Information fusion in large-scale biometric identification systems

Sam Afzal-Houshmand

Kongens Lyngby 2018

Technical University of Denmark
Department of Applied Mathematics and Computer Science
Richard Petersens Plads, building 324,
2800 Kongens Lyngby, Denmark
Phone +45 4525 3031
compute@compute.dtu.dk
www.compute.dtu.dk

Summary (English)

The increased use of biometrics due to interest and acceptance have necessitated the development of accurate and efficient systems. This development is relevant to deal with large-scale systems whose usage have a trajectory to be increased further in the futures and are already in place in some capacity. These large-scale system consists of match candidates where there are a plethora of false match-candidates associated to each true match-candidate in the identification mode of biometric systems i.e. finding an identity within the biometric system and possibly who that identity belongs to. A good methodology to address the issue of accurately identifying true match-candidates is information fusion of biometric data from multiple biometric modalities. The issue that arises with the introduction of an information fusion in biometric systems is the huge workload. This motivates the idea of finding an intelligent way of applying information fusion in large-scale biometric systems that reduces the workload significantly while retaining the same (or even better) accuracy compared to an information fusion application on a full-scale system. The proposed intelligent way of applying information fusion in this project is a multi-stage multi-modal hierarchical k-stage system. This system pre-selects a shortlist of the best match-candidates, which are denoted by comparison scores from a given biometric modality, hierarchically in k-stages using different modalities at each stage. It is noted that the pre-selection is conducted on the shortlist denoted by the previous stage with the exception of the 1.level pre-selection which is performed on the full-scale system (i.e. the full-scale database or full-scale list) and the final level where the final selection (also called final decision) is made which is the level where the selection/ final match is conducted i.e. the decision determining if the claimed identity is within the system and who it belongs too. It is also stressed that a different biometric modality is used at each level. The

assumption with this system is that it removes false match-candidates while retaining true match-candidates in concordance with significantly reducing the workload by removing false match-candidates from the system thus reducing the number of attempts (biometric identification attempts) to acquire the true match-candidate i.e. reduce the number of necessary biometric identification decisions.

The goal of the thesis is to investigate the effects of information fusion on large-scale biometric system. The evaluation methodology was to establish a baseline consisting of individual biometric modalities and information fusion on the full-scale generated biometric system, and compare those baselines analytically to evaluations for configurations of the k-stage system that reduces the full-scale biometric system in size by multiple modalities over multiple levels. It is noted that a k-stage system configuration are a certain combination of modality orderings and pre-selection sizes. From the analytic comparison between baselines and k-stage configurations, it was possible to establish a model that will in terms of accuracy vs. efficiency denote the best k-stage system configurations.

The evaluations are based on ISO/IEC standard evaluation techniques such as Detection Error Trade-offs (DET) and Cumulative Match Characteristic (CMC) along with some common evaluation methodologies such as loss of genuines, possible number of biometric identification decisions and score distributions (for validation). Along with those evaluation techniques a workload reduction evaluation, whose metric for workload that has been proposed by Drozdowski et al. [DRB18a], is utilized in this project to evaluate the workload reduction caused by the k-stage system applications. Subsequently, those evaluations techniques help illustrate the workload against accuracy trade-offs which represent the effects of efficiency vs. accuracy of the k-stage system compared to basic full-scale fusion techniques on large-scale biometric systems.

Summary (Danish)

Den øgede brug af biometriske systemer på grund af interesse og accept har nødvendiggjort udviklingen af nøjagtige og effektive systemer. Denne udvikling er relevant for at håndtere storskala systemer, hvis brug ser ud til at øges yderligere i fremtiden og er allerede på plads i en vis kapacitet. Disse storskala systemer består af mathc-kandidater, hvor der er en overflod af falske match-kandidater forbundet med hver sande match-kandidat i identifikationsmetoden for biometriske systemer dvs. at finde en identitet inden for det biometriske system og eventuelt hvem denne identitet tilhører. En god metode til at løse problemet med nøjagtig identificere af sande match-kandidater er informationsfusion af biometriske data fra flere biometriske modaliteter. Det problem, der opstår med indførelsen af en informationsfusion i biometriske systemer, er den enorme arbejdsbyrde. Dette motiverer ideen om at finde en intelligent måde at anvende informationsfusion på i store biometriske systemer, der reducerer arbejdsbyrden betydeligt, samtidig med at den samme (eller endnu bedre) nøjagtighed sammenlignet med en informationsfusion applikation på et fuldskala system. Den foreslåede intelligente måde at anvende informationsfusion på i dette projekt er et multi-trin multimodalt hierarkisk k-niveau system. Dette system vælger en liste over de bedste matchkandidater, som er angivet ved sammenligningsscorer fra en given biometrisk modalitet, hierarkisk i k-trin ved hjælp af forskellige modaliteter i hvert trin. Det bemærkes, at forhåndsudvælgelsen udføres på shortlisten angivet ved den foregående fase med undtagelse af 1.level-forvalg, der udføres på fuldskalaen (dvs. fuldskala-databasen eller fuldskala-listen) og det endelige niveau, hvor det endelige valg (også kaldet endelig afgørelse) er lavet, hvilket er det niveau, hvor udvælgelses- / slut beslutningen udføres, dvs. beslutningen om, hvorvidt den påståede identitet er inden for systemet og hvem den tilhører. Det understreges også, at der anvendes et andet biometrisk modalitet på hvert niveau. Forudsætningen med dette system er, at det fjerner falske match-kandidater, samtidig med at der holdes sande kampkandidater i overensstemmelse med væsentligt at reducere arbejdsbyrden ved at fjerne falske match-kandidater fra systemet, hvilket reducerer antallet af forsøg (biometriske identifikationsforsøg) for at finde den sande match-kandidat dvs. reducere antallet af nødvendige biometriske identifikationsbeslutninger.

Formålet med afhandlingen er at undersøge virkningerne af informationsfusion på et stort biometrisk system. Evalueringsmetoden var at etablere en basis bestående af individuelle biometriske modaliteter og informationsfusion på det fuldskalede genererede biometriske system og sammenligne denene basis analytisk med evalueringer for konfigurationer af k-scenesystemet, der reducerer det fuldskala biometriske system i størrelse ved flere modaliteter over flere niveauer. Det bemærkes, at en k-fase systemkonfiguration er en bestemt kombination af modalitetsbestillinger og forudvalgsstørrelser. Fra den analytiske sammenligning mellem baselinier og k-scenekonfigurationer var det muligt at etablere en model, der med hensyn til nøjagtighed vs effektivitet vil betegne de bedste k-scenesystemkonfigurationer.

Evalueringerne er baseret på ISO / IEC standard evalueringsteknikker som Detection Error Trade-offs (DET) og Kumulative Match Karakteristik (CMC) sammen med nogle fælles evalueringsmetoder såsom tab af genuines, muligt antal biometriske identifikationsbeslutninger og scorefordelinger (til validering). Sammen med disse evalueringsteknikker er en arbejdsbyrdsreduktionsevaluering, hvis beregning af arbejdsbelastning, som er foreslået af Drozdowski et al. [DRB18a], bruges i dette projekt til at evaluere den arbejdsbyrde reduktion forårsaget af k-stage system applikationer. Derefter hjælper disse evalueringsteknikker med at illustrere arbejdsbyrden mod nøjagtighedsafvejninger, som repræsenterer virkningerne af effektivitet vs. nøjagtigheden af k-scenesystemet sammenlignet med grundlæggende fuldskala-fusionsteknikker på storskala biometriske systemer.

Preface

This thesis was prepared at DTU Compute and Center for Advanced Security Security Research Darmstadt (CASED) in fulfillment of the requirements for acquiring an M.Sc. in Engineering at DTU. The author of this thesis is a student of the Digital Media Engineering study at DTU. It was supervised by Professor Christian D. Jensen, Professor Christoph Busch and Doctor Christian Rathgeb and Pawel Drozdowski.

The thesis dealt with information fusion in large-scale biometric system in the identification scenarious, focusing on multi-stage multi-modal biometric information fusion applications.

Sam Afzal-Houshmand

Lyngby, 22-July-2018

Sam Afzal-Houshmand

Acknowledgements

I would like to thank my supervisors Professor Christoph Busch, Pawel Drozdowski and Doctor Christian Rathgeb of Hochschule Darmstadt in Germany for providing the guidance, feedback, data, software and concept this project was based on.

Contents

Chapter 1

Introduction

Biometrics is being used more worldwide due to interest and acceptance as it is reflected in its inclusion in smartphones, computers, forensics etc. This has led to the increased popularity and usage of large-scale biometric systems [tse] [Aad]. Generally in biometric recognition systems, the biometric samples (images of faces, iris, fingerprints etc.) of people are compared and matchers (classifiers) indicate the level of similarity/dissimilarity between any pair of samples by a score.

The current conventional retrieval methods requires exhaustive 1:N comparisons in the identification mode i.e. search to see if sample identity (biometric reference of a particular biometric data subject) is within the biometric system (biometric reference database) and possibly what the identity is of the sample if found in the biometric database. Several issues are apparent which encompass the huge workload required (computational cost) and the significant risk of false-positives in large-scale contexts [Dau00]. One possible solution to this issue is increasing the discriminative power by fusing information from multiple biometric sources. In the case of fusion, more sources of data need to be processed during full-scale 1:N search in the biometric system which motivates a more intelligent way of doing the fusion which is the overall goal of this project.

This project focuses on the issue of accurately and efficiently establishing the identity of a person based only on their biometric data as it is done during the authentication in biometric identification systems. The challenge is that by only using biometric data for the identification, a possible worst case scenario requires an exhaustive 1:N database search as opposed to the 1:1 comparison between an individual's data and the stored reference in the verification mode of a biometric system.

Many biometric system deployments allow data from different biometric characteristics such as images of irides, face, fingerprint etc. There are many information fusion techniques, which can be implemented at various stages of the biometric system work-flow to be used for the combination of such data. By fusing information from multiple sources, the discriminative power of a biometric system can be increased and, thus, be an important factor in the alleviation concerning the issue of accuracy and efficiency in large scale biometric information systems.

The goal in this project is to investigate various existing methods of biometric information fusion on large-scale multi-modal biometric datasets, along with proposing and testing a new solution as well which will intelligently reduce the workload while retaining comparable (or better) accuracy. The proposed idea is to investigate trade-offs between biometric performance and workload reduction in a hierarchical k-stage multi-modal biometric information fusion system. This approach is based on the concept of k-stage system in which biometric data (as denoted by comparison scores) is organized in a hierarchical way and different heuristics is implemented for the retrieval and pre-selection of match-candidates. There is a possibility of, at each level, pre-selecting a subset (i.e. shortlist) of most likely candidates and proceed only with those to the next level with a different modality until a final decision level where an unused modality is used to decide the identity.

The outcomes from the experimentation of basic fusion techniques on a full scale system compared to the proposed k-stage system, showcased tendencies that helped build a model/approach to denote a varied range of optimal configurations of the k-stage system (i.e. specific orderings of modality and associated pre-selection sizes) addressing issues of accuracy vs. efficiency of information fusion in large-scale biometric systems.

CHAPTER 2

Biometric Systems Fundamentals

Establishing fundamentals about biometric systems is a necessity to move onto the central aspect of this project which is the proposed k-stage system. Those encompass knowledge regarding some general evaluation methods used in biometrics and accepted as ISO standard i.e. Cumulative-match characteristic (CMC) scores and Detection error trade-offs (DET).

The content of this chapter is comprised of theories regarding generic biometric systems, theories of each modality used for the project (iris, fingerprint and face), theories of fusion techniques, workload reduction and multibiometrics. After moving on from the individual modalities there is a discourse about information fusion in biometric systems.

2.1 Generic Biometric Systems

This section establishes the fundamentals and operational details of a generic biometric system.

2.1.1 Workflow

The basic workflow of a biometric system, regardless of the biometric characteristic(s) can be generalized by the ISO/IEC standard on biometric testing and reporting (see figure 2.1)[ISO11].

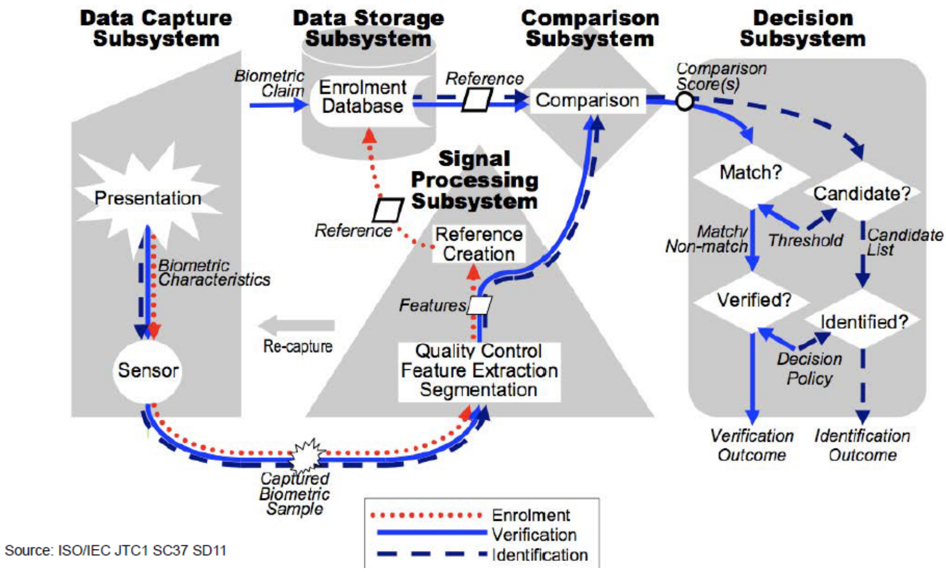


Figure 2.1: The ISO standard generic biometric system [ISO11]

A general walkthrough of the key steps is outlined as:

- **Data Capture:** Acquiring a sample from a subject through a sensor.
- **Signal Processing:** The transforming of the acquired sample into a standardized biometric template form for the given biometric characteristic(s).
- **Segmentation:** The distinguishing of the biometric characteristic signal from the rest of the acquired sample.
- **Feature Extraction:** Process of obtaining a feature set from a sample. The concept is for that feature set to have low intra-class variation (i.e. remain largely invariant in different samples from the same subject) and high inter-class variation (i.e. have enough discriminatory power to reliably distinguish between different subjects).

- **Quality Control:** It happens, occasionally, that the poor quality of an acquired sample or a segmentation error can make the template unusable. Automated quality assessment can be implemented (both for raw images and produced templates) to control that.

- **Comparison and Decision:** The comparing of a new template against existing records of enrolled templates. The results are then used to determine the final outcome of a query.

2.1.2 Operation Modes

There are two modes a biometric system can operate in which determine the flow of information in the system and how the outcome decision is made. For the sake of practicality, it is sensible to only consider the open-set scenario where there may be attempts from users not enrolled in the system. The two modes are typically called the verification scenario and the identification scenario which can be described as:

- **Verification** The subject has to present a claim to an identity. Thereafter, a biometric sample is acquired from the subject, and then the sample is transformed to a template and compared against the enrolled template of the claimed identity. This system requires a 1:1 template comparison to reach a decision.

- **Identification** The system is presented with a sample acquired from the subject and has to ascertain whether the subject has previously been enrolled in the system and, possibly, what their identity is then. This scenario can result in a worst case N template comparisons in order to reach a decision, essentially, comparing the new template against every enrolled template or, in other words, a 1:N comparison search in the case of a naïve approach.

With denoting the probability of a false-accept in a verification trial as Pv the probability of not getting a false-accept in any given verification attempt is $(1 - Pv)$. For the identification scenario, the probability of a false accept in identification trials after an exhaustive search through the database of N unrelated templates can be denoted as Pi. In the identification scenario, $(1 - Pv)$ must happen N independent times which means that the probability of not getting a false-accept in any of those N identification attempts within this scenario is $(1-Pv)^N$. Therefore, in the identification scenario the probability of making at least one false-accept among those N identification attempts is: $Pi =$

Information	U	N	P	C	A	E
DNA	Yes	Yes	Yes	Poor	Poor	*****
Gait	Yes	No	Poor	Yes	Yes	***
Keystroke dynamics	Yes	Yes	Poor	Yes	Yes	****
Voice	Yes	Yes	Poor	Yes	Yes	****
Iris	Yes	Yes	Yes	Yes	Poor	*****
Face	Yes	No	Poor	Yes	Yes	****
Hand geometry	Yes	No	Yes	Yes	Yes	****
Fingerprint	Yes	Yes	Yes	Yes	Fair	****

Figure 2.2: Comparative study of biometric modalities in terms of universality(U), uniqueness (N), permanency (P), collectability (C), acceptability (A) and performance (E) evaluated by number of stars indicating the performance of that modality's EER [MEA12].

$1-(1-Pv)^N$. Thus, with observing that the approximation that $Pi \approx NPv$ for small $Pv << 1/N << 1$ when searching a database of size N an identifier need to be N times better than a verifier when searching a database of size N to achieve comparable odds against a possible false-accept [Dau00].

2.2 Modalities

This project requires an effort to understand state-of-the-art algorithms and techniques for each biometric modality that is going to be used in order to correctly verify or reject the hypotheses that are stated for the project. The chosen modalities for this project which is face, fingerprint and iris where chosen due to factors of availability of data/software, usability and frequency of use in the biometric community. Specifically, the common evaluations of modalities were utilized which are denoted in terms of universality(U), uniqueness (N), permanency (P), collectability (C), acceptability (A) and performance (E) denoted by some measure of accuracy (see figure 2.2).

The system in this project is a multi-biometric fusion approach. Multi-biometric encompass information from different biometric characteristics which is introduced in this section.

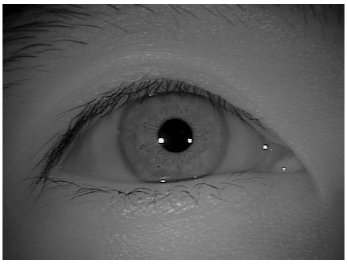


Figure 2.3: Example of an Iris biometric image sample

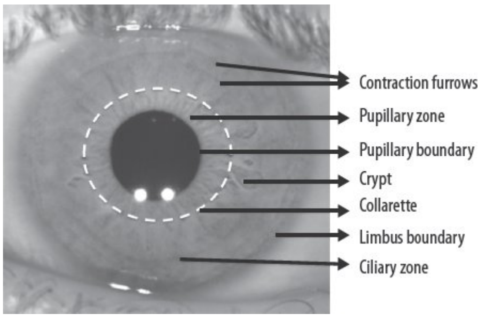


Figure 2.4: A sample image highlighting the different biometric features of the iris characteristic [Ros10]

2.2.1 Iris

A wide variety of iris feature extraction and comparison score computation software is available. In this project the Osiris system [Oth11] was decided to be used due to accessibility and usability. The iris' complex pattern can contain many distinctive features such as arching ligaments, furrows, ridges, crypts, rings, corona, freckles and zigzag collarette which is, generally speaking, the texture and shape of the iris (see figures 2.3 and 2.4) [Dau04].

The state of the art in iris recognition uses a binary code for the representation of iris where the code can be used for distinction of individual irides. For the encoding, the parameters of interest are the max-min diameters of the iris to pupil and the contours parameters. In turn, the contours parameters is obtained by a segmentation of the iris. The contour parameters for iris and pupil are coordinates (x_r, y_r, ϕ) where x_r and y_r is the coordinate of the radius relative to the estimated center and ϕ is the angle between 0 to $2 * \pi$ which is used by the Osiris system [Oth11].

The Osiris software is utilized for extracting the features of iris data. This software takes an image input (grey-scaled iris image) and applies normalization to gain a binary representation whereafter an encoding using Gabor filters is applied i.e. the normalized image is filtered by three complex Gabor filters in order to extract features that characterize the iris texture. Next, a threshold is applied to the result relative to 0 in order to form 6 binary images. Osiris saves the whole binary image as iris-code. The information of texture is carried by the phase of Gabor filters, which is encoded on two bits. The matching is done using the Hamming distance between two iris codes. A matrix of application points is used to indicate which pixels should be considered during the matching. The matching was done by using the Hamming distance between two iris codes generated by the Osiris feature extraction program [Oth11]. The overall structure of the processing chain of the Osiris system can be seen in figure 2.5.

2.2.2 Fingerprint

The quintessential parameter when comparing fingerprint is minutia (see figure 2.6). Minutiae are the features of the fingerprint consisting of: ridge ending, ridge bifurcation, short ridge (independent ridge), island, ridge enclosure, spur, crossover or bridge, delta and core [oB15][C.B16] (see figure 2.7).

This project utilizes an implementation of fingerjetFX from NFIQ 2.0 to extract minutia and then a separate software (i.e. MCCSDK) for minutiae cylinder code extraction (MCC) and comparison by extracted finger templates given by MCC [C.B16].

A known representation method for fingerprints are using the minutiae feature which can be transformed to the the minutia cylinder code (MCC) format representation of the finger-print template [CFM10]. MCC encode the neighborhood of each minutiae into a fixed-length bit vector which is invariant with respect to rotation and translation. The bit vectors are indexed by means of locality sensitive hashing(LSH)[oB15]. Analytic tools such as NFIQ 2.0 and fingerjetFX can be used for the purposes of extracting Minutiae and software such as MCCSDK can be used to extract those MCC vectors from fingerprint minutiae data to compare them[nis][C.B16][Dp11]. Then, a similarity score is calculated using the vectors by local similarity of the vector matrices which is a value in the range 0 to 1. The similarity between two cylinders (C_a and C_b) is simply defined by a vector correlation measure (see equation 2.1):

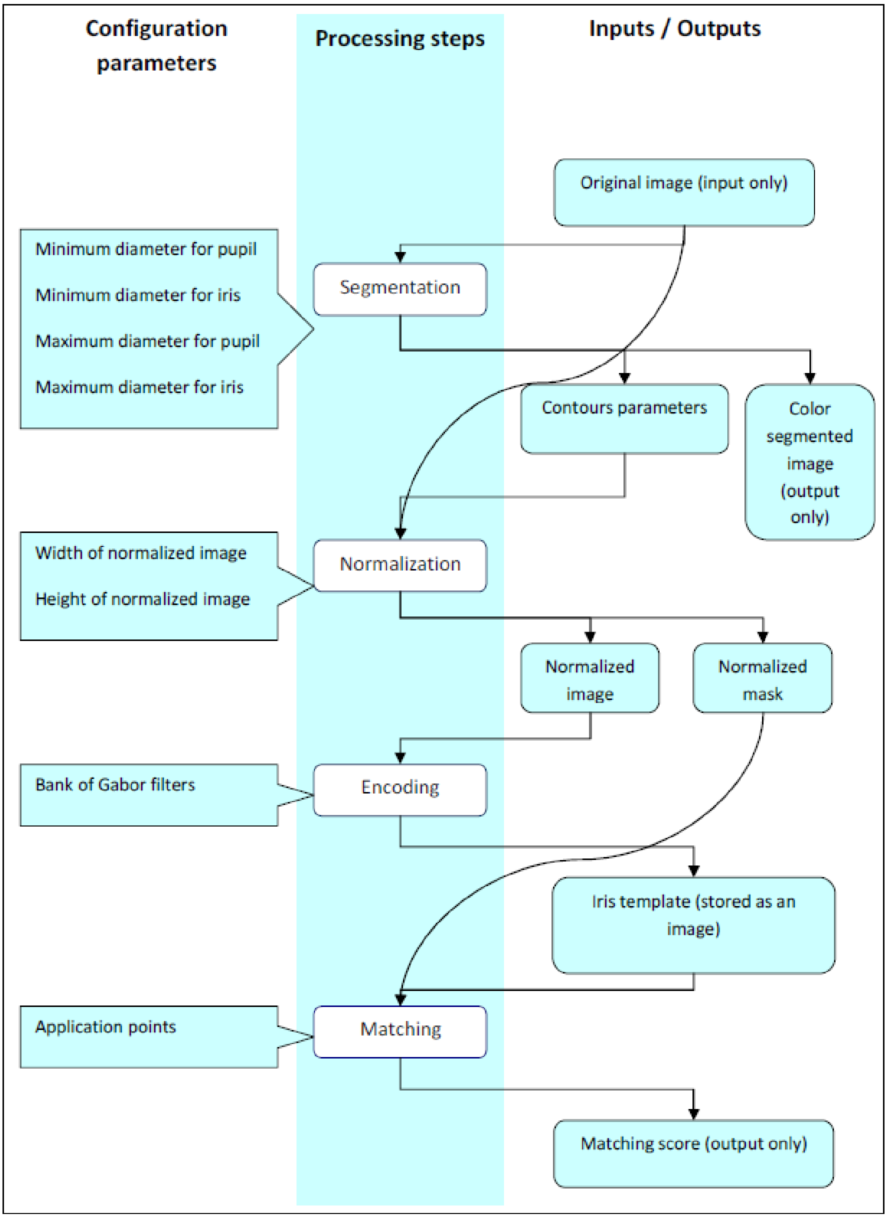
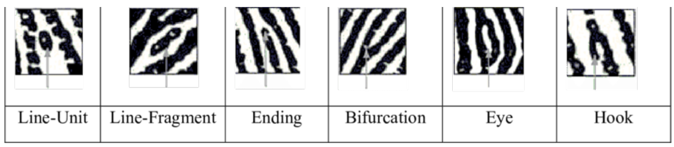


Figure 2.5: Overall structure of the Osiris system [Oth11]

Figure 2.6: Example of a fingerprint biometric image sample

(a) A sample image of the finger modality highlighting features [BO15].



(b) A sample image of the finger modality highlighting features on specifying level [BO15]

Figure 2.7: Images of the a sample fingerprint images and its features highlighted.

$$y(a,b) = 1 - \frac{\left\| c_{a|b} - c_{b|a} \right\|}{\left\| c_{a|b} \right\| + \left\| c_{b|a} \right\|} \tag{2.1}$$

When comparing two minutiae templates a single value (global score) denoting their overall similarity from all minutiae points is obtained from the local similarities. At a preliminary step n_r minutiae pairs are selected starting from those with highest local similarity then a relaxation approach is applied to iteratively modify local similarities on the basis of the compatibility among minutiae global spatial relationships. At the final step, the global score is calculated as the average of the relaxed similarities values of the n_p pairs with the largest efficiency selected from the n_r pre-selected pairs [oB15]. A simplified overview of how the fingerprint system works can be seen in figure 2.8.

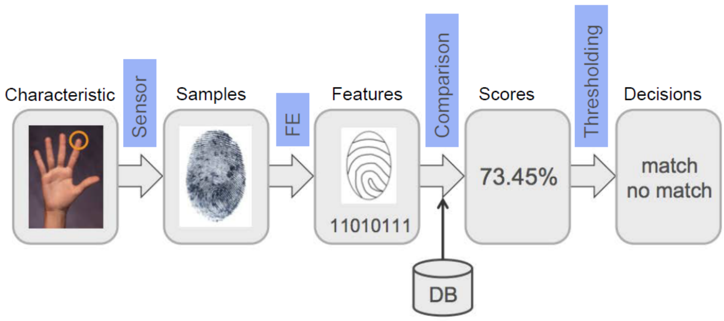


Figure 2.8: A simple overview of a typical finger comparison system [RJ03]

Figure 2.9: Example of a face biometric image sample [lfw07]

2.2.3 Face

Face is another biometric that is used for recognition (see figure 2.9). Though, comparatively the face biometric modality is still widely discussed in regards to what features are distinctive. Therefore, many face recognition methods use all the features of the face (contours, shape, texture) and use complex algorithms to establish distinguishable recognition [SKP15]. There are a plethora of face recognition/extraction software, however, this project utilizes Google's facenet software which uses the machine learning framework of tensorflow in combination with some customary python implementation to make it comparable to the used dataset [San].

A deep neural network (known as facenet) uses convolutional layers which has been trained on large sets of face data. This network is used for the feature

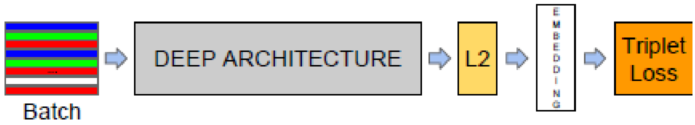


Figure 2.10: The facenet network consist of a batch input layer and a deep CNN (convolutional neural network) followed by L_2 normalization, which results in the face embedding. This is followed by the triplet loss during training [SKP15].

extraction of the face feature which is encoded as a 128 float vector for each image sample. Facenet use the image pixels as features and, then, embeds into the same vector so the distance between these vectors can be used to distinguish between them. The facenet software uses a pre-trained model that have been trained on a large database (i.e. CASIA-Celeb faces) where the faces from the Labeled Faces in the Wild (LFW) database [lfw07] used in this project is used for validation with approximately 98.6% accuracy [OP15].

Thus each face is represented by a 128 dimensional float vector [IEE18], which is ideal for large scale clustering and recognition[SKP15]. Comparison scores were calculated by squared Euclidean distance between a pair of embedding representations of the face images which was constructed externally from the facenet software as a python script/jupyter notebook. The formula for the squared Euclidean distance is denoted as equation 2.2:

$$d^2 = (p_1 - q_1)^2 + (p_2 - q_2)^2 + ... + (p_i - q_1)^2 + ... + (p_n - q_n)^2 \tag{2.2}$$

where p and q is any two points/vectors.

An overview of the facenet system and how it some of its components is build can be seen in figure 2.10 and figure 2.11 and a sample of how the features are extracted in figure 2.12.

2.3 Information fusion

Information fusion can be defined as the reconciliation of evidence presented by multiple sources of information necessary for making a decision. In the context of biometrics, a multi-biometrics system combines information presented by multiple biometric sensors, algorithms, samples, units or traits [Ros07][RJ03][ARJ06].

Several algorithms can be used for the information fusion in most multi-biometric

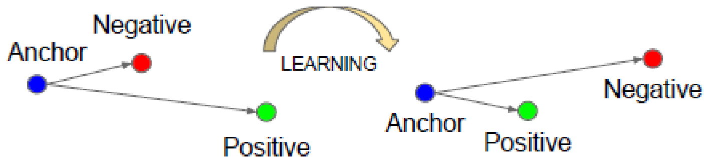


Figure 2.11: The triplet loss functionality minimizes the distance between anchor and a positive, both of which have the same identity, and maximizes the distance between the anchor and a negative of a different identity [SKP15].

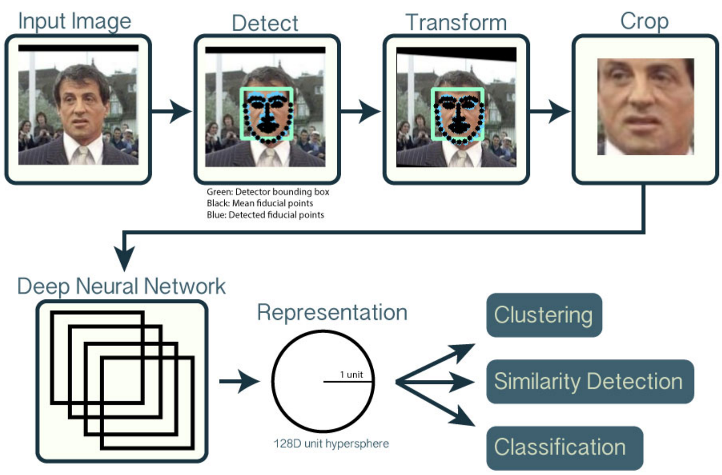


Figure 2.12: An example of how the features of a face is detected and extracted [DTK$^+$13].

systems such as an approach that relies on getting the consensus ranked lists, where the initial ranked lists (obtained after matching input and templates) can be integrated by several methods e.g. the highest rank method, Borda count method, logistic regression method, Bayesian method, fuzzy method, or Markov chain method[ARJ06][OA17].

Different levels of fusion can be categorized into two very broad categories i.e. pre-classification fusion and post-classification fusion. Fusion technique schemes based on pre-classification include fusion at the sensors (i.e. raw data) and the feature levels in the typical biometric system pipeline. Conversely, the post-classification schemes include fusion at the match score (comparison score), rank and decision levels in biometric system pipelines[Ros07][RJ03].

