Aurora Face Recognition Engine Validation Report B:
Simple Verification Statistics

Timur Almaev

psxta4@nottingham.ac.uk

Michel Valstar

michel.valstar@nottingham.ac.uk

Computer Vision Laboratory
School of Computer Science
University of Nottingham

2nd of October 2015

A. Executive summary

This report provides an independent validation of Aurora's latest face recognition technology. A team from the Computer Vision Lab of the School of Computer Science, University of Nottingham has performed the evaluation study as an independent 3d party. The results obtained demonstrate that Aurora's own verification figures are correct, attaining excellent performance. Equal Error Rates, putting equal emphasis on False Acceptance and False Rejection rates, are as low as 1.1%.

B. Introduction

The purpose of this document is to provide results of independent validation of the Aurora's latest to date face recognition engine in terms of its verification capability in a way that is both illustrative and common in the industry to enable easy comparison with other companies on the market. The results are presented with TAR (True Acceptance Rate) vs FAR (False Acceptance Rate) and FAR vs FRR (False Rejection Rate) ROC graphs as well as tables of Equal Error Rate (ERR) and FRR values for different target levels of FAR.

Software

The figures presented in this report reflect performance of the most Aurora face recognition engine called stingray-v3. The engine is provided to the University of Nottingham in the form of a "black box" with no source code disclosed. However additional actions were taken to ensure the validity of the results, please see section D for more details.

Dataset

Evaluation has been performed using Aurora's own proprietary dataset of infrared (IR) images, referred to as the master dataset later in the text. It is composed of in total 96,377 facial images of 11,538 different subjects. The number of images per subject varies from 2 to 97.

Report Structure

The remainder of this report is organised as follows: section C provides evaluation figures obtained on a number of subsets of the master database provided by Aurora. The purpose of section C is to provide baseline to compare against results obtained in section D. Section D provides results obtained using a subset of the master dataset constructed in the University of Nottingham. Apart from that section D also includes additional actions to ensure validity of the results.

C. Aurora's own subsets of data verification

In this section of the report the system is tested on the listed below 4 different mutually overlapping subsets of the master dataset all provided by Aurora. The purpose of this section is to first of all ensure that the results obtained in the University of Nottingham are identical to those obtained by Aurora, and second is to provide a baseline for further experiments on custom subsets of data, please see section D for more details.

- **F-FaL**. In this subset of the data a number of last images were selected from the start and end of each person's temporally ordered image set for maximum temporal variance between images. Images that require a lot of padding for alignment were removed. F-FaL is composed of 522 subjects.

- **F-Random**. This subset of the data is the same as F-FaL, with the exception that all of the images were selected randomly rather than from the beginning and end of a set. F-Random is composed of 522 subjects.

- **HF-FaL**. This subset of the data is the same as F-FaL, but in addition images that require even a small amount of padding for alignment removed. HF-FaL is composed of 490 subjects.

- **HF-Random**. This subset of the data is the same as HF-FaL, with the exception that all of the images and subjects were selected randomly. HF-Random is composed of 490 subjects.

Every subset of the data contains 11 images of each subject and is stored in a separate XML file. During verification every image of every subject is compared against each other for every subset of the data.

