Computer Vision Laboratory
cvl.cs.nott.ac.uk

The University of Nottingham
UNITED KINGDOM · CHINA · MALAYSIA

Aurora Face Recognition Engine Validation Report A: **Simple Identification 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 identification figures are correct, attaining excellent performance. False Non-Identification rates are as low as 2.6% for rank-1 detection, and as low as 0.74% for rank-50 detection.

B. Introduction

The purpose of this document is to provide results of independent validation of Aurora's latest face recognition engine in terms of its person identification capability; in a way that is both illustrative and common in the industry, thus enabling easy comparison with other companies on the market. The results are

presented with tables of False Non-Identification Rates (FNIR) for ranks 1 and 50 as well as complete Cumulative Match Curve (CMC) graphs covering ranks from 1 to 101.

Software

The figures presented in this report reflect performance of the most recent face Aurora 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 secondly 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 the first and last images were selected from the start and end of each person's image set for maximum temporal variance between images. Images that require a lot of padding for alignment were excluded from this set.

- **F-Random**. In this subset of the data all of the images and subjects were selected randomly, all other selection criteria followed the F-FaL set.

- **HF-FaL**. This subset of the data is the same as F-FaL, but with a lower exclusion criterion; images that require only a small amount of padding for alignment were also removed.

- **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.

Every subset of the data contains 2 images of 10660 subjects. One image of each subject is used for the subject enrolment to the system and the other one is used for identification among other subjects. The image data are recorded in two XML files, one for enrolment and another one for identification.

Table 1 below summarises misidentification rates shown by the stingray-v3 engine for Rank 1 errors and <= Rank 50 errors. The graphs that follow demonstrate a detailed breakdown of how CMC curve changes as the tolerated rank error grows for each subset of the data individually.

Subset of data	**Rank 1**	**Rank 50**
	MUGSHOT (ENROL)	*MUGSHOT (ENROL)*
	MUGSHOT (SEARCH)	*MUGSHOT (SEARCH)*
F-FaL	0.0356	0.0137
F-Random	0.0311	0.0098
HF-FaL	0.0351	0.0124
HF-Random	0.0259	0.0074

Table 1. FNIR (False Non-Identification Rate) figures per subset of the data using stingray-v3 engine for ranks 1 and 50.

