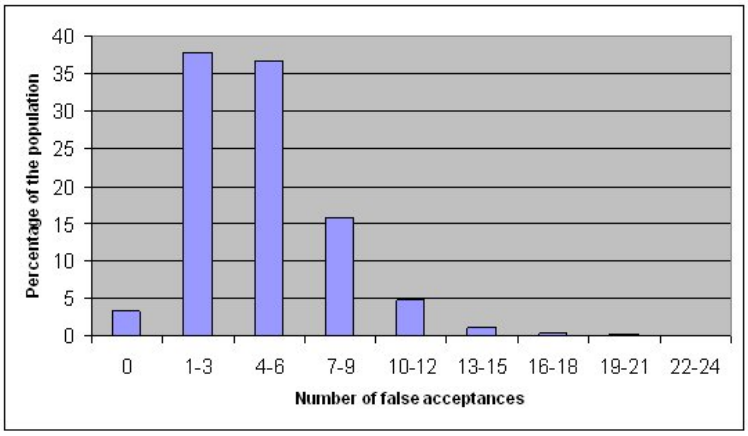


**Fig. 1.** The frequency distribution for the number of false acceptances in the test set II (1 % FAR, 14 % FRR)

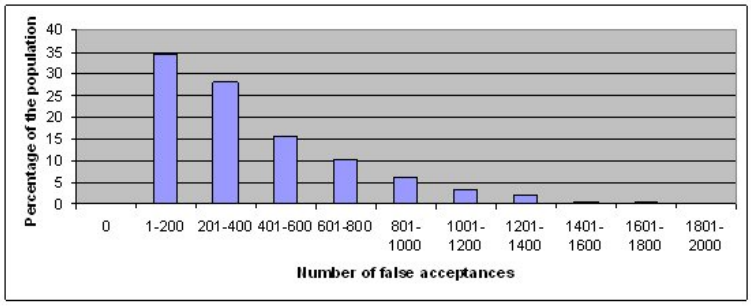


**Fig. 2.** The frequency distribution for the number of false acceptances in the test set III (1 % FAR, 14 % FRR)

At the operating point of 1 % FAR, 97 % the subjects in the data set II generated one or more false acceptances while 99.99 % of the subjects in the data set III generated more than one false acceptance. We repeated the experiment at the operating point of 0.1 % FAR. The results showed that majority of the subjects in the data set II did not generate any false acceptances while 92 % of the subjects in the data set III generated more than one false acceptance.

Additional information about the experiment can be found in MSc thesis of M. Kosmerlj [13].

## 3 The Effect of Additional Human Inspection

Based on our discoveries of look-alikes that might be able to pass a computerized face recognition environment, a natural next step would be to investigate whether an additional human guard would detect these false acceptances by the computerized system. In the previous experiment the computerized face recognition system compared normalized images without hair while in a real-world situation the people passing a control post will have hair. Therefore it is natural to investigate how god a human guard will be at recognizing human faces, both faces with hair, and faces without hair. This way we could see whether humans' face recognition process is affected by the presence of hair or not. If an impostor is able to find someone he or she resembles, this person may alter his hair style, colour etc., to amplify the similarities between her and the target person.

### 3.1 Experimental Results

The data set was a subset of the data set used in the experiment in Sect. 2. From this data set we chose the images of the persons that generated high number of false acceptances and the images of their look-alikes. Only subjects from the CVL-, XM2VTS- and the AR Face face databases are used since the two other databases did not include more than one image of each subject.

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

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

The participants were presented with several image-pairs that could be composed by two images of the same individual taken at different times, or of an image-pair composed by one image of one individual and the other image consisting of an image of his or her look-alike. Each participant was given a maximum

time of 10 seconds to evaluate the images. Each participant had to mark the image-pairs as either being of the same individual or of someone else.

The analysis of the experimental results reveals, as shown in Fig. 3, that there is a significant difference between the number of false acceptances when the hair is removed and when it is not. The results show an increased false acceptance rate for the images where the hair is removed compared to where it is not. When looking at false rejections however there seem to be no significant difference in this error rate.

When looking at the distribution of the false acceptances and false rejections in the two experiments, we observed that there were only 3 image-pairs that have not been evaluated wrongly from one or more participant in the experiment when the hair was removed, while there were 18 image-pairs that was never evaluated wrongly from the image-pairs where the hair was present. We can also observe that most of the image-pairs have been falsely evaluated more than once when the hair was removed.

Additional information about the experiment are provided in the Master's thesis of Tom Fladsrud [14].

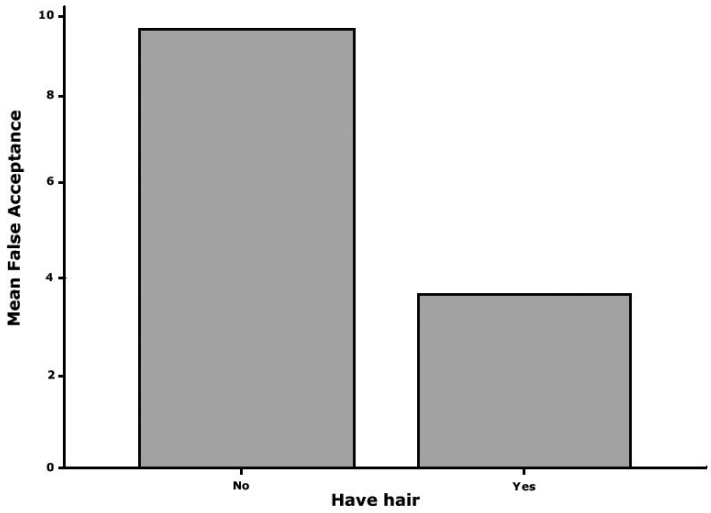


**Fig. 3.** The histogram shows a graphical overview of the false acceptances of the two groups with and without hair

## 4 Discussions

Our adversary model in a border control setting may contain some unrealistic assumptions. For example, we assumed that the adversary is able to obtain

access to several face databases of people in different countries, which may not be a realistic assumption. However, the adversary can probably buy photos and personal information from companies that sell personal data, or they may collect personal data on the Internet. Some of the stolen passports will probably be detected in border control assumed that the revocation routines are effective and in place. On the other hand, people whose passports have been stolen might not detect it right away.