Figures and tables below provide a summary of results per subset of the data. For each of the subsets there are two FAR vs FRR ROC curve graphs of different scales as well as a TAR vs FAR ROC curve graph. There is also a table indicating levels of FRR for different values of FAR.
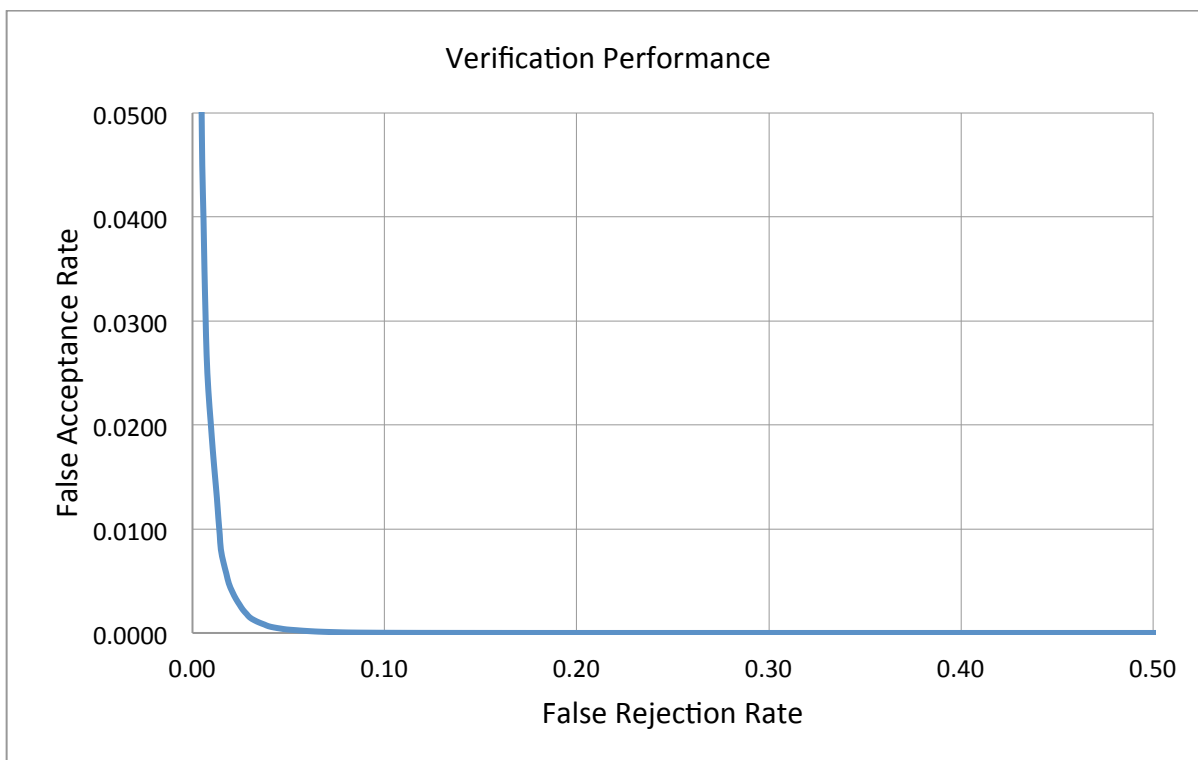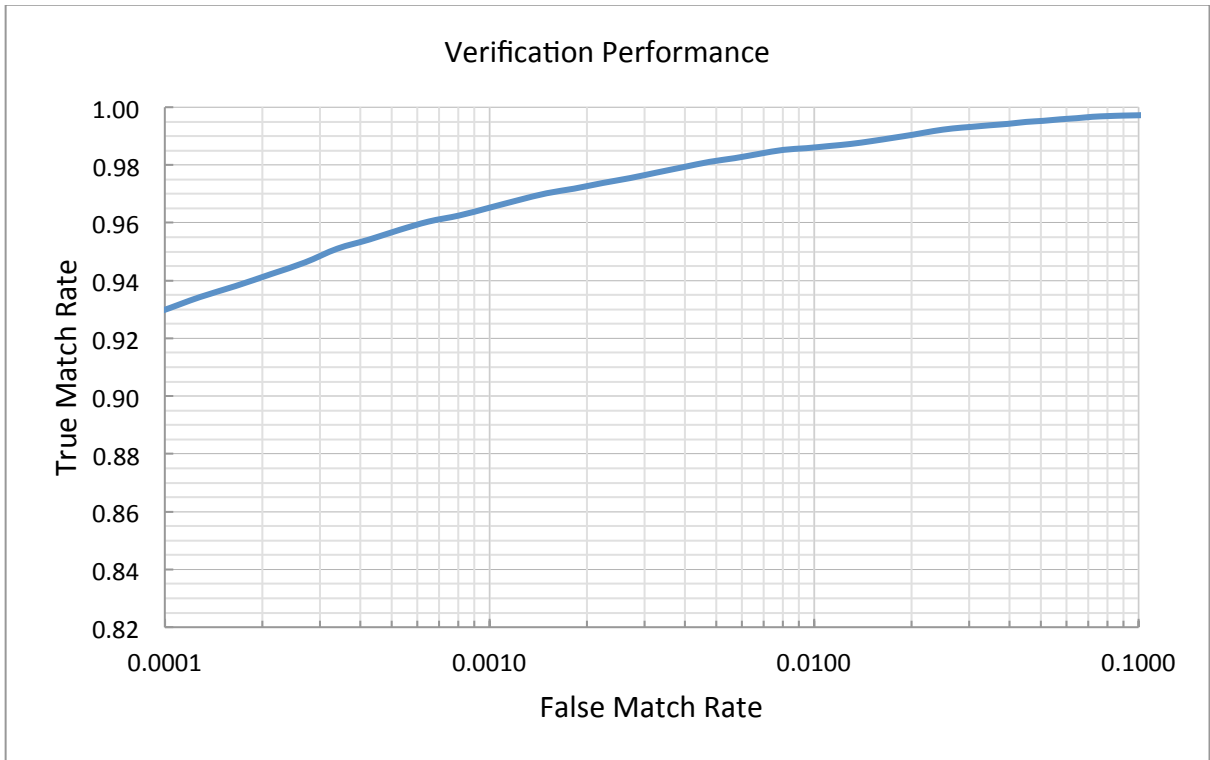
Figure 1a. ROC curve for the F-FaL dataset.
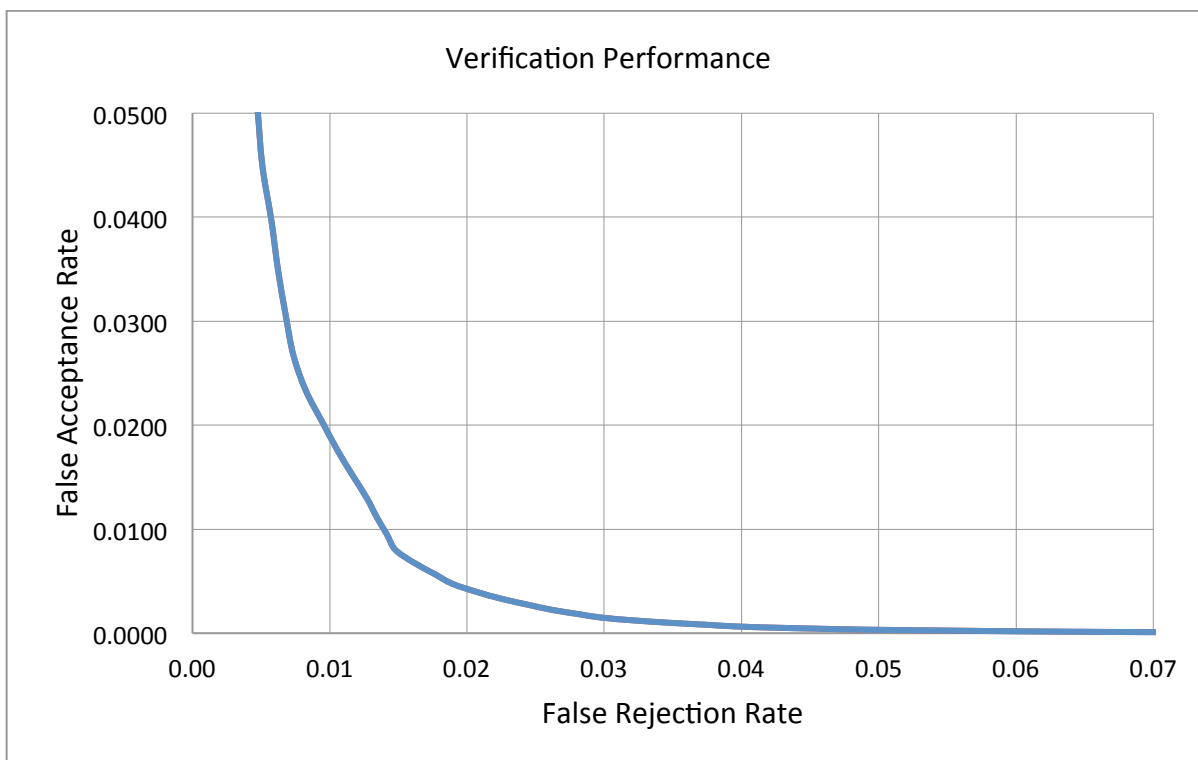
Figure 1b. ROC curve for the F-FaL dataset.