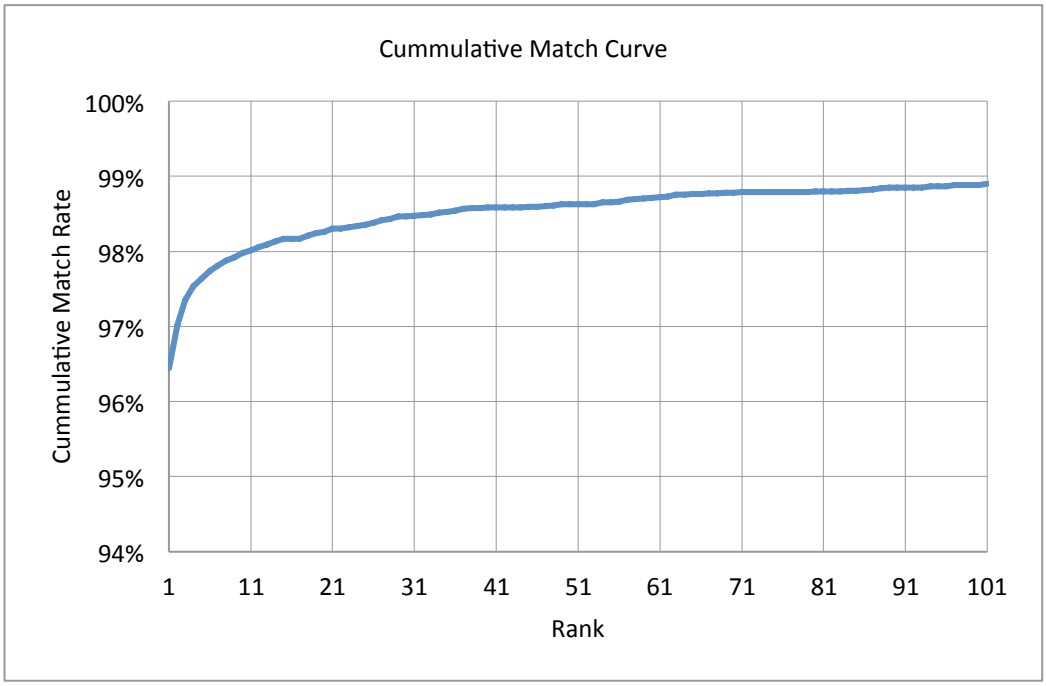

Figure 1. CMC curve for the Aurora's F-FaL subset of the data.

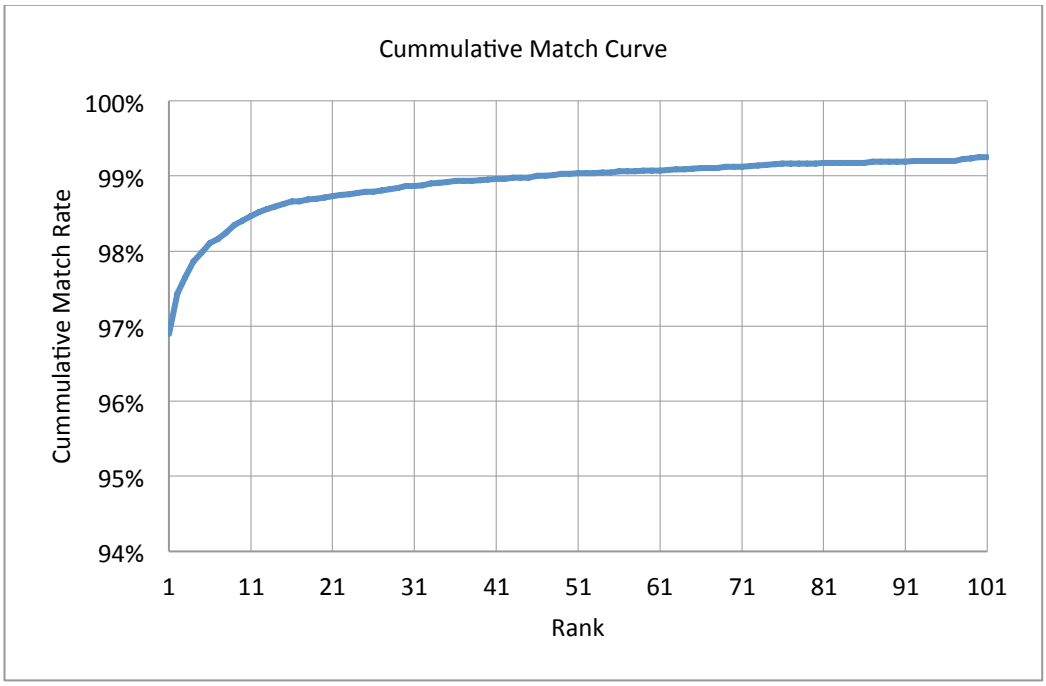

Figure 2. CMC curve for the Aurora's F-Random subset of the data.

