Evaluation of automatic face recognition for automatic border control on actual data recorded of travellers at Schiphol Airport

L.J. Spreeuwers*, A.J. Hendrikse*, K.J. Gerritsen†

*Faculty of Electrical Engineering, Mathematics and Computer Science
University of Twente
P.O. Box 217
7500 AE Enschede, Netherlands
l.j.spreeuwers@utwente.nl
† Dutch Ministry of Internal Affairs

Abstract: Automatic border control at airports using automated facial recognition for checking the passport is becoming more and more common. A problem is that it is not clear how reliable these automatic gates are. Very few independent studies exist that assess the reliability of automated facial recognition for border control. In this evaluation study the reliability of automated facial recognition for automatic border passage was investigated. To investigate the quality of the images and face recognition, during 2 weeks data of real passengers were acquired at Schiphol Airport using 2 different automatic gates of about 950 passengers for both gates. This data alone already makes the evaluation study of great value. The evaluation experiment consisted of comparing live images of every passenger to the digital photographs stored on their passports. Every live image is compared to every digital passport photograph. In this way we can estimate both the False Accept Rate as well as the Verification Rate. In spite of the critical analysis in this study, the prospects for automatic border passage using face recognition are very good. We expect that, provided that the quality of the live images acquired by the gates is improved and if possible the quality of the digital photographs stored on the passport, excellent recognition results can be obtained with Verification Rates (VR) of above 99% at a False Accept Rate (FAR) of 0.1% or even lower.

1 Introduction

This study attempts to provide insight into the reliability of automatic border control/ passport checks by special gates using automatic face recognition at e.g. airports. These gates could partly replace human passport checks and possibly increase throughput. A point of concern is that there are very few objective studies on the reliability of these gates. This study addresses exactly the question of the reliability. It was carried out in the framework of the No-Q project [rMoSJ10] of the Dutch Ministry of Justice by the University of Twente. The official report of the evaluation study [SH11] can be requested from the Dutch Ministry of Security and Justice.

The construction of an automatic gate includes more than just automatic face recognition. Important factors are for example the ”protocol” used by the gate (e.g. first show passport, then walk to camera to make a picture, etc”), opening and closing of the gate, user-friendliness/ease of usage etc. Factors influencing the face recognition are e.g. the illumination, placing and quality of cameras etc.

This study only addresses the performance of the face recognition process. Data from 2 gates at Schiphol were used for the evaluation of face recognition algorithms of 5 different vendors. These data were actual passenger data using the gates in a normal way, which makes these data extremely valuable, as there are few datasets like these available.

There are three major factors influencing performance of the automatic face recognition:

1. Quality of the facial image digitally stored on the passport (the reference image)
2. Quality of the live recorded facial image by the gate (the probe image)
3. Quality of the face recognition algorithm used.

All three are investigated in this study. Beforehand, we already knew that there are some quality issues with the digital passport photographs.

In the subsequent sections, first the data used in the evaluation study is described. Next the evaluation setup is explained. Consequently, the results and finally conclusions are presented.

2 Gates and vendors

Two gates for automatic border passage and passport checking were used for the acquisition of the data used in this evaluation study. The gates are denoted gate 1 and 2.

In this study 5 vendors submitted face recognition algorithms for evaluation. In the study they are denoted vendor A-E. Some of the vendors submitted more than 1 algorithm. In that case, the algorithms are denoted A1, A2 etc. If a vendor submitted a single algorithm, it is simply denoted algorithm A, B etc. All vendors were given the opportunity to submit multiple algorithms. The vendors did not (officially) have knowledge about the used gates and the configuration. Several vendors complained that they had very little insight in the errors made. Unfortunately, because there was not enough data available and because of privacy issues, we could not make data available to them.

Gates and face recognition algorithms were available for this evaluation under the condition that results are published anonymously, hence no further details can be given here.