2.4 Evaluation

As a basis the score distribution between genuine and impostor scores for each modality is constructed as a histogram in order to specify values of the various evaluation points. The evaluation of fusion techniques focuses on the trade-off between workload and accuracy. By biometric standards Detection error trade-offs (DET) can be used to denote the accuracy of a system, while the Cumulative match characteristic (CMC) denotes the recognition rate at certain ranks where ranks can present the size of the database required to achieve that recognition rate.

2.4.1 Accuracy and Efficiency

The metrics used for accuracy in the evaluations are typically error rates and recognition rates which can be attained by various techniques. Furthermore, there are techniques that denote recognition rate at certain system sizes. Thresholds for accuracy and efficiency are defined to determine acceptable levels of biometric performance [C.B16].

2.4.1.1 DET

Ranked detection error trade-off (DET) characteristic curves are a performance metric. Each point on the DET curve exhibits the false non-match rates associated with a certain threshold value for false match rates. The curve spans the whole range of possible threshold values which is normally the range of the match scores. The curves describing DET can be gained using an implementation using side-kit [NMR+17] when fed two lists of genuine and impostor scores respectively [DR13].

2.4.1.2 CMC

The curves for Cumulative match characteristic (CMC) curves decides the rank of a given probe in the given biometric system. Essentially, from going by the highest correlated comparison score it counts how many identification attempts was made before reaching the identity with a given rank (or size of biometric system database). Ideally the highest score should be the actual match meaning the rank is 1. From that, the probability rate of reaching the correct identity of a probe with a given rank can be calculated. Calculating the CMC ranks can be done using the ISO standard algorithm [iso06] (see figure 2.13) which can be plotted and visualized in diagrams.

F.2 Algorithm for generating CMC

A suggested procedure for efficiently generating this data is as follows (assuming a single template per individual).

a) Determine the identification rank of each attempt as follows:

   1) Look up the genuine similarity score for that attempt

   2) Count the number of the similarity scores for this attempt (against non-self templates and the self-template) that are

      i) greater than the genuine score: x

      ii) equal to this genuine score: y

   3) If ($y = 1$) the attempt has identification rank ($x + 1$), otherwise the rank is defined by a range of values ($x + 1$), …, ($x + y$)

b) For each rank r (of interest)

   1) Count the number attempts with rank r or less. Attempts that have a range of ranks are counted according to the fraction of these values at rank r or less.

   2) Dividing by the total number of attempts gives the probability that the true template/model for a test sample will be found somewhere within the r most similar templates in the enrolment database. This probability is plotted against r on the CMC graph.

Figure 2.13: The ISO standard CMC algorithm [iso06]

The CMC denotes the recognition rate at a certain rank where rank can be normalized to represent the percentage of the database that is being evaluated. This can be used to determine how many biometric identification attempts at a given rank (or database size) going by best comparison score is necessary in a given data-set to achieve 100% probability of accurate matches of scores for identities.

2.4.2 Workload

The workload definitions are scalable from the baselines in terms of pair-wise computations. The idea is to take the number of comparisons (biometric identification attempts) from the 1:N search comparisons and relate it to the given tests i.e. number of comparisons (or identification attempts) required by the system.

The workload reduction in this project is evaluated as suggested in the work by Drozdowski et al. [DRB18a]. There is not yet an ISO mandated workload evaluation methodology and, therefore, this project uses the one outlined in this section. In that paper the workload of a system, denoted by $F = \frac{W_p}{W_b}$ where W_p represents the proposed workload by the reduction and W_b represent the baseline workload, is calculated by multiplying the number of enrollees (N), the penetration rate (p) and the computational cost of a single one-to-one comparison (C) and, if applicable, adding the pre-selection cost (c) denoted in the formula $W = N * p * C + c$. In that paper, the evaluation of workload reduction in a given system is decided by three parameter i.e.:

- $TP_{0.01}$ - the true-positive identification rate measured at a false-positive identification rate of 0.01%
- F - the fraction of the required modality baseline workload per lookup
- τ - a metric [Pro13] defined as the Euclidean distance from the optimal operating point ($TP_{0.01} = 0$ and $F \approx 0$) calculated as: $\tau = \sqrt{(TP_{0.01} - 1)^2 + F^2}$

It is observed for this metric (i.e. F) that the number of comparisons (i.e. identification attempts) for the modality in combination with the bit-wise (or pair-wise) sizes of templates by a given modality can be necessary information which can be seen in tables such as the one in figure 2.14 in order to compute the workload and, subsequently, the workload reduction. The number of biometric identification attempts is the total number of subject-to-subject comparisons combinations found in the remaining biometric system (or full-scale biometric system) as match candidates where some combination would be the same subject being compared to themselves (i.e. genuine or true match-candidates) and other combinations would be subjects compared to other subjects (i.e. impostor or false match-candidates).

Device	Size in Bytes
Fingerprint	200 – 2,000
Speaker	2,000 upwards (text dependent) 4,000 - 50,000 (text independent)
Finger Geometry	14
Hand Geometry	9
Face	100 – 3,500
Iris	512
Vascular	256 - 1,000

Figure 2.14: Table showing typical template sizes in bytes for different modalities [Bus06]

CHAPTER 3

Related works

In order to dive deeper into the theories relevant for this project, a wide variety of related works is surveyed and discussed in this chapter. Including all works on fusion and workload reduction is not possible so this section discusses these topics more generally and point out some key works and surveys. Communities in biometrics have conducted intense work in the field of fusion in biometric verification/ identification scenarios on score level, decision level and more [DR13][SUM+05][HHF+10]. This chapter works through some general fusion techniques and how they have been analyzed throughout different works.

Generally, there is a lot of different fusion methods but most relevant to this work is the methods and techniques associated with decision level fusion and score level fusion. Thereafter, works that have been focused in evaluating workload reduction is discussed as it is relevant to the intended evaluation for this project. Finally, various works that have focused on multi-level (also called multi-stage) systems such as the proposed k-stage system fusion approach that is evaluated in this project is discussed since these techniques have various elements that is applicable to this project e.g. shortlists from pre-selection over multiple levels. Lastly, a summary of the surveyed works is presented and discussed.

3.1 Basic information fusion methods

This section includes a discourse and discussion of general works in the field of information fusion and analysis of some specific works in the subcategories of the field of topic which are consequently relevant to this project.

Fusion of biometric information is utilized to consolidate data from multiple sources in order to improve the discriminative power of a system [ARJ06]. Several key categories of this data consolidation [Dau00] can be distinguished as:

- **Image level fusion:** The consolidating the raw data of same subject.
- **Feature level fusion:** The consolidating of the feature vectors of various biometric instances from the same subject.
- **Score level fusion:** The combining of the scores yielded by multiple comparators from each modality.
- **Rank level:** The combining of multiple lists of candidate identities produced by multiple comparators.
- **Decision level:** The combining of the decisions yielded by multiple comparators.

Information fusion schemes have repeatedly been shown to improve the systems' biometric performance [Ros07] [DH17]. Relevant to this project, it is more sensible to dive into the score-level fusion and decision level fusion as they are the most accessible and comparable with the proposed k-stage system (as described in chapter 1) and can be used to produce baselines that can be compared with a proposed fusion technique that works more intelligently.

3.1.1 Decision level fusion - State of the art

The decision level with majority voting fusion approach that is used in this project is similar to the methodologies denoted in the paper about information fusion in biometrics done by A.Ross et al [ARJ06]. This paper denotes fusion at the decision level where each sensor can capture multiple biometric data input and the resulting feature vectors individually classified into two classes i.e accept or reject. A majority vote scheme, such as that employed in the work by Zuev and Ivanon [ZI96] was suggested to be used to make the final decision [RJ03].

3.1.2 Score level fusion - State of the art

Several works have been done in the evaluation of various score level fusion techniques. Some works have been focused on surveying known techniques and evaluating them in multi-modal biometric systems such as sum-rule bases score level fusion and support vector machines (SVM) based scores[HHF+10]. The works concerning sum-rule bases score-level fusion lay the foundation for the score-level fusion that is used in this project, which is a simple min-max normalization to 0-1 range of scores and sum-based fusion with an alternative version using averages of the score-level fusion.

3.2 Multi-biometric information fusion system - State of the art

Many works have been made diving into the topic of multi-modal biometric systems as well as surveying the history of works to current state-of-the-art. Works by Fierrez et. al. have surveyed the fundamentals of multiclass and multimodal fusion works focusing on multi-classifier systems that exploit the input measures quality. Ultimately, this paper states the importance of multi-classifier systems in the context of multi-modal contexts [J.F18a] [J.F18b]. The work by R. Dwivedi et. al. addresses some of the issues with biometric-based information system such as low-performance due to low intra-class variations, data outliers and invasion of privacy by proposing a hybrid fusion techniques using protected modalities (i.e. iris, finger, face) that uses a combination of decision level fusion and, then, score-level fusion using two different combinations of modalities. This work, however, doesn't address the issues of workload [RD18]. Further works has been done by Iovane et. al. to encrypt information fusion techniques (this work used prime numbers and face biometrics) to address the increasing interest of secrecy[IBMN18].

Generally, some of the more interesting works have been made to investigate which biometrics are most appropriate for multi-model identification that focused on biometric traits which shows indication that the modalities for this project was appropriate choices [S.B18]. Additionally, several projects throughout history and most recently have experimented with the performance of multimodal authentication which indicates that it is indeed more accurate that single modality biometric identification systems [KD18].

3.3 Evaluation methods

The metrics presented in the work on multi-iris indexing and retrieval by Drozdowski et al. is applicable to this project, which includes metrics for workload reduction, DET and CMC evaluations. The work in this paper present a multi-iris indexing system for efficient and accurate large-scale identification where the system is based on Bloom filters and binary search trees. The evaluation approach used for several fusion strategies for the system described in this work is applicable to the one that is used in this project. Essentially, taking the workload (comparisons by bit, pairs etc.) against biometric performance where this accuracy is the measure that can be gained from the error rates denoted by DET [DRB17]. The system in that paper was evaluated empirically on a combined database from numerous publicly available datasets. That system was tested in an open-set identification scenario and maintained its biometric performance [DRB18a].

3.3.1 DET and CMC - state of the art

For the DET and CMC evaluations which was also present in the work by Drozdowski et al. there are a broad selection of works on how to relate DET and CMC. Through different academic literature, the matching accuracy of a biometric system is quantified through measures such as the Receiver Operating Characteristic (ROC) curve and Cumulative Match Characteristic (CMC) curve [DR13]. It should be noted that Detection Error Trade-off (DET) can be considered an associated contrast of the ROC serving similar functionalities. The ROC curve when measuring the verification performance, is based on aggregate statistics of match scores corresponding to all enrolled biometric samples. Contrarily, the CMC curve when measuring the identification performance, is based on the relative ordering of match scores corresponding to each biometric sample in the closed-set identification scenario. A study was conducted to determine whether a set of genuine and impostor match scores generated from biometric data can be reassigned to virtual identities, such that the same ROC curve can be accompanied by multiple CMC curves. The reassignment was accomplished by modeling the intra- and inter-class relationships between identities based on the "Biometric Menagerie" phenomenon explained in the paper by Martin et al [GDR98]. The outcome of that study suggests that a single ROC curve can be mapped to multiple CMC curves in closed-set identification scenarios, and that presentation of a CMC curve should be accompanied by a ROC (or contrarily the DET) curve when reporting biometric system performance, in order to better understand the performance of the matcher algorithm [DR13].

3.4 Workload reduction- state of the art

A central component of this project is workload reduction. Several approaches are possible when doing workload reduction but most relevant for this project would be the method for calculating workload by Drozdowski et al [DRB18b].

3.4.1 Workload reduction approaches

There is a lot of different approaches to workload reduction and a lot of works done in the field, however, there are some general theorems that are key aspects of the topic of workload reduction.

3.4.1.1 Serial combination of algorithms

The serial combination of algorithms approach encompass a multi-level method that uses different algorithms at each level to create a short-list of most likely template match-candidates from the possible match-candidates provided by its previous level. This can continue until a final level where an identification match by final decision is made based on the most probable template match from the match-candidates that hasn't been sorted out. The probability for templates can be evaluated by a multitude of methods e.g. highest rank score, majority voting etc. One approach to the serial combination of algorithms methodology is to use computational efficient algorithms on the larger lists of match-candidates, and more accurate (and slower) algorithms on the smaller lists of match-candidates [RBBB15].

3.4.1.2 Classification/binning

With the classification (or binning) the template database is split into several subsets with low intra-class variation and high inter-class variation. Then, the class of the given probe template is determined and actual comparisons are performed only with database templates of that class [RBBB15].

3.4.1.3 Indexing

Indexing techniques want to decrease the system load in terms of running time and/or growth of space requirement in correlation with input size growth. These indexing techniques utilizes probabilistic and hierarchical data structures to reduce the search space [RBBB15].

3.4.1.4 Hardware acceleration

Harware acceleration methods address the big workload that is in the identification scenario efficiently by the utilization of many CPUs/threads or by using a GPU. The processes can work on disjoint parts of the database. At the end the results are aggregated. This approach doesn't reduce workload but is merely distributing it [RBBB15].

3.4.2 Calculating workload metric

Suggestions on how to calculate workload can be found for various modalities from a wide selection of sources. One article [DRB18b] suggests a formula for the total system workload in a single lookup during the identification scenario (ω) is derived from a set of stated requirements in that paper: S - the number of subjects enrolled, ρ - the penetration rate and τ - the cost of a single step (i.e. one comparison). In summation the workload formula is: $\omega = S * \rho * \tau$ which is one of the methods for workload reduction [ISO11], however, it should be noted that the metric for workload is yet to be a ISO-standard.

Other methodologies for workload reduction and how to document them have been done in various studies such as Indexing techniques based on minutiae for fingerprint, face indexing based on linear subspace approximation and texture/color indexing for iris [IK17].

The metric to evaluate workload reduction in this project is similar to the method presented by Drozdowski et al [DRB18b] using the number of comparisons (biometric identification attempts) known with the bit and pair information for the templates of each modality along with the evaluated error rates that would be gained.

3.5 Hierarchical multi-level biometric system - State of the art

The two-stage approach used in the work of Gentille et al. is similar to the multi-level pre-selection that is going to be used in the proposed k-stage system in this project. In that work they use a short-length iris code at one level to pre-screen a large database of irides which will reduce the number of full-length iris-code comparisons to a fraction of the original total length. Due to the short-length code being smaller, the process will be much faster than a full scale comparison where candidates is chosen at the first level, and at the second level a full comparison of full-length iris-codes is made from the remaining candidates [GRC09]. Other works include a machine learning approach to multi-modal hierarchical information fusion which support the credibility of such an approach [YE13].

Conceptually, the k-stage system in this projects is a serial combination of algorithms based on representations of the same or multiple modalities. The method for the k-stage system is conceptually similar to the method used in the paper by Gentille et al. [GRC09]. This paper describes an identification system based on a two-tier indexing scheme. The paper introduces an scheme for the iris modality that uses short-length iris-code to pre-screen a large database and, thus, reduce the number of full comparisons needed. In that scheme, a shorter representation of the biometric modality is used to pre-align the probe (single iris-code in that case) to each gallery sample (entire database of iris-codes in that case) to then generate a short-list of match candidates. This short-list is then compared to the probe using the more expensive full representation of the modality at the second tier. Since the majority of non-match candidates are sorted out at the first tier with minimal effort, the result is a faster recognition system.

3.6 K-stage system theorem

A multi-biometric k-stage system has not been applied yet. Something similar has been applied with ear and palm [PRSB15] and voice [BRB+14], however, the specific approach with the modalities finger-face-iris has not been applied yet. This project suggest a newer multi-stage approach that will be applied in correspondence with some of the existing methods. This approach is a concept of k-stage system in which the data is organized in a hierarchical way and different heuristics is implemented for the retrieval and pre-selection. There is a possibility of, at each level, pre-selecting a subset of most likely candidates and proceed only with those to the next level with a different modality. Here, there is possibility to test what order of hierarchy organization yields best results and what other dependencies might be discovered. In the end, there will be a large system algorithm/parameter space, and the results in terms of biometric performance, workload etc. will necessitate to be visualized and analyzed. This will be reported with typical metrics using DET and CMC curves as well as workload metrics.

3.7 Summary

This chapter presented the current state-of-the-art in multi-biometric fusion which also encompass its subsidiaries of the individual modalities, biometric performance in multi-level systems in relation to accuracy trade-offs and workload reduction. The used methods in this work is a decision level fusion, score level fusion and a proposed k-stage hierarchical fusion. With a standardized method set for the evaluation these can be compared in order to denote findings that may enlighten the investigation of information fusion in large-scale biometric systems.

Due to inconsistency found through the survey, it has become necessary for the sake of this project to work of some assumptions based on some of the works such as a standardized way of reporting workload reduction, detection error trade-offs and cumulative matching characteristic within the confines of the ISO standards.

The challenges with the k-stage system with the proposed 2-stage systems is that this concept was not used before in a multi-biometric system and also not with 3 modalities. This requires experimentation which was conducted in this project.

Chapter 4

Proposed k-stage system

In this section the k-stage system for this project is discussed. In the beginning the general concepts and techniques for the proposed k-stage system is discussed. Thereafter, the setup which encompass the conceptualization/processing approach and data preparation is presented. Furthermore, the concepts behind the basic fusion techniques is presented as these are necessary for the evaluation of the k-stage system. Finally, the hierarchical k-stage system is discussed as a concept.

Essentially, the k-stage system work by a multi-level system that will pre-select a number of match candidates for the claimed identity by different modalities over multiple levels. After the pre-selections and a final shortlists have been generated a final selection (final decision) using the final unused modality is made to select the correct match which is exponentially fewer since it is a fraction of possible number of biometric identification decisions (comparisons) necessary compared to the 1:N search. The reduction is of course depending on the sizes of the pre-selections along with the modality templates remaining.

4.1 Fusion for k-stage system approaches

Fusion on a multitude of levels is possible in biometrics including decision level and score level. There is two distinct approaches to do fusion in the context of this project namely decision level fusion and score level fusion. For the sake of the k-stage system there is a necessity to do a baseline evaluation of basic fusion techniques. Once the baseline is established the CMC information can be used to filter out the databases by best scores from the ranks (percentage of database with 99% accurate matches and forth). The case in this project is multi-modal fusion, which entails that it is based on the prefix that biometric information is gained for multiple modalities from the same subjects. Essentially, from a single person information about each biometric characteristic (iris, finger and face) is gained, whereafter, feature extraction and comparison score for all possible combinations (comparison score compared to themselves and others which is also known as genuine and impostors) can be calculated from that information against the rest of the enrolled database. The overlapping feature of the different biometric information sources is the labelling of subject source which enables distinction of genuine and impostor scores between the modalities.

4.1.1 Decision level fusion

Fusion of information on the decision level takes place when each biometric system independently makes a decision about the identity of the claimant and then a final decision regarding the identity using methods such as majority voting is made. The fusion specifically takes place at the rank level i.e. ranking of plausible identities based on comparative evaluation. Different methods can be used to combine the ranks such as the highest rank method where each possible identity is assigned the best of all ranks computed by the different systems. The ties are broken randomly to arrive at a strict ranking order and the final decision is made based on the consolidated ranks. A distinguishing of impostor and genuine can be made by majority voting i.e. at least two out of three is considered genuine and vice versa for impostor. Then DET for each modality in the fusion scenario can be made.

Score is set to a threshold using the DET and score distributions for each modality where everything beneath the threshold is considered genuine and everything above is impostor through the decision level fusion. Illustrations of the concept and overview of basic decision level fusion can be seen in figures 4.1 and 4.2.

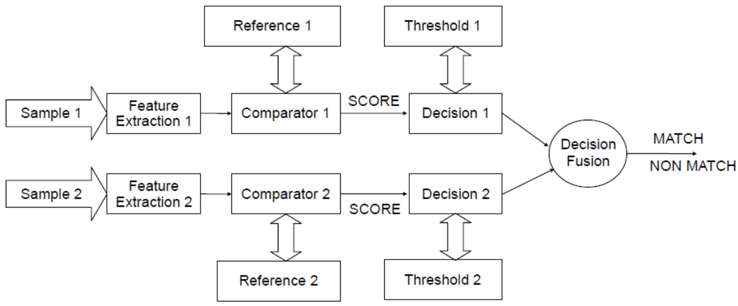


Figure 4.1: The concept of a multi-biometric decision level fusion [RJ03].

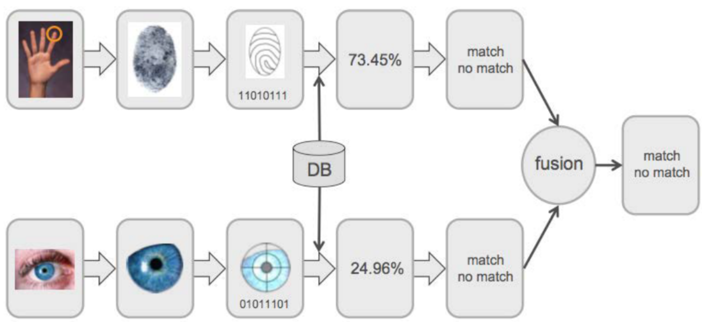


Figure 4.2: The overview of a multi-biometric decision level fusion [RJ03].

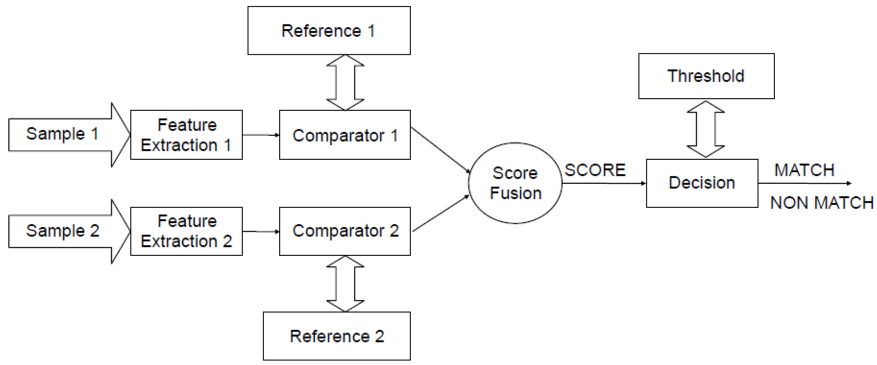


Figure 4.3: Concept of basic score-level fusion without normalization [RJ03].

4.1.2 Score-level fusion

In the multi-modal scenario, the score-level fusion encompass the normalization of the scores from the different modalities so they are in the same space. Once the normalization has been applied, a summation (sum-based, average etc.) is used to gain the fused scores where the fuse score can be evaluated.

There are multiple scores (one for each modality). The fusion is first applied after performing a normalization that ensures the multiple scores are in the same spectrum (min-max to range 0-1 normalization) and that they are all for the same similarity measure i.e. either similarity or dissimilarity. Thereafter, an average or summation can be made of the multiple scores from different modalities. Thus, a cumulative score for different modalities is made for each combination of subjects (genuine and impostor). Then, these scores can be fed to the DET to gain the spectrum from genuine and impostor scores from the averaged normalized similarity scores from the different modalities. Illustrations of the score level fusion concept with and without normalization, and an overview of the score-level fusion system can be seen in figures 4.3, 4.4 and 4.5.

4.2 K-stage system concept

The k-stage system in the context of a biometric system is visualized in figures 4.6 and the pseudo-code (see algorithm 1) for gaining the k-stage filtered database that was ready for any selection task e.g. score for genuine and impostors used for DET, rankings for CMC, workload reduction and more. The

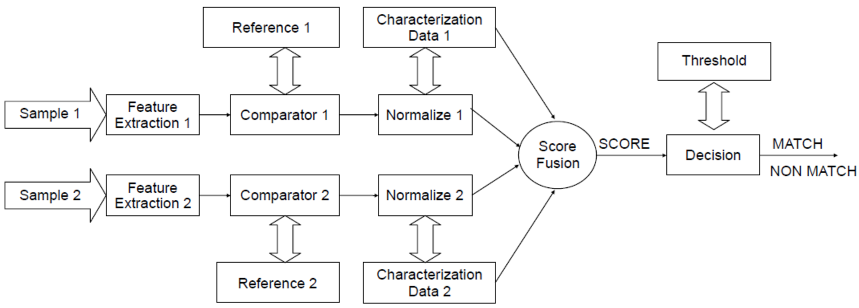


Figure 4.4: Concept of basic score-level fusion with normalization [RJ03].

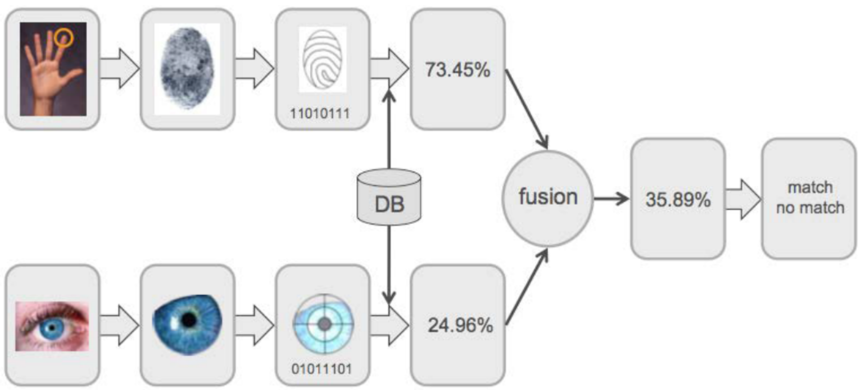


Figure 4.5: Overview of basic score-level fusion without normalization [RJ03].

total database that stands as the basis of the system (gained from baselines) consist of items with a key made of the labelling of template comparison (i.e. subject_x_image_x_subject_y_image_y) and a value which was a list for each modality (i.e. [finger, face, iris]).

Algorithm 1 K-stage configurations database

```
Databases = load([finger_db, face_db, iris_db])
total_database = [merged_items]
for d in Databases: do
    for k,v in d: do
        total_database[k].add(v)
sorted_database = sorted(total_database[modality_1])
database_lvl1 = sorted_database[pre_selection_lvl1]
sorted_database_lvl1 = sorted(database_lvl1[modality_2]
database_lvl2 = sorted_database_lvl1[pre_selection_lvl2]
database_lvl3_selection = items
for k, v in database_lvl2 do
    database_lvl3_selection[k] = v[modality_3]
return database_lvl3_selection
```

The return/outcome of the k-stage algorithm with given parameters for a configuration was a reduced database where the remaining unused modality could be used for selection for the various evaluations method i.e. DET, CMC and workload reduction among other things.

In the context of the k-stage system scenario, multiple modalities across several levels was used instead. The important note is that each match candidate possess multiple comparison scores representing each modality, so at each level there is a pre-selection of match-candidates before the identification match e.g:

- Level 1: Face modality - pre-select 100 best candidates
- Level 2: Fingerprint modality - From the 100 candidates pre-select best 10 candidates
- Level 3: Iris modality - from the 10 candidates select the best possible match via exhaustive search

This will mostly be in context of large-scale open-set identification scenario. The baseline evaluations was benchmarking biometric performance/ workload of those parameters separately and in simple fusion scenarios i.e. decision level fusion and score-level fusion.

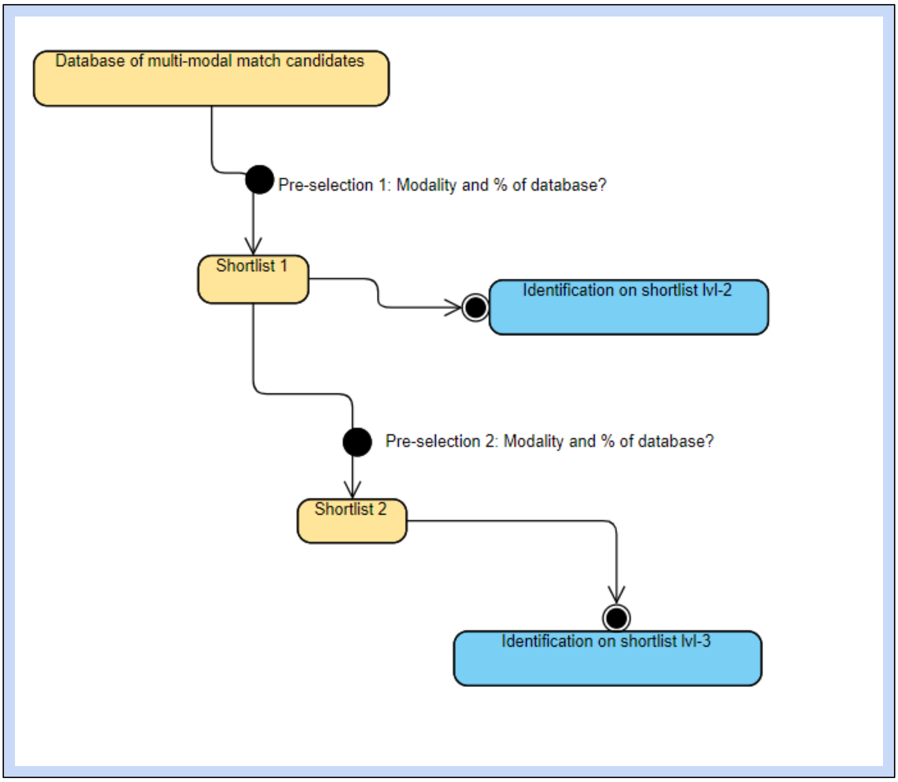


Figure 4.6: A visualization of the concept for k-stage system fusion technique. This process can continue for k modalities.

The necessary tools for biometric data processing are freely available as open-source software or industry trial software. For the most part scripts was implemented for different algorithms/heuristics and data visualization/analysis in the programming language python. An implementation of the ISO standard CMC algorithm was constructed for calculation and visualization. For the DET, a script implementation provided by Hochscule-Darmstadt, which takes lists of genuine and impostor scores as input, was utilized. Furthermore, an implementation of workload computation was constructed based on the discussed equations for the calculations and visualizations. One of the useful implementations where the pickle saving of extracted information which made evaluations by different methods implementation easy.

CHAPTER 5

Experimental Setup

In this chapter, the experimental setup for this project is discussed which encompass the approach to establish baselines which in turn is necessary for creating and evaluating the proposed k-stage system. Firstly, the problem disposition for the experiment is presented and reiterated to establish the purposes of these experiments. Thereafter, the setup for the k-stage system experiment which encompass the data preparation (i.e. raw data pre-processing, feature extraction and comparison score), algorithmic implementation and evaluation is discussed. Hereafter, the establishment of how to gain the baselines is discoursed since they are necessary to create and evaluate the proposed idea. The baselines is an establishment of how well each modality perform individually and together in some basic fusion techniques (i.e. score-level fusion) in the exhaustive search identification scenario of biometric systems (i.e. basic multi-biometric decision-level fusion and score-level fusion). Afterwards, the actual implementation of the proposed k-stage system is discussed which is based upon the concepts and results from the baseline implementation. Furthermore, the evaluation techniques (DET, CMC, Workload reduction and etc.) is also discussed. Additionally, there is also some discourse and discussion about the data and software used for feature extraction, match score computations and subsequent analysis of computed match-scores.

For the k-stage system in this first experiment, databases of single modalities were combined to create a multi-biometric dataset of 2000 subjects. The three modalities used for this was iris, fingerprint and face. This database was generated by assigning images for each modality into items containing three values each (for each modality) and an index. The index showed which two of the 2000 subject had been compared and which image number had been used since every subject would contain 2 templates for each modality. It is noted that in total there is 2000 genuine identification references (2 sample per subject) and with the inclusion false identification attempts totals 331884 possible biometric identification decision in this experiment. The number of possible identification decisions (comparisons) are estimated by using some simple data computations (counting) on the look-up dictionary to find the total number as there is some loss of data by failed feature extractions etc.

In summation, this first experiment for the k-stage system is total 331884 match candidates where 720 is genuine and the rest is impostor. The split between baseline and experiment for that experiment is as described in table 5.6.

5.1 Problem disposition

The k-stage system is the proposed multi-level pre-selection method of this project. At each level there is a pre-selection of match candidates before an final identification match using different modalities.

The purposes of the experimentation of the k-stage system is to possible draw conclusions that may address the issues of:

How should the modalities be ordered?

What fraction of candidates should be selected at each level?