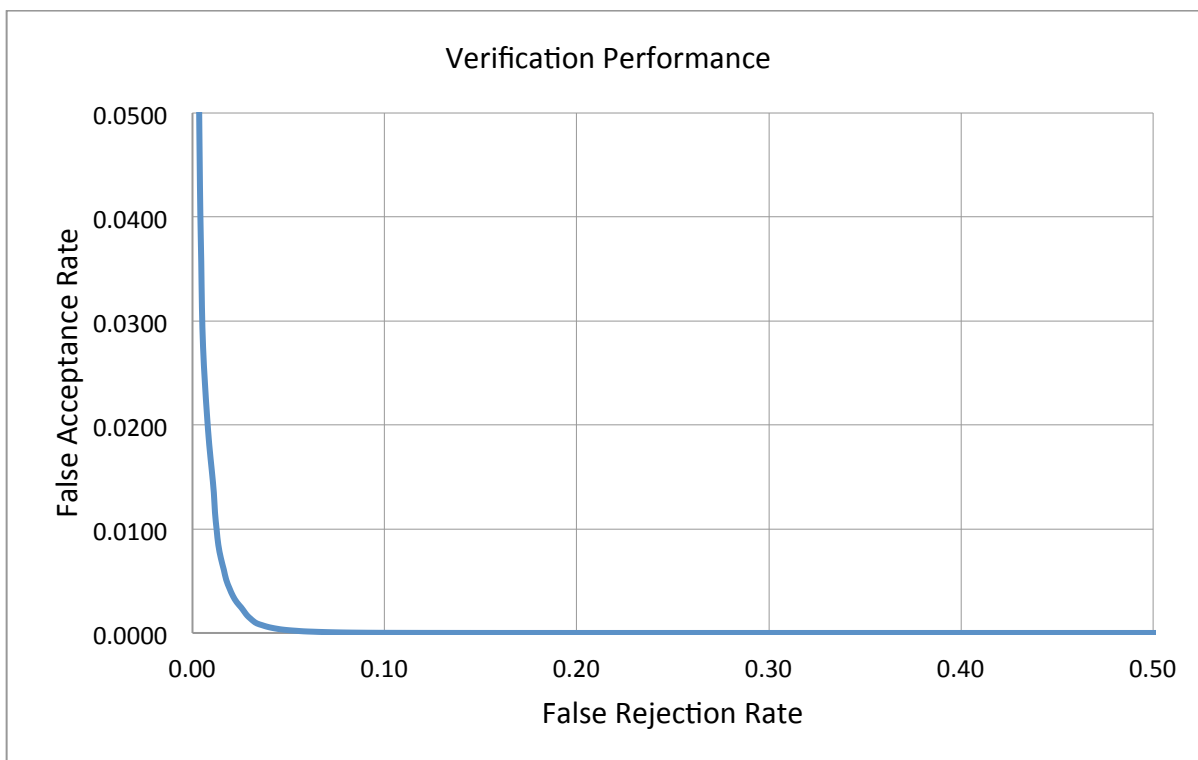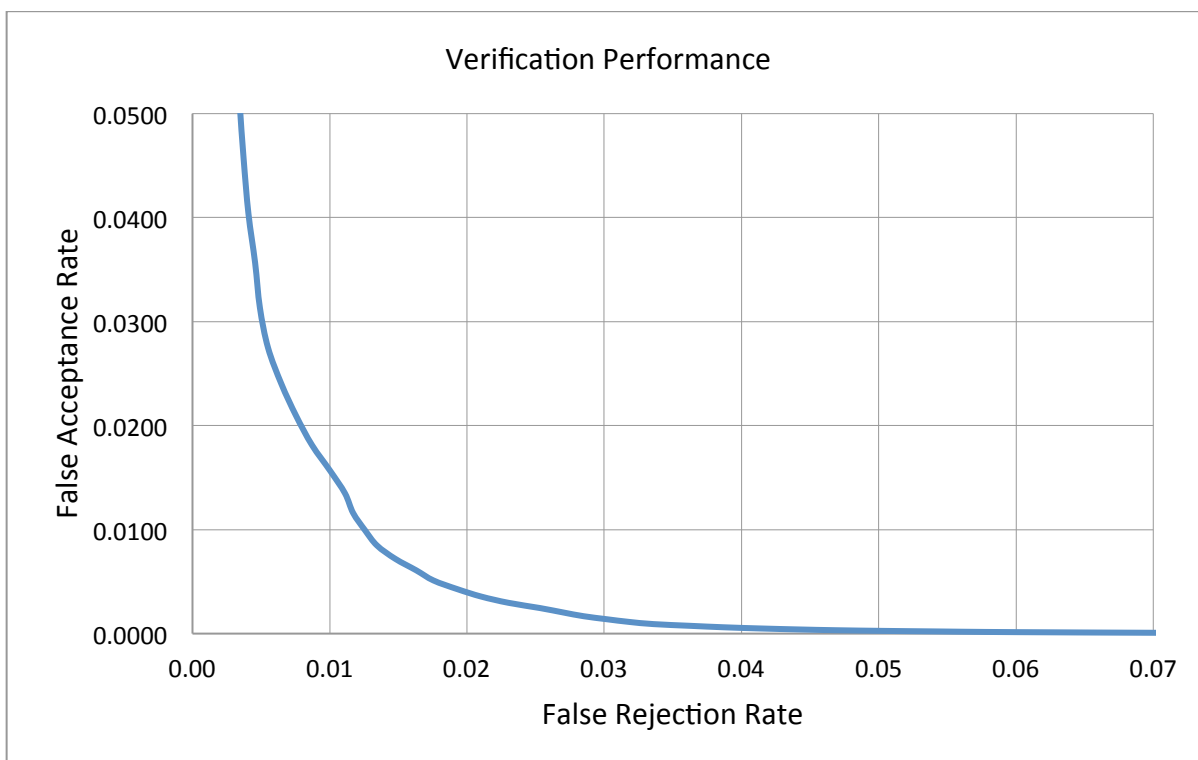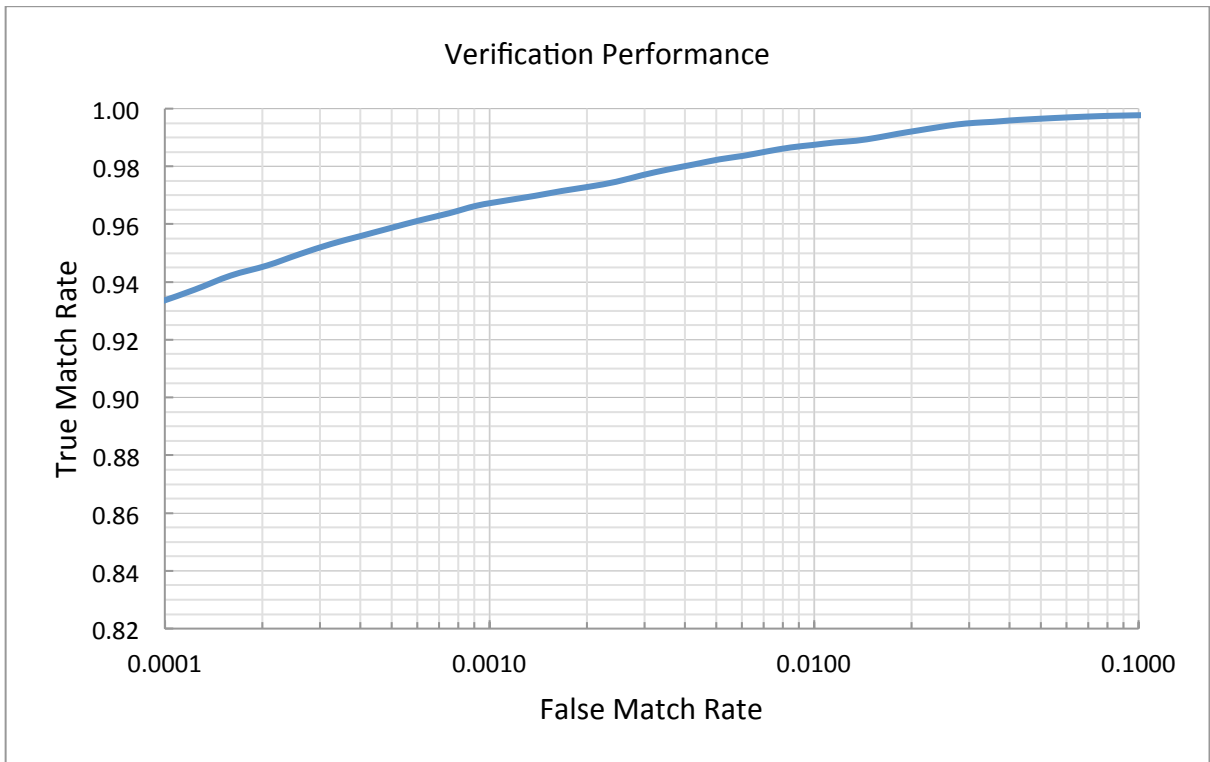**Figure 1c.** ROC curve for the F-FaL dataset.