3 Data

The data for this evaluation study were acquired at Schiphol Airport, Amsterdam, The Netherlands in June and July 2010. The data were acquired using two different fully operational gates for automatic border passage. These gates will be called gate 1 and gate 2 from now on. The gates were placed next to the ”normal” border passage/passport check desks. Normal passengers were asked to take part in this evaluation study. They would then sign an agreement that their anonymised passport data would be used in this study, but that no data would be made public. They then presented their passport to the reader on one of the automatic gates, would position themselves in front of the built-in camera and the face recognition software in the gate would try to match the digitally stored photograph on the passport with the live recorded image. If a match was found, the gate would open and the passenger could proceed (and receive a small present to thank him for his cooperation). If the match failed, a fixed number of retries was attempted where new live images were recorded (in this case 2). If after the retries there is still no match, the subject is labelled by the face recognition software as an imposter. In normal operation the digitally stored photo on the passport and the live photographs are not stored. For this evaluation study, the gates were adapted in order to store the images. In addition to the digitally stored image from the passport and the live images, a scan of the printed photograph of the passport was stored. Three corresponding images (passport digital image, live image and scanned passport image) will be called a *triplet* in this study. The figures 1 and 2 show examples of the triplets recorded by respectively gate 1 and gate 2. From these figures, it can been seen that the live scans made by gate 1 are gray images, whereas those made by gate 2 are colour images.

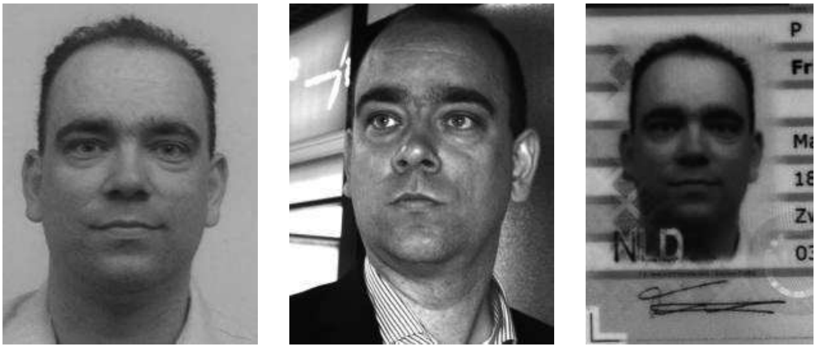
From gate 1 a total of 1434 triplets or 4302 images were recorded from 959 subjects. From gate 2 a total of 993 triplets or 2979 images were recorded of 964 subjects.

Gate	#subjects	#triplets	#images
gate 1	959	1434	4302
gate 2	964	993	2979

Table 1: Number of subjects, triplets and images recorded by the gates

The data set contains a mix of Dutch ($\approx 80\%$) and European travellers ($\approx 20\%$) and some from outside Europe. A small number of the recorded triplets were not from passengers, but from people from the Dutch Immigration Service (IND), people employed at Schiphol or otherwise involved in the No-Q project. The images shown in figures 1 and 2 were selected from these triplets and the imaged persons have given their consent to use their images in this study.

Some of the triplets are not genuine triplets (i.e. the live image shows the face of a different subject than depicted on the digitally stored image on the passport). These are denoted *impostors* as opposed to *genuines*. In total 75 of the triplets of gate 1 were identified as impostors and 15 triplets of gate 2. Most of the imposter triplets were employees at Schiphol or persons otherwise involved with the No-Q project who were asked to test the gates' reaction on imposter attempts. A few imposter attempts were caused by partners who accidentally swapped their passports. A single imposter attempt was caused by an identical twin who were asked to swap their passports as a test.



digital image | live image | scanned image

Figure 1: Examples of acquired images from gate 1

digital image | live image | scanned image

Figure 2: Examples of acquired images from gate 2

3.1 Quality of images

The two gates use different techniques to acquire the images from the passport and the live images. Therefore, image quality was assessed separately for the two gates.

Each individual image was annotated using a number of quality labels: *blur, eyeglasses, illumination, pose, scan lines, quality (normal, low, very low)* and *failed scan*. They are detailed below.