5.2 Datasets

Listed in the table (see table 5.1) is some of the potential datasets for the various biometric characteristics that are going to be used in this project. Samples from each dataset can be found in figures 5.1, 5.2 and 5.3.

Modality	Dataset
Face	Labeled Faces in the Wild (LFW) [lfw07]
Fingerprint	NIST Sd14 [nis]
Iris	CASIA-Thousand Iris [fBC05]

Table 5.1: Table of potential datasets for different biometric characteristics

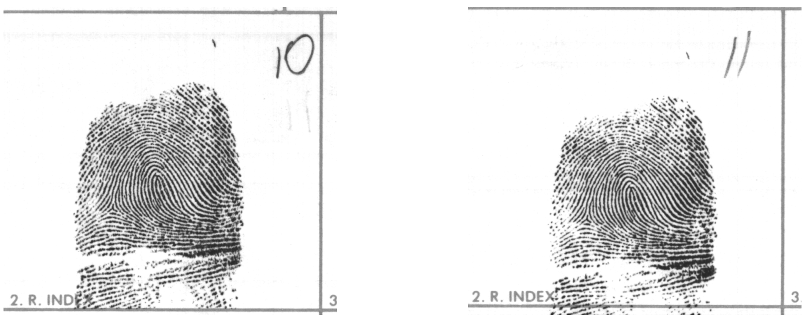


Figure 5.1: Raw images for the finger modality from the same subject

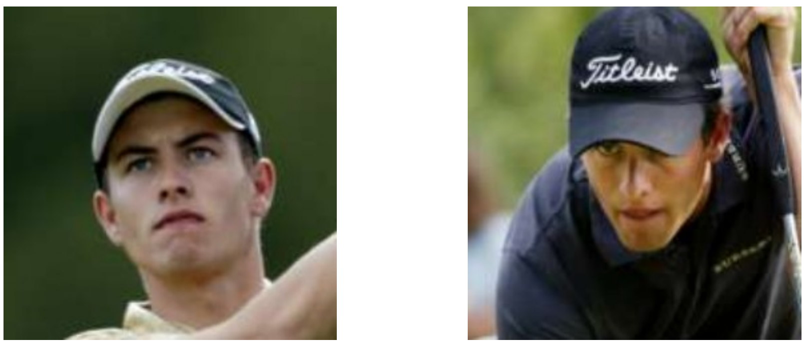


Figure 5.2: Raw images for the face modality from the same subject

The raw datasets had some pre-processing in the case of the fingerprint images as they needed to be cropped to be comparable with the minutiae extraction software which was subsequently verified by a sanity test plotting the minutiae points onto said images after the extraction had been applied. The face images had to be aligned which was done via an alignment software from the facenet system. No pre-processing of significance on the raw data was done for the iris images before applying the software for feature extraction since the Osiris and custom implementations provided many of the usual pre-process options i.e. iris segmentation and normalization (un-rolling).

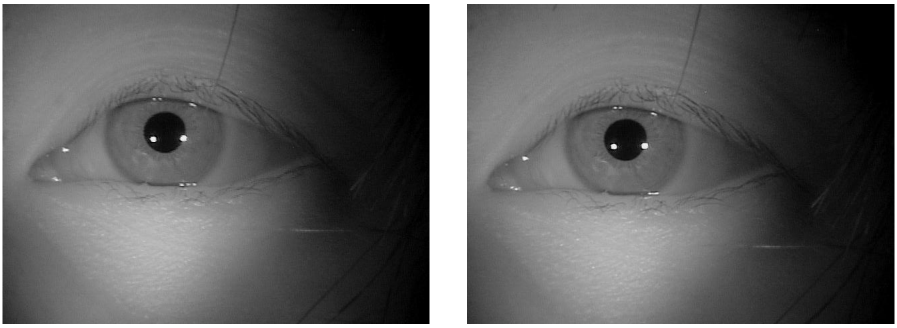


Figure 5.3: Raw images for the iris modality from the same subject

Beyond the pre-processing the images had to be assigned to 2000 subjects where each subject had biometric information for a single subject for each modality that is distinguishable from subject to subject with 2 sample for each modality

In the beginning, the largest publicly available datasets were used to perform experiments and gain proof of concept. Then, for large scale testing, the project moved unto synthetic datasets where real data is not available (for the finger and iris modality).

5.3 Software

Before conducting any experiment of any kind there was a preparation process that involved obtaining and cleaning the provided biometric datasets, and the obtaining and installation of the various software used for feature extraction and comparison score computation which in some cases required some modification to the provided installation files/processes. Different software is used for each modality listed as:

- **Iris**: Osiris software extract iris codes from iris images (among other things) and also encompass a match-score calculation for pairs of binary iris-codes using Hamming distance [Oth11].

- **Fingerprint**: An FingerJetFX implementation by NFIQ 2.0 was used to generate MCC templates (in the .iso or .txt format). The match scores were calculated by using a MCCSDK program (.exe) which compares ISO templates in a customized python wrapper [Dp11].

- **Face**: The Facenet software was utilized with a modified validation script (Python), so that it would save the encoded embedding (a 128 number float) for each face image. Subsequently, a squared Euclidean distance was calculated for each possible combination of embedded representations for face images [SKP15].

The key aspects of interest that require attention for implementation and input is the processes that extract features and compute comparison scores with the given software. The typical approach is that the raw images for the given modality is feed to the implemented software for feature extraction and preprocessing. Gained from this is feature templates that can be feed to another system to compute comparison scores. Sample of the feature extractions is included in this section. For the iris modality see figures 5.4, 5.5, 5.6, for face see figures 5.7, and for finger see figure 5.8. The specifics of these systems in regards to their theoretical fundamentals are as discussed in the theory section (see chapter 2).

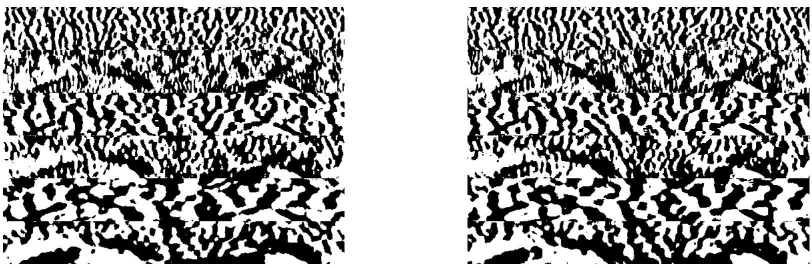


Figure 5.4: Iris code images

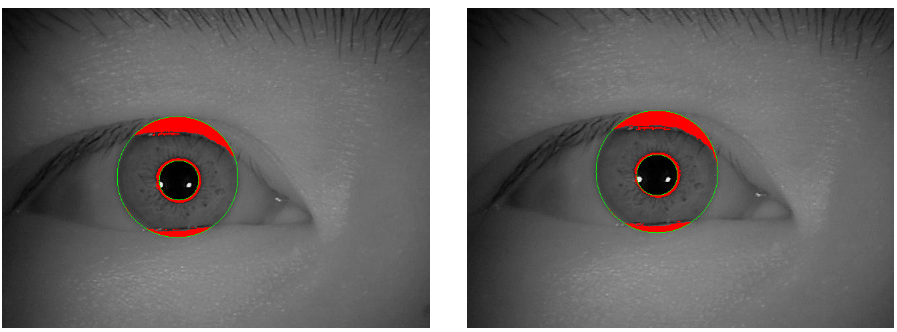


Figure 5.5: Iris segmentation images

The templates were fed to various systems and comparison scores were calculated for the given algorithm for the given modality. The aspect is that any two templates could be compared and with labelling it is possible to know which

Figure 5.6: Iris mask images

Figure 5.7: Aligned face images

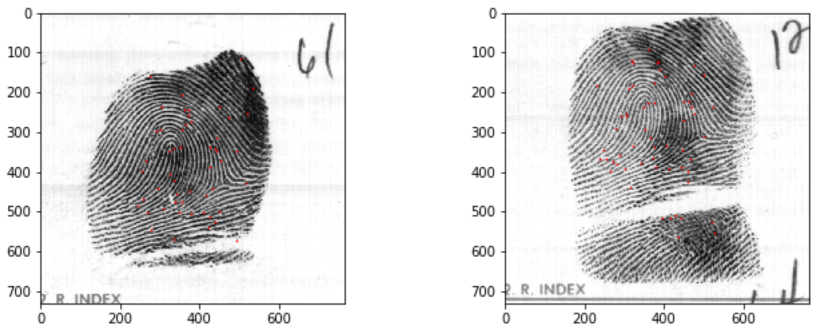


Figure 5.8: Finger sanity pictures i.e. plotted minutiae points (red dots) onto finger-print images

subjects had been compared for the given modality and which image from that given subjects had been utilized i.e. see tables 5.2, 5.3 and 5.4. Similar approach as applied to the all the modalities together in the fusion context (see table 5.5).

Finger Match-candidate label	Finger Comparison-score
subject_1528-subject_270-feature_R_index_4096_01-feature_R_index_93_02	0.0808
subject_1528-subject_270-feature_R_index_4096_01-feature_R_index_94_02	0.0807

Table 5.2: Lookup table of finger database

Iris Match-candidate label	Iris Comparison-score
subject_1-subject_1-S5000L01.jpg-S5000L02.jpg	0.3409
subject_1-subject_2-S5000L01.jpg-S5000R01.jpg	0.4574

Table 5.3: Lookup table of iris database

5.4 Implementation of k-stage system

The approach is to generate DET curves and CMC in order to evaluate a baseline that will be used for the k-stage system. For that, all possible combinations of modalities and pre-selection sizes are exhausted for validation. It is noted that the range of pre-selection sizes were chosen intelligently by the utilization of the baseline CMC evaluations. The k-stage system itself is fairly simple from a technical standpoint as the requirement is to remove certain items given certain parameters which is indices (given a pre-selection size) for a certain value in the item list. This system works by best score notion where it is of course anticipated that genuine scores have the best scores given the modality in use at the level that is current by point of pre-selection, were afterwards, final selection is fairly simple using the remaining unused modality.

The implementation pipeline is illustrated as figure 5.9.

5.4.1 Part 1: Baselines

The evaluation is conducted using 768 enrolled and 512 impostor data subjects where the rest of the generated multi-biometric dataset from the 2000 data subjects is used for validation which means that it is done k-fold times for DET and CMC. The project will apply comparative scores in the evaluation. From

Face Match-candidate label	Face Comparison-score
subject_1-subject_77-feature_Aaron_Eckhart_0001-feature_AJ_Lamas_0001	2.1139
subject_2-subject_77-feature_Aaron_Guiel_0001-feature_AJ_Lamas_0001	1.7220

Table 5.4: Lookup table of face database

Total match candidate label	Face-score	Iris-score	Finger-score
subject_1-subject_6	1.9575	0.44866	0.0878
subject_6-subject_6	0.52362	0.36136	0.11200

Table 5.5: Lookup table of all modalities together in one database of comparison scores

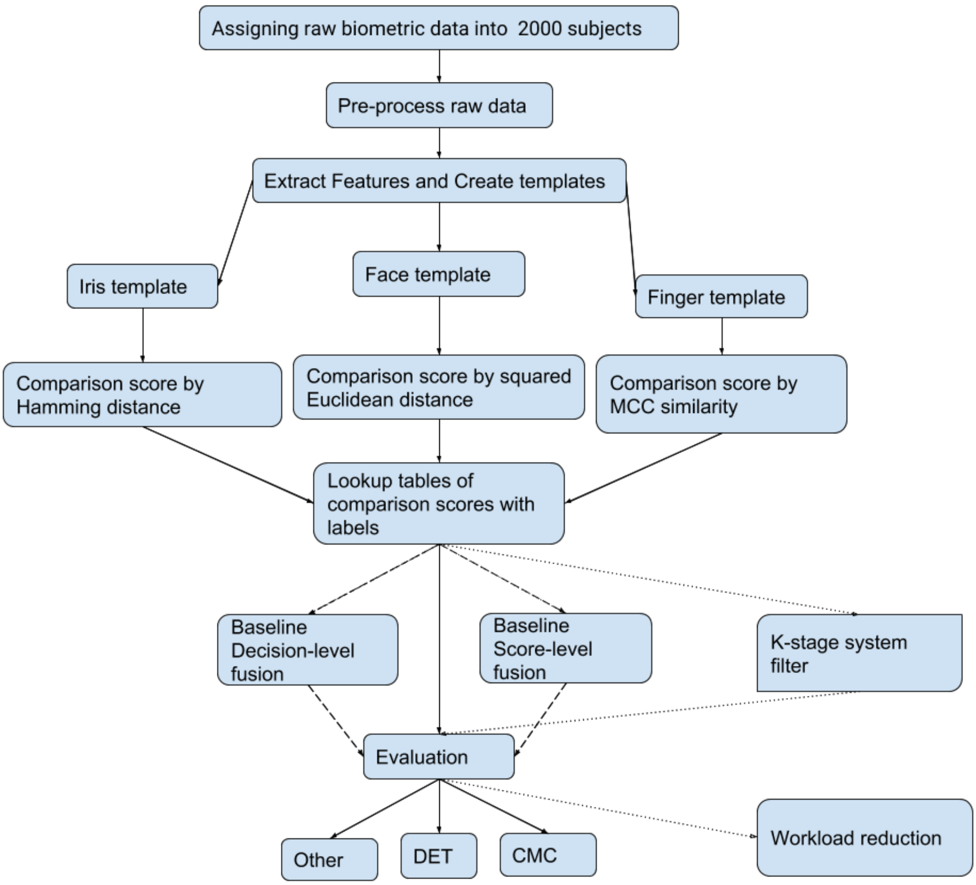


Figure 5.9: An overall overview of the implementation process for the whole experiment

training set there is 768 data subjects used for genuine and 512 data subject used for impostor from the sample (see table 5.6).

Once all match scores from all extracted features for each modality is generated, the information is structured into look-up tables for each modality so that

Total	Genuine	Impostor	Training
2000	768	512	720

Table 5.6: Table illustrating the split of subjects for the evaluation

genuine and imposter scores could easily be assigned as an instance. Furthermore, these look-up tables are merged into a super list where they are identified by which match candidate they belong to. A genuine score is when a match comparison was conducted against same subject e.g.

- Face picture 1 of subject 1 was compared to face picture 2 of subject 1.

An impostor score is a comparison of samples of two different subjects e.g.:

- Face picture 1 of subject 1 was compared to face picture 2 of subject 2.

To create the DET curve where both being disjoint sets of the same enrolled database is used.:

- The genuine scores of 768 subject
- The impostor scores of 512 subjects

5.4.2 Part 2: K-stage-system

The idea of the k-stage is to use the CMC rank from each modality to create a shortlist of match candidates by pre-selection. Then by some confidence interval (set to a threshold) an arbitrary shortlist using another modality is created and then another instance using another modality until the final match (final decision) is made.

For example:

1. A [x:n1] shortlist of match candidates from the finger modality using the CMC rank
2. Then a [x:n2<n1] shortlist of match candidates found from the previous shortlist using another modality, essentially, sorting out the candidates from previous layer and shortening by best confidence intervals for those candidates still in the list for another modality.

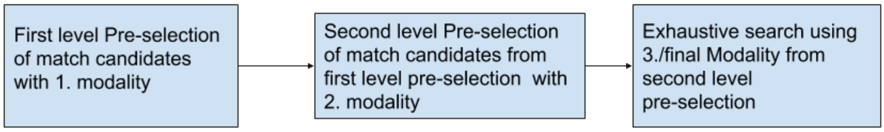


Figure 5.10: Hierarchical concept of k-stage system. The modalities are ordered so a 1. level pre-selection is made using the scores of one (1st) modality and then a 2. level pre-selection is made on that 1. level pre-selection using another (2nd) modality and finally a selection of the best scores from the 2. level pre-selection is made using yet another (3rd) modality for the selection task.

3. Next step is essentially a final match via an exhaustive search on the shortlist of match candidates denoted by the previous level.

A concept illustration of the pre-selection with different modalities aspect of the k-stage system can be seen in figure 5.10. The theoretical details of the CMC algorithms is as presented in chapter 2.

In technical terms, for the implementation the k-stage system is as simple as repeating the same methodology applied to the baseline evaluation, to 1) obtain a super list of match candidates which denotes three values of comparison-scores representing each modality. 2) the discarding based on pre-selection sizes is simply done by choosing the index on the match candidates sorted by the values representing the modality in use for the given level. This discarding is done over multiple levels using different modalities for the sorting which is simply done by setting the index for the three values to match the comparison score for the modality in question when sorting by values. 3) the size of the discarding is done by setting the range by indices for how many items should be included in the shortlists of match candidates (i.e. a certain fraction of N). This shortlist is used as a basis when moving on to the next level until the level with the final selection is reached. 4) In the final level there is a shortened list of match candidates where the remaining unused modality is used for the final selection in whatever evaluation task necessary e.g. DET and CMC.

The setup for the k-stage system also make it easy to test out different configurations of biometric algorithms due to already having applied different extraction methods for each modality in the baseline. The parameters of pre-selection sizes and orderings is simply constructed by changing the indices for the discarding of items and changing the value index to match the given modality in the value list at any given level.

In this project, a k-stage system configuration is denoted as a unique ordering of modalities and pre-selection sizes for each level of modality in that ordering. Therefore, each configurations were distinguished by the ordering of modalities and the pre-selection size at each level (i.e. for each modality) exempt from the modality for selection which was used for evaluation of workload, DET and CMC. For example, a representation such as fi-ir-fa_0_1 represents finger with 10% pre-selection on the 1. level creating shortlist_1 and iris with 25% pre-selection on shortlist_1 creating shortlist_2 and final selection with face on shortlist_2 (note that the numbers at the end refer to a list containing pre-selection sizes based of the baseline CMC evaluations ranging from low-high). The approach for this experiment is to exhaust (almost) every combination of orderings with every combination of pre-selection sizes selected from the CMC evaluations.

Chapter 6

Baseline Results

This section includes the presentation of results from the baseline evaluations. With the comparison scores calculated from the extracted features for each modality, an evaluation using DET and CMC curves on a training set of the scores was computed. The biometric information are distributed among 2000 subjects where the subjects are split into disjoint training and test set. From the training set two disjoint sets of enrolled genuine and enrolled impostor sets are established. In correlation the CMC ranking was also computed using the ISO/IEC standard CMC algorithm.

Throughout this section there is, firstly, some general analysis of the gained comparison scores from the feature extractions and subsequent comparison score computation for each modality which is denoted by score distribution. Afterwards, the results for the DET evaluations of the baselines for each modality individually and together in basic fusion techniques (i.e. decision level fusion and score level fusion) are presented and discussed. Thereafter, the results for CMC evaluations for each modality and the basic fusion techniques are presented and discussed. Additionally, the possible sources of errors that may have an influence on the results are discussed.

6.1 Comparison scores

Match scores (or similarity/dissimilarity scores) for all possible template combinations from each dataset of each modality were calculated and collected.

6.1.1 Iris

Using the Osiris software, extracted iris-codes from CASIA dataset for the iris modality can be used in combinations to calculate similarity/match scores. The similarity is calculated by Hamming distance between two given templates from the datasets where the templates are extracted iris-codes from each enrolled iris image. This actually means that the match scores is a dissimilarity score having a value between 0 and 1.

6.1.2 Finger

The fingerprint modality uses minutiae points which is a three-dimensional matrix that represents the position (x,y) of minutiae point and a direction (angle value) at each row (i.e. each row in the matrix represents a minutiae point). The extracted minutiae positions were checked by plotting them atop the sample images there were extracted from. Similarity score were calculated by a local matrix similarity measure which is simply defined by a vector measure as defined in its equation and approach as seen in chapter 2. The match score is a similarity value in the range between 0 and 1.

6.1.3 Face

From the facenet software a vector (or array) of 128 floats is the feature representation of a given face image sample representing a complex composition evaluation of face features. The similarity scores between any given two of these extracted feature representation is calculated by a squared euclidean distance which is a dissimilarity measure in an unbound range where 0 means complete similarity.

6.2 Score Distribution

From the entire training set the score distributions of impostor and genuine scores could be generated. Specifically, histograms showing the distribution of genuine and impostor score of each of the face (see figure 6.1), finger (see figure 6.2) and iris (see figure 6.3) modalities. The score distribution functions as an indicator for the dataset that is being worked which can be used for the discussion of the outcomes from the experiments.

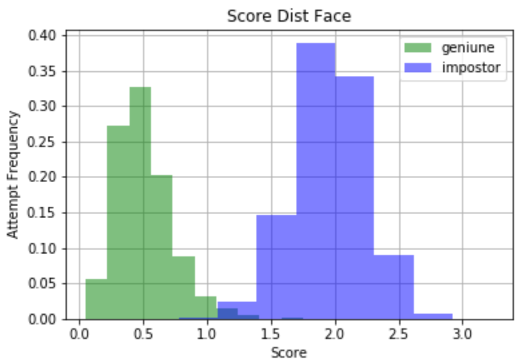


Figure 6.1: Score distribution for face modality full scale

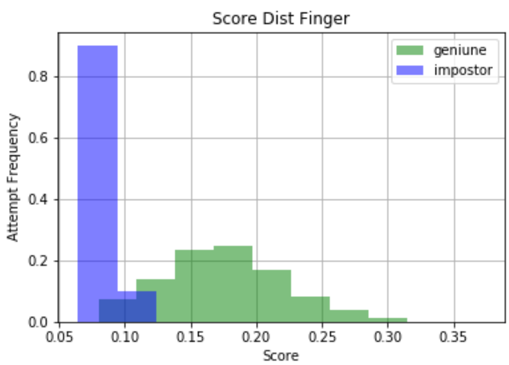


Figure 6.2: Score distribution for finger modality full scale

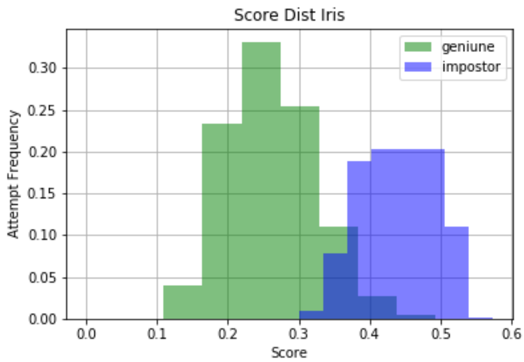


Figure 6.3: Score distribution for iris modality full scale

6.3 DET Evaluation

The software for the DET evaluation is based on a side-kit implementation from Hochschule-Darmstadt [NMR+17]. The parameters that have been used as input are a list of genuine scores and a list of impostor scores. The DET provides a spectrum of false-negative rate against false-positive rate. This establishes an accuracy and denotes information about the trade-offs between false-positives and false-negatives. Furthermore, the evaluations of the DET outputs the equal error rate (EER) which is an algorithm that can be used to predetermine the threshold values for the given system's false acceptance rate and its false rejection rate. When these two rates are equal the common value is what is called the equal error rate. This value indicates that the proportion of false acceptance is equal to the proportion of false rejections. To quantify it, it can be said that the lower the equal error rate value is, the higher accuracy of the biometric system. It is noted that the biometric system evaluated is in the identification scenario. A more interesting metric denoted by the DET could be the false non-match rate at a certain low threshold of false-match rate (e.g. 0.01%) as the diagram denotes those rates againts one another.

6.3.1 Modalities

The DET evaluation was generated for each modality i.e. iris, finger and face. These evaluations are supposed to work as a baseline so they can be compared to the performance of basic fusions and the k-stage system. From two disjoint datasets from the enrolled training set genuine scores and impostor scores were extracted for each modality. The genuine score is the match score of two different images from the same subject of the same biometric source e.g. two different images of the same subject's finger print or same iris or same face. Impostor is the match score of two different images from different subjects i.e. two images of different subjects' finger prints, different iris and different face. The DET curves show a spectrum of trade-offs between false non-match rate and false match-rate.

6.3.1.1 Iris baseline result

The EER is 1.57 % and it can be seen that the false non-match rate increases in a extreme fashion exponentially from a low false-match rate of 0.10 and lower towards 0 (see figure 6.4).

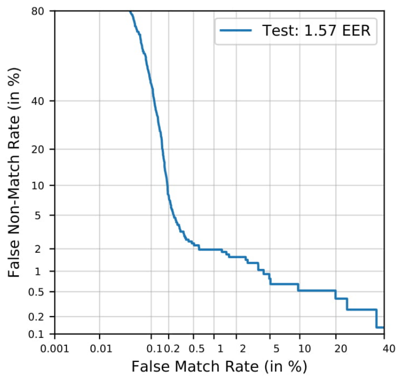


Figure 6.4: DET for Iris modality

6.3.1.2 Finger baseline result

The EER is 4.86 % and it can be seen that the false non-match rate increase significantly exponentially (extreme jump) from a false match rate of 0.1 and lower towards 0 (see figure 6.5).

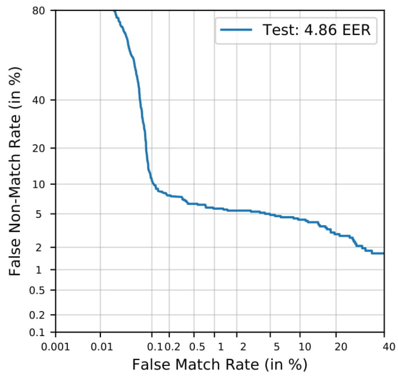


Figure 6.5: DET for the Finger modality

6.3.1.3 Face baseline result

The EER is 0.968 % and it can be seen that the false non-match rate increase intensely in a exponential way from a higher false-match rate of 0.5 and lower towards 0 (see figure 6.6).

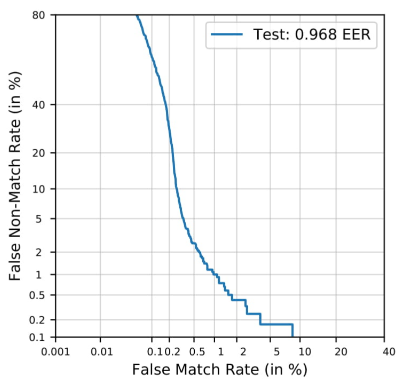


Figure 6.6: DET for face modality

6.3.2 Fusion

The results for the baselines show a comparable low DET in the identification mode for a naïve brute force exhaustive search.

6.3.2.1 Decision-level fusion basic result

The results from a basic decision level fusion with majority voting can be seen in figures 6.7 and 6.8. The technique to produce the outcomes for decision making fusion differentiated from the score-level fusion and singular modality fusion. The specialized method needed a separated calculations to compute false-matches and false-non matches and generate their rates as opposed to the input of raw data denoting genuine and imposter results.

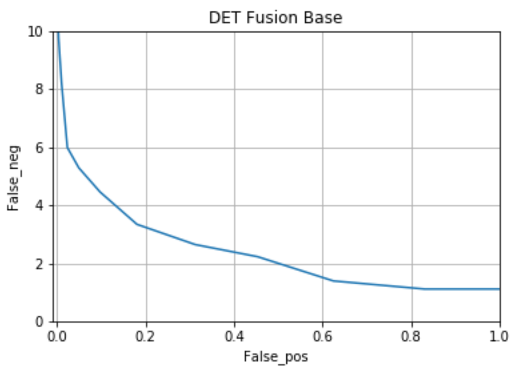


Figure 6.7: Specialized diagram created by a novel implementation of a Decision level fusion with majority voting DET for baselines in percentage - contains a 100 evaluation points

6.3.2.2 Score-level fusion basic results

The results from basic score-level fusion with score normalization (see figure 6.9) shows the curve for score level fusion compared to the curves for the baselines for each modality independently all together in one illustration. The score normalization used in this project is min-max normalization to a 0-1 range with a transformation to reflect dissimilarity measure.

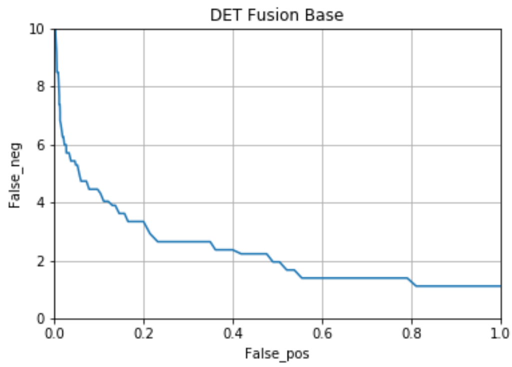


Figure 6.8: Specialized diagram created by a novel implementation of Decision level fusion with majority voting DET for baselines in percentages-contains a 1000 evaluation points

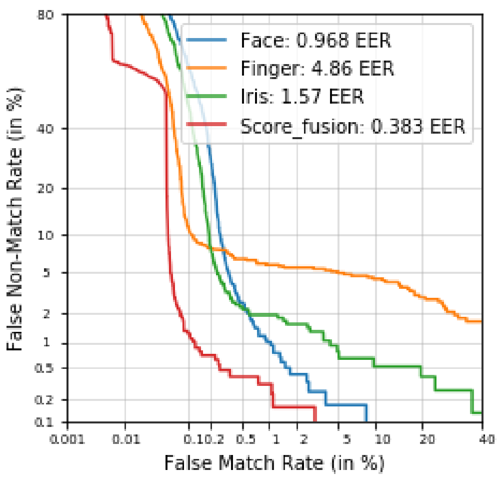


Figure 6.9: Baseline DET for basic score level fusion with normalization collected along the baselines for each modality.

6.4 CMC

These rankings (in the relative form) denote the probability for error rate of recognition against how big of a percentage of the enrolled database is under that error rate. Essentially, the plot denotes how much of a database (rank in percentages) is necessary to attain a certain rate for probability of recognition i.e. probability rate of correct recognition/ correct matching of identities with scores. The maximum rank denotes how much of the database is necessary to attain a 100% probability of correct matching.

6.4.1 Modalities

The CMC rankings for each modality was generated using the CMC algorithm in an identification scenario. For each modality it is denoted how much of the enrolled database is necessary to attain a certain probability rate of correct recognition. The purpose of generating these CMC rank curves for each modality is to compare the individual modalities against the fusion. The purpose for the CMC rank curves in regards to the k-stage fusion technique is to use the rankings as an indicator for setting the pre-selection spectrum whereafter the accuracy between the reduced searches can be compared to the exhaustive search. The idea is to evaluate a range of pre-selection spectrum which can be gained from the CMC rankings curves.

6.4.1.1 Iris baseline CMC rank

At rank 1% of the enrolled database it is possible to gain little bit more than 95% probability of recognition. The maximum rank is just beneath 75%. It is clear that the lower ranks are more dense than the higher ranks (see figure 6.10).

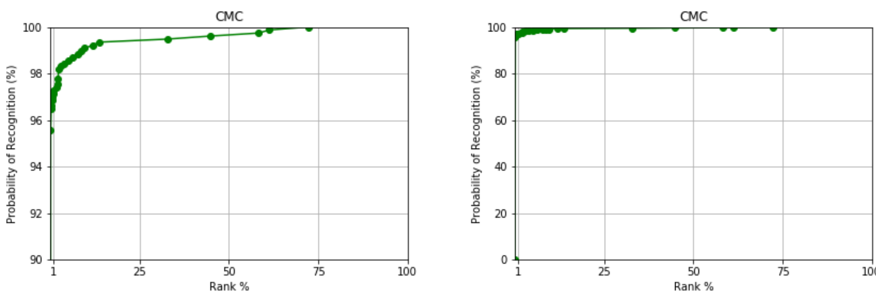


Figure 6.10: CMC for iris modality in different zoom illustrations

6.4.1.2 Finger baseline CMC rank

At rank 1% of the enrolled database it is possible to attain a probability rate of 93%. The maximum rank is 95% for 100% probability of recognition. It is clear that the lower ranks are more dense than the higher ranks (see figure 6.11).