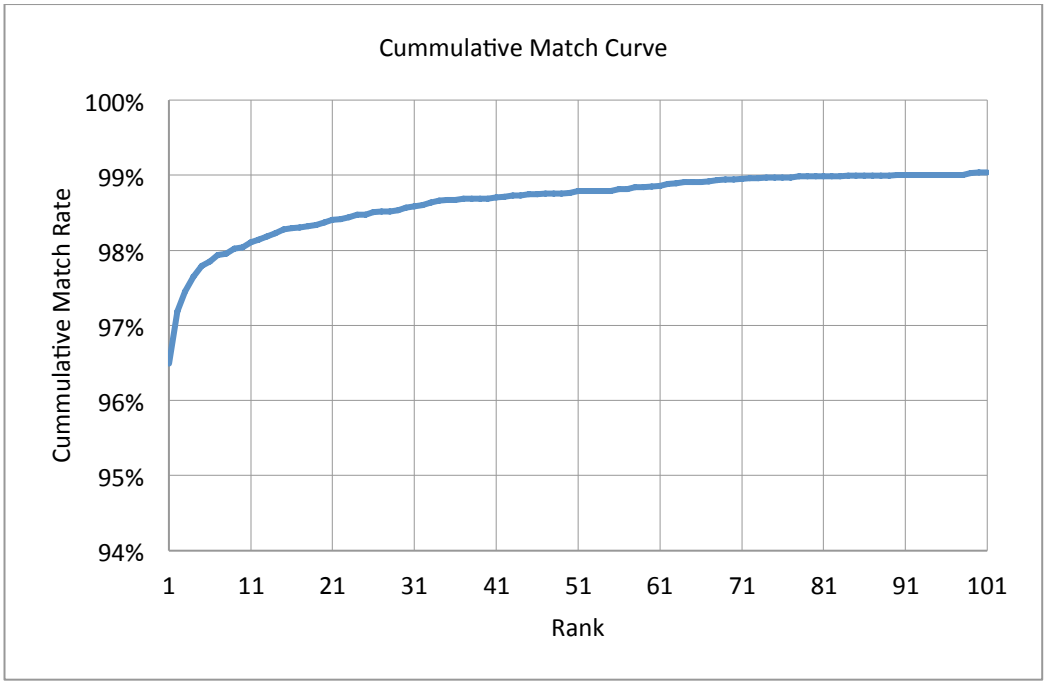

Figure 3. CMC curve for the Aurora's HF-FaL subset of the data.

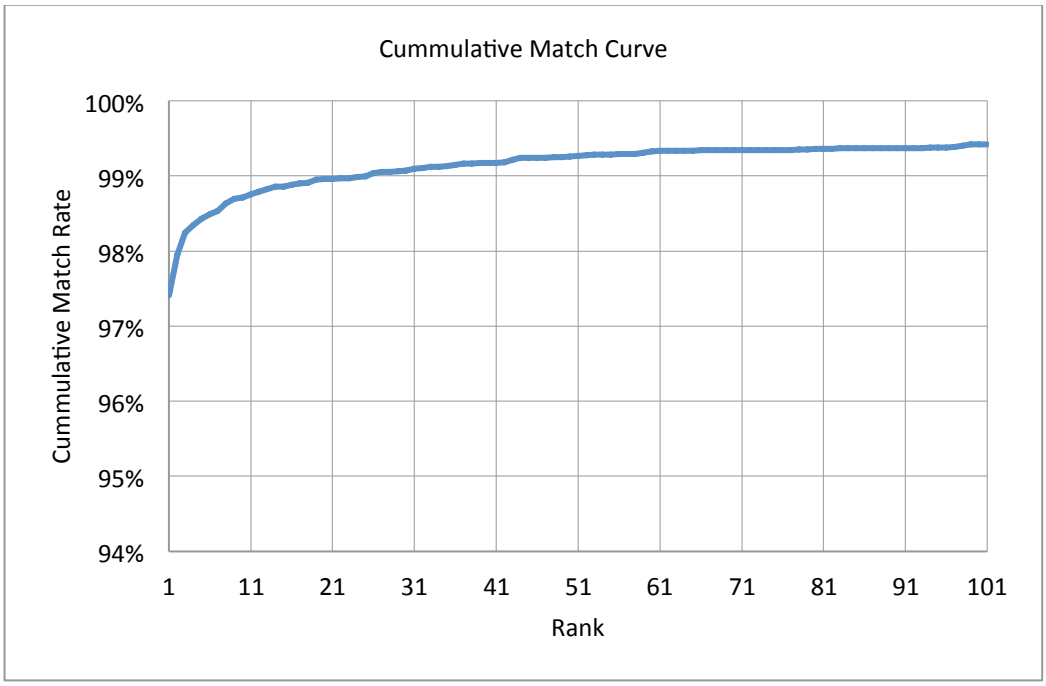

Figure 4. CMC curve for the Aurora's HF-Random subset of the data.