Blur The subject moved his head during the acquisition of the live image. This results in unsharp live images. In very rare cases, the digital passport images are slightly unsharp due to an improperly focused camera. Most of the scanned passport images are somewhat unsharp. The latter two types of blur are not considered here.

Eyeglasses The subject is wearing eyeglasses on the image. The reason for using this quality label is that some face recognition algorithms may have problems properly recognising subjects with glasses (e.g. if the glasses he is wearing on the live image are different from those on the digital passport photograph, or he is wearing glasses on only one of the images).

Illumination Poor illumination includes low contrast images, over exposure and shadows. Poor illumination can be found in the live images and in the passport images as well.

Pose This refers to images that show faces in a non frontal pose. Images showing faces with a rotation of 20 degrees or more around the vertical axis (looking to the left or to the right) and those with 30 degrees or more around the horizontal axis (looking up or down) are assigned this label.

Scan lines It turned out that many of the digital passport images contain vertical lines of constant colour across the full height of the image.

Quality An overall quality was assigned to each image. Three different quality levels were distinguished: *normal, low* and *very low* quality.

Failed scan In rare cases, the live scan failed and returned an empty image.

The results of the annotations are shown in table 2. Note that multiple labels can be assigned to a single image.

Label	Affected images	#Images gate 1	#images gate 2
Blur	Live	221	131
Eyeglasses	All	494	379
Illumination	All	121	39
Pose	All	69	93
Scan lines	Digital	264	135
Normal quality	All	4086	2886
Bad quality	All	162	47
Very bad quality	All	54	7
Failed acquisition	Live	0	13

Table 2: Annotation results for images acquired by gate 1 resp. 2

3.2 Discussion and conclusions

Firstly, many of the acquired live images of both gates show motion blur (about 15% for both gates). Poor illumination occurs more often for live images from gate 1 (about 9%) than from gate 2 (about 4%). Most illumination problems are due to over exposure or too dark images. A significant number of digital passport images also show illumination problems, generally low contrast. Pose problems are present in about 5% of the live images recorded by gate 1 and in 9% of the images recorded by gate 2. The scan lines problem was observed in digital passport images from both gate 1 (18%) as well as gate 2 (13.5%). It is therefore likely that the lines are present in the digitally stored photographs on the passport. The lines are probably caused by defective image scanning devices that are used to scan the paper photographs (which were printed from a digital original). Low quality images include images with not completely (but almost completely) visible faces (but no occlusions), images distorted by the camera lens (especially gate 2) and some of the images with more severe illumination problems. Very low quality images include live images with partly occluded faces (by a hand, passport etc.), faces of which a large part is not visible, faces that are very dark or show severe motion blur or pose deviation from frontal. Except for the blur and failed acquisition, all of the above quality issues occur not only for the live images, but also for the digital images stored on the passports, albeit less frequently. The following quality issues were observed with the digital passport images:

- Photograph cut out around the face and pasted on a white background. This leads to very sharp edges around the hair.
- Non frontal pose.
- Wrong eye colour due to compression or red eye correction.
- Cracks in photo or dust/hairs on photo.
- Colour smudges or stains on photographs.
- Poor contrast. This is the most frequent issue.
- Scan line failures.
- Very high contrast and over exposure.
- Compression artifacts resulting in blocks/posterising. A significant number of the images show moderate to heavy compression artifacts.
- Distorted aspect ratio (i.e. the face is vertically stretched or compressed).
- Combinations of the above artifacts.

All these deficiencies were on passports from different countries except for the bad scan lines which were mostly found on the Dutch passports. Because the handling of photographic material has changed since, new Dutch passports don't show this deficiency anymore.

The scanned passport images are generally of lower quality as well, because they are partly covered by (sometimes shiny) marks on the passport and the images on the passport are already of suboptimal quality. For this evaluation study we actually do not use these images, so this is not a real problem.