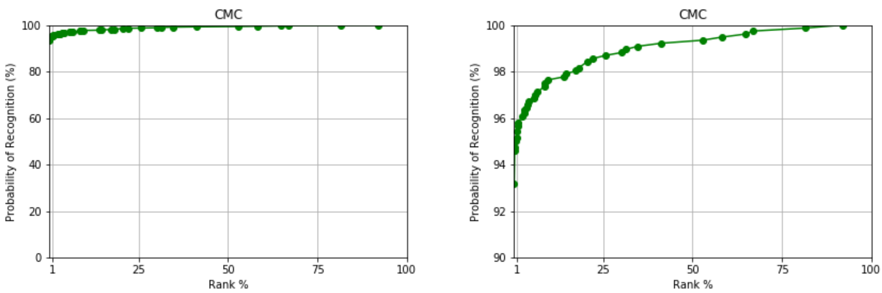


Figure 6.11: CMC for finger modality in different zoom illustrations

6.4.1.3 Face baseline CMC

At rank 1% of the enrolled database it is possible to attain a probability rate of 55%. There is a low density at the lowest level of rank, and the highest density between 2% to 5% of the enrolled database, and again a lower density at the higher ranks of the enrolled database (see figure 6.12).

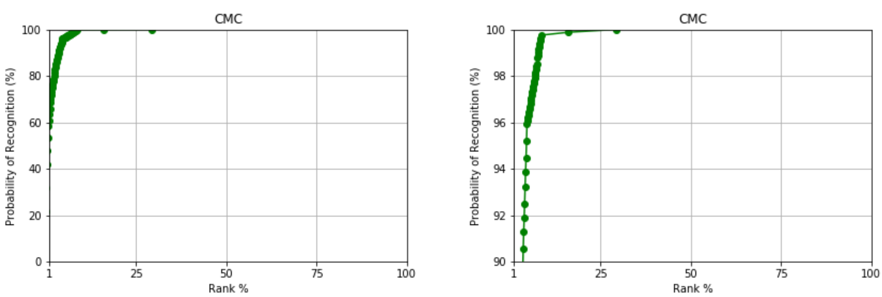


Figure 6.12: CMC for face modality in different zoom illustrations

6.4.2 Fusions basic CMC results

CMC of Baselines show a very low rank. As it can be seen the CMC for each fusion techniques have been somewhat maintained compared to the modalities i.e. see figure 6.13 and 6.14.

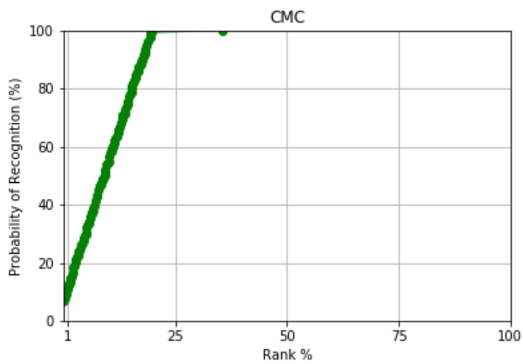


Figure 6.13: CMC for the basic decision level fusion for the baseline

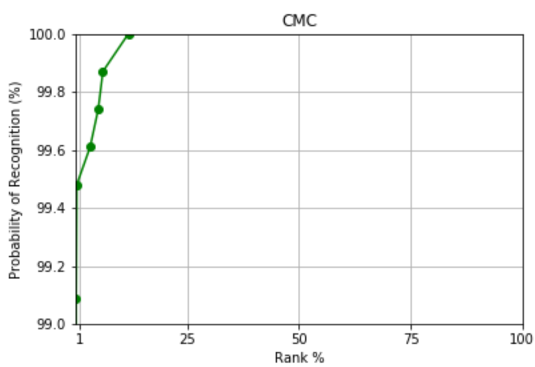


Figure 6.14: CMC for the basic score level fusion for the baseline zoomed into the 99-100% recognition rate.

6.5 Sources of Error

Due to the nature of the data-sets and software used there may be some inaccuracies in the baseline for each individual modality. The dataset is raw with minimal to none pre-processing (depending on the modality) and some images used for extraction is of poor quality. Furthermore, there is always disputes to be made regarding the distance measures used for similarity/dissimilarity, though the most commonly recognized were used as there exist many different distance metrics which may augment the biometric performance slightly. In regards to the software it may fail its task which can be caused by many factors from bugs/glitches to platforms OS inadequacies. However, the important factors is the performance of the basic fusion (which is all modalities together) in regards to the k-stage configurations. Furthermore, the validity of the approach to evaluate the k-stage fusion is of most significance.

6.6 Summary

The idea is to use these spectrums of CMC curves for each modality and the basic fusion(s) to set a wide range of reduction parameters for the pre-selection at each level using different modalities to be evaluated using DET. An array of thresholds can be set using the DET evaluations for each individual modality to create thresholds for genuine and impostors for the DET in the case of decision level fusion. Contrarily, score-level fusion uses the average/sum of multiple normalized scores to select genuines and impostors. It is noticeable that each evaluation for the DET in the baseline are similar in shape with deviations in accuracy.

Specifically, the score fusion evaluations denotes comparable results with a slightly better accuracy as denoted by the DET curve with all graphs included. The CMC for the score level basic fusion indicate a very good rank-1 probability of successful matching with the maximum rank for 100% of successful matching being low at under 25% of the total size of the entire database of match candidates.

An overview of the relevant results from the baseline experiment evaluation can be seen in table 6.1.

The intervals for the ranks are extracted from the CMC evaluations so the range of pre-selections areas are determined by the recognition rate as denoted by the y-axis against the rank denoted by the x-axis.

	EER	**Rank**
Finger	4.46%	25-95%
Face	0.968	5-30%
Iris	1.57%	10-74%
Score Fusion	0.383	1-10%
Decision Fusion	0.82%	20-30%

Table 6.1: Table denoting the key results from the baseline evaluation for each modality (finger, iris and face). For the sake of completeness the basic fusion techniques (Decision-level fusion with majority voting and Score-level fusion with normalization). These results include the EER rate from the DET evaluation and the optimal pre-selection sizes from the CMC rank evaluation where the range is within a 99-100% confidence interval as it pertains to the CMC curve.

Chapter 7

K-stage fusion Results

This section have the results for some of the experimentation done with a k-stage level system to conduct subject matches based on biometric information in an identification scenario. The tests were done with variations of modality ordering and pre-selection sizes based closely on the CMC rank evaluations from the baselines. Furthermore, the accuracies can be anticipated from the DET evalaution for the baselines.

Throughout this section the results for the evaluations (DET and CMC) of a wide variety of configurations for the k-stage system as conceptualized in the experimental setup section (see chapter 5) is presented and discussed. The configurations were set by the parameters of modality orderings and pre-selection sizes which in this experiment was done exhaustively for virtually every combinations of ordering and pre-selection sizes which denoted a large plethora of results that was subsequently analyzed to identify patterns and draw conclusions regarding the k-stage fusion. The best and most interesting findings and results are included in this section which is in terms of accuracy or other behaviours that may be of interest. Additionally, this section includes the workload reduction of these findings, which was computed and illustrated as discussed in the biometric fundamental theories chapter 2. Furthermore, other results that could be calculated from the shortened list of candidates by the k-stage system is included in order to garner clues about the behaviour of the applied k-stage algorithm.

This section lastly includes analysis of the results of the k-stage configurations as well as the baselines which is further accompanied with a summary of this part of the project. This functions to close out the central experimentation of this project and conclusions that could be drawn which will be discussed in depth in subsequent discussion and conclusion chapters.

7.1 Prediction for the biometric algorithms in the k-stage system

Given the results and subsequent findings from the baseline evaluations it is possible to predict how different configurations of the k-stage system are going to perform given the knowledge available about each modality and the baseline fusion methods.

Looking at the DET, it seems that using the face modality for the final selection will denote the most accurate results and using the finger modality will denote the worst, while iris is suitable but a bit worse than face in this case.

Reasoning, based on the CMC evaluations for the first experiment indicate that larger scale non-match candidate discarding with minimal loss of real match-candidates is suitable for the finger modality where it is noted that this is done on the 1.level of pre-selection as suggested in the setup for the k-stage system i.e. a high level of discarding is optimal for the 1.level. This is a combination of the minimal deviation in recognition rate over a large sparse rank area in concordance with a relatively worse accuracy performance by the finger modality in relation to the other modalities. The least suitable for this purpose seem to be the face modality due to the inverse behaviour in relation to the finger modality.

Furthermore, the CMC also denotes which pre-selection sizes denote the most accurate matching. Obviously, it is best to use a rank that denotes 100% recognition rate but since there is a large difference in ranks in some cases where the range of recognition rate is between 99% and 100%, it is sensible to test out configurations with different pre-selection sizes within the 99-100% recognition rate range where it is expected that the larger ranks denote more accurate results. However, there is the plausible factor of many accepted non-match candidates included in the higher spectrum of pre-selection sizes i.e. above 50% of the match-candidates or rank 50%. Generally, it presumed that the CMC doesn't change shape after having been through the k-stage system for any of the given modalities, however, deviations might occur due to effects of possible loss of genuine match scores.

The assumption is that the k-stage system removes a large amount of potential false match-candidates whilst keeping true-matches via its pre-selection. Due to the differences in the accuracy of garnered score for each modality it can be assumed that different orderings will denote different results e.g. since face baseline is the most accurate (denoted by its baseline DET) it is assumed this is the most suitable modality for the final selection level. It is also assumed that higher 1. rank (or lower ranks) with a reasonably lower max rank CMC curves (as the the one for the finger modality baseline) allow for smaller shortlists without too much loss of true-positive matches. The spectrum of CMC is most interesting for the shortlist size as, for example, the baseline for finger modality have a relatively higher maximum rank but has 99% accuracy or above starting at 10% meaning the difference between recognition rate 99-100% is quite sparse in terms of rank. That difference may indicate what kind of loss can be expected by going from a shortlist of the size for the maximum rank to a much lower rank of 10% without losing too much recognition accuracy (i.e. from 100% to 99% accuracy). It is assumed that well performing configurations have lower CMC maximum rank indicating how well the algorithm managed to distinguish genuines and impostors. It is assumed there is little to no loss of genuine identities. Therefore, the pre-selection sizes are presumed to be directly correlated to the CMC curves in terms of their performance in regards to accuracy as a measure of recognition rate.

The general prediction would be that configurations ending on the face modality would perform most accurately and that the modalities with lower maximum rank can expect to smaller pre-selection sizes where smaller pre-selection sizes may also be applied to the modalities with significantly sizable 1.rank CMC performance and higher low rank performance such as the finger modality. The assumption is that the k-stage system will remove false-positives and retain comparable and maybe even better performances in terms of accuracy where it is a given that the workload reduction is in effect significantly. Specifically, the performance of the k-stage system is compared in regards to the basic fusion evaluation which is what the different configuration are analyzed against in terms of better or worst performance in regards to accuracy.

In summation, it is assumed that k-stage filter will remove false-matches while retaining real matches. In simple points:

- The most accurate modality in terms of DET (Face) is most suitable for final selection (or final decision) and vice versa for the 1st level and 2nd level.
- The pre-selection sizes effect on accuracy in terms of recognition rate and rank is directly aligned to the baseline CMC, so if the pre-selection size is within 99% recognition rate it is very accurate.

- Configurations with low max rank are better for recognition rate against workload.

The purpose is to denote a spectrum of results for the k-stage system configurations. Subsequently, the configurations with low workload and high recognition rate (low false-negative rate or low EER) are the best configurations of this system in terms of accuracy vs. efficiency.

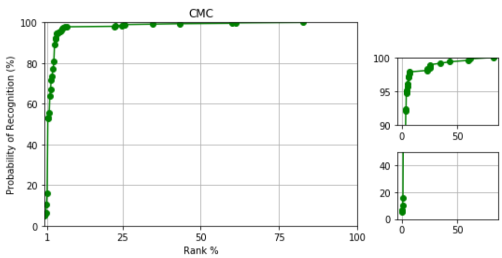
7.2 K-stage hierarchical results

The results in this section showcase the DET and CMC results from various configurations of the hierarchical k-stage system fusion method with different ordering of modalities and different pre-selection sizes. This k-stage system uses the setup illustrated and described figures 5.9 and 5.10 in chapter 5.

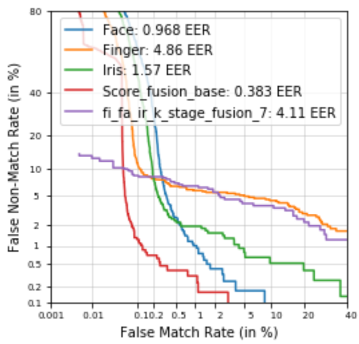
A lot of different configurations for the k-stage system were evaluated in comparison to the baseline DET evaluations of each modality and the basic score level fusion with normalization. The pre-selection sizes were settled by using the CMC from the baselines for each modality. The number of possible identification attempts required can be calculated by applying the pre-selection sizes on number of biometric identification decisions (subject comparisons) necessary for the exhaustive search 1:N scenario, which is in the case of this k-stage experiment around 331884 identification attempts. Furthermore, the number of possible identification decisions (subject comparisons) in conjunction with the bit-size/pair-size for the given template for each modality can help calculate the workload (i.e. F) that is used to compute the workload reduction for each configuration. The bit-wise/pair-wise workload reduction is detailed in the workload part of this section, whereas the pair-wise workload reduction is presented at each configuration of interest. The CMC for each configuration was also done to establish the spectrum of ranks of probabilities for successfully matching probes with their real identity which help indicates how much of the shortlist denoted by the k-stage system can be adjusted for further inquiries. Additionally, there might be some loss of genuine identities due to configurations working by best score in the identification scenario and irregularities might denote imposter score to be better than genuine when discarding match-candidates which prompts that the loss of genuines compared to the full sized database being denoted as another evaluation factor.

7.2.1 Configurations with ordering: finger-face-iris

For the ordering of finger-face-iris the following results were found to be of interest.



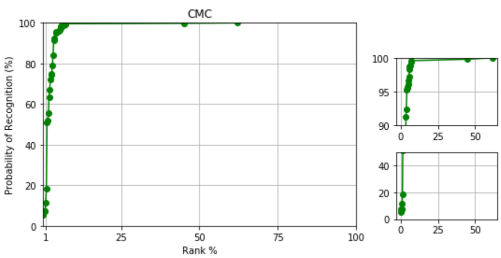
(a) CMC for finger-face-iris ordering and pre-selection sizes finger=10%, face=60% and final selection iris



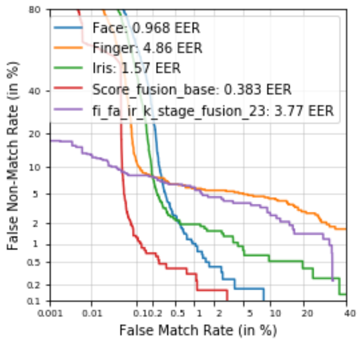
(b) DET for finger-face-iris ordering and pre-selection sizes finger=10%, face=60% and final selection with iris.

Figure 7.1: K-stage fusion with finger-face-iris and pre-selection sizes finger=10%, face=60% and final selection with iris. Of interest is the basic score fusion (red graph) and the k-stage fusion (purple graph) where the comparative performance shows that the k-stage configuration is worse. This k-stage system configuration has a 4.11 EER.

A medium-good performance for this configuration of the k-stage system (see figure 7.1) as it is comparable with the baseline DET evaluations but shows a slightly worse performance as also indicated by the higher EER score. This configurations requires a number of possible identification attempts equating to 6% of the baseline number of possible identification attempts in the exhaustive search scenario i.e. 19913 identification attempts. The CMC rank of this configurations indicate that the number of identification attempts can be reduced further without compromising the accuracy as it shows a very high density around 99% recognition rate at a lower rank that become spread out on the ranks between 99% to 100%, however, it drastically decreases to poor probability of successful matching at the lowest ranks. This configurations causes 4.91% loss of genuine identity matches compared to the total amount of genuine identity matches in the database. For the Workload reduction, the DET is used where $TP_{0.01}$ can be identified and computed as 87% and F can be identified as the number of possible identification attempts required for this configuration which is 6%. So given the equations discussed in the workload section of chapter 2 the distance to optimal workload reduction metric is calculated to be $\tau = 0.14$.



(a) CMC for finger-face-iris ordering and pre-selection sizes finger=95%, face=35% and final selection with iris



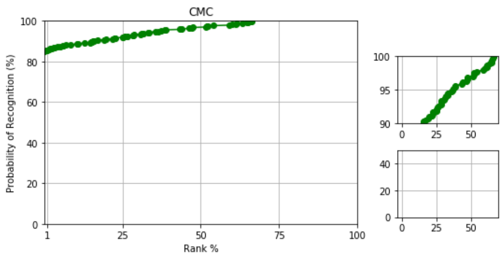
(b) DET for finger-face-iris ordering and pre-selection sizes finger=95%, face=35% and final selection with iris

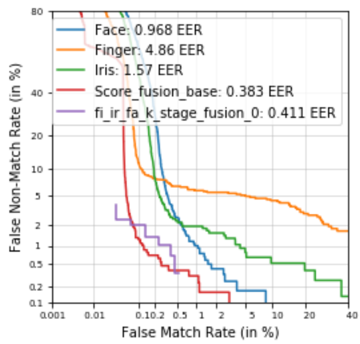
Figure 7.2: K-stage fusion with finger-face-iris and pre-selection sizes finger=95%, face=35% and final selection with iris. This k-stage system configuration has a 3.77 EER. Of interest is the basic score fusion (red graph) and the k-stage fusion (purple graph) where the comparative performance shows that the k-stage configuration is worse.

A comparable medium-good performance for the configurations of the k-stage system (see figure 7.2) that is close to the baseline DET evaluations which is also indicated by the EER score which is comparatively a little bit higher by the basic fusion curve. This configurations requires a number of possible identification attempts equating to 33% of the number of possible identification attempts in the exhaustive search scenario i.e. 109521 identification attempts. The CMC rank of this configuration indicate that the number of identification attempts can be reduced further without compromising the accuracy as it shows a relatively high density with 99% recognition at a low rank that becomes sparse between 99% to 100% on the ranks with a steep drop to worse around its lowest ranks in terms of probability of successful matching. This configurations causes 0.47 % loss of genuine identity matches compared to the total amount of genuine identity matches in the database. For the Workload reduction, the DET is used where $TP_{0.01}$ can be identified and computed as 87% and F can be identified as the number of identification attempts required for this configuration which is 33%. So given the equations discussed in the workload section of chapter 2 in this report the distance to optimal workload reduction metric is calculated to be $\tau = 0.35$.

7.2.2 Configurations of ordering: finger-iris-face

For the ordering of finger-iris-face the following results were found to be of interest.



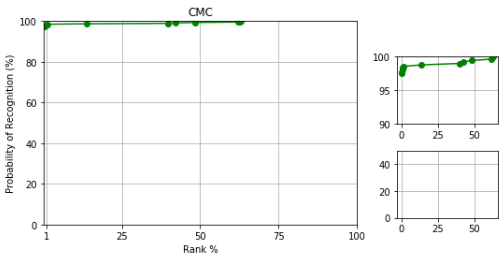


(a) CMC for finger-iris-face and pre-selection sizes finger=10%, iris=10% and final selection with face

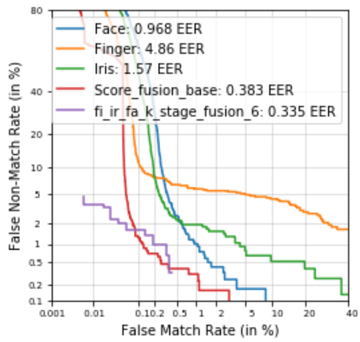
(b) DET for finger-iris-face and pre-selection sizes finger=10%, iris=10%, and final selection with face

Figure 7.3: K-stage fusion with finger-iris-face and pre-selection sizes finger=10%, iris=10%, and final selection with face. This k-stage system configurations has a 0.411 EER. Of interest is the basic score fusion (red graph) and the k-stage fusion (purple graph) where the comparative performance shows that the k-stage configuration is slightly worse.

A good performance for this configurations of the k-stage system (see figure 7.3) which is also indicated by its relatively small EER score compared to the baseline DET evaluations. This configurations requires a number of possible identification attempts equating to 1% of the number of possible identification attempts in the exhaustive search scenario i.e. 3318 identification attempts. The CMC rank of this configurations indicate that the number of identification attempts will be problematic to reduced further without compromising the accuracy as it shows a relatively medium maximum rank with a steady drop to increasingly bad probability of successful matching around its lower ranks from that point that is still relatively highly accurate (i.e. 85%) which is not in that desired 99% to 100% area. This configurations causes 49.65% loss of genuine identity matches compared to the total amount of genuine identity matches in the database. For the Workload reduction, the DET is used where $TP_{0.01}$ can be identified as 94% and F can be identified as the number of identification attempts required for this configuration which is 1%. So given the equations discussed in the workload section of chapter 2 in this report the distance to optimal workload reduction metric is calculated to be $\tau = 0.11$.



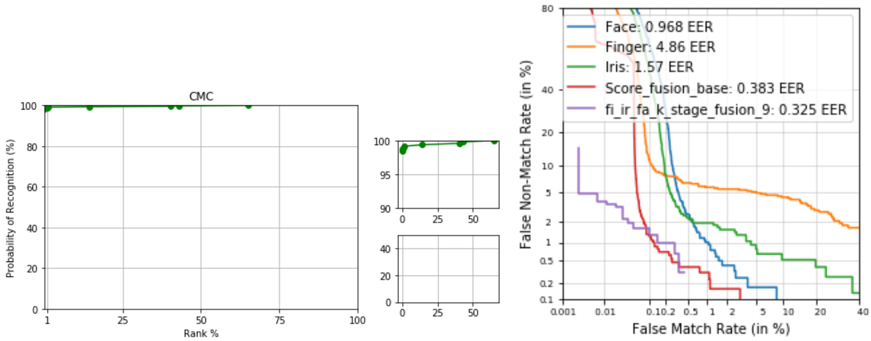
(a) CMC for finger-iris-face ordering and pre-selection sizes finger=10%, iris=50% and final selection with face



(b) DET for finger-iris-face ordering and pre-selection sizes finger=10%, iris=50% and final selection with face.

Figure 7.4: K-stage fusion with finger-iris-face and and pre-selection sizes finger=10%, iris=50% and final selection face. This k-stage system configuration has a 0.335 EER. Of interest is the basic score fusion (red graph) and the k-stage fusion (purple graph) where the comparative performance shows that the k-stage configuration is slightly worse.

A good performance as the configuration of the k-stage system (see figure 7.4) shows a better performance to the baselines in the DET which is also indicated by its relatively lower EER score compared to the baselines. This configurations requires a number of possible identification attempts equating to 5% of the number of possible identification attempts in the exhaustive search scenario i.e. 16594 identification attempts. The CMC rank of this configurations indicate that the number of identification attempts will be non-problematic to reduce further without compromising the accuracy as it shows a relatively medium sized maximum rank with a very low steady drop that never goes below 95% accuracy. This configurations causes 5.41% loss of genuine identity matches compared to the total amount of genuine identity matches in the database. For the Workload reduction, the DET is used where $TP_{0.01}$ can be identified as 96% and F can be identified as the number of identification attempts required for this configuration which is 5%. So given the equations discussed in the workload section of chapter 2 in this report the distance to optimal workload reduction metric is calculated to be $\tau = 0.06$.



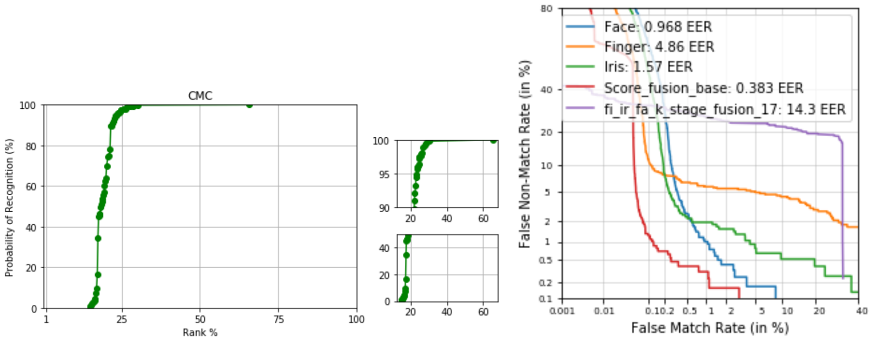
(a) CMC for finger-iris-face and pre-selection sizes finger=25%, iris=50% and final selection with face

(b) DET for finger-iris-face and pre-selection sizes finger=25%, iris=50%, and final selection with face

Figure 7.5: K-stage fusion with finger-iris-face and pre-selection sizes finger=25%, iris=50%, and final selection with face. This k-stage system configuration has a 0.325 EER. Of interest is the basic score fusion (red graph) and the k-stage fusion (purple graph) where the comparative performance shows that the k-stage configuration is better.

A very good performance by this configuration of the k-stage system (see figure 7.5) which is also indicated by its relatively lower EER score compared to the baseline DET evaluations. This configurations requires a number of possible identification attempts equating to 13% of the number of possible identification in the exhaustive search scenario i.e. 43145 identification attempts. The CMC rank of this configurations indicate that the number of identification attempts will be non-problematic to reduced further without compromising the accuracy as it shows a relatively medium maximum rank with a very low steady drop to a probability in the 98% area at the lowest ranks. This configurations causes 2.59 loss of genuine identity matches compared to the total amount of genuine identity matches in the database. For the Workload reduction, the DET is used where $TP_{0.01}$ can be identified as 95% and F can be identified as the number of identity attempts required for this configuration which is 13%. So given the equations discussed in the workload section of chapter 2 in this report the distance to optimal workload reduction metric is calculated to be $\tau = 0.14$.

A bad performance for this configurations of the k-stage system (see figure 7.6) relative to the baseline DET evaluations which is also indicated by its EER



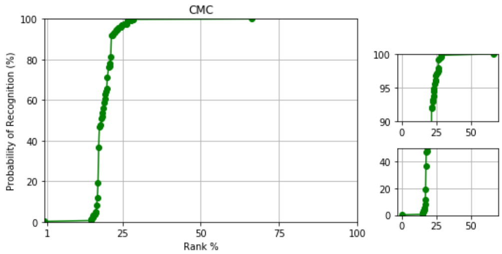
(a) CMC for finger-iris-face and pre-selection sizes finger=50%, iris=75% and final selection with face

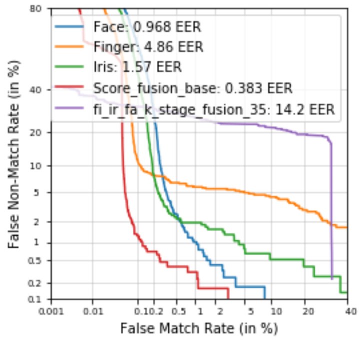
(b) DET for finger-iris-face and pre-selection sizes finger=50%, iris=75,% and final selection with face

Figure 7.6: K-stage fusion with finger-iris-face and pre-selection sizes finger=50%, iris=75% and final selection with face. This k-stage system configuration has a 14.3 EER. Of interest is the basic score fusion (red graph) and the k-stage fusion (purple graph) where the comparative performance shows that the k-stage configuration is significantly worse.

score. This configurations requires a number of possible identification attempts equating to 38% of the number of possible identification attempts in the exhaustive search scenario i.e. 126115 identification attempts. The CMC rank of this configurations indicate that the number of identification attempts will be non-problematic to reduced further without compromising the accuracy as it shows a relatively medium-small maximum rank, however, there is a steep drop within a 1% deviation to very bad probability of successful matching around its lower ranks indicate that reduction lower than the 20% rank area will be highly problematic. This configurations causes 1.17% loss of genuine identity matches compared to the total amount of genuine identity matches in the database. For the Workload reduction, the DET is used where $TP_{0.01}$ can be identified as 72% and F can be identified as the number of identification attempts required for this configuration which is 38%. So given the equations discussed in the workload section of chapter 2 in this report the distance to optimal workload reduction metric is calculated to be $\tau = 0.47$.

The configuration of the k-stage system DET evaluation (see figure 7.7) shows a worse performance compared to the baseline DET evaluations as also indicated by its EER score. This configurations requires a number of possible identification attempts equating to 71% of the number of possible identification attempts





(a) CMC for finger-iris-face ordering and pre-selection sizes finger=95%, iris=75% and final selection with face

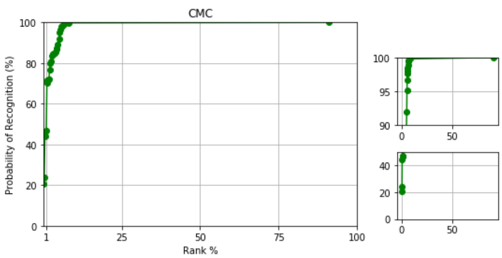
(b) DET for finger-iris-face ordering and pre-selection sizes finger=95%, iris=75% and final selection with face

Figure 7.7: K-stage fusion with finger-iris-face and pre-selection sizes finger=95%,iris=75% and final selection with face. This k-stage system configuration has a 14.2 EER. Of interest is the basic score fusion (red graph) and the k-stage fusion (purple graph) where the comparative performance shows that the k-stage configuration is significantly worse.

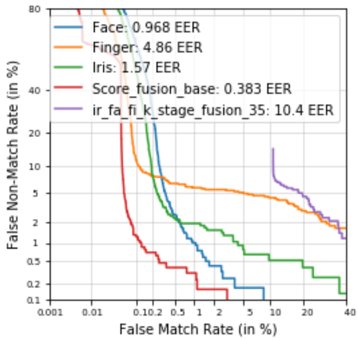
in the exhaustive search scenario i.e. 235637 identification attempts. The CMC rank of this configurations indicate that the number of identification attempts will be non-problematic to reduce further without compromising the accuracy as it shows a relatively medium sized maximum rank in terms of density, however, there is a stark drop beyond a 1% deviation around the 25% rank to an extremely poor and borderline non-existent performance at the lowest ranks in terms of probability of successful matching indicating reduction lower than that would be highly problematic. This configurations causes 0% loss of genuine identity matches compared to the total amount of genuine identity matches in the database. For the Workload reduction, the DET is used where $TP_{0.01}$ can be identified as 74% and F can be identified as the number of identification attempts required for this configuration which is 71%. So given the equations discussed in the workload section of chapter 2 in this report the distance to optimal workload reduction metric is calculated to be $\tau = 0.76$.

7.2.3 Configuration of ordering: iris-face-finger

For the ordering of iris-face-finger the following results were found to be of interest:



(a) CMC for iris-face-finger ordering and pre-selection sizes iris=75%, face=60% and final selection with finger



(b) DET for iris-face-finger ordering and pre-selection sizes iris=75%, face=60% and final selection with finger.

Figure 7.8: K-stage fusion with iris-face-finger and pre-selection sizes iris=75%, face=60% and final selection with finger. This k-stage system configuration has a 10.4 EER. Of interest is the basic score fusion (red graph) and the k-stage fusion (purple graph) where the comparative performance shows that the k-stage configuration is worse.

A bad performance as showcased by the DET for the configuration of the k-stage system (see figure 7.8) which is also indicated by its EER score compared to the baseline DET evaluations. This configurations requires a number of possible identification attempts equating to 45% of the number of possible identification attempts in the exhaustive search scenario i.e. 149347 identification attempts. The CMC rank of this configurations indicate that the number of identification will be non-problematic to reduce further without compromising the accuracy as it shows a relatively small maximum rank in terms of density within a 99% to 100% area with a stark drop beyond a 1% deviation around the 5% rank area to an extremely poor and borderline non-existent performance at the lowest ranks in terms of successful matching. This configurations causes 0% loss of genuine identity matches compared to the total amount of genuine identity matches in the database. For the Workload reduction, the DET is used where $TP_{0.01}$ can be identified as 50% and F can be identified as the number of identification attempts required for this configuration which is 45%. So given the equations

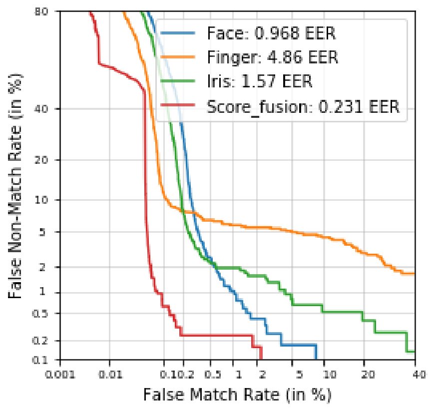


Figure 7.9: This DET showcase the baseline associated to the finger-face configurations in the 2 modalities configurations.