The validity of the proposed indicator for the true false acceptance rate is also 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 experiment in Sect. 2 showed that the majority of the subjects generated at least one false acceptance (except for the data set II at the security level of 1 % FAR). There might be several reasons for such a high number of false acceptances in data. 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 data 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: people who do not belong to the target population, from which the training data set is selected, will probably generate higher number of false acceptances compared to people who belong to the target population.

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 the people having one or more look-alikes is probably much lower than the experiment demonstrated.

The eye coordinates of the passport photos in the data set III were generated automatically, which means that the 16 % of the eye coordinates were not correct. This has probably affected the number of false acceptances in the passport data set. On the other hand, this data set contained much higher number of subjects than the data set II, and this may indicate that the probability of finding look-alikes increases as the number of subjects in data set increases.

As we can see from the results of our second experiment, the false acceptance rate on the image-pairs where the hair is removed are significantly higher than for those where the hair is present. The hair is a feature that can be easily manipulated, indicating that there is in fact a great opportunity for an impostor to circumvent both the system and the human guard using simple and cheap methods. When combining this with facial make-up and the influence eyebrows, the colour of the eyes and beard have on human face recognition performance we see that using a human supervisor to increase the security may be insufficient. A better solution to achieve higher security would then be to employ multi-modal biometric systems [15–18].

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

It should be noted that only 45 of the 60 image-pairs in the experiment where the hair and background was removed were actually composed of face images of different persons, while 15 image-pairs were composed of images of the same person to control the results. This produces a average false acceptance rate of 21,36 %. Combining this with the observation that most of the face image-pairs where evaluated wrong by more than one individual, we have an indication that the human supervision does not provide sufficient additional security. It would in this context also be interesting to, in a further study, see how alteration of hair combined with manipulation of other features such as eyebrows, teeth and beard would affect the human recognition performance.

## 5 Concluding Remarks

Automatic identity verification of a passport holder by use of a face recognition system may not give significant additional security against identity theft in a border control setting unless additional security measures are used. For example, in order to increase the ability of a face recognition system to correctly verify an individual's identity and increase its robustness against spoofing attacks, the users that generate many false acceptances could have the higher verification threshold than the users with fewer false acceptances, or the face could be combined with some other biometric identifier.

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 border control application with non zero-effort impostors. The more representative measure for the true acceptance rate would be, for example, the percentage of the impostors who have at least 20 look-alikes in the target population.

In a border control the face recognition will be supervised by a human guard. However as observed in our experiment in Sect. 3, a human guard may provide insufficient additional security, especially because hair, which is a feature that is easy to manipulate, plays such a significant role in human evaluation of faces.

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.

## 6 Acknowledgments

The face images used in this work have been provided, among others, by the Computer Vision Laboratory, University of Ljubljana, Slovenia [8], Computer Vision Center (CVC) at the U.A.B. [10], Centre for Vision, Speech and Signal Processing at the University of Surrey [9] and the Gjøvik University College [11].

## References

1. ICAO: Biometrics Deployment of Machine Readable Travel Documents. ICAO TAG MRTD/NTWG. Technical Report, Version 1.9. Montreal. (May 2003)
2. United States General Accounting Office: Technology Assessment: Using Biometrics for Border Security. (November 14, 2002)
3. 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)
4. P.J. Phillips, P. Grother, R.J Micheals, D.M. Blackburn, E Tabassi and J.M. Bone: FRVT 2002: Evaluation Report. (March 2003)
5. D. Blackburn, M. Bone and P. J.Phillips: Facial Recognition Vendor Test 2000 Evaluation Report. (February, 2001)
6. John Daugman: Phenotypic versus Genotypic Approaches to Face Recognition. Face Recognition: From Theory to Applications. NATO ASI Series. Series F: Computer and system sciences. (**163**)
7. A. J. Mansfield and J.L. Wayman: Best Practices in Testing and Reporting Performance of Biometric Devices. Version 2.01. (August 2002)
8. Faculty of Computer and Information Science, University of Ljubljana, Slovenia: (CVL FACE DATABASE)
9. 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)
10. A.M. Martinez and R. Benavente: The AR face database, CVC Tech. Report #24. (1998)
11. (Gjvik University College (http://www.hig.no))
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)
13. M. Kosmerlj: Passport of the Future: Biometrics against Identity Theft? MSc thesis. Gjvik University College, NISlab. (June 30, 2004)
14. T. Fladsrud: Face Recognition Software in a border control environment: Non-zero-effort-attacks' effect on False Acceptance Rate. MSc thesis. Gjvik University College, NISlab. Master's thesis (June 30, 2005)
15. D. Maltoni, D. Maio, A.J., Prabhakar, S.: Handbook of Fingerprint recognition. Springer Verlag, New York, USA (June 2003)
16. Anil K. Jain, Arun Ross and Salil Prabhakar: An Introduction to Biometric Recognition. IEEE Transactions on circuits and systems for video technology **14** (January 2004)
17. Arun Ross and Anil Jain: Information Fusion in Biometrics. Pattern Recognition Letters **24** (2003) 2115–2125
18. Anil K. Jain and Arun Ross: Multibiometric Systems. Communications of the ACM **47** (January 2004)