Summarising, we conclude that:

1. The images captured by the two gates show several artifacts that could have serious impact on recognition performance. Blurred images, non-frontal images and poorly illuminated images are the most frequent problems. Live images recorded by gate 2 often show quite severe lens distortion effects. These errors could have been avoided easily by using a more professional camera and illumination setup.

2. A large number of the digital facial images stored on passports (around 5%) shows quite severe quality deficiencies that could have serious impact on the recognition

performance. Most occurring deficiencies are poor contrast, compression artifacts and scan line failures.

4 Evaluation setup

4.1 Division of data into subsets

There are two data sets available from the two gates with respectively 1434 triplets of 959 subjects (gate 1) and 993 triplets from 964 subjects (gate 2). First, all imposter data were removed (75 resp. 15 triplets) and also the triplets with failed live scans were removed (13 for gate 2) and the triplets with bad quality and very bad quality images (216 resp. 54 for gate 1 and 2).

Next the remaining data of each of the two gates are subdivided into four sets: a *clean set*, a *realistic set*, a *challenging set* and a *last in series set*. The clean set only contains images of normal quality and without blur, illumination and pose or scan line problems. The realistic set contains in addition the images with illumination and scan line problems. The challenging set in addition also includes blurred images and images with faces with large pose variations. The last in series set always takes the last triplet of a series of triplets of the same subject (if there are multiple subsequent recordings of the same subject). Multiple subsequent triplets of the same subject are recorded if the subject was not recognised using previous first triplets. Normally, the last in a series is expected to be the best quality of the series. This selection is made by the face recognition software running on the gates. It has to be noted that the face recognition used in this selection was unknown to the vendors of the face recognition algorithms tested. In the last in series set, the bad and very bad quality images are included. The subsets are summarised in table 3.

Set	Description	#gate 1	#gate 2
Clean	No blur, pose, illum., scan lines	741	513
Realistic	No blur, pose	862	687
Challenging	All except bad quality images	952	911
Last in series	Last triplet from sequence	982	967

Table 3: The four subsets and the number of triplets

Two experiments are carried out for each algorithm available: a ROC test and a FAR stability test. They are described in the subsequent sections.

4.2 ROC test

For the ROC test, all live images are compared to all digital passport images. Each comparison results in a score. The score is compared to a threshold. If the score is above the threshold and the comparison is an imposter pair (i.e. the live image and the digital passport image are from different subjects), this results in a *False Accept*. The number of false accepts divided by the total number of imposter comparisons gives the *False Accept Rate* or *FAR*. Likewise, the number of genuine comparisons (both images are from the same subject) where the score is above the threshold divided by the total number of genuine comparisons gives the *Verification Rate* or *VR*. The number of genuine comparisons where the score is below the threshold divided by the total number of genuine comparisons gives the *False Rejection Rate* or *FRR*. By definition: $VR = 1 - FRR$. By varying the threshold and plotting the VR against the FAR, we obtain the *Receiver Operating Curve* or *ROC*. The ROC shows the trade-off between the FAR and the VR, i.e. between a high recognition rate and the probability that an imposter is accepted.

A common choice for the FAR is 0.001 (0.1%), i.e. 1 out of 1000 impostors is allowed. A FRR of more than a few percent (i.e. several out of 100 genuines are not recognised and must be rechecked by a human) will lead to delays and longer queues. The threshold resulting in the selected FAR is a parameter setting for the face recognition algorithm that is fixed in normal operation of the gate and can be optimally chosen using the ROC.

For each of the subsets of the data and for each of the face recognition algorithms, the ROC has been determined. The results are presented in section 5.

4.3 FAR stability test

The ROC gives the trade-off between FAR and VR and allows us to choose a threshold for the score for a certain desired maximum FAR. However, if the data set that is used to determine this threshold is too small or deviates in some way from the actual data, the true FAR can be different from the chosen FAR. In order to investigate the stability of the FAR, we used the clean subset and split it into two equally large parts. One half was used to determine the ROC and the threshold for a FAR of 0.001. The second half of the data were then used to measure the FAR. We repeated this experiment 50 times by randomly splitting the clean data set in two equally large parts. We report the average measured FAR and its standard deviation and the maximum and minimum of the measured FAR for both gates and all face recognition algorithms. The results are presented in section 5.