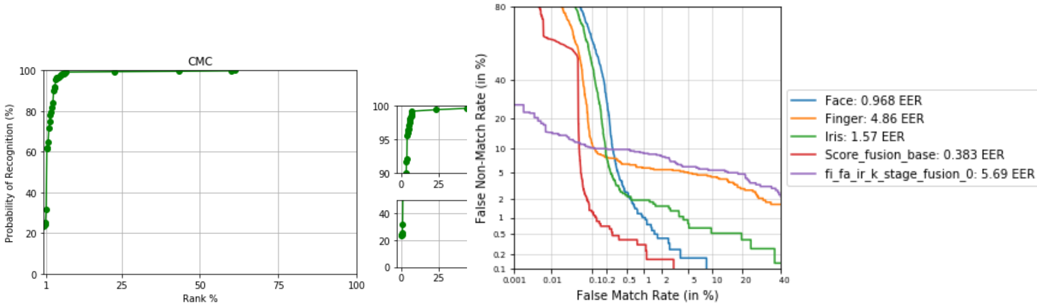
discussed in the workload section of chapter 2 in this report the distance to optimal workload reduction metric is calculated to be $\tau = 0.67$.

7.2.4 Configurations ordering: finger-face

The same tests/experimentation was applied to every combination for any two combination of the modalities. That test showed generally similar result that was somewhat worse that the experiment with two levels of pre-selection, however, no configurations denoted better results than the baselines for this two-level experiment whereas the three-level experiment denoted results for some configuration which performed better. From these tests the finger-face ordering denoted the best results in terms of accuracy seen on the DET curve compared to its related baseline (see figure 7.9).

This configuration was with a 10% pre-selection of finger (see figure 7.10).

This configuration has a 10% reduction of identification attempts compared to the 1:N search scenario i.e. 33188 identification attempts. In this case the loss of genuine match scores equates to 4.94%. The CMC curve show that there is room for further reduction without compromising the accuracy as it has a highly accurate recognition at a relatively low maximum rank in the 99% to 100% area with a stark drop around the 5% rank to poor recognition at the lowest ranks which indicate that further reduction beyond that point would be



(a) CMC for iris-face-finger ordering and pre-selection sizes iris=10% and final selection with face

(b) The finger-face combination with 10% total pre-selection and selection with the face modality

Figure 7.10: K-stage fusion with finger-face and pre-selection size finger=10% and final selection with face. Of interest is the basic score fusion (red graph) and the k-stage fusion (purple graph) where the comparative performance shows that the k-stage configuration is worse.

highly problematic. As it can be seen this configuration is comparable at best to the other configurations that incorporates three-levels. Furthermore, this is the best performing configuration of two-modality combinations.

7.2.5 Summary plots

All DET plots summarized as a single entity providing a comprehensive overview of the results presented throughout this section in figure 7.11.

Furthermore, the summary of all their corresponding CMC plots can be seen as figures 7.12 and 7.13.

A selection of the results for the best/most interesting performing k-stage system configuration can be seen in table 7.1 which include the parameters, EER, workload reduction and loss of genuine. In depth presentation and interpretation of some of the more interesting results from the k-stage experiments are included in this section showing various configurations of the k-stage system evaluated with the their graphs, computations and analysis. This is followed by a presentation of workload reduction with an associated evaluation of workload as a factor against accuracy.

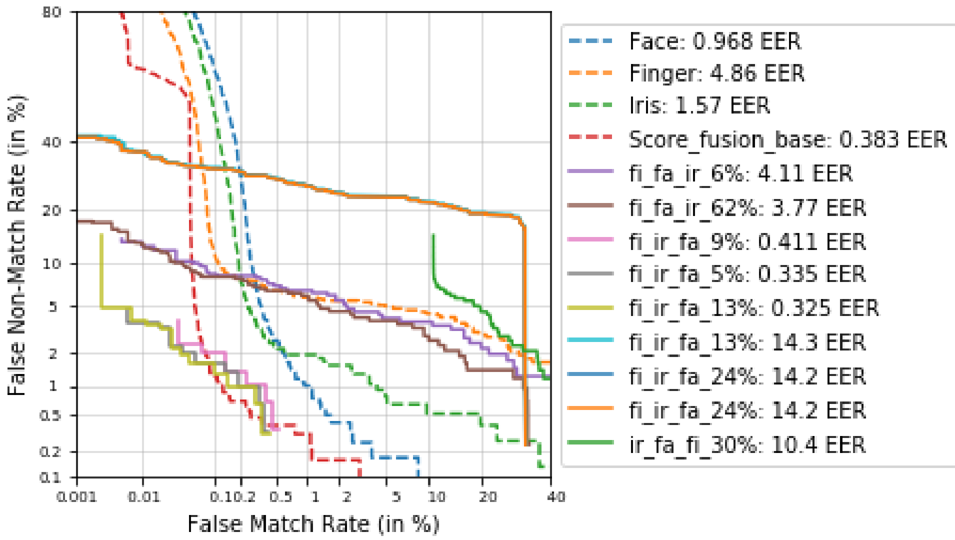


Figure 7.11: All the DET plots used for discussing findings summarized into one entity. Of interest is the basic score fusion (red graph) and the k-stage fusion graphs. The labelling indicate what the order of modalities is e.g. fi-fa-ir is finger-face-iris and the total pre-selection size for the given configuration i.e. first-level pre-selection times second pre-selection size.

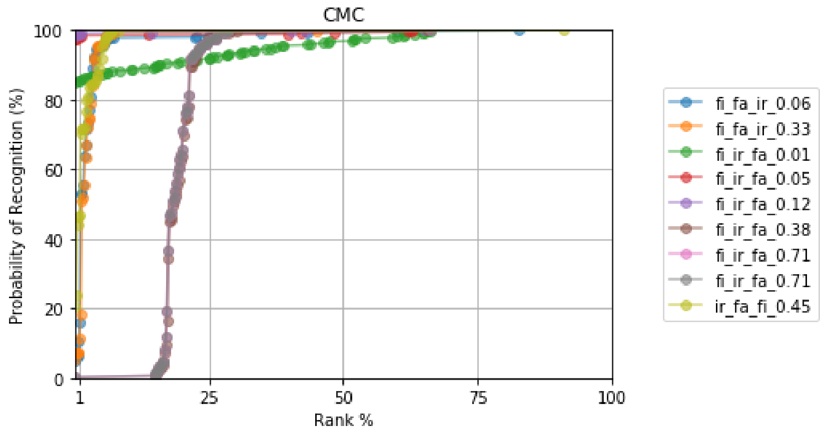


Figure 7.12: CMC summary plots corresponding to the DET plots from the chosen k-stage configurations

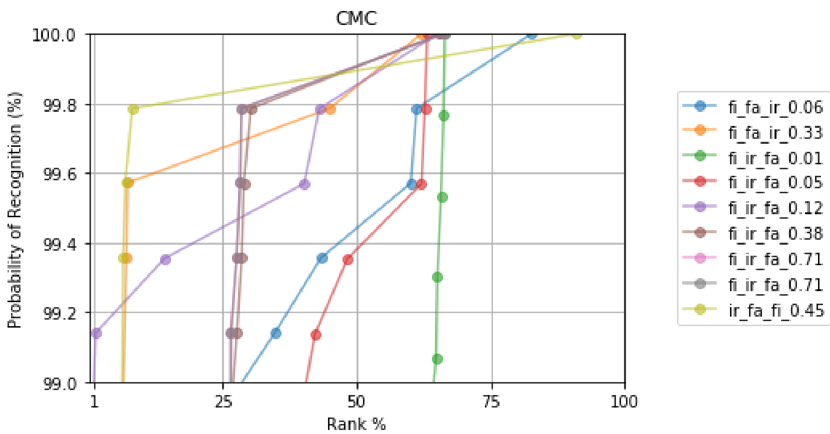


Figure 7.13: CMC summary plots corresponding to the DET plots from the chosen k-stage configurations

1. lvl	2.lvl	FS	EER	W(F)	W(τ)	L(G)	Rank(LB)	Rank(UB)
Finger 10%	Face 60%	Iris	4.11 %	0.87%	1.2	4.910 %	10%	80%
Finger 95%	Face 35%	Iris	3.77 %	4.82%	4.9	0.470 %	5%	65%
Finger 10%	Iris 10 %	Face	0.411 %	0.01%	0.9	49.65%	60%	70%
Finger 10%	Iris 50%	Face	0.335%	0.03%	1.0	5.410%	5%	70%
Finger 25%	Iris 50%	Face	0.325%	0.06%	1.0	2.590%	1%	70%
Finger 50%	Iris 75%	Face	14.3 %	0.2%	0.7	1.170%	25%	70%
Finger 95%	Iris 75%	Face	14.2%	0.39%	0.8	0%	25%	78%
Iris 75 %	Face 60%	Finger	10.4 %	38%	38.2	0%	3%	95%
Iris 10%	Finger 75%	Face	46.3 %	0.01%	0.9	1.1%	40%	95%
Face 10%	Iris 75%	Finger	8.35 %	41%	44.2	0.4%	5%	50%
Face 10%	Finger 95%	Iris	12.6 %	5.2%	5.5	1.2%	25%	80%
Finger 10%	...	Face	5.69 %	0.002%	1.5	4.94%	5%	60%
Iris 25%	...	Face	10.9 %	0.001%	1.2	5.1%	10%	75%

Table 7.1: Table of some of the best/most interesting configurations with their results from the k-stage experiment denoting their parameters with modality and pre-selection for 1. level pre-selection (***1.lvl***), 2. level pre-selection (***2.lvl***) and Final selection (***FS***). Furthermore, along with the configurations' parameters their associated results is denoted for EER, reduced workload (***W(F)***), the denoted distance τ (***W(τ)***) from reduced workload against accuracy as True-negative rate at 0.01 % and loss of genuine identities (***L(G)***). Additionally, the denoted pre-selection range from the CMC within a 99-100% confidence interval range (***Rank***) at the selection level is denoted, with columns for the lower bound (***LB***) and the upper bound (***UB***). The color coded rows indicate what configuration row relates to as denoted in the various DET, CMC and workload reduction graphs.

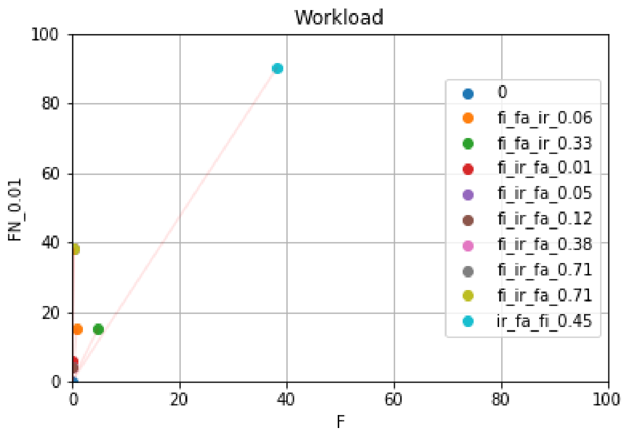


Figure 7.14: Workload reduction presentation using False-non match rate at false-match rate 0.01% as accuracy.

7.3 Workload Reduction for k-stage system

The bit-wise comparisons workload reductions for these configurations of interest are denoted as discussed in [DRB18a] which has been discussed earlier in this report. The template sizes where each retrieved from the documentations of the respective software utilized along with common knowledge about computations e.g. a float is 32 bit. The workload reductions are denoted as a graph showing workload (F) as a function of accuracy (denoted as False Negative at rate at False Positive rate 0.01 %) which can be seen in figure 7.14, 7.15 and 7.16.

7.4 Summary of k-stage experiment

This experiment has denoted a plethora of results which denotes what kind of orderings and what kind of combination of pre-selection sizes is best suitable for a given task in terms of accuracy and speed. It also denotes which configurations cause incomparable results i.e. which configurations to avoid using. Conclusively, some configurations denote comparable accuracy performance or even better at very low losses of genuine at a fractional number of possible identification attempts (biometric identification decisions/ subject comparisons) compared to the number of identification attempts required for the exhaustive search of the whole database without the k-stage system reduction. As discussed

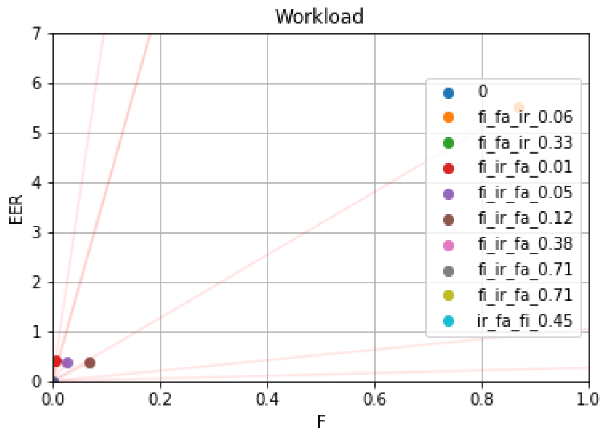


Figure 7.15: Zoomed Workload reduction presentation with EER for accuracy.

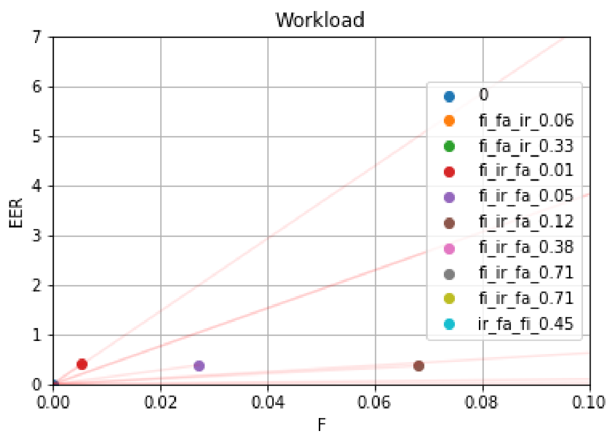


Figure 7.16: Highly Zoomed Workload reduction presentation with EER for accuracy.

it also present configurations that are incomparable to the 1:N scenario.

The k-stage algorithm in this experiment was set up to exhaust every possible ordering of the used modalities with a selection of pre-selection size combinations where the available pre-selection sizes were set for each modality independently based upon their baseline CMC evaluations. This setup results in a way of creating different shortlists of match candidates. Thereafter, the final evaluation for CMC and DET can be determined by using the last remaining unused modality. So the biometric algorithm is predicted by the behaviors found in the baselines which has been broadly confirmed by the k-stage experiment.

The pre-selection sizes affect accuracy as they are denoted in the baseline CMC ranks meaning that by a given thresholds (i.e. when accuracy drops beneath 99%). The recognized patterns is that the face modality is best for selection since it is most accurate, and finger is the best for smaller pre-selection sizes (i.e. higher discarding) based on its CMC and workload. Correlated, these are also the best in terms of workload reduction as long as they follow the rule established by their given CMC rank baseline for threshold of rank versus accuracy of choosing match-candidates. The general tendency for workload reduction is that the higher reduction means less accuracy but there are plenty of deviations of this tendency by some configurations which include the aforementioned positive parameters which fulfills the assumptions made about the k-stage system which was discussed in the beginning i.e. successfully discarding non-match candidates without sacrificing accuracy and increasing workload reduction.

In summation, the findings can be summarized as:

- **Selection with face:** The discoveries throughout the results as well as the predictions that could be made from the baselines indicate that orderings where the face modality is the selection modality denoting the most accurate results.
- **Large reduction with finger:** The findings indicate that the larger scale discarding (smaller pre-selections sizes) is best adaptable with the finger modality since it has the best chance of removing false candidates.
- **Pre-selection size thresholds:** The pre-selection sizes are very much correlated with the CMC that can be found in the baselines meaning that highly accurate (99% over) ranks correspond to similar sized pre-selection (i.e. pre-selection sizes correlated to highly accurate ranks denote good results). Pre-selection sizes corresponding to less accurate ranks denote similarly bad results.
- **Loss of genuine:** The loss of genuine is mostly correlated to low pre-selection sizes with highly accurate modalities or modalities with high

ranks for high accuracy (99%) such as iris.

- **Workload reduction:** Findings from the workload reduction showcase a pattern of highly accurate results are denoted from configurations that has a lot of workload reduction (due to low pre-selection sizes) which verifies the assumption of the removal of false candidates with fusion techniques whilst maintaining high accuracy.

A more in-depth discussion of the impact of the k-stage system experiment as compared to baselines will denote how this experiments and its results can be applied to a larger and broader scale. Furthermore, a discussion will provide the necessary clues to conclude the general results from the k-stage system experiment proposed in this paper denoting the abstract implications of such a system in a larger perspective and how to apply it in any context irreverent of the parameters for modality ordering and pre-selection sizes.

Chapter 8

Large-scale Dataset experiment

This chapter presents a second experiment on a larger scale was conducted using synthetic datasets where necessary to generate 100000 subject instead of the 2000 which was used for the first experiment. Due to the magnitude of the experiment a subset from the dataset was used for creating false-match candidates.

The approach to this experiment is completely the same as discoursed earlier in this report for the first k-stage experiment, using the same methodologies and techniques for the evaluation with first getting some results for the baselines and then compare it to results after applying the k-stage system filter. It is noted that the same software, techniques and implementations are used for the feature extraction, comparison scores and evaluations tasks i.e. DET, CMC, workload etc.

8.1 Datasets

The purpose of the second experiment is to move the k-stage experiment to a larger-scale to further modify the eventual model and to investigate whether some additional key findings can be found by upping the scale for the k-stage system. This mean that it was necessary to move onto larger data sets that would allow the generation of 100000 subjects for experimentation (see table 8.1).

Iris	Face	Finger
SIC-Gen: A Synthetic Iris-Code Generator [Dro17]	IMDB Wiki dataset [RR15]	SFinGe Finger Generator [Cap04]

Table 8.1: Datasets for the second experiment that is larger than the ones of the first experiment allowing the generation of 100000 test subjects

In the case of the iris modality and finger modality, there are no available real datasets of the necessary magnitude which necessitates the utilization of synthetic biometric information generators. The software for generating the synthetic datasets for these modalities where provided by Da-Sec. This software generates binary codes representing irides where the templates are binary codes which are available in both bmp and txt format. These templates can be used for comparison computations by Hamming distance which can be done via the Osiris system or a customized distance Hamming implementation.

The face modality the publicly available database for faces IMDB-Wiki was used for the this second experiment as the database was of sufficient size to create the necessary amount of subjects for this experiment. The face modality follows the exact same methodology as the one used in the first experiment except it was applied to another dataset i.e. a txt file of 128 float values used as template and can be used for comparison score computation with a squared euclidean distance measure.

The finger modality is fully generated by the sFinGe software which also enables the generation of fully usable finger images but also iso templates that represent generated finger-print images. These templates can be processed with the MCCSDK software that computes comparison score from two iso templates.

8.2 Software

In large parts, the same software that was used for the first experiment was used for the second. Specifically, the methodologies presented in first experiment was used in the second i.e. Hamming distance, squared euclidean and MCC comparison. However, the methodologies in some cases had to be adjusted as they had been specified for the given dataset in addition to the somewhat reduced non-match candidates generated for the second experiment which was done for practicality purposes as opposed to the exhaustive method in the first experiment. Specifically, the feature extraction of the face modality was affected by the modification necessary for the validation script and the iris comparison scores were affected by the customized implementation of the Hamming distance calculation on the synthetically generated feature extraction templates as opposed to letting the raw iris images run through the standardized Osiris software which included comparison score computations.

8.3 Modification to experiment setup

Besides the modifications to comparisons score computations implementations and the software for the feature extraction, the experiment setup is majorly the same. The key difference is the generating of non-match candidates. In the first experiment every possible combinations of 2000 subjects where exhausted in the first experiment, while the second experiment used a sizable subset of randomized combinations of 100000 subjects due to reasons of practicality in terms of computation workload for the feature extraction and comparison calculations.

8.4 Baselines Results

Similar to the first experiment, a generated dataset representing subjects (100000 in this experiment) with each subject containing 3 comparison scores representing each modality (finger, iris, face) was made. This dataset was split into sets for training and testing like in the first experiment. Furthermore, a score level fusion was applied to the full scale dataset.

8.4.1 Score distributions

This section shows the score distributions for the modalities used for the large-scale dataset experiment as can be seen in figures 8.1, 8.2 and 8.3.

8.4.2 Large-scale dataset base DET evaluations

Evaluation with DET was done to each modality and the basic score-level fusion to establish the baselines as can be seen in figure 8.4. This is in the context

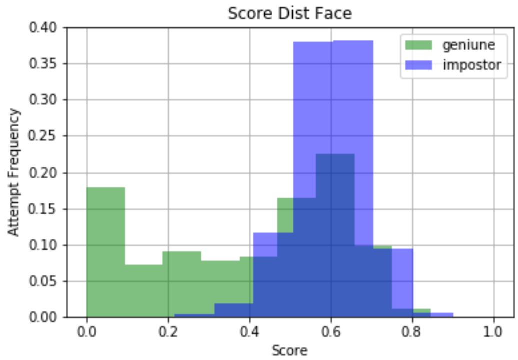


Figure 8.1: The score distribution graph for the face modality in the large-scale dataset experiment

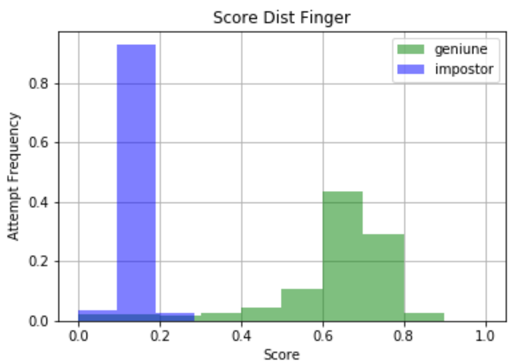


Figure 8.2: The score distribution graph for the finger modality in the large-scale dataset experiment

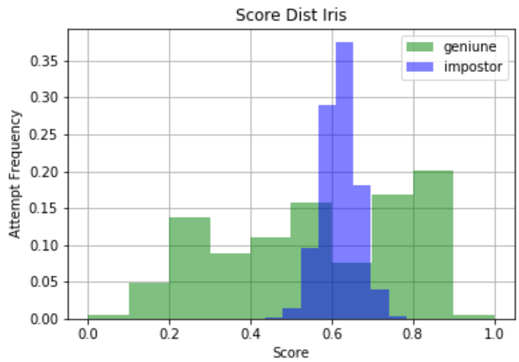


Figure 8.3: The score distribution graph for the iris modality in the large-scale dataset experiment

of second k-stage experiment with the large-scale dataset based on the feature extraction and comparison score computation for the second experiment.

8.4.3 Large-scale dataset base CMC evaluations

Evaluation with CMC was done to each modality and the basic score-level fusion to establish the baselines as can be seen in figures 8.5. As was done in the first experiment the baseline CMC evaluations are used for setting the pre-selection sizes.

8.5 K-stage experiment Results

This section include some of the best and most interesting results from the second k-stage experiment using the large-scale datasets.

8.5.1 k-stage results for large-scale dataset experiment

The DET and CMC results from the large-scale experiment can be seen in the figures in this section (see figures 8.6 8.7,8.8, 8.9).

The associated CMC plots to the chosen configurations denoting the best/most

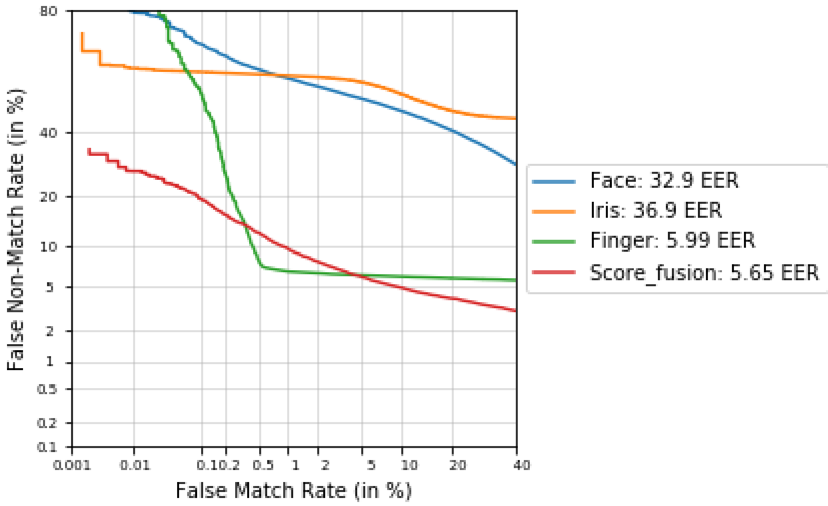


Figure 8.4: The DET curves summarized into one diagram for the baselines used for the second experiment with the large-scale datasets.

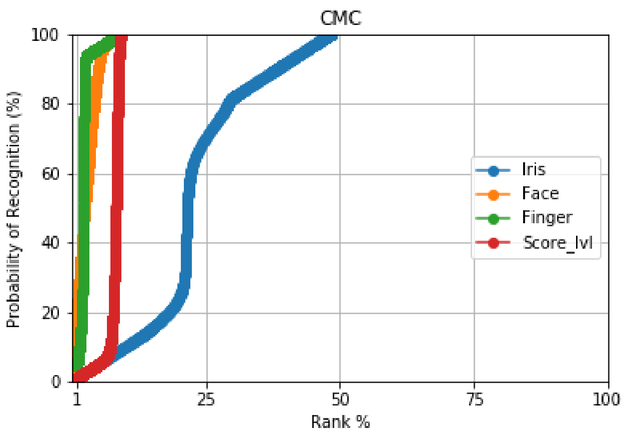


Figure 8.5: The CMC curves summarized into one diagram in full scale for the second experiment using the large-scale datasets.

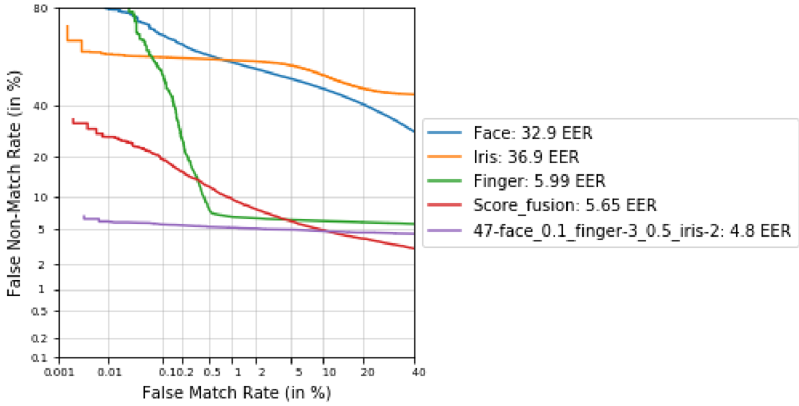


Figure 8.6: This configuration have 10% pre-selection with the face modality on the 1.level. It has 50% pre-selection with the finger modality on 2.level. Final selection was done with the iris modality. It can be seen that this configuration denote the lowest False-non rate at false-match rate 0.01 % and that this configuration denote a 4.8 EER.

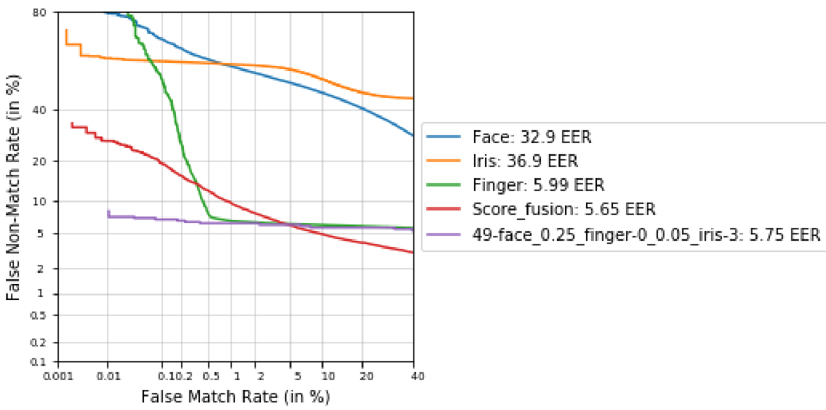


Figure 8.7: This configuration have 10% pre-selection with the face modality on the 1.level. It has 75% pre-selection with the iris modality on 2.level. Final selection was done with the finger modality. It can be seen that this configuration denote the lowest False-non rate at false-match rate 0.01 % and that this configuration denote a 5.75 EER.

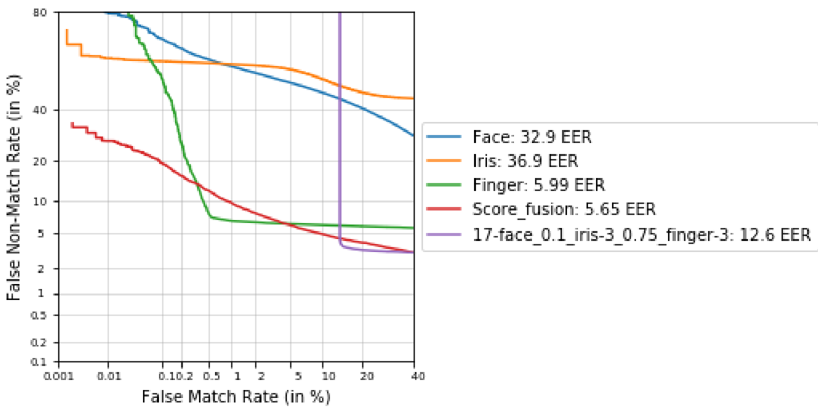


Figure 8.8: This configuration have 10% pre-selection with the face modality on the 1.level. It has 75% pre-selection with the iris modality on 2.level. Final selection was done with the finger modality. It can be seen that this configuration denote the highest False-non rate at false-match rate 0.01 % and that this configuration denote a 12.6 EER.

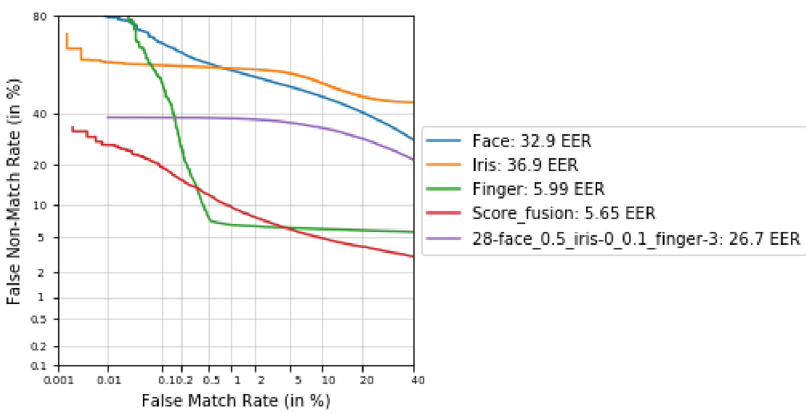


Figure 8.9: This configuration have 50% pre-selection with the face modality on the 1.level. It has 10% pre-selection with the iris modality on 2.level. Final selection was done with the finger modality. It can be seen that this configuration denote the 2nd lowest False-non rate at false-match rate 0.01 % and that this configuration denote a 26.7 EER.

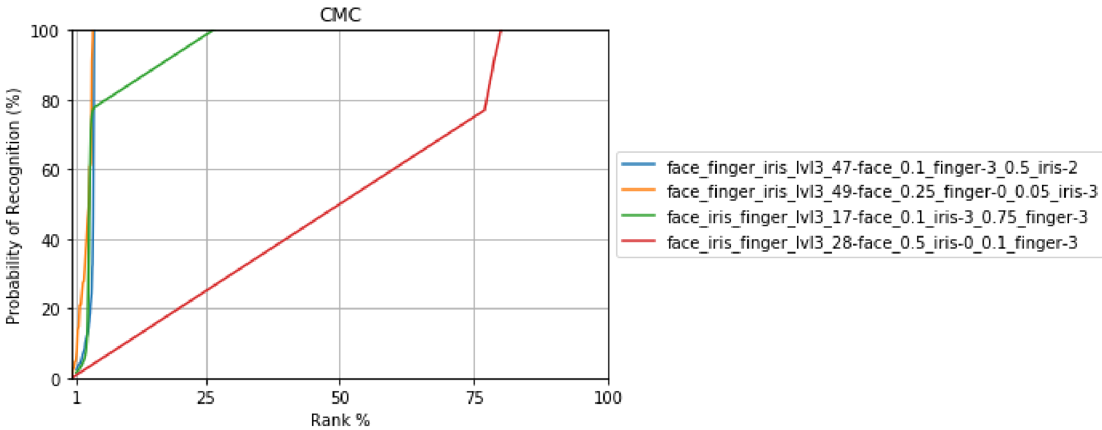


Figure 8.10: This plot present a summary illustration of the CMC curves associated to the configurations with the best/most interesting results.

interesting results in the k-stage experiment on the large-scale datasets presented in this section by their DET plots can be seen in figures 8.10 and 8.11. It is noted that the CMC retains is shape is somewhat met, however, it is noted that loss of genuine identities does have an effection on the retention of CMC curve between baseline and k-stage configuration.