A. Nottingham custom subset of data verification

In order to ensure the validity of the evaluation process a new subset of the data was drawn from the master dataset, which we refer to as *Nottingham*. The new subset contains 2 randomly selected images of each of the 11538 subjects in the master dataset, one of each is used for enrolment and the other one is for identification.

Similar to the previous subsets the enrolment images were kept separate from identification images in two different XML files. This time however image names as well as their corresponding subject IDs in both XMLs were replaced with random unique strings different for enrolment and identification. In addition to that images have also been saved on the disk using the same encoded file names. The mapping between the real and the encoded image names and subject IDs was saved in separate XMLs and loaded locally during performance figure computation in order to correctly match subjects. This ensures that there is no way for the software to use any knowledge about the dataset structure during evaluation.

Table 2 below and the graph that follow provide a summary of the results obtained with the Nottingham subset of the data. The figures obtained correspond to the results from the previous section.

Subset of data	**Rank 1**	**Rank 50**
	MUGSHOT (ENROL)	*MUGSHOT (ENROL)*
	MUGSHOT (SEARCH)	*MUGSHOT (SEARCH)*
Nottingham	0.03328	0.01205

Table 2. FNIR (False Non-Identification Rate) figures for the Nottingham custom subset of the data using stingray-v3 engine for ranks 1 and 50.

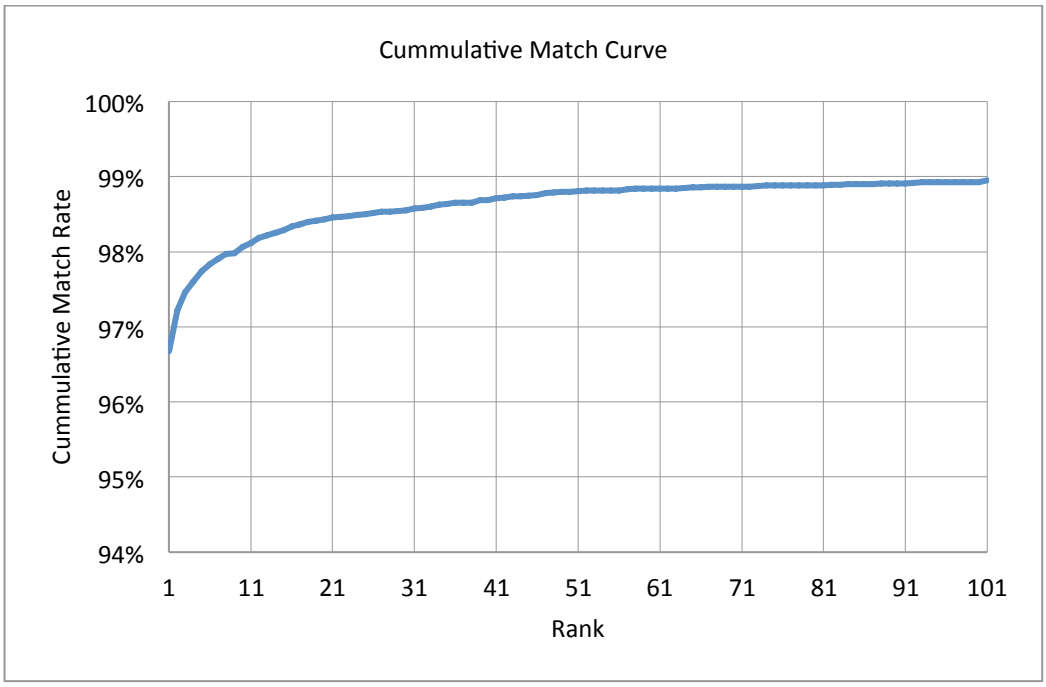

Figure 5. CMC curve for the Nottingham subset of the data.