4.4 Accuracy considerations

The FAR, FRR and VR are estimated from data. If we have N triplets of N distinct subjects, we have N genuine comparisons and N*(N-1) imposter comparisons, i.e. for each subject there is exactly one genuine comparison and there are N-1 imposter comparisons. In the data set of gate 2 there are hardly any subjects with more than 1 triplet. In the data set of gate 1 the situation is a bit more complicated, because for about 20% of the subjects there are 2 or 3 triplets. For a subject with K triplets, there are K*K genuine scores and K*(N-K) imposter scores.

Some rough calculations then show that for gate 2 there are around 500-1000 genuine comparisons for the clean set to the challenging set and around 250000-1000000 imposter comparisons. If we require a reasonably accurate estimate of the FAR or FRR we need at least 10 comparisons. This means we can only estimate a FAR of above 10/250000=0.04% to 10/1000000=0.01% accurately and FRR of above 10/500=2% and 10/1000=1%. Thus we should take into account that an estimated FAR below 0.01-0.04% and a FRR below 1-2% might be inaccurate.

Likewise for gate 1 we can roughly estimate the number of genuine comparisons for the clean set as: 0.8*750+0.2*750*3=1050 to for the challenging set: 0.8*950+0.2*950*3=1330. The number of imposter scores for the gate 1 data set becomes: 0.8*750*749+0.2*750*(750-3)=561450 for the clean set to for the challenging set: 0.8*950*949+0.2*950*(950-3)=901170. Thus for the gate 1 data set the accuracy boundaries are around 10/1000=1% to 10/1300=0.8% for the FRR and 10/500000=0.02% to 10/1000000=0.01% for the FAR.

These calculations show that with the given data sets we cannot estimate the FAR accurately below around 0.05% and the FRR below around 1%-2%.

The estimation of the stability of the FAR using the described method has one disadvantage: the data used were all recorded in a small interval of time during about 2 weeks. If circumstances change (e.g. illumination, background), the data may change which may cause the FAR (and FRR) to change. With the given data, however, the described method provides at least some insight in the stability of the FAR. In the ideal case, the gate should be tested regularly using a similar approach as in this evaluation study: record data from a number of passengers and measure the FAR and FRR for the threshold setting in the gate.

5 Results

5.1 ROC tests

The ROC's for all algorithms for the four subsets of the data from gate 1 and 2 were determined by performing an all-to-all comparison experiment. This means all live images of a subset were compared to all digital passport images of the same subset.

Some of the images could not be processed by some of the algorithms (Failure to Enrol=FTE). Table 4 shows the Failures To Enrol (FTE's). In the determination of the ROC's these images were not included, which means that for algorithm B and D, which have the most FTE's the actual FRR may be up to 4% (algorithm B) and 8% (algorithm D) higher (the Verification Rate VR is 4%, resp. 8% lower) than shown in the ROC's.

From the ROC's we determined the VR at FAR=0.1%. The results for all algorithms and data sets for the two gates are shown in table 5.

Firstly, we can observe that overall the algorithm of vendor B performs best. However, we have to take into account that due to FTE's, the actual verification rate can be up to 4% lower for gate 1 and up to 1% lower for gate 2. This means that the results are much closer to those of algorithm C1, especially for the important last in series sub set.