8.5.2 Workload reduction results

This section includes an illustration of the workload reduction of the chosen configurations that denoted the best/most interesting results for the second experiment that was done on the large-scale dataset as seen in figures 8.12, 8.13, 8.14, 8.15.

8.6 Summary of large-scale dataset experiment results

As it can be seen from the DET evaluations there is a clear advantage in terms of accuracy with the highly accurate modalities where the accuracy for the modality is determined by the shape of its baseline DET with a critical criteria of

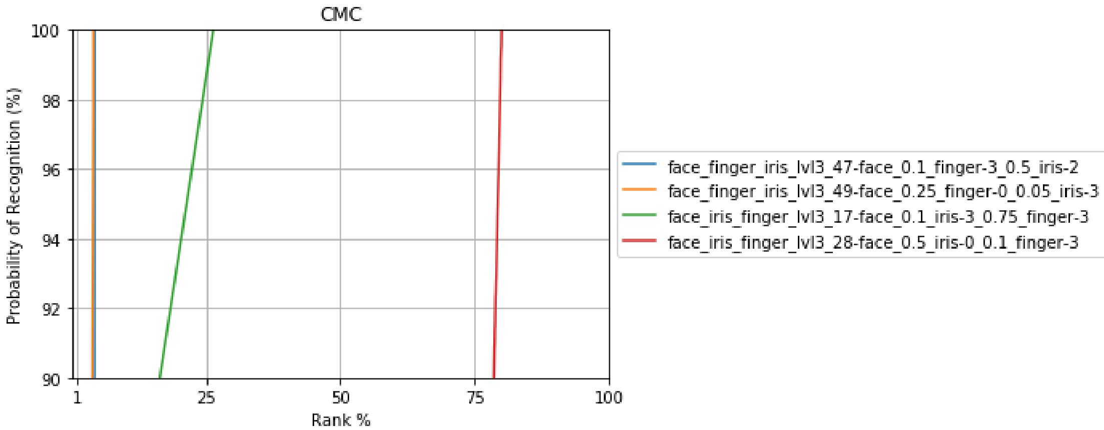


Figure 8.11: This plot present a summary illustration of the CMC curves associated to the configurations with the best/most interesting results on a zoomed scale.

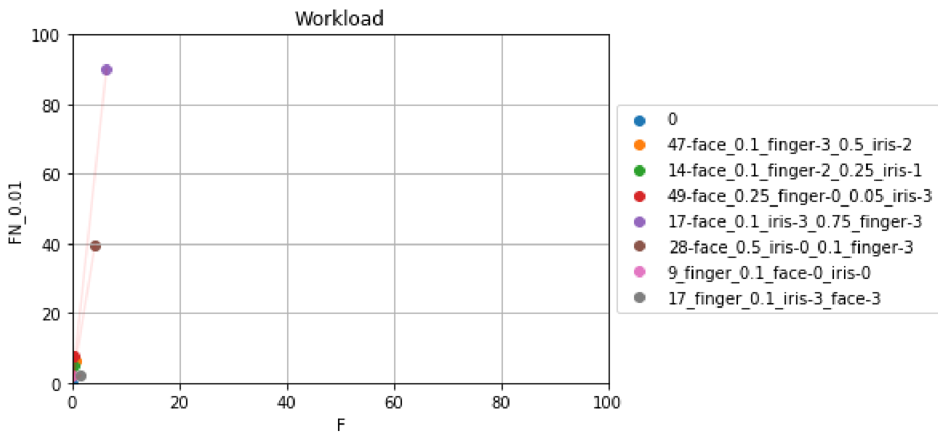


Figure 8.12: The full scale illustration of workload vs. accuracy for k-stage configurations in the large-scale dataset experiment. The false non-match rate at false match-rate 0.01% is used as the accuracy measure in this diagram.

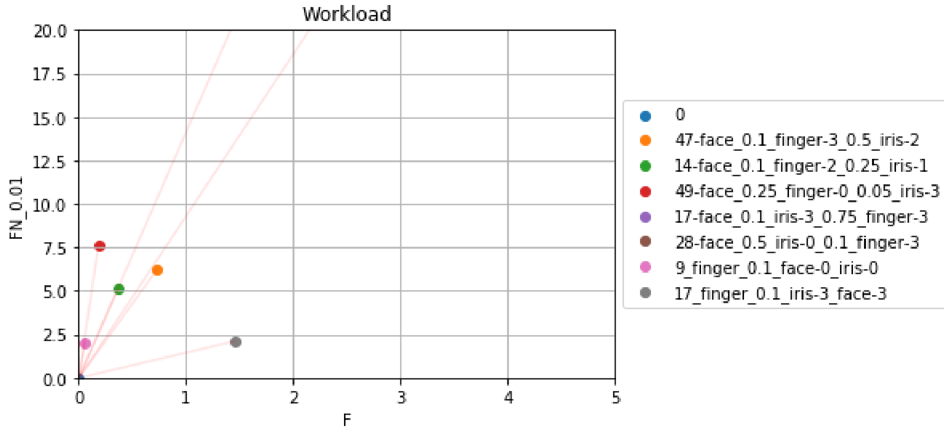


Figure 8.13: The zoomed scale illustration of workload vs. accuracy for k-stage configurations in the large-scale dataset experiment. The false non-match rate at false match-rate 0.01% is used as the accuracy measure in this diagram.

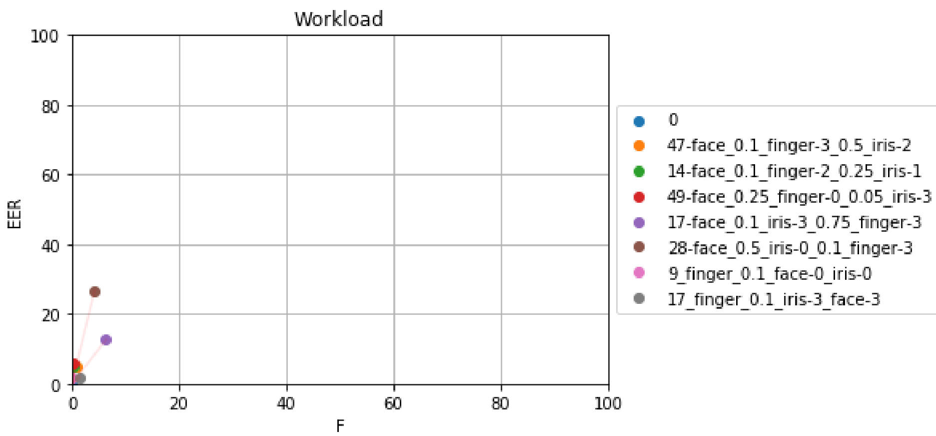


Figure 8.14: The full scale illustration of workload vs. accuracy for k-stage configurations in the large-scale dataset experiment. The EER is used as the accuracy measure in this diagram.

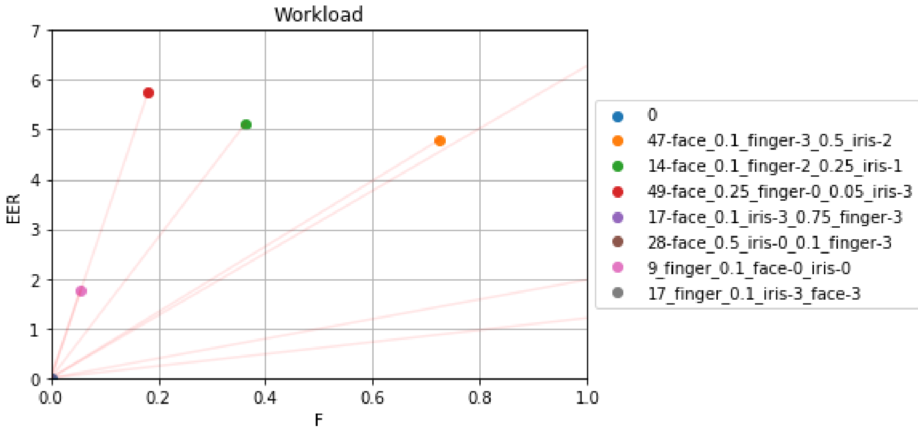


Figure 8.15: The zoomed scale illustration of workload vs. accuracy for k-stage configurations in the large-scale dataset experiment. The EER is used as the accuracy measure in this diagram.

false-non match rate at false-match rate 0.01% denoting the accuracy performance of a modality or any kind of evaluations. Specifically, a tendency showcase that the iris modality which has the lowest error rate in correspondence with the critical criteria denote the best results in terms of accuracy from the DET curves. This is similar and follow the trends of the first k-stage experiment with the highest accurate modality being best for final selection tasks in the k-stage system. Workload and loss of genuines are mostly the same between the two experiments. The important results from the synthetic data in summarized in table 8.2.

8.7 Discussable sources of errors and inconsistencies

Due to the datasets in the case of iris and finger being synthetic there might be some inconsistencies as how the modalities perform in the second experiment as compared to the first most notably with the iris and face modality due to having higher inaccuracies. Additionally, the dataset for the face modality is also different which is another factor. This might be due to some of the software being fine-tuned to work optimally the given dataset, specifically, in the case of the face modality the pre-trained model for the neural network has been tuned to the LFW dataset. Another indignation of the face modality might have been that

1. lvl	2.lvl	FS	EER	W(F)	W(τ)	L(G)
Face 10%	Finger 50%	Iris	4.8 %	0.73%	5.3	11.84%
Face 10%	Finger 25%	Iris	5.12 %	0.36%	4.1	4.37%
Face 10%	Finger 5%	Iris	5.75%	0.18%	6.6	0.70%
Face 10%	Iris 75%	Finger	12.6 %	6.37%	89.2	42.06%
Face 50%	Iris 10 %	Finger	26.7 %	4.25%	38.7	12.81%
Iris 10%	Finger 50%	Face	28.5%	0.03%	26.5	23.53 %
Finger 10%	...	Face	1.75%	0.05%	1.0	24.32%
Finger 10%	...	Iris	1.75%	1.45%	1.8	24.32%
Iris 10%	...	Face	2.98%	0.05%	28	24.17%

Table 8.2: Table of some of the best/most interesting configurations with their results from the k-stage experiment denoting their parameters with modality and pre-selection for 1. level pre-selection (***1.lvl***), 2. level pre-selection (***2.lvl***) and Final selection (***FS***). Furthermore, along with the configurations' parameters their associated results is denoted for EER, reduced workload (***W(F)***), the denoted distance τ (***W(τ)***) from reduced workload against accuracy as True-negative rate at 0.01 % and loss of genuine identities (***L(G)***).

the neural network wasn't allowed to run long enough training for the validation on the much larger database as compared to the first experiment. The Osiris system was also circumvented in the second experiment to save computation time which might have an effect.

The sheer size of the dataset of course also plays a factor. Most noteworthy is the choice to not exhaust every possible combination of possible match candidates due to effects of practicality which can be addressed in future works.

However, the most important aspect is the relational performances of the modalities as denoted by DET and CMC. This help establish how the modalities relate to one another as can be predicted by the evaluation and how the score-level fusion systems as baselines perform as compared to the k-stage level systems which helps further develop the overall model for evaluating large-scale biometric information fusion systems and how it relates to a k-stage system that aim to improve accuracy and efficiency of the system. This is based on the condition that the evaluation is equitable for each modality in the large-scale k-stage experiment.

Chapter 9

Discussion

This chapter moves unto the analysis of the results gained from the k-stage experiments as compared to the baseline evaluations and the other theories discussed in this project. Thereafter, a broader perspective is applied to the analysis of the results and the project as a whole with the addition of the future works which is all utilized for a summing conclusion on the whole project.

9.1 K-stage System Analysis

This section is the analysis on some of the more interesting results from the k-stage system experiments including recognized tendencies and interesting findings. It is noted that the results of the k-stage experiments is all in comparison to their baseline, with specific focus on the performance of the k-stage fusion compared to the baseline score-level fusion. The baseline score-level fusion is relatively highly accurate but workload heavy as it include a full search of a combination of all templates. The assumption is that the k-stage system has reduced the workload significantly compared to the score-level fusion workload while retaining comparable/better accuracy as denoted by DET and relatively small-to-none loss of real identities. The CMC curves for those configurations indicate the recognition rate of the remaining database for evaluating purposes of validation, further analysis and/or further computations on the k-stage short-list of the full database. The assumption is that the CMC curves retain their shape between the evaluations of the baselines and the k-stage configurations which has been mostly actuated through the first and second k-stage experiment, however, deviations might occur as an effect of loss of genuine identities.

9.1.1 Analysis approach

The strategy for the k-stage system experiment was to evaluate all possible orderings of the three modalities finger, face and fingerprint with choosing the thresholds intelligently based of baseline CMC evaluations instead of brute-forcing every possible combinations of selected pre-selection sizes. The selected pre-selection sizes are established as three lists (for each modality) of 4 number values where the values are the percentage of the database that will be selected (i.e. pre-selection). This way it is possible to discard non-match candidates through setting a limit range list by the index on the sorted database of match candidates using the rank from the CMC evaluations for the given modality. The values themselves are absolute values selected by evaluating the CMC curves from the baselines for each modality. The evaluation script was set up in a way where it would exhaust every combination of thresholds for each possible ordering combination of modalities.

9.1.1.1 Final Selection

Evaluation of the k-stage system configurations consist simply of a DET and CMC evaluation to see what accuracy (indicated by its related EER score and False non-match rate at false match rate at 0.01 % also denoted as $FN_{0.01}$) and rank the k-stage filtered database, after going through levels of pre-selection, denotes which is all gained by the evaluations curves. Of significance is the shape of the DET curves, specifically with the criteria of high accuracy at a low false-match rate threshold or other error rate measures such as EER. This will denote the appropriate modality for final selection as those tasks are highly dependent on high accuracy for good performance whilst discarding only present a risk of possible removal of genuine identities.

9.1.1.2 Pre-selection

Each match candidate consist of three match scores e.g. Finger-x, Face-y and Iris-z. The idea for each ordering of modalities is that the the two first modalities is used for pre-selection in multiple different levels whereafter the final unused modality is used for the selection of match candidates (i.e. final decision/final selection) associated to the selection tasks (i.e. DET, CMC and workload) in the evaluation of baselines for the modalities individually (see figure 9.1).

For example, one of the configurations started with the scores for the finger modality with threshold 10% used to discard 90 % of the match candidates, and of the database that is left after that discarding (i.e. level-1 shortlist) another discarding using the face modality at threshold 40% meaning another 60% of match candidates is discarded from the 10% remaining database from the first discarding (i.e. level-2 shortlist), whereafter a selection (i.e. final decision) of genuine and imposter matches is selected using the iris modality on the remaining scores for match-candidates.

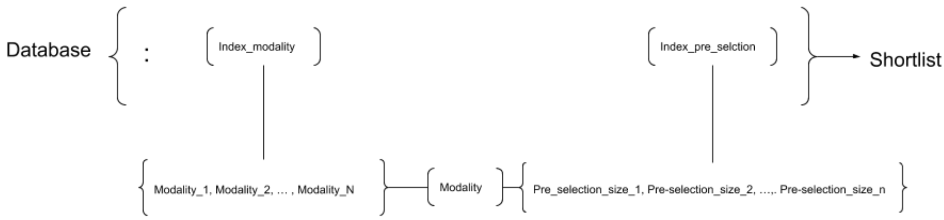


Figure 9.1: Example that showcase the process of a pre-selection

The settings for the pre-selection sizes for each configuration denotes how much of the database is kept (i.e. shortlist) and thus how many of the match candidates is discarded. As these percentages are arbitrary they are applied as an index on the database in the form of a dictionary where keys have the label of match candidate i.e. match scores based on two given candidates that are different (also known as non-match candidates/impostor scores) or themselves (also known as match candidates/genuine scores) and values which is a triple pair list/tuple of scores where each score is for the finger, face, iris modality respectively. Sorting (by best) the list of score for the given modality and then setting the range via indices by the chosen pre-selection size discards the number of match candidates chosen in the given configurations.

The CMC ranks indicate what size of pre-selection is appropriate for each modality by what kind of CMC curve each configuration denotes. As it was learned the modality with higher rank-1% scores allowed a sizable pre-selection without losing significant accuracy denoted by loss of true-positives. Similarly, the assumption the higher the lower rank scores the better performance will be garnered for the configurations with smaller pre-selection sizes in terms of high recognition rate at higher density areas inside a 99-100% confidence interval for recognition rate was met.

The specifics of the modalities differ between experiment, however, the inter-relational behaviour is still the same in regards to the pre-selections.

9.1.1.3 Reduction of identification attempts

The number of biometric identification decisions (number of subject comparisons/ biometric identification attempts) with the k-stage system is significantly reduced from the the 1:N exhaustive search scenario. For example, in the context of the first experiment conducted in this project it is noted that in the exhaustive scenario it is necessary to do 331884 possible identification attempts with the 2000 subject experiment. Generally, as computational speed varies from machine to machine, it is more reasonable to compare the k-stage system denoted shortened list of match candidates (i.e. shortlist of match-candidates) computational/comparison speed as a bit-wise/pair-wise factor in relation to the number of identification attempts in the 1:N scenario i.e. exhaustive 1:N vs. 1:(N-N+Shortlist).

As an example, a configuration from the first experiment with pre-selection sizes finger=10%, iris=50% and final selection with face the N sized database of match candidates has been reduced by 95% or conversely 0.05 % of the database is pre-selected for the shortlist of match candidates over two levels with two different modalities at each level. This means that the shortlist from k-stage system needs 16594 possible identification attempts compared to the 331884 identification attempts possibly required from the exhaustive search scenario with the 2000 subject experiments, whilst that configuration denotes comparable accuracy score given by its DET evaluation which is lower than the baselines (see figure 7.4). This is a significant increase in comparison speed as the number of identification attempts needed have been reduced to a fraction of its original size whilst keeping a relatively comparable accuracy as denoted by the DET evaluations and their associated accuracy score (e.g. EER scores or false non-match rate at false match rate at 0.01%). Furthermore, the rank of this configuration is used for validation by comparing whether the CMC has remained consistent between baseline and k-stage fusion configuration. The ranks can also be used to impose further optimization or provide a work-space spectrum which can be used for great variety of tasks in the field of identification scenarios.

The second experiment showcased similar tendencies for a larger scale. The number of biometric identification attempts for the large scale is increased i.e. 832803 biometric identification attempts, however, it could be larger, but due to modifications to promote practicality the size wasn't as big as it could be i.e. a subset of non-match candidates were generated instead of exhaustively due to computer processing time. Nonetheless, comparable tendencies in regards to the reduction of biometric identification attempts could be identified.

9.1.2 Analysis Findings

The overall assumption is that the k-stage systems remove false-matches whilst retaining the genuine matches.

In summary, there are some orderings of modalities from the k-stage system experiment that denote better results which is noticeable when there is a significant difference between the modalities denoted by their baseline evaluations. The baselines provides some indication of how various orderings may effect the final outcome. Unsurprisingly, higher larger databases denotes better results, however, certain combinations of smaller thresholds denoting smaller databases also works. An interesting tendency was that better results could be gained from starting with smaller pre-selection sizes on the first levels and larger thresholds on the later levels i.e. the smaller the database become the larger the pre-selection size becomes. Interesting is the results showing that a too large database creates a lot of false-positives due the sheer amount of non-match candidates apparent in the database. The CMC also seems to be corresponding to the tendencies denoted in the DET plots. More notably, the configurations' corresponding CMC evaluation functions is utilized as a validation by whether they have maintained their shape between baseline and k-stage experiment. Throughout the evaluation of different configurations there is certain tendencies that could be identified such as which orderings yielded great or bad results for the k-stage system. For example, in the first experiment with 2000 subjects it could be identified which orderings denoted quality outcomes e.g.:

- Good: finger-iris-face
- Medium: finger-face-iris, iris-face-finger
- Bad: Every other ordering denotes mostly incomparable results in relation to the baselines

The second experiment related different performances in terms of ordering, however, the effects of the baselines for which orderings denote good and bad results are still the same. Essentially, the baseline evaluations provide the information required to determine good orderings in terms of good results denoted by accuracy vs. efficiency.

9.1.2.1 Accuracy performance

Some interesting outcomes where that some configurations that were virtually similar deviated a lot in accuracy performance exemplified by such configurations as figure 7.6 that is virtually the same as figure 7.5, however, due to the difference of pre-selection size being smaller at each level (more discarding) for different modalities there is a visual difference in the occurrence of false matches and subsequently the overall accuracy. This showcase how pre-selection on a multi-level system can affect the final outcome as opposed to a single level system selection.

9.1.2.2 Efficiency performance

The pair-wise workload (denoted by biometric identification attempts) also has an affect on the outcome and overall evaluation. This can be exemplified in the results from the first experiment denoted in figures 7.4 and 7.7. These two configurations share the same orderings but varies in terms of pre-selection sizes. It is worth noting that a larger amount of comparisons doesn't equate higher accuracy which is likely caused by the amount of imposters that follows with a larger database. The results from one of the configurations denoted in figure 7.4 showcase how large scale reduction on the first level and smaller scale reduction on the second level can reduce the number of identification attempts by 95% while retaining high accuracy as denoted by its associated DET and CMC. However, increasing the number of identification attempts too much as can be seen in figure 7.7 have a negative effect on the accuracy and have an affect the CMC spectrum which might be caused by the inclusion of many impostors in that configuration due to the sheer size of pair-wise workload. This finding indicates that there is a turning point for each ordering of modalities where the number of comparisons affects the accuracy negatively. Conversely, there are also examples where too much reduction affects accuracy negatively such as with the configuration with some of the other orderings in the first experiment.

The actual workload reduction of each configuration was also denoted to evaluate which configurations might be most advantageous in terms of efficiency. The metric is affected by a factor of accuracy (true-positives or correctly matched genuine) and a factor of number of identification attempts/ comparisons i.e.

reduction percentage configuration denoted in regards to total database size. Therefore, in some cases the best workload reduction denotes the higher accuracy percentage of genuine matches and the lower ratios of number of identification atttempts in terms of percentage of total database kept, ultimately, resulting in smaller distance to the optimal operating point τ. An example of a poor performing configuration is the configuration from the first experiment with the ordering of iris-face-finger that has significantly worse performance in terms of accuracy and workload due to the modality of selection (finger) and the modality of (larger) reduction in that case. Similar tendencies for good configuration and bad configurations in terms of pair-wise workload and accuracy can be found within samples from the second k-stage experiment.

9.1.2.3 Risk of data loss

Additionally, a quantified computation for each configuration have been included in k-stage experimentation analysis to denote how many genuine probes were discarded by the k-stage system as a percentage loss. This shows how much loss of genuine matches was made by the k-stage system as a percentage compared to the total amount of genuine matches in the test database. This loss should be kept at a minimum (optimally at 0%). The trade-off between loss of genuine and reduction of comparisons is an interesting factor of the k-stage system as depending on the task at hand it might make sense to utilize a high reduction which can come at a cost of high loss e.g. the result from figure 7.5 has a balanced reductions between levels starting with a large reduction in the first level and a smaller one on the second which denotes a number of comparisons that equates to 13% of the total search of the system with a 2.88% loss of genuine identities.

9.1.3 Specific findings of experiment(s)

The specific findings vary greatly between experiment but they are highly correlated to their respective baseline showcasing the inter-relational effects of fusion in each case which showcase similar tendencies.

For the first experiment, it can be seen that the ordering of finger-iris-face denotes the best performance but within certain thresholds of shortlist sizes at a minimal loss of real matches, while finger-face-iris denotes comparable results in terms of accuracy, however, the finger-iris-face is faster due to a higher level of reduced identification attempts necessary at a minimal loss of real matches. The other orderings denote relatively incomparable results rendering them unusable for any substantial task in an identification scenario. These evaluations showcase the prediction that are possible when the associated baseline is obtained. Similar inter-relational tendencies could be identified in the second experiment where the iris modality proved most suitable for final decision as it had the lowest false non-match rate at false-match rate 0.01% compared to the other modalities.

It is apparent that without the maximum ranks denoted by CMC it is unavoidable to get some loss when trying to discard non-candidates. Furthermore, a larger selection of match-candidates does indeed mean that there is a better chance of not having discarded any genuine matches, however, it obviously also means that a greater portion on non-match candidates might falsely be accepted even with the utilization of the maximum ranks.

The performance of the different configurations can be predicted by the baselines established earlier for the exhaustive search 1:N identification scenario. And as could be anticipated by the baselines the k-stage configurations denote comparable results when the rules for pre-selection and final decision denoted by DET and CMC is upheld. Furthermore, the factor of workload can be incorporated as an evaluation factor to be considered when contemplating accuracy vs. efficiency.

9.1.4 Summary of Analysis

From the results and subsequent analysis, it is possible to draw some conclusions regarding the inter-relations of the different configurations, and in a larger perspective the overall fusion of the information system as it relative to the k-stage system. In summation, the key findings indicate:

- The best performing modality in terms of DET as it relates to a criteria (e.g . False-negative rate at 0.01 % false-positive rate or EER) is most accurate in terms of selection. Conversely, it can be argued that the least accurate modality is good for 1.level pre-selection but this is also dependent on factors such as workload and recognition rate.
- Pre-selection is directly correlated to CMC where reducing the rank based on 99% - 100% recognition rate range will reduce the workload significantly.
- Reducing with highly accurate modalities will retain most genuines i.e. reduce the loss of genuines
- For a hierarchical multi-stage system such as the one in this project, it is

essential to include a highly accurate, in terms of DET, to have comparable results from a k-stage system

- The assumption that the k-stage system removes false match-candidates while reducing the workload significantly, resulting in comparable (or better) performance in terms of accuracy and significantly more efficient (reduced workload), was showcased in configurations that follows the the prerequisites in the model i.e. accurate modality for selection and pre-selection correlated to CMC ranks curves.
- Workload reduction is highly dependent on the template sizes of the modality so if possible it is always a good idea to remove the larger template modalities on the lowest levels before the selection.

These findings will help put the experiment and its results into a broader perspective and develop a useful for making critical conclusion on a large biometric information fusion system as it relates to the k-stage system.

9.2 Perspective on abstract aspect of k-stage experiment results

Through the two experiments showcase different results in terms of the specifics for the performance of each individual modality, the two experiments showcased similar patterns in terms of inter-relational behavior between baselines and k-stage system as denoted by the used evaluation techniques i.e. DET, CMC, workload etc. This enables the possibility of drawing general conclusions regarding the k-stage system and information fusion on large scale, without specifying the modalities individual performance and, rather, specifying the inter-relations and behaviour. Essentially, it is possible to draw conclusions regarding behaviours of the k-stage system based on baseline evaluations but the specific performance for the baselines may vary depending on dataset, software among other things.

The standing question is whether it is possible to build a model which will predict the good k-stage configurations based on evaluations (i.e. CMC/DET etc.). The answers to those questions is found by looking at the predictions and assumptions made regarding the k-stage system, specifically, after obtaining information from the baselines and compare those expectations to the actual outcome through the k-stage experiments. As it stands it is quite possible to predict good orderings based on DET curves and great pre-selection range based on the CMC curves. The metric for workload reduction seems quite accurate as it correlates highly with the configurations' parameters as all of these outcomes are closely associated with each other. Furthermore, the workload metric is a great utility for illustrations of the diagrams denoting workload vs. accuracy (as EER or $FN_{0.01}$) as a representation of accuracy vs. efficiency trade-offs from different k-stage configurations. However, there are some deviations as could be expected with every experiment for the individual evaluations but if all evaluations are combined in the final analysis it is clear what patterns can be identified i.e. a pre-selection with high recognition rate within a confidence interval, with a highly accurate modality in terms of DET and with a high workload reduction denote the best results in terms of accuracy vs. efficiency. These results relies on the prefix that the conditions are equatable between each modality. By all those notations it is possible to predict the outcomes of a k-stage system if every evaluation method is combined together for the analysis that identifies a pattern.

The specificities of how the modalities should be ordered and the sizes of the pre-selections at each level for the different modalities can vary vastly from experiment to experiment. The key lies in the datasets and software used to extract the features from the biometric sources and compute the comparison scores for possible match candidates along with plausible other factors such as operating system (OS). Therefore, the important conclusions to draw from the experiment in this project is the the effect of the methodology for the experimentation and the subsequent analysis approach to that experiment.

9.2.1 Effects of Accuracy

A baseline had to be established to verify the performance of the k-stage system where the basic score-level fusion proved especially useful for comparative purposes as it relate to the overall purposes of this project which is to find a intelligent way of applying fusion to large-scale biometric information system since fusion has proven to be more accurate as denoted by other works and this project. It was learned from the DET evaluations of the baselines for the individual modalities and the basic fusion methods (decision-level fusion with majority voting and score-level fusion with min-max normalization) that the biometric modality that denote the best performance in terms of accuracy, where accuracy is defined by lowest false non-match rate against false match rate at 0.01% ratio (denoted as $FN_{0.01}$) or best EER (given that the curves generally follows a similar pattern), indicate the best modality to used for the final selection in the k-stage system. The CMC evaluations plays the major factor in choosing the modality for large scale pre-selection as it can be recognized that baseline evaluations with high recognition rate for low rank evaluations for a given modality is suitable for large-scale reduction.

9.2.2 Effects of Reduction

In association with the DET evaluations, it was learned from the CMC evaluations that the pre-selection sizes denoting accurate results can be accurately determined using the CMC curves from the baselines. In this case it is most sensible to evaluate the complete spectrum of significant ranks in relation to the CMC curves for the baselines in the k-stage system as the aspect of significant factors of a configuration can be dependent on the selection task at hand. Regardless, it was learned that the accuracy of a system denoted by DET was highly correlated to the pre-selection which were directly correlated to the CMC evaluations of the baselines meaning that pre-selection sizes based on ranks with high (99-100%) recognition rate were accurate and vice versa. With the objective of removing as much as possible without compromising accuracy it is therefore desired to have extremely high recognition rate for low ranks. Furthermore, it was learned that higher recognition rate denoted more sparse variance between the ranks, hence, the 99-100% spectrum was applied as the sparseness could result in ranks differentiating in over 50% of the entire database size in ranks for 1% differentiation in recognition rate. The CMC for the k-stage system configurations may indicate how good the given configuration for recognizing genuine and imposter candidates is and if it can be validated that the shape of the CMC has been maintained between baseline and k-stage configuration. Additionally, k-stage configuration CMC may also indicate if the system may be reduced further without major consequence, so the system with minimum rank of 100% score and 100% recognition rate creating a 100% efficient system.

9.2.3 Effects of Efficiency

Besides the general factors for accuracy in terms of rank and recognition rate, the workload of each configuration is a key evaluation factor. There are several presentations of workload and the related workload reduction in this experiment. It is noteworthy that each configurations workload is independent from the accuracy as it is associated with number of comparisons (identification attempts) and template sizes, however, in a broader context it is often interesting to have the factors of a system denoting accuracy in terms of recognition rate as it stands to rank against the workload of a given configuration since what is weighted more as important vary from task to task i.e. efficiency vs. accuracy. The workload is presented both from a comparative (pair-wise) and bit-wise aspect in these experiments. Furthermore, it was also evaluated with workload as a factor against accuracy where accuracy could be $FN_{0.01}$ or EER rate. In the case of comparative (pair-wise) workload reduction it is fairly straightforward to anticipate configuration performance in terms of workload reduction against accuracy by simply associating it with the CMC evaluation. On the other hand, the bit-wise workload reduction against accuracy evaluation is more dependent on the ordering of modalities, in accordance with the pre-selection sizes based on CMC, as template sizes vary dramatically for the different given modality. Therefore, the spectrum denoting workload against accuracy present a new interpretation of the configurations where configurations can be chosen by what is weighted more for the given tasks.

Additionally, to provide a complete overview of the performance of each configurations it is appropriate to denote the loss of genuine scores for each configurations which shouldn't optimally happen but can be unavoidable depending on the configuration as that is useful information for any use of a given configuration.

9.2.4 Combining Effects - Summary

The assumptions for a k-stage system can be achieved by using this evaluation technique to anticipate the outcomes of a k-stage system configuration and set the parameters accordingly. Subsequently, evaluating the outcomes as denoted throughout the discussion and results sections will provide an overview and specific results that help draw conclusion on whether the given anticipation had been successfully achieved.

9.3 Proposed Evaluation Model

Throughout the results and subsequent analysis it is possible to establish a model for evaluation of large scale information fusion systems in biometrics. In simple terms:

1. Baseline evaluations
2. Select modality of final selection based on DET curve with a criteria for EER or False-negative rate at a 0.01% False-positive rate.
3. Select the lesser accurate modality for 1.level pre-selection in combination with workload of given template.

4. Use pre-selection size denoted by CMC curve of given modality where rank is in a 99%-100% recognition rate (rank size can be very sparse in this interval).
5. Pre-select a shortlist of best scores based on previous established parameters
6. Pre-selecti shortlist from previous shortlist using another modality based on same parameters
7. Evaluate k-stage filtered shortlist of database using the selection modality

These steps/model supposedly denote the best configuration in terms of performance when it comes to accuracy vs. efficiency i.e. low error rate and low workload. A conceptualization of the model is illustrated in figure 9.2.

It can be considered to remove modalities based on their workload reduction potential is combination with their accuracy, but this project's results have showed a somewhat worse performance in terms of accuracy as denoted by DET with the removal of levels/modalities.

9.4 Future Works

It has been shown that it is possible to create a model for a large-scale multimodal information fusion biometric system using biometric evaluations such as DET, CMC, Workload reduction etc. as presented in this project. That model uses a combination of all evaluations to determine ordering and pre-selection sizes in a information fusion biometric system in terms of best performance in an accuracy vs. efficiency context.

Moving unto even larger scales are another possibility in future works of this project, where the utilization of high-powered computers/machines would be a necessity. The obvious addition is to adding more biometric (different) characteristics and even more combinations of pre-selection sizes on a broader scales based on baseline evaluations to expand the investigation model for k-stage system configuration. These works will only establish for those specific cases but the model will fundamentally be the same. Another possibility is to add more levels to the multi-stage k-stage filter. The model itself can be further developed by adding more evaluation methodologies that are conventional for biometric works and even some that may not be considered conventional in the broader scale of biometrics.

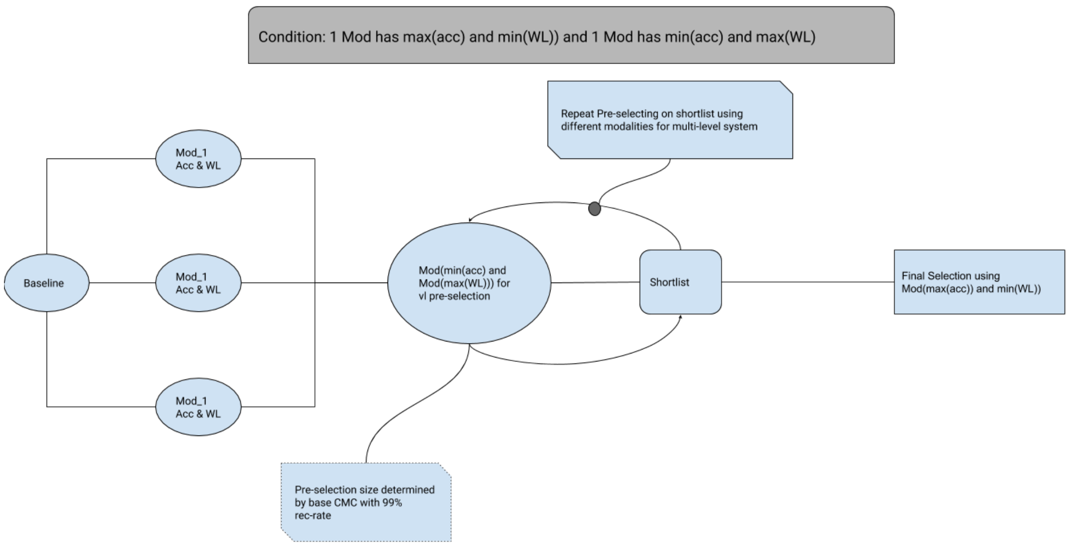


Figure 9.2: The key abbreviations used in the figure denotes acc(accuracy), Mod(modality), WL(workload), lvl(level) and rec-rate(recognition rate for CMC). It is important to note that this model is appropriate for getting the best configuration in terms of accuracy and efficiency in the k-stage system, however, the choices are relevant to the given condition in the title. If the condition is not met, depending the context in which the k-stage system is used the factors of accuracy, efficiency and loss of genuines might be weighted differently when it comes to choosing the modality for the given model. Generally, a highly accurate modality for the final selection is prioritized while effects of workload can be debatable as they do not only depend on the pre-selection sizes but the size of the templates for the given modality and factors of accuracy vs. efficiency may depend on the task at hand. Another context dependent factor that could be considered would be the loss of genuines.

Chapter 10

Conclusion

The objective of this project was to investigate information-fusion in large-scale biometric system and evaluate its effect on efficiency vs. accuracy. Furthermore, this project suggested and evaluated a multi-stage hierarchical system called the k-stage system which pre-selected shortlist based on the previous level's shortlist (or whole database on the first level) of match candidates across multiple levels using different modalities. An approach/model to evaluate such a system is presented in this project with key evaluation techniques such as DET, CMC and workload reduction is presented and discussed. Two experiments where conducted of different scales to evaluate the identified patterns. These experiments denoted results that showcased the spectrum of effects on a biometric system when applying information fusion and, specifically, the k-stage system in terms of accuracy and efficiency when compared to an established baselines (also discoursed in this project) and, subsequently, the associated trade-offs.

The experiments on the k-stage system implied that the criteria for a configuration with good accuracy requires a highly accurate modality utilized for the final selection denoted by evaluations from DET, CMC etc. The pre-selection has to be highly correlated to the CMC that could be established for the given modalities' baseline. Obviously, the best workload reduction can be found by modalities that have smaller templates (in terms of bit-size) and fewer necessary identification attempts as denoted by lower pre-selection sizes which is an effect of CMC with high accuracy/recognition rate at a lower rank. The inter-relating effects of these can overlap e.g. biggest workload reduction is also highest accurate. The assumption that the k-stage system with properly adjusted parameters removes non-match candidates while retaining its accuracy or improving it was also met for some configuration. This was associated with configurations that denoted highly accurate results while having low workload. There was some other factors presented that could be of interest for any evaluation task with the k-stage system which was the effect of distribution or loss of genuine identities. The sensitivities will of course vary depending on the task the system is used, however, the objective of the project was to denote a spectrum to help evaluate trade-offs which has been done and it has been established that configurations following a certain pattern denote the best results as termed by accuracy vs. efficiency.

The concluding model has eliminated a lot of ambiguity regarding the system via experimentation and analysis establishing patterns. However, there are still some ambiguity that was especially introduced with the second experiment but that mostly relate to the specificities of the used modalities which isn't the focus of this project. Regardless, the future for this topic should aim to test even more pre-selection combinations and modalities to specify the model to a modality based level.

In summation, the project has successfully discoursed a spectrum of trade-offs for information fusion in large-scale biometric systems and established a model that denote how to evaluate and gain the best performing configurations for the proposed k-stage system in terms of efficiency vs. accuracy.

Appendices

APPENDIX A

Related works

This section include some of the more in-depth investigations of the surveyed related works on a specific level. In the report the fields of fusion and workload approaches is discussed in general, because it is not possible to provide a detailed survey on all fusion approaches and workload reduction methods for face/fingerprint/iris. Therefore, in this section some of the surveyed works that are associated with this project but not as relevant as the works included in the report itself is included.

A.1 Various modality information fusion - state of the art

There are quite a few fusion techniques proposed by various studies utilizing different biometric sources and analyses techniques for fast and accurate identification such as:

- For fingerprint, pre-classifying the modality into three standard types i.e. whorls, loops or arches. Then, each search can be restricted to about one-third of the full-size effectively reducing the computational cost[Dau00].

- A two-stage recognition system utilized for the iris data. A short-list of match-candidates is generated from a reduced iris code representation in the first stage, thereafter, that short-list is matched to the full-length representation of the iris code in the second stage[GRC09].

- For iris biometric data, an Iris indexing scheme based on Bloom filters and binary search trees. With a statistical model, the system is shown to be theoretically scalable for arbitrarily many enrollees [DRB18a].

- For face, there is work that focuses on feature level fusion which also incorporates the hand modalitiy. The work denotes three techniques which is fusion of PCA (principel component analysis) and LDA (linear discriminant analysis) coefficients of face; fusion of LDA coefficients corresponding to the (R,G,B) channels of a face image; fusion of face and hand modalities [RG05].

- Embedding score coherence in the fusion process using static weights for different biometric sources as a single representation of the fused information [DRBK17].

Other more well known techniques have also been proposed by several studies such as: simple-sum, min-score, max-score, mather weighting and user weighting[DRBK17][ARJ06][Ros07].

A.1.1 Gaussian distribution Score-level fusion

Another paper approaches score level fusion with a model of the joint distribution of all scores by a (semiparametric) Gaussian copula model, with the resulting correlation matrix, subsquently, being structured. The correlation matrix from this showed results that had many zeros and many correlations having a common value. It studies semiparametric estimation of constrained euclidean parameters where the restrictions are divided into two cases: the parameter has to be in the image of a continuously differentiable function of a lower dimensional parameter and the parameter has to belong to the zero set of a continuously differentiable function of the parameter. From a biometric perspective it also proposes a semiparametric likelihood ratio-based score level fusion strategy by modelling the marginal individual likelihood ratios non-parametrically and the dependence between them by parametric copulas. The dependence parameter is estimated by pseudo-likelihood estimation with discourse regarding its convergence[Sus16].

A.2 Empirical cross-entropy

Work has been done to derive empirical cross-entropy (ECE) as a measure of accuracy of a forensic speaker recognition systems according to other equivalent measures such as U_{log} or normalized cross-entropy (NCE). ECE has been used in a verification scenario where it can be interpreted as the average information needed by the fact finder, over cases and after evidence analysis in order to know whether the recovered and control speech samples come from the same source or not. ECE also measures the information loss due to non-perfect likelihood rations (LR) calibration which allowed ECE plot as a representation which presents average information supplied by evidence analysis in court with a clear separation role. Thus in turn allows transparent reporting of the performance of the system in terms of such information theoretical magnitude [RGR18]. The ECE could be inspiration for an alternative evaluation methodology for the match-scores.

Appendix B

First Experiment Results

This Appendix includes some more results from the second and first experiment. Over thousands of configurations and, subsequently, results where gained for each experiment so a two step vetting was conducted to chose the best/most interesting results to highlight in the report and to include here in the appendix.

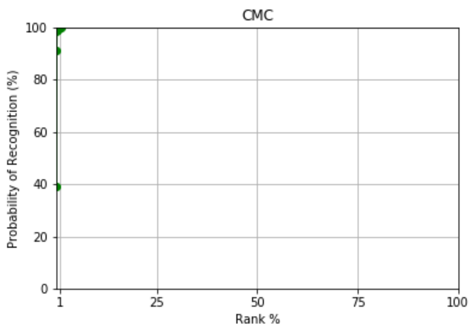
B.1 Pre-selection sizes for first experiment

A lot of different configurations for the k-stage system were evaluated. All results are included in this appendix section.

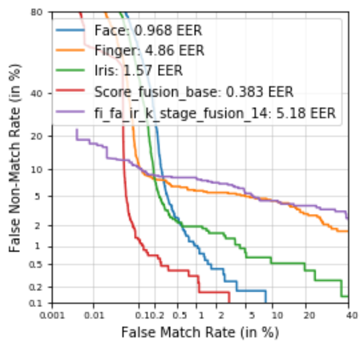
Based on the baseline CMC, threshold index (index start at 0) for the threshold combinations:
$pre\_selection\_iris = [0.1, 0.25, 0.5, 0.75]$
$pre\_selection\_finger = [0.1, 0.25, 0.5, 0.95]$
$pre\_selection\_face = [0.1, 0.2, 0.35, 0.6]$

B.2 Accuracy

This section include a plethora of results gained from the first experiment highlighting the variety and the discernible patterns that could be found between correlated configurations i.e. configurations of specific orderings and pre-selection sizes. The labelling for each configuration refers to the modalities finger (fi), face (fa), iris (ir) and index values in parenthesis ordered accordingly to the ordering of modalities (x_1, x_2, x_3) referring back to the pre-selection size lists going by indexing a typical array. Note that the last pre-selection size index can be ignored since that pre-selection doesn't take place due to that being the final selection level.



(a) CMC for finger-face-iris ordering and threshold combination (1,2,1)



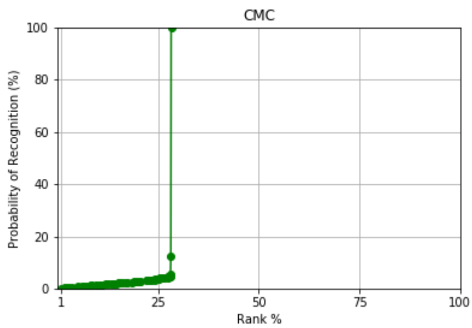
(b) DET for finger-face-iris ordering and threshold combination (1,2,1)

Figure B.1: K-stage fusion with finger-face-iris and threshold combination (1,2,1) denoting a medium performance

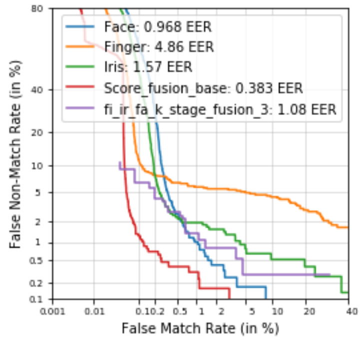
With finger-face-iris ordering and threshold ordering combination (1,2,1), the comparative DET error rate in concordance with each singular modality and the basic score fusion base. The CMC shows a relative low rank probability, however, the first rank is comparatively low at approximately 40% (see figure B.1).

With finger-iris-face ordering and threshold ordering combination (0, 1, 0), the comparative DET error rate in concordance with each singular modality and the basic score fusion base. The CMC shows a relative low 1.rank probability, however, it increases dramatically at around the 25% rank (see figure B.2).

Similar tendencies between the figures. With their ordreing and threshold ordering combination, the comparative DET error rate in concordance with each singular modality and the basic score fusion base. The CMC shows a relative

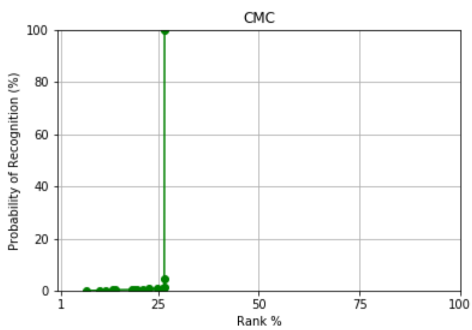


(a) CMC for finger-iris-face ordering and threshold combination (0, 1, 0)

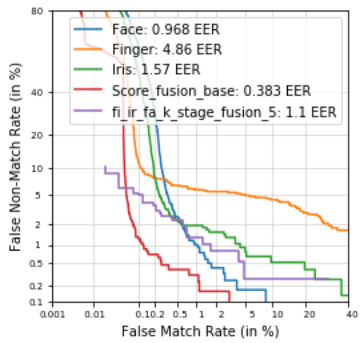


(b) DET for finger-iris-face ordering and threshold combination (0, 1, 0) denoting a Medium-good performance

Figure B.2: K-stage fusion with finger-iris-face and threshold combination (0, 1, 0)



(a) CMC for finger-iris-face ordering and threshold combination (0, 2, 0)

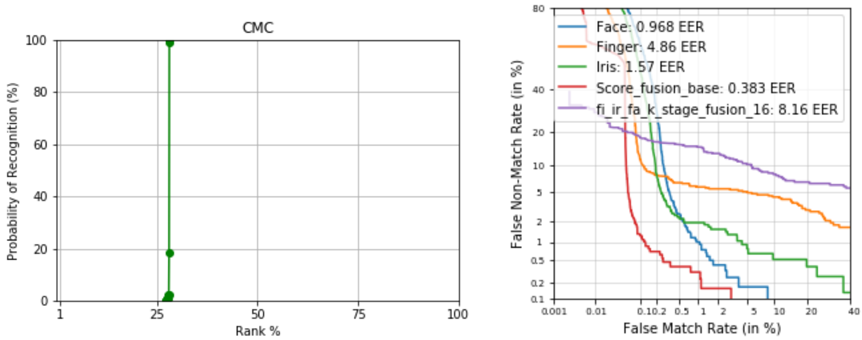


(b) DET for finger-iris-face ordering and threshold combination (0, 2, 0)

Figure B.3: K-stage fusion with finger-iris-face and threshold combination (0, 2, 0) denoting a Medium-good performance

low 1.rank probability, however, it increases dramatically at around the 25% rank (see figure B.3, B.4, B.5, B.6, B.7).

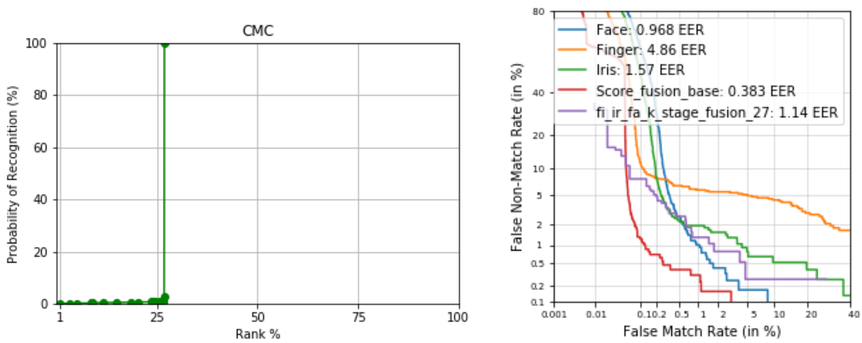
Same tendency for the figures. Comparative DET error rates regardless of ordering and threshold combinations with some deviations at around EER 5%. The CMC score show a very low rank score, however a steep increase from the 1.rank CMC that starts very low around 0% (see figure B.8, B.9, B.10, B.11, B.12, B.13, B.14, B.15, B.16, B.17, B.18, B.19, B.20).



(a) CMC for finger-iris-face ordering and threshold combination (1, 3, 1)

(b) DET for finger-iris-face ordering and threshold combination (1, 3, 1)

Figure B.4: K-stage fusion with finger-iris-face and threshold combination (1, 3, 1) denoting a Medium-bad performance



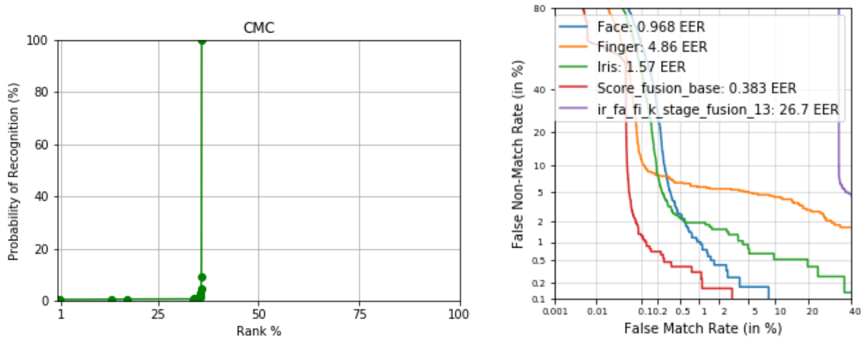
(a) CMC for finger-iris-face ordering and threshold combination (3, 0, 0)

(b) DET for finger-iris-face ordering and threshold combination (3, 0, 0)

Figure B.5: K-stage fusion with finger-iris-face and threshold combination (3, 0, 0) denoting a Medium-good performance

Comparative DET curve for the k-stage system with finger-iris-face ordering and threshold combinations (0,2,2) with DET for each modality and the base score level fusion. The CMC is very low at 1.rank and forward till approximately 25% whereafter there is a dramatic steep increase (see figure B.21).

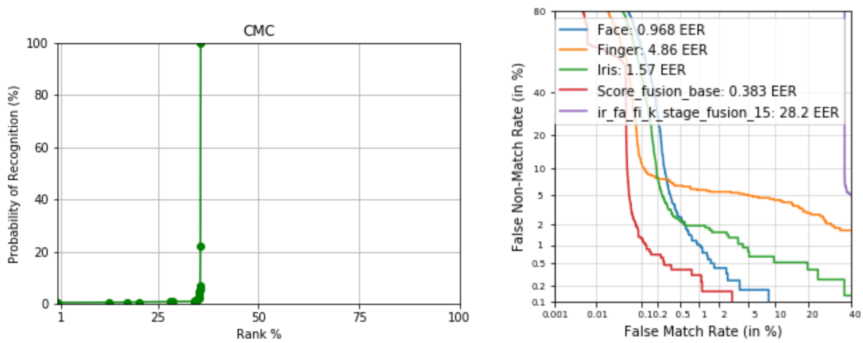
The figures show the same tendency. The DET have steady decrease and a relatively higher EER, however, it is still comparative with the other modalities and base score level fusion. The CMC have a non-existent 1.rank probability, however, a stark increase is visible at around 25% rank (see figures B.22, B.23,



(a) CMC for iris-face-finger ordering and threshold combination (1, 1, 3)

(b) DET for iris-face-finger ordering and threshold combination (1, 1, 3)

Figure B.6: K-stage fusion with iris-face-finger and threshold combination (1, 1, 3) denoting a Bad performance



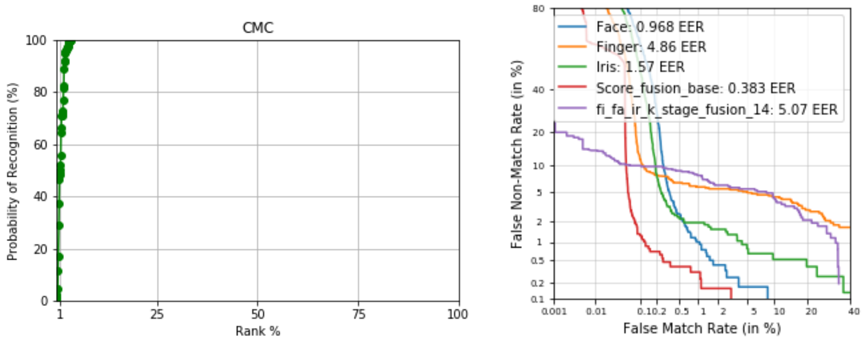
(a) CMC for iris-face-finger ordering and threshold combination (1, 2, 2)

(b) DET for iris-face-finger ordering and threshold combination (1, 2, 2)

Figure B.7: K-stage fusion with iris-face-finger and threshold combination (1, 2, 2) denoting a Bad performance

B.24, B.25, B.26, B.27, B.28, B.29).

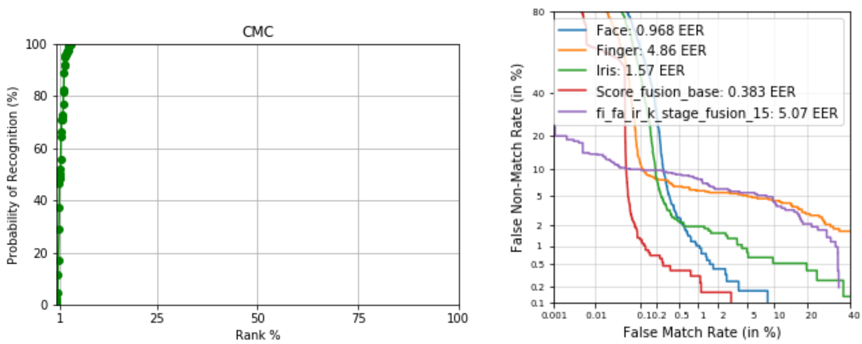
The figures show similar tendencies. The curvature for the k-stage system show similar behaviour i.e sharp increase at a certain point. It is comparative with the baselines for the each modality and basic score level fusion, however, it is relatively worse with a higher EER. The CMC have a poor 1. rank probability but a relatively low maximum rank at around 30%. Furthermore, there is a sharp increase starting at around 26 % (see figures B.30, B.31, B.32, B.33, B.34, B.35, B.36, B.37, B.38).



(a) CMC for finger-face-iris ordering and threshold combination (1, 2, 1) with level-3

(b) DET for finger-face-iris ordering and threshold combination (1, 2, 1)

Figure B.8: K-stage fusion with finger-face-iris and threshold combination (1, 2, 1) with level-3 pre-selection denoting a Medium performance

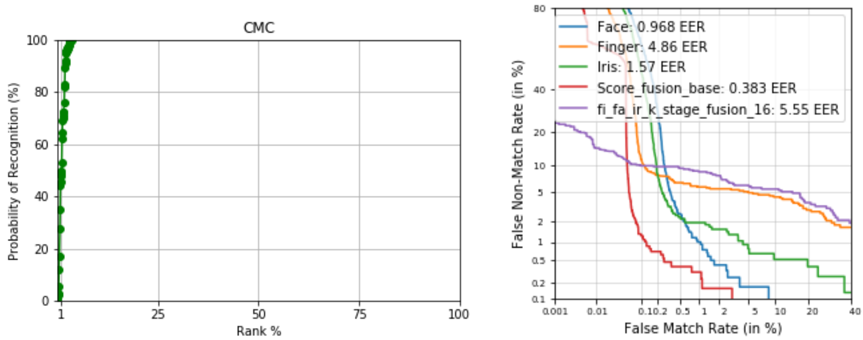


(a) CMC for finger-face-iris ordering and threshold combination (1, 2, 2) with level-3

(b) DET for finger-face-iris ordering and threshold combination (1, 2, 2).

Figure B.9: K-stage fusion with finger-face-iris and threshold combination (1, 2, 2) with level-3 pre-selection denoting a Medium performance.

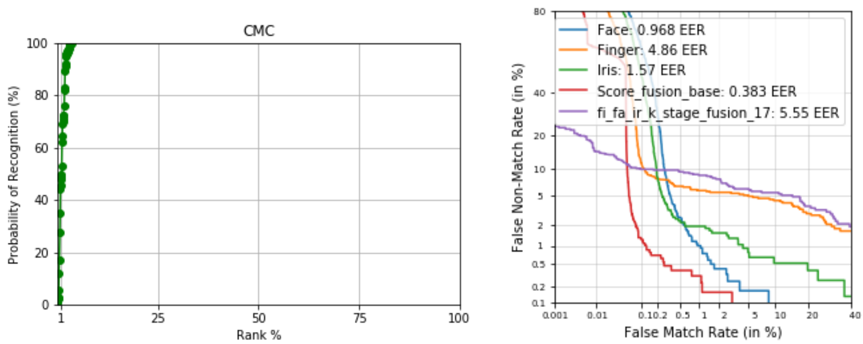
The figures show the similar tendencies. The curvature of the DET for the k-stage configuration show a steady decrease. The curve is comparative to the baselines, however, still with a relatively higher EER score. The CMC have a very low 1. rank probability which is kept until around 25% rank whereafter a steep increase till 100% probability with a rank at around 26 % (see figures B.39, B.40, B.41, B.42, B.43).



(a) CMC for finger-face-iris ordering and threshold combination (1, 3, 1) with level-3

(b) DET for finger-face-iris ordering and threshold combination (1, 3, 1).

Figure B.10: K-stage fusion with finger-face-iris and threshold combination (1, 3, 1) with level-3 pre-selection denoting a Medium performance.

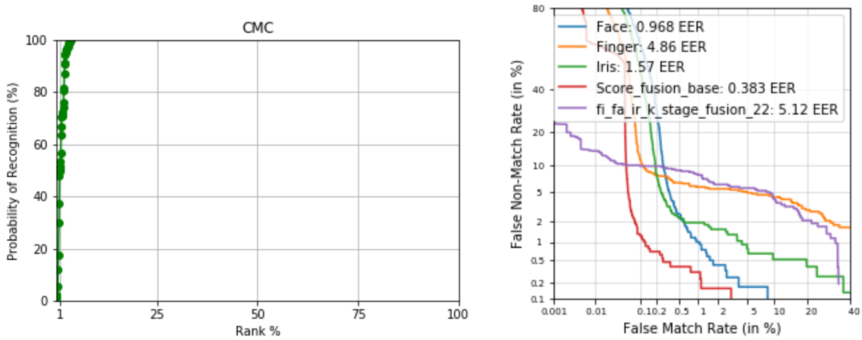


(a) CMC for finger-face-iris ordering and threshold combination (1, 3, 3) with level-3

(b) DET for finger-face-iris ordering and threshold combination (1, 3, 3).

Figure B.11: K-stage fusion with finger-face-iris and threshold combination (1, 3, 3) with level-3 pre-selection denoting a Medium performance.

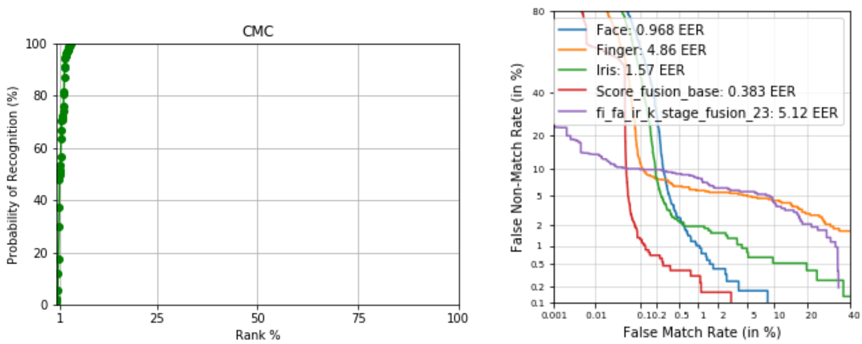
The figures show the similar tendencies. The curvature of the DET for the k-stage configuration show a steady decrease. The curve is comparative to the baselines, however, with a relatively lower EER score. The CMC have a very low 1. rank probability which is kept until around 25% rank whereafter a steep increase till 100% probability with a rank at around 26 % (see figures B.44,



(a) CMC for finger-face-iris ordering and threshold combination (2, 2, 0) with level-3

(b) DET for finger-face-iris ordering and threshold combination (2, 2, 0).

Figure B.12: K-stage fusion with finger-face-iris and threshold combination (2, 2, 0) with level-3 pre-selection denoting a Medium-good performance.



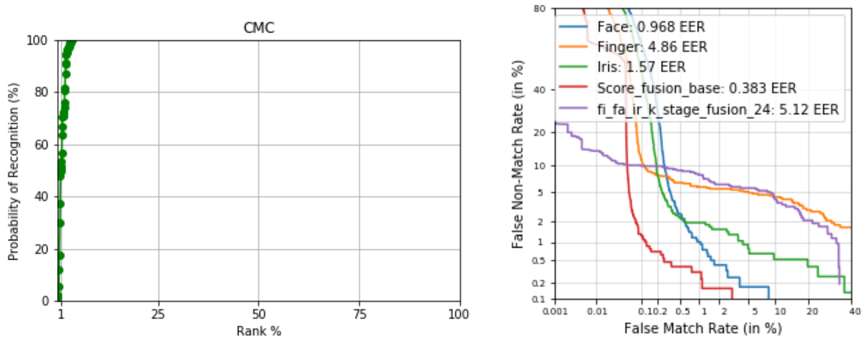
(a) CMC for finger-face-iris ordering and threshold combination (2, 2, 1) with level-3

(b) DET for finger-face-iris ordering and threshold combination (2, 2, 1).