FAR	FRR
0.000000	0.198989898990
0.000005	0.143573667712
0.000010	0.115395332637
0.000050	0.080215952630
0.000100	0.070045280390
0.001000	0.034796238245
0.010000	0.013897596656
0.020000	0.009543712992
0.050000	0.004702194357
0.100000	0.002751654476
0.200000	0.001985370951
0.400000	0.000452803901
0.800000	0.000000000000
EER	**0.012775110597**

Table 1. EER and FRR values for different levels of FAR of the F-FaL dataset.

Figure 2a. ROC curve for the F-Random dataset.


Figure 2b. ROC curve for the F-Random dataset.

Figure 2c. ROC curve for the F-Random dataset.

FAR	FRR
0.000000	0.201219087426
0.000005	0.134238941135
0.000010	0.110623476141
0.000050	0.077603622431
0.000100	0.066248693835
0.001000	0.032810867294
0.010000	0.012504353884
0.020000	0.007906652734
0.050000	0.003448275862
0.100000	0.002264019505
0.200000	0.001811215604
0.400000	0.000522466040
0.800000	0.000000000000
EER	**0.011692737093**

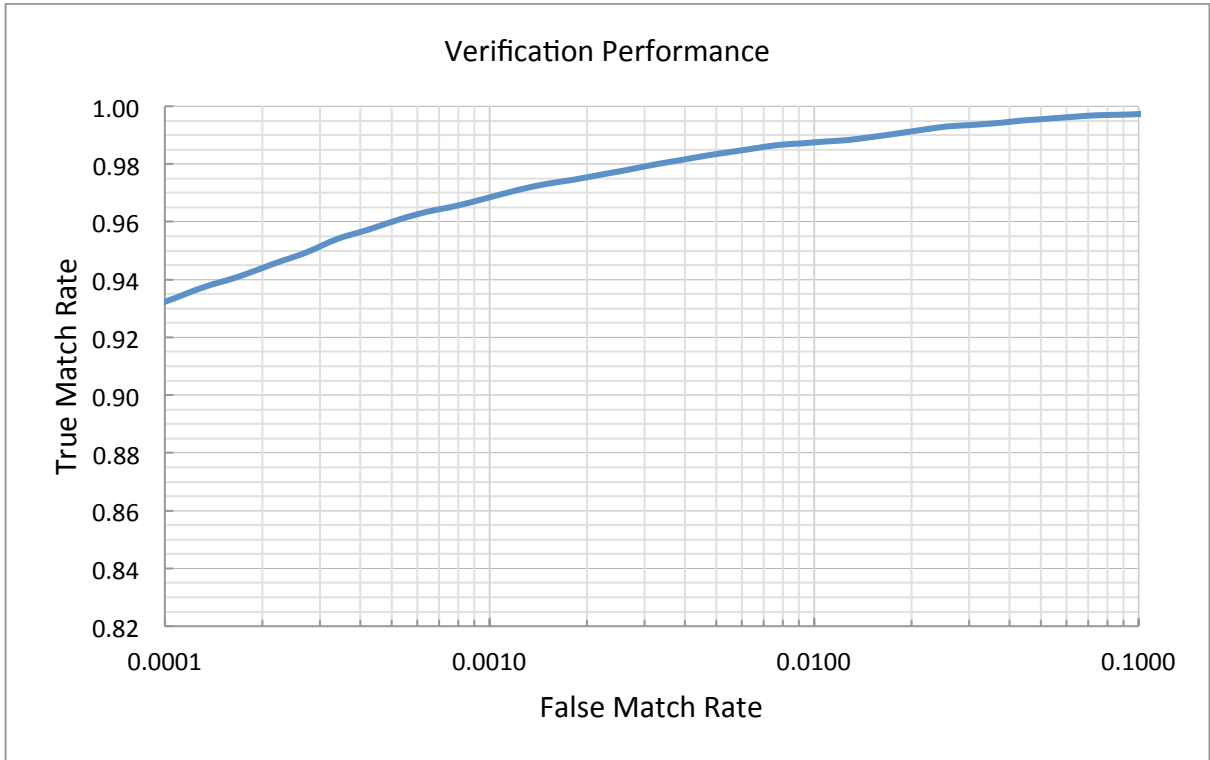**Table 2.** EER and FRR values for different levels of FAR of the F-Random dataset.
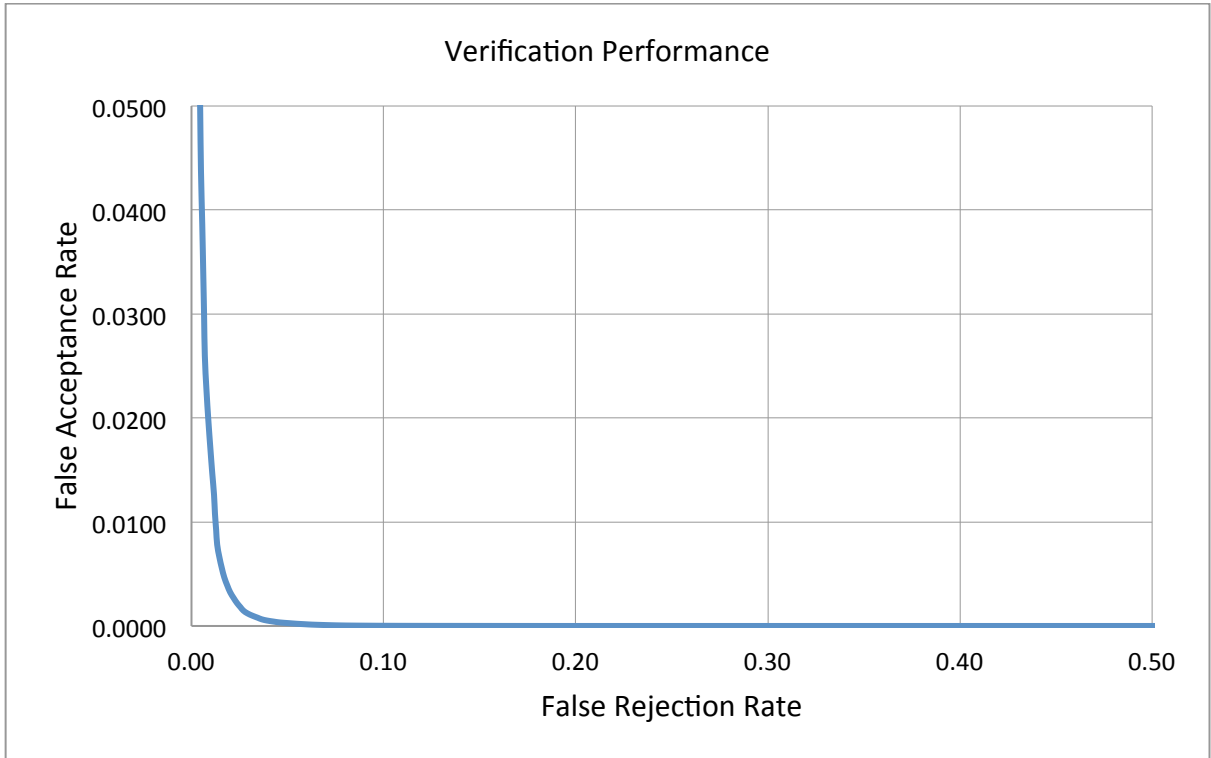
Figure 3a. ROC curve for the HF-FaL dataset.
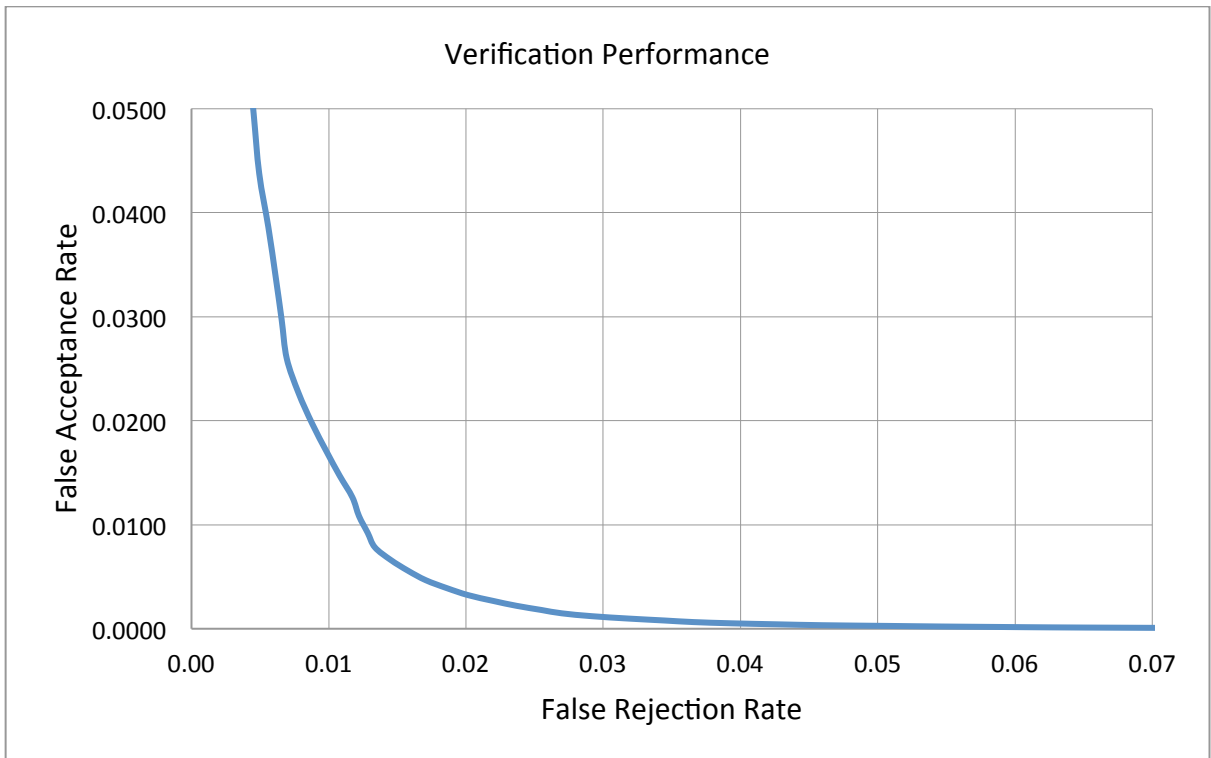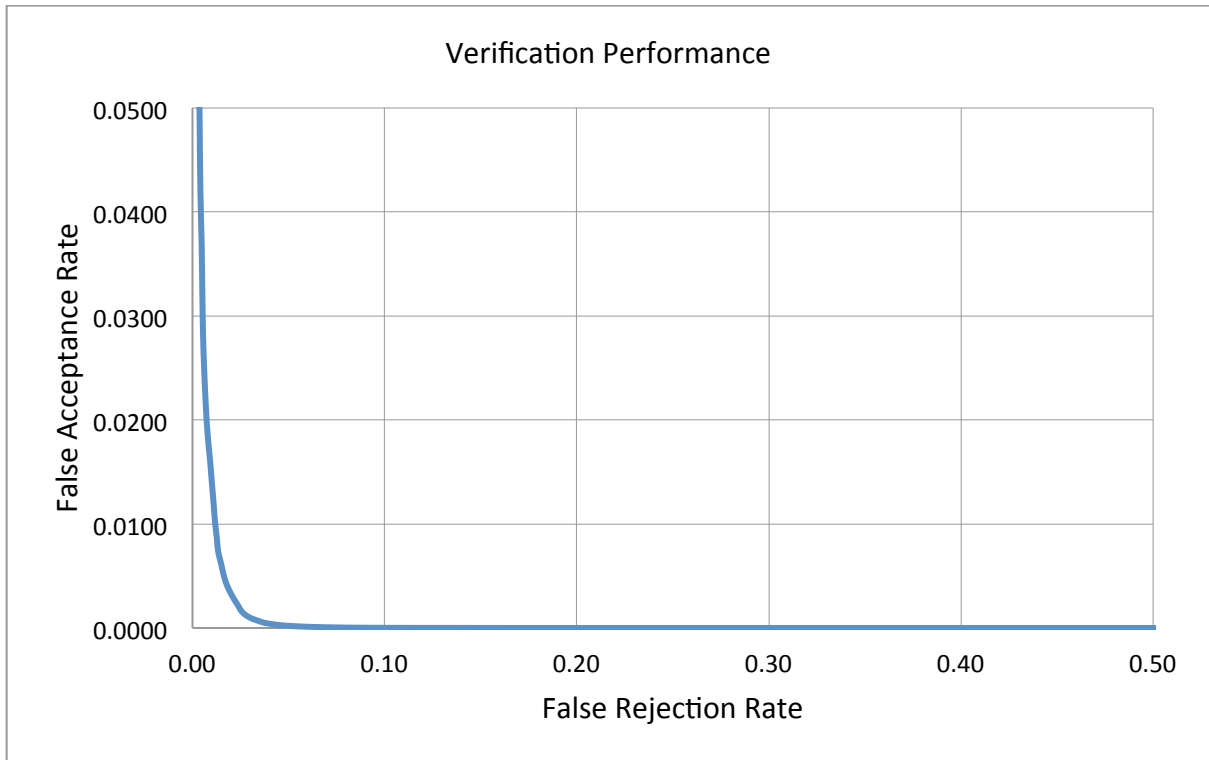