Algorithm	gate 1 # FTE live	gate 1 # FTE digital	gate 2 # FTE live	gate 2 # FTE digital
A1-A4	0	0	0	0
B	39	1	5	4
C1-C4	3	0	3	0
D	72	4	15	3
E	0	0	0	0

Table 4: Failures To Enrol (FTE) of live images and digital passport images

	VR gate 1				VR gate 2			
Algo-rithm	**clean Clean**	**rea-listic**	**challen-ging**	**last in series**	**clean Clean**	**rea-listic**	**challen-ging**	**last in series**
A1	83.6%	80.5%	79.4%	90.4%	93.9%	93.9%	92.4%	92.4%
A2	84.4%	81.4%	80.6%	91.4%	94.9%	94.7%	93.0%	92.9%
A3	87.4%	84.1%	83.3%	93.1%	95.9%	95.3%	94.2%	93.9%
A4	87.7%	84.9%	83.7%	92.8%	95.4%	95.1%	93.5%	93.4%
B	**97.1%**	**98.1%**	**97.5%**	**99.2%**	**99.7%**	**99.6%**	**98.6%**	**98.5%**
C1	88.3%	85.7%	84.5%	95.3%	97.1%	97.0%	95.4%	95.4%
C2	76.6%	74.5%	73.2%	86.8%	91.3%	91.0%	89.2%	89.1%
C3	81.7%	78.7%	77.2%	90.1%	94.2%	94.0%	92.3%	92.4%
C4	86.0%	82.6%	81.6%	93.3%	95.6%	95.3%	93.7%	93.7%
D	58.3%	55.2%	53.9%	65.8%	74.5%	74.1%	71.4%	71.5%
E	86.5%	83.8%	82.4%	92.5%	94.9%	94.9%	93.5%	93.4%

Table 5: Comparison of all algorithms for FAR=0.1% for the different subsets. **Note: Algorithms B has many FTE's and the actual VR is up to 4% lower for gate 1 and up to 1% for gate 2!**

Secondly, in general the performance on gate 2 is significantly better than on gate 1. It is likely that this difference is caused for a large part by the better illumination of the live images of gate 2 (see table 2).

Interestingly, the last in series sub set for gate 1 gives significantly better performance than the other sub sets for all algorithms, but for gate 2 the performance for the last in series sub set is slightly worse. Apparently, repeated attempts help to remove poor live recordings for gate 1, while for gate 2 they do not contribute to obtaining better live images.

5.2 FAR stability tests

For the FAR stability tests, the clean sets were divided into two equally large subsets. One of them is used to determine the threshold for a FAR of 0.1% while the other is used to measure the FAR on an independent set. This was repeated 50 times for random divisions. The resulting average, standard deviation, minimum and maximum measured FAR's are reported in table 6.

From the results we can see that the average of the measured FAR is very close to the FAR chosen in the ROC. We can also observe that one can expect a maximum value for the FAR of about 0.15% instead of the selected 0.1% (for some algorithms even up to 0.18%). This suggests, that if we want to ensure the maximum operational FAR is below 0.1%, we should set a somewhat higher threshold for the matching scores. This however, will lead to lower verification rates. We did not investigate what FAR should be chosen in the ROC in order to ensure a measured FAR of 0.1%.

Another important consideration is that the data used in this test were all acquired in a period of approximately 2 weeks. If the gates are operated during a longer period of time, there may occur larger differences between the FAR used to select the threshold using the ROC and the actually measured FAR. It is, therefore, advisable to repeat FAR measurements regularly.

Algo-	**Measured FAR gate 1**				**Measured FAR gate 2**			
rithm	**Mean**	**St.Dev**	**Min**	**Max**	**Mean**	**St.Dev**	**Min**	**Max**
A1	0.101%	0.018%	0.064%	0.145%	0.102%	0.020%	0.065%	0.148%
A2	0.100%	0.023%	0.058%	0.160%	0.100%	0.017%	0.054%	0.147%
A3	0.100%	0.020%	0.062%	0.145%	0.098%	0.024%	0.062%	0.186%
A4	0.099%	0.018%	0.070%	0.147%	0.100%	0.017%	0.068%	0.158%
B	0.101%	0.015%	0.070%	0.138%	0.102%	0.017%	0.061%	0.141%
C1	0.102%	0.017%	0.069%	0.151%	0.101%	0.014%	0.076%	0.130%
C2	0.101%	0.016%	0.062%	0.138%	0.103%	0.016%	0.068%	0.130%
C3	0.100%	0.018%	0.067%	0.133%	0.099%	0.015%	0.068%	0.129%
C4	0.102%	0.013%	0.077%	0.130%	0.101%	0.016%	0.071%	0.138%
D	0.128%	0.029%	0.085%	0.181%	0.130%	0.022%	0.065%	0.171%
E	0.096%	0.022%	0.054%	0.154%	0.103%	0.023%	0.066%	0.161%