Figure B.13: K-stage fusion with finger-face-iris and threshold combination (2, 2, 1) with level-3 pre-selection denoting a Medium performance.

B.45).

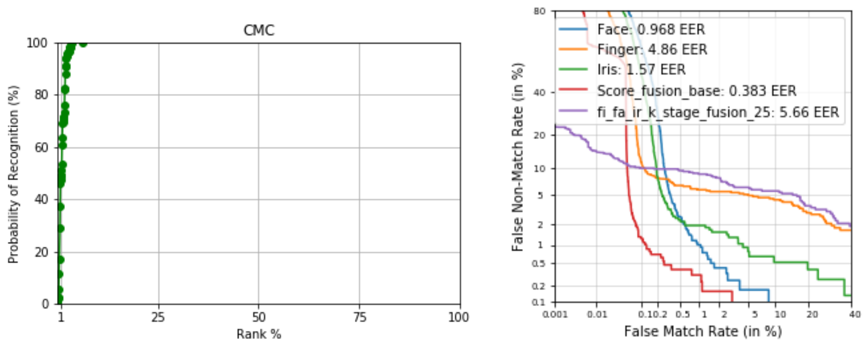
The figures show the similar tendencies. The curvature of the DET for the k-stage configuration show a similar behavior with a steep increase at some point. The curve is comparative to the baselines, however, still with a relatively lower



(a) CMC for finger-face-iris ordering and threshold combination (2, 2, 3) with level-3

(b) DET for finger-face-iris ordering and threshold combination (2, 2, 3).

Figure B.14: K-stage fusion with finger-face-iris and threshold combination (2, 2, 3) with level-3 pre-selection denoting a Medium performance.

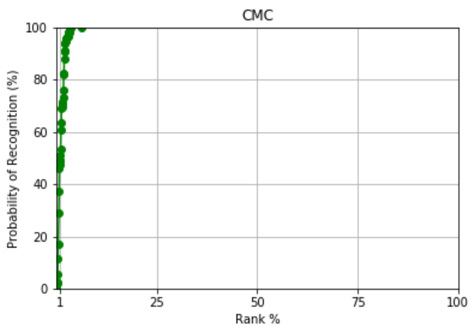


(a) CMC for finger-face-iris ordering and threshold combination (2, 3, 2) with level-3

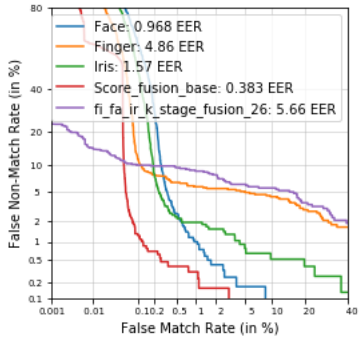
(b) DET for finger-face-iris ordering and threshold combination (2, 3, 2).

Figure B.15: K-stage fusion with finger-face-iris and threshold combination (2, 3, 2) with level-3 pre-selection denoting a Medium performance.

EER score except for some which is relative small deviations. The CMC have a very low 1. rank probability which is kept until around 25% rank whereafter a steep increase till 100% probability with a rank at around 26 % (see figure B.46, B.47, B.48, B.49).

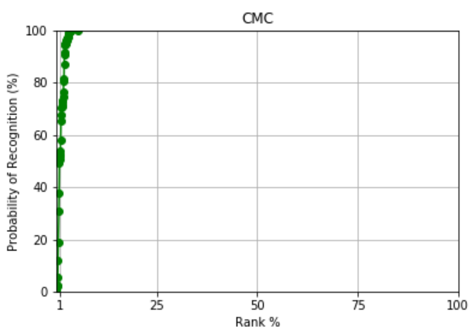


(a) CMC for finger-face-iris ordering and threshold combination (2, 3, 3) with level-3

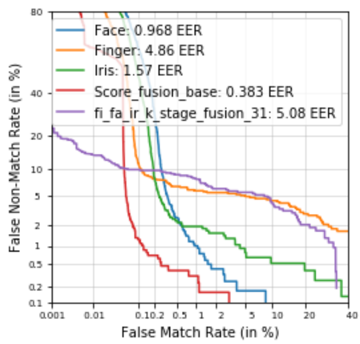


(b) DET for finger-face-iris ordering and threshold combination (2, 3, 3).

Figure B.16: K-stage fusion with finger-face-iris and threshold combination (2, 3, 3) with level-3 pre-selection denoting a Medium performance.



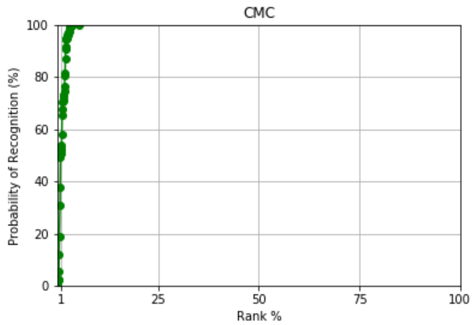
(a) CMC for finger-face-iris ordering and threshold combination (3, 2, 2) with level-3

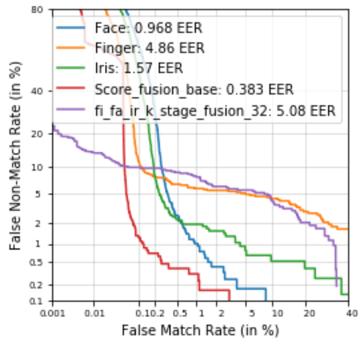


(b) DET for finger-face-iris ordering and threshold combination (3, 2, 2).

Figure B.17: K-stage fusion with finger-face-iris and threshold combination (3, 2, 2) with level-3 pre-selection and a Medium performance.

The figures show the similar tendencies. The curvature of the DET for the k-stage configuration similar behaviour with a sharp increase at some point. The curve is comparative to the baselines, however, still with a visibly higher EER score. The CMC have a very low 1. rank probability which is kept until around 30% whereafter a steep increase to 100% probability to around 31% (see figure B.50, B.51, B.52).

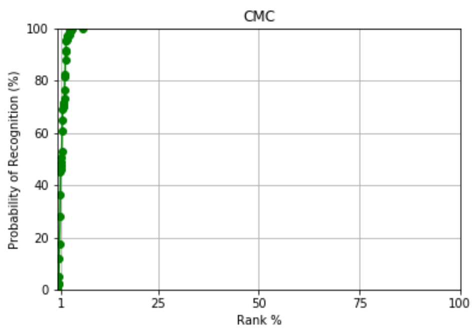


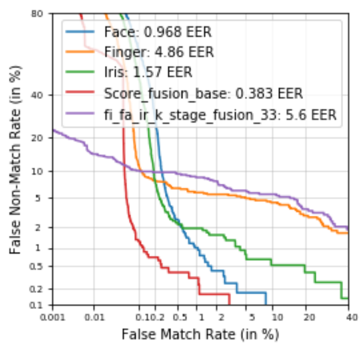


(a) CMC for finger-face-iris ordering and threshold combination (3, 2, 3) with level-3

(b) DET for finger-face-iris ordering and threshold combination (3, 2, 3).

Figure B.18: K-stage fusion with finger-face-iris and threshold combination (3, 2, 3) with level-3 pre-selection denoting a Medium performance.

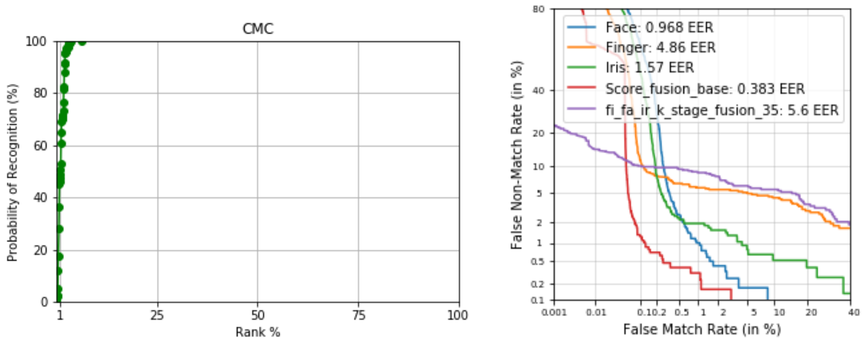




(a) CMC for finger-face-iris ordering and threshold combination (3, 3, 0) with level-3

(b) DET for finger-face-iris ordering and threshold combination (3, 3, 0).

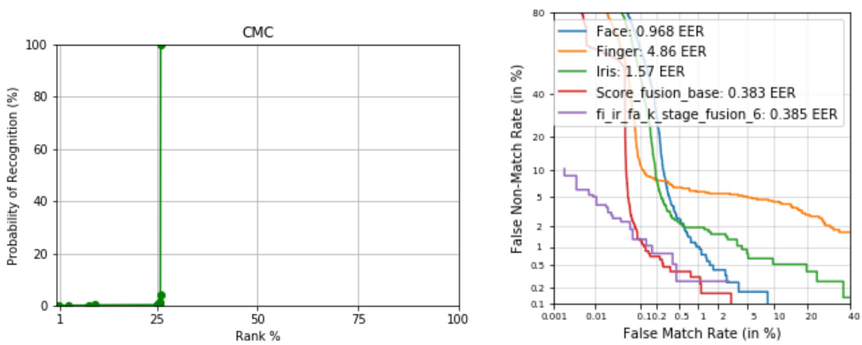
Figure B.19: K-stage fusion with finger-face-iris and threshold combination (3, 3, 0) with level-3 pre-selection denoting a Medium-good performance.



(a) CMC for finger-face-iris ordering and threshold combination (3, 3, 2) with level-3

(b) DET for finger-face-iris ordering and threshold combination (3, 3, 2).

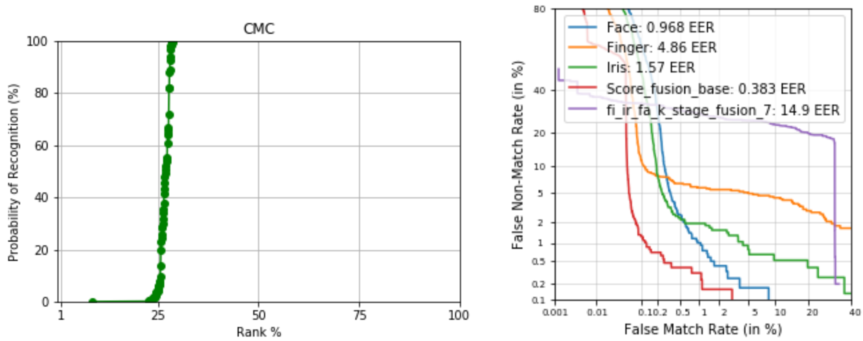
Figure B.20: K-stage fusion with finger-face-iris and threshold combination (3, 3, 2) with level-3 pre-selection denoting a Medium performance



(a) CMC for finger-iris-face ordering and threshold combination (0, 2, 2) with level-3

(b) DET for finger-iris-face ordering and threshold combination (0, 2, 2).

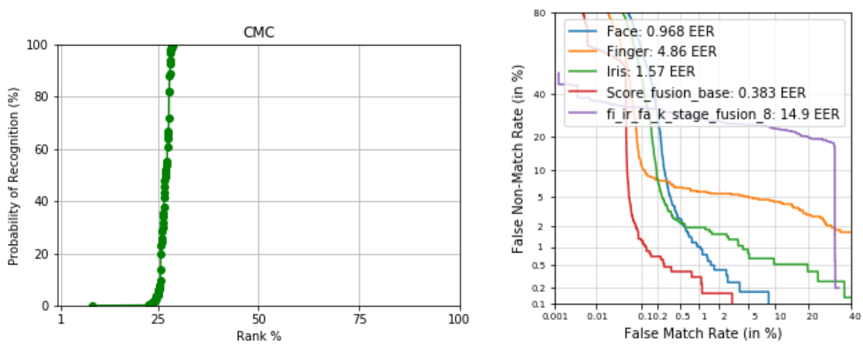
Figure B.21: K-stage fusion with finger-iris-face and threshold combination (0, 2, 2) with level-3 pre-selection denoting a Good performance.



(a) CMC for finger-iris-face ordering and threshold combination (0, 3, 0) with level-3

(b) DET for finger-iris-face ordering and threshold combination (0, 3, 0).

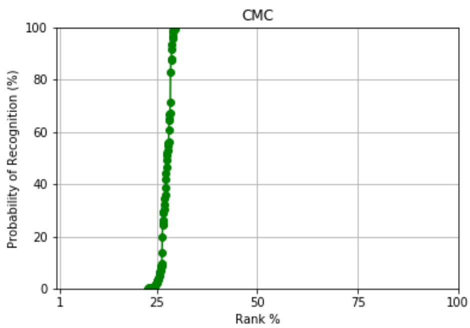
Figure B.22: K-stage fusion with finger-iris-face and threshold combination (0, 3, 0) with level-3 pre-selection denoting a Medium performance.

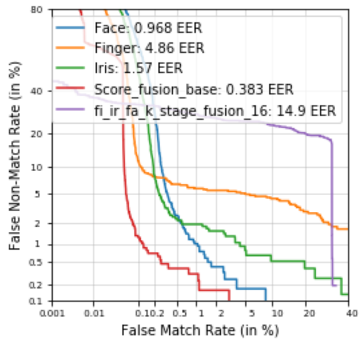


(a) CMC for finger-iris-face ordering and threshold combination (0, 3, 3) with level-3

(b) DET for finger-iris-face ordering and threshold combination (0, 3, 3).

Figure B.23: K-stage fusion with finger-iris-face and threshold combination (0, 3, 3) with level-3 pre-selection denoting a Bad performance.

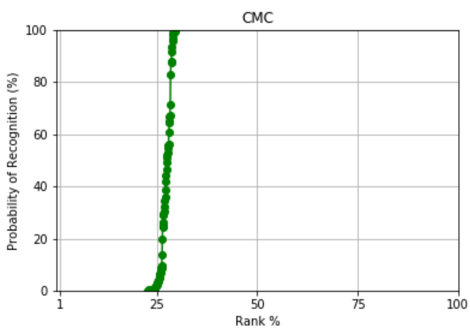


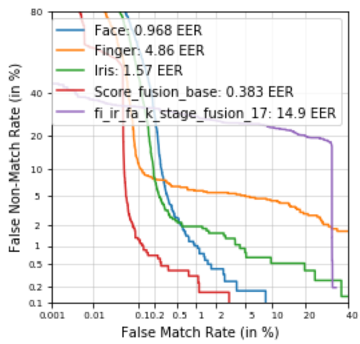


(a) CMC for finger-iris-face ordering and threshold combination (1, 3, 1) with level-3

(b) DET for finger-iris-face ordering and threshold combination (1, 3, 1).

Figure B.24: K-stage fusion with finger-iris-face and threshold combination (1, 3, 1) with level-3 pre-selection denoting a Medium-bad performance.

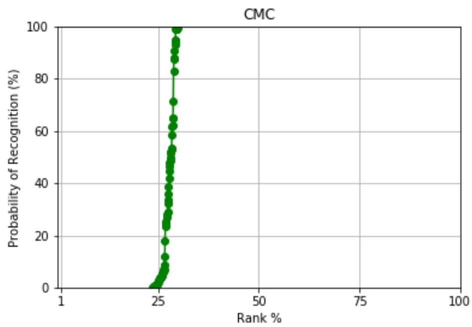


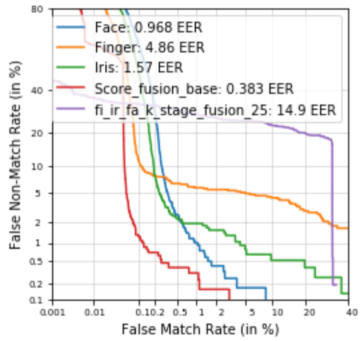


(a) CMC for finger-iris-face ordering and threshold combination (1, 3, 3) with level-3

(b) DET for finger-iris-face ordering and threshold combination (1, 3, 3).

Figure B.25: K-stage fusion with finger-iris-face and threshold combination (1, 3, 3) with level-3 pre-selection denoting a Medium-bad performance.

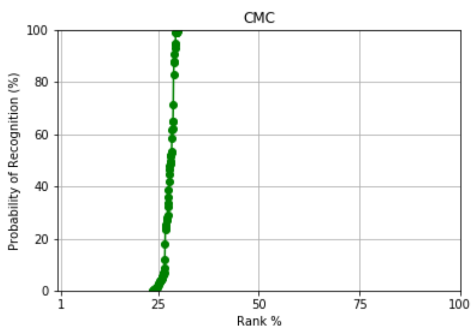


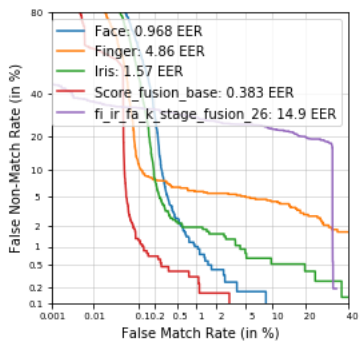


(a) CMC for finger-iris-face ordering and threshold combination (2, 3, 2) with level-3

(b) DET for finger-iris-face ordering and threshold combination (2, 3, 2).

Figure B.26: K-stage fusion with finger-iris-face and threshold combination (2, 3, 2) with level-3 pre-selection denoting a Bad performance.

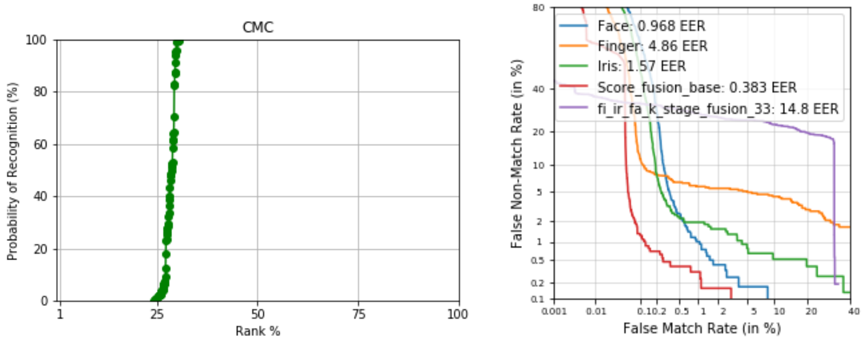




(a) CMC for finger-iris-face ordering and threshold combination (2, 3, 3) with level-3

(b) DET for finger-iris-face ordering and threshold combination (2, 3, 3).

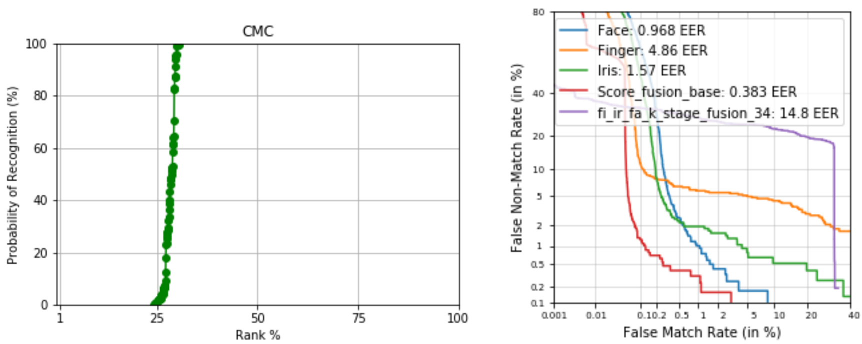
Figure B.27: K-stage fusion with finger-iris-face and threshold combination (2, 3, 3) with level-3 pre-selection denoting a Bad performance.



(a) CMC for finger-iris-face ordering and threshold combination (3, 3, 0) with level-3

(b) DET for finger-iris-face ordering and threshold combination (3, 3, 0).

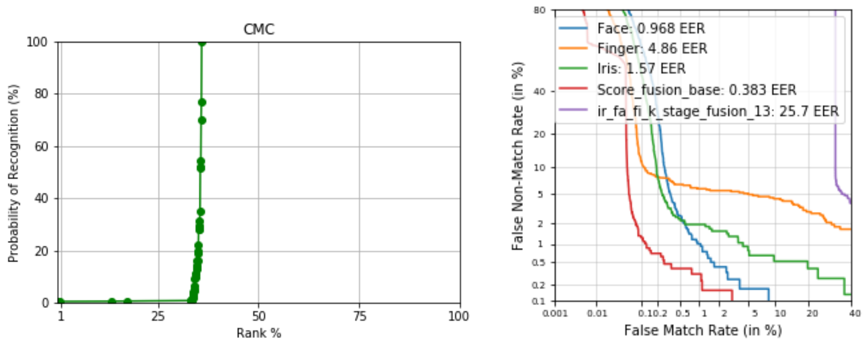
Figure B.28: K-stage fusion with finger-iris-face and threshold combination (3, 3, 0) with level-3 pre-selection denoting Medium performance.



(a) CMC for finger-iris-face ordering and threshold combination (3, 3, 1) with level-3

(b) DET for finger-iris-face ordering and threshold combination (3, 3, 1).

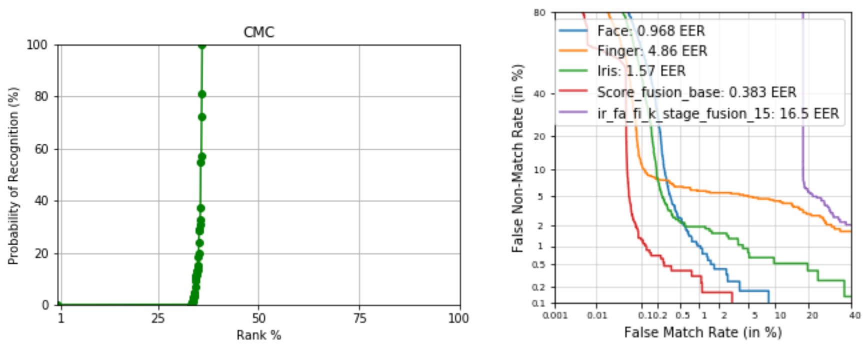
Figure B.29: K-stage fusion with finger-iris-face and threshold combination (3, 3, 1) with level-3 pre-selection denoting a Medium-bad performance.



(a) CMC for iris-face-finger ordering and threshold combination (1, 1, 3) with level-3

(b) DET for iris-face-finger ordering and threshold combination (1, 1, 3).

Figure B.30: K-stage fusion with iris-face-finger and threshold combination (1, 1, 3) with level-3 pre-selection denoting a Bad performance.

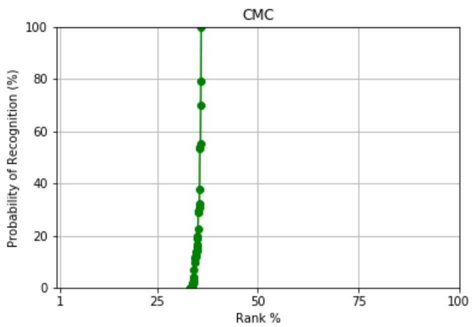


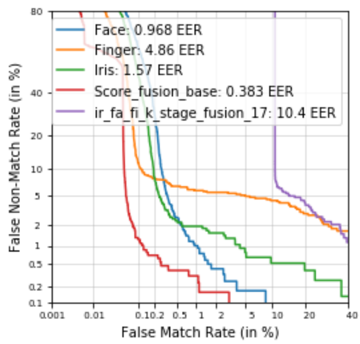
(a) CMC for iris-face-finger ordering and threshold combination (1, 2, 2) with level-3

(b) DET for iris-face-finger ordering and threshold combination (1, 2, 2).

Figure B.31: K-stage fusion with iris-face-finger and threshold combination (1, 2, 2) with level-3 pre-selection denoting a Bad performance.

,

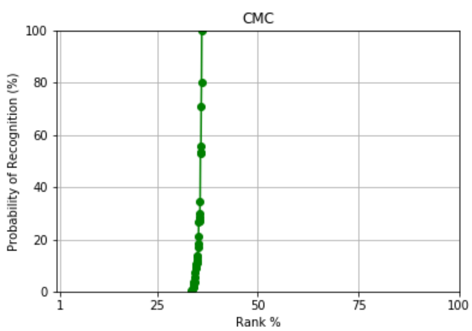


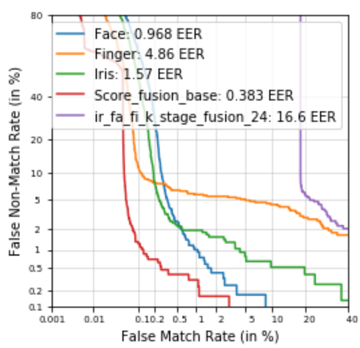


(a) CMC for iris-face-finger ordering and threshold combination (1, 3, 3) with level-3

(b) DET for iris-face-finger ordering and threshold combination (1, 3, 3).

Figure B.32: K-stage fusion with iris-face-finger and threshold combination (1, 3, 3) with level-3 pre-selection denoting a Medium-bad performance.

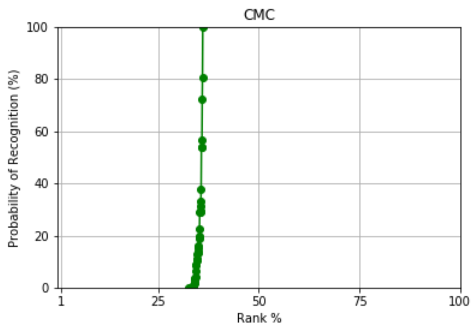




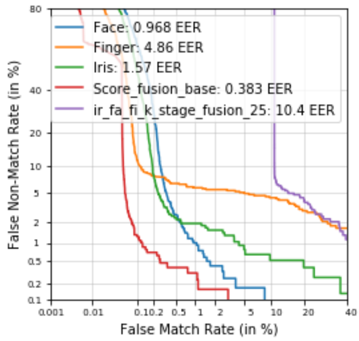
(a) CMC for iris-face-finger ordering and threshold combination (2, 2, 3) with level-3

(b) DET for iris-face-finger ordering and threshold combination (2, 2, 3).

Figure B.33: K-stage fusion with iris-face-finger and threshold combination (2, 2, 3) with level-3 pre-selection denoting a Bad performance.

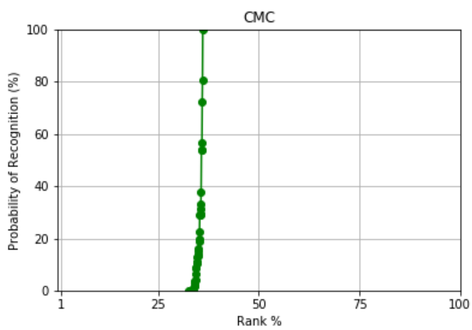


(a) CMC for iris-face-finger ordering and threshold combination (2, 3, 2) with level-3

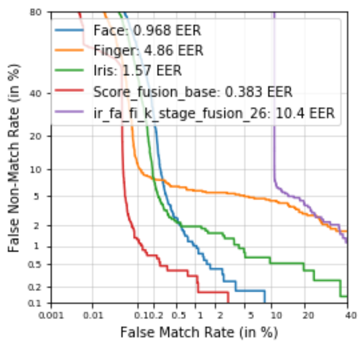


(b) DET for iris-face-finger ordering and threshold combination (2, 3, 2).

Figure B.34: K-stage fusion with iris-face-finger and threshold combination (2, 3, 2) with level-3 pre-selection denoting a Bad performance.

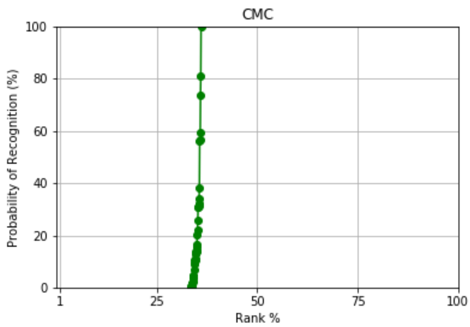


(a) CMC for iris-face-finger ordering and threshold combination (2, 3, 3) with level-3

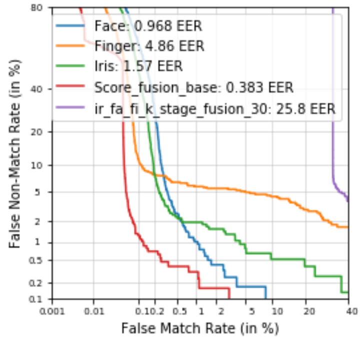


(b) DET for iris-face-finger ordering and threshold combination (2, 3, 3).

Figure B.35: K-stage fusion with iris-face-finger and threshold combination (2, 3, 3) with level-3 pre-selection denoting a Medium-bad performance.

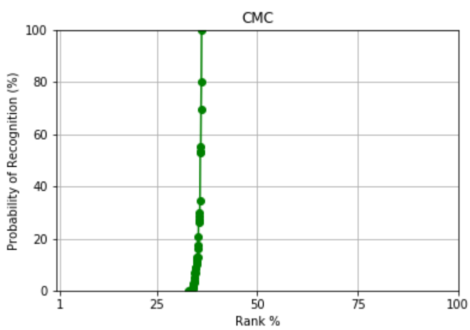


(a) CMC for iris-face-finger ordering and threshold combination (3, 1, 3) with level-3

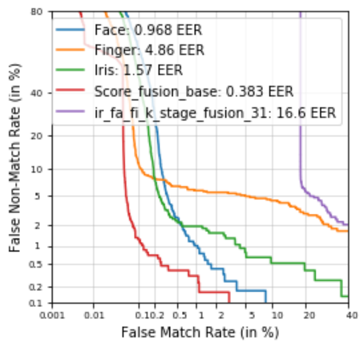


(b) DET for iris-face-finger ordering and threshold combination (3, 1, 3).

Figure B.36: K-stage fusion with iris-face-finger and threshold combination(3, 1, 3) with level-3 pre-selection denoting a Bad performance.

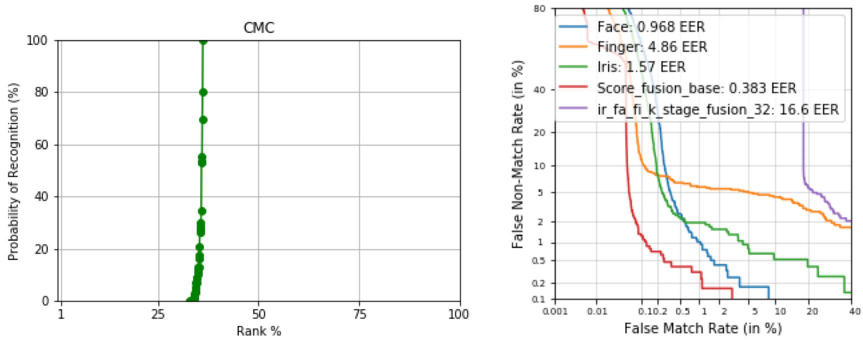


(a) CMC for iris-face-finger ordering and threshold combination (3, 2, 2) with level-3



(b) DET for iris-face-finger ordering and threshold combination (3, 2, 2).

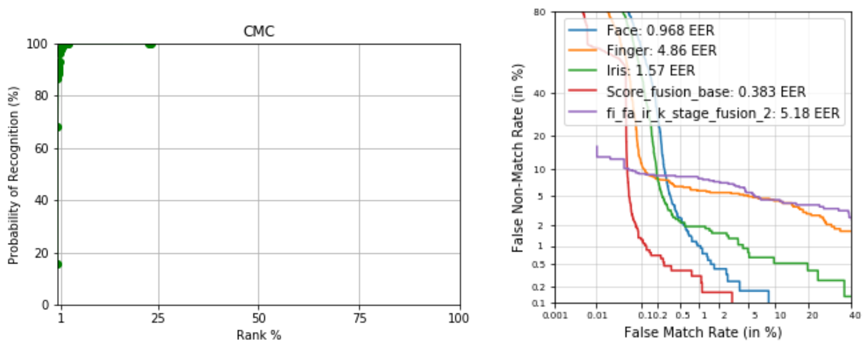
Figure B.37: K-stage fusion with iris-face-finge and threshold combination (3, 2, 2) with level-3 pre-selection denoting a Bad performance.



(a) CMC for iris-face-finger ordering and threshold combination (3, 2, 3) with level-3

(b) DET for iris-face-finger ordering and threshold combination (3, 2, 3).