Figure 3b. ROC curve for the HF-FaL dataset.

Figure 3c. ROC curve for the HF-FaL dataset using stingray-v3 engine.
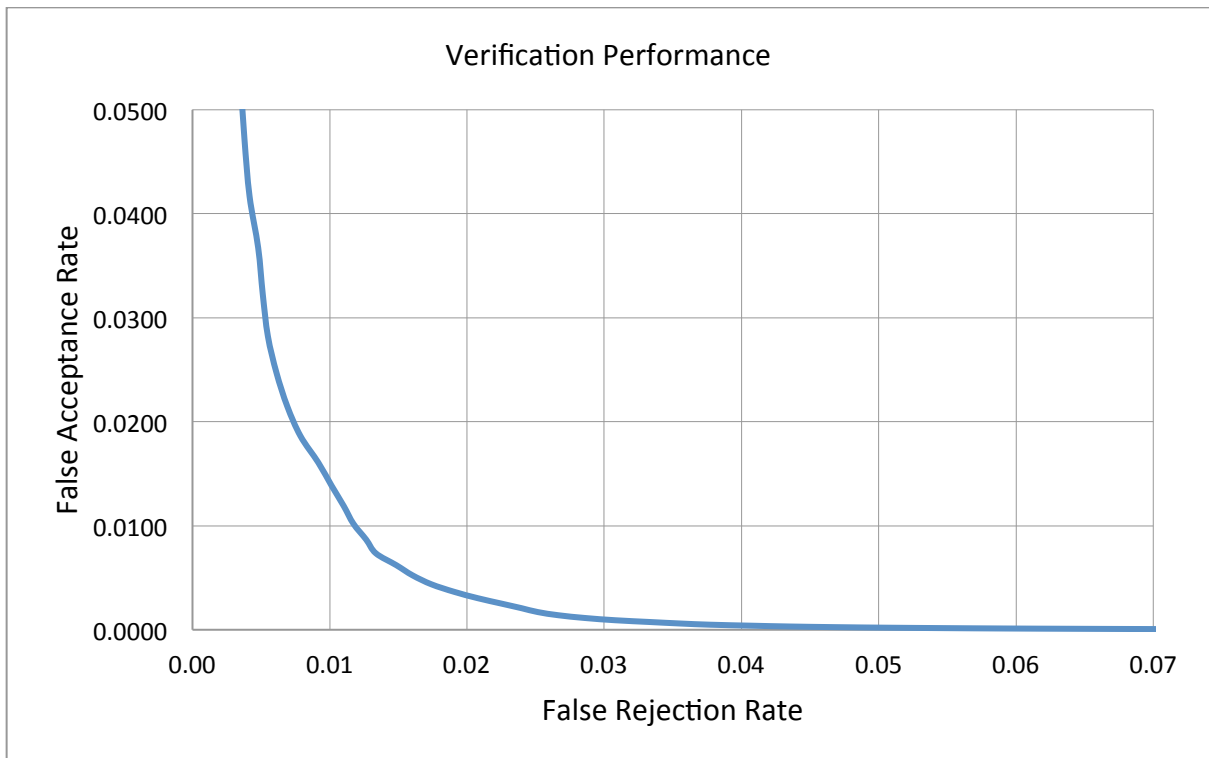
FAR	FRR
0.000000	0.215324675325
0.000005	0.153803339518
0.000010	0.123821892393
0.000050	0.078775510204
0.000100	0.067606679035
0.001000	0.031465677180
0.010000	0.012430426716
0.020000	0.008571428571
0.050000	0.004452690167
0.100000	0.002671614100
0.200000	0.001929499072
0.400000	0.000482374768
0.800000	0.000000000000
EER	**0.011947817407**

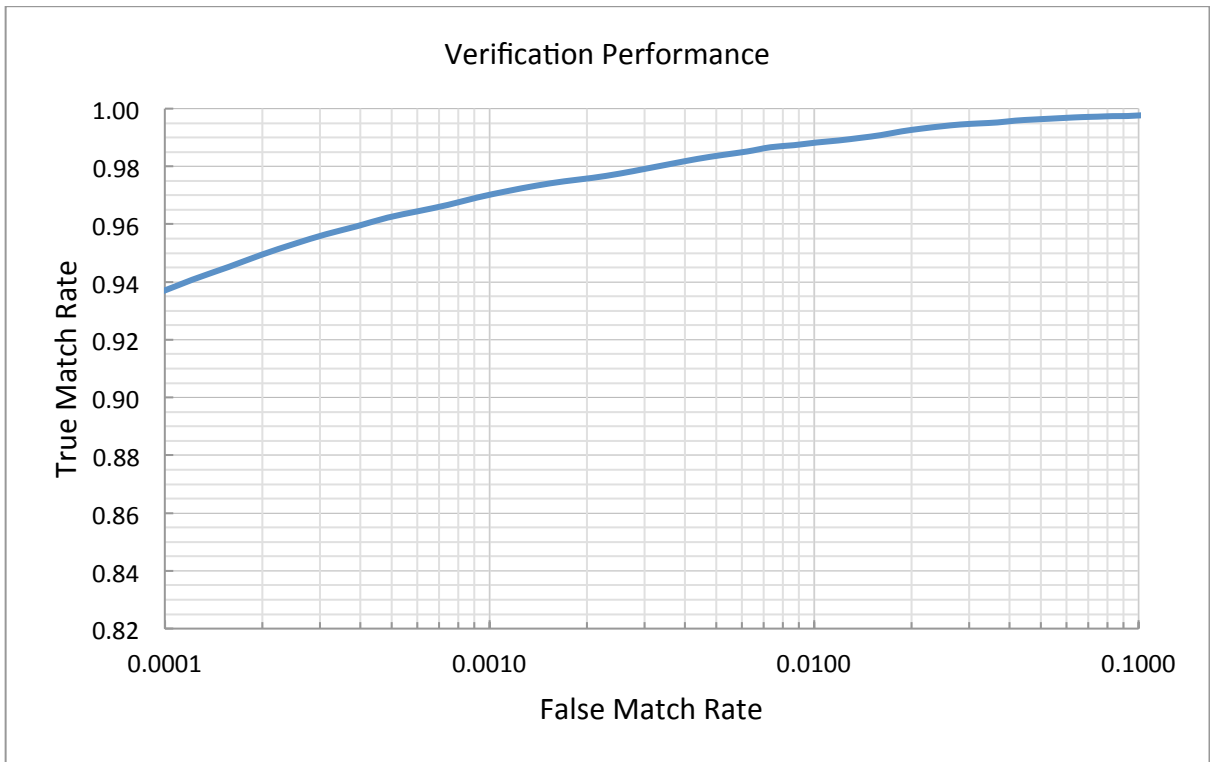**Table 3.** EER and FRR values for different levels of FAR of the HF-FaL dataset.

Figure 4a. ROC curve for the HF-Random dataset.

Figure 4b. ROC curve for the HF-Random dataset.

**Figure 4c.** ROC curve for the HF-Random dataset using stingray-v3 engine.

FAR	FRR
0.000000	0.184155844156
0.000005	0.136252319109
0.000010	0.112059369202
0.000050	0.075027829314
0.000100	0.063302411874
0.001000	0.029536178108
0.010000	0.011799628942
0.020000	0.007272727273
0.050000	0.003636363636
0.100000	0.002337662338
0.200000	0.001558441558
0.400000	0.000259740260
0.800000	0.000000000000
EER	**0.011280196711**

**Table 4.** EER and FRR values for different levels of FAR of the HF-Random dataset.

D. Nottingham custom subset of data verification

To ensure the validity of the evaluation process a new subset of the data was drawn from the master dataset, which we will refer to as *Nottingham*. The new subset contains 539 randomly selected subjects with 11 random images each.

Unlike the subsets of the data used in section C, each of which is defined in a single XML file, the Nottingham subset is defined in two XMLs. Both refer to an identical set of images, however the two sets are saved with different filenames and their corresponding subject IDs are generated randomly. During evaluation the system goes through the first XML matching images to every image from the second XML. The mapping between the original and the encoded image file names and subject IDs were saved in separate XMLs and loaded locally during performance figures computation in order to correctly match subjects to each other. This approach ensures that there is no way for the software to use any knowledge about the dataset structure during evaluation.
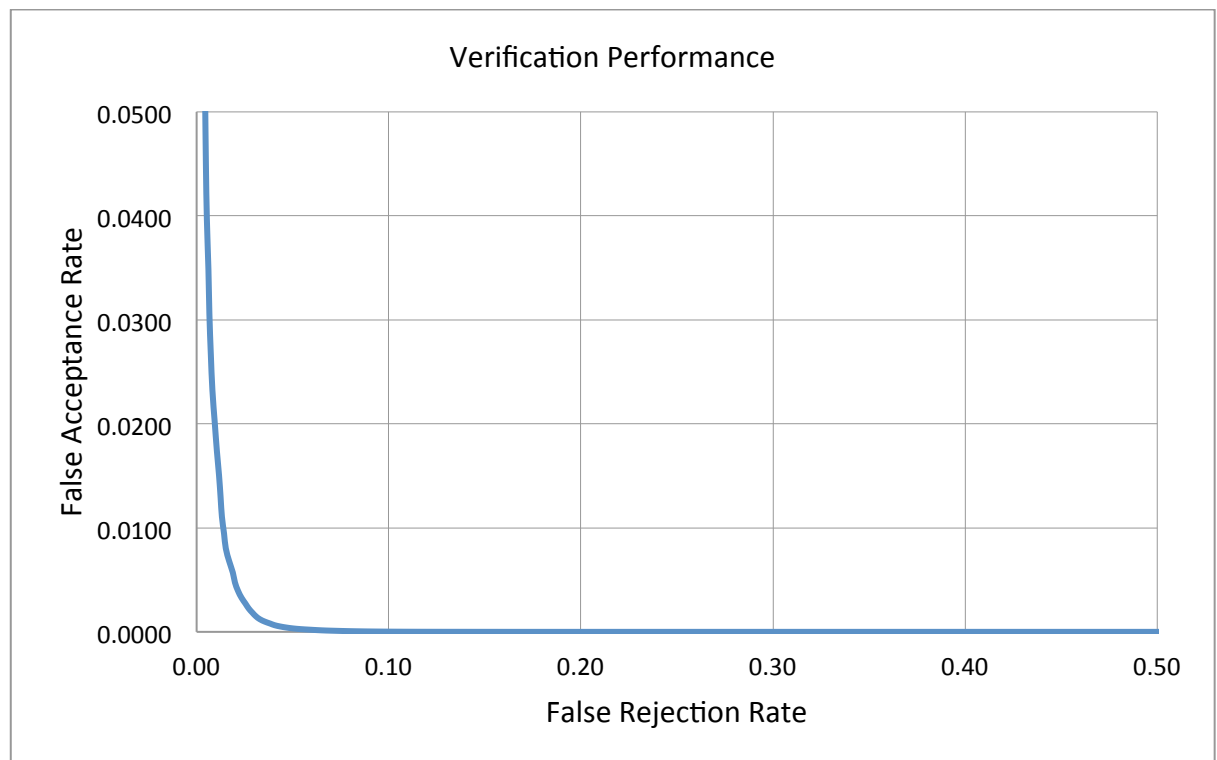
Figure 5a. FRR vs FAR ROC curve for the Nottingham dataset.
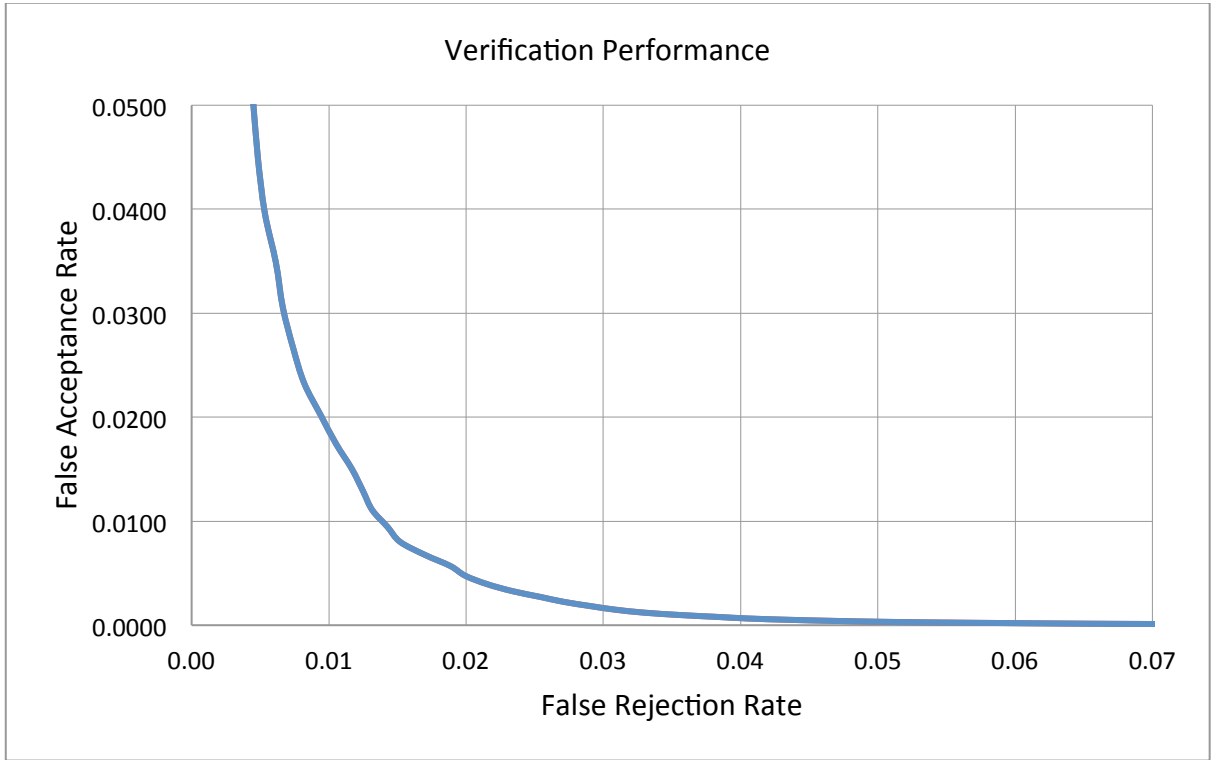
Figure 5b. FRR vs FAR ROC curve for the Nottingham dataset.
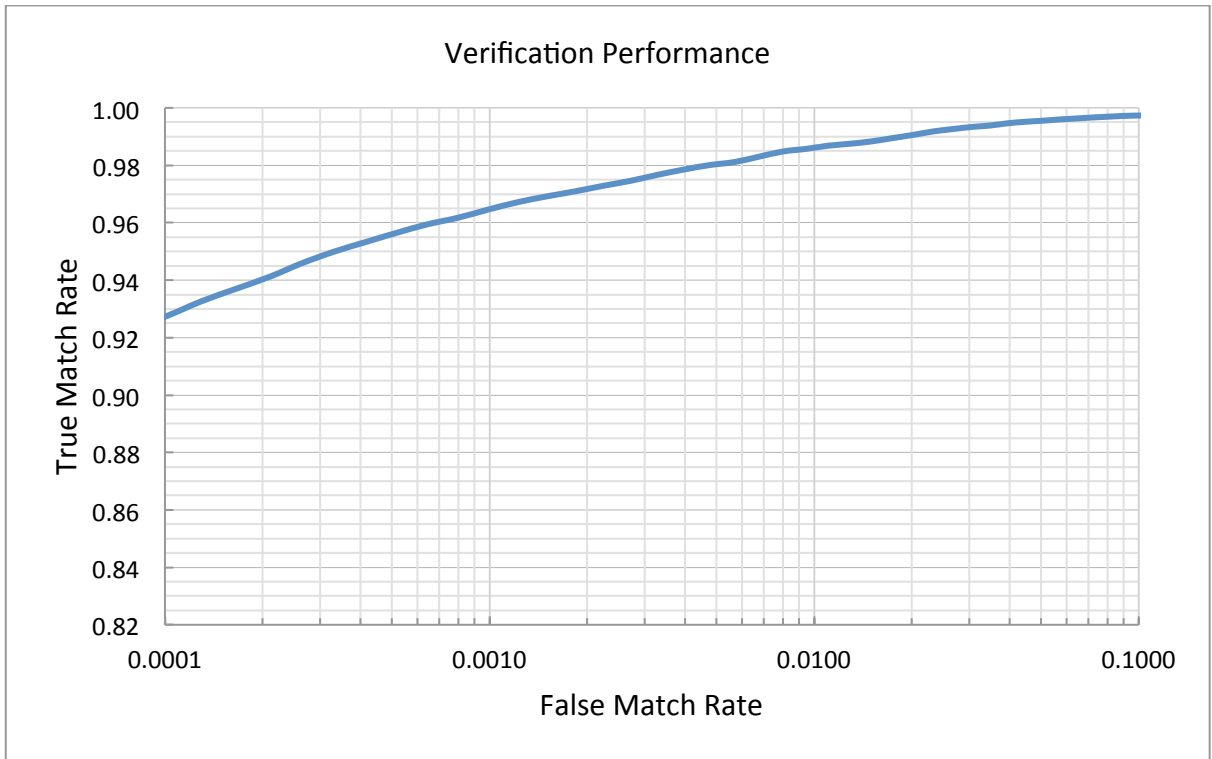
Figure 5c. FAR vs TAR ROC curve for the Nottingham dataset.

FAR	FRR
0.000000	0.211941305448
0.000005	0.132906054984
0.000010	0.114218249283
0.000050	0.085815483218
0.000100	0.072659807725
0.001000	0.035216731321
0.010000	0.013830325519
0.020000	0.009478832855
0.050000	0.004452690167
0.100000	0.002631135099
0.200000	0.001821555068
0.400000	0.000438522516
0.800000	0.000000000000
EER	**0.012597658527**

Table 5. EER and FRR values for different levels of FAR for the Nottingham dataset.