Table 6: FAR stability test for the clean subset from gate 1 and 2

6 Conclusions

In this evaluation study we investigated the reliability of automatic facial recognition for automatic border passage using specialised gates.

Concerning the quality of the digital passport images: it is shocking to see the poor quality of the images. About 5% of the images of the data set contains serious deficiencies. Major issues include: poor or too high contrast, compression artifacts, dust and hairs on photographs and cracks, bad scan lines, non frontal pose, colour smudges and stains and blurred images. The introduction of a better approach to capture and store the images on the passport could easily reduce the fraction of poor quality digital passport images from currently 5% to close to 0%.

Two different gates were used to acquire 2 separate data sets. Gate 1 acquires gray images and gate 2 colour images. The live images of especially gate 1 suffered from poor illumination: many images showed large areas with over exposure and many images were very dark. Both gates used slow shutter times resulting in a significant number of images with motion blur. The lens of the camera in gate 2 caused significant deformation of the acquired facial images. Better illumination control and better quality cameras can significantly improve the quality of the live recorded facial images by the automatic gates. These recommendations were taken into account with the implementation of automatic border control at Schiphol Airport.

A total of 11 algorithms of 5 vendors were evaluated using 2 tests: an ROC test and a FAR stability test. The ROC test showed that algorithm B performed best on data from both gate 1 and gate 2. Algorithm C1 followed and then A3 and E). Algorithm D performed poor on all data. The results show that at a False Accept Rate (FAR) of 0.1% a Verification Rate (VR) of up to 97.5% can be achieved (the measured VR was even higher, but there were some Failures To Enrol (FTE's) which have to be taken into account). In practice this means that 1 out of 1000 impostors is not recognised as an imposter and can pass, while 3 out of 100 genuines are stopped, because they are not recognised as genuine owners of their passports. The best results were obtained for gate 2. Gate 1 results were worse probably because of the poor illumination of the faces. The FAR stability test showed that if the threshold, that is used to compare matching scores to (above threshold means positive recognition, below negative), is selected for a fixed FAR using the ROC, the actual FAR may be up to 50% higher. If this is unacceptable, a higher threshold should be chosen.

There are two important considerations concerning the accuracy of this evaluation study. The first is about the size of the data sets. The number of samples allows us to calculate the FAR down to about 0.05% and the VR up to 98-99% (FRR of down to 1-2%). For reliable estimates of FAR, VR and FRR beyond these limits, larger data sets are required. A second consideration concerns the fact that all data were acquired within a period of two weeks. If the gates are operated during a longer period of time, it is likely that the characteristics of the live recorded images will change over time (due to changes in illumination, background), which will influence the FAR and VR. It is therefore advisable to repeat FAR and VR measurements regularly.

If the live image acquisition is improved significantly, we can expect Verification Rates of above 98% at False Accept Rates below 0.1%. In our opinion, these are very acceptable numbers for operating an automatic border passage gate. If a better process is used to acquire and store the digital photographs on the passport, the results will further improve to above 99% Verification Rate.

References

[rMoSJ10] Press release: Ministry of Security and Justice. Electronic border crossing starts today. `http://www.government.nl/documents-and-publications/press-releases/2010/05/25/electronic-border-crossing-starts-today.html`, 2010.

[SH11] L.J. Spreeuwers and A. Hendrikse. No-Q evaluation project report. Internal Report by University of Twente/Dutch Ministry of Security and Justice, June 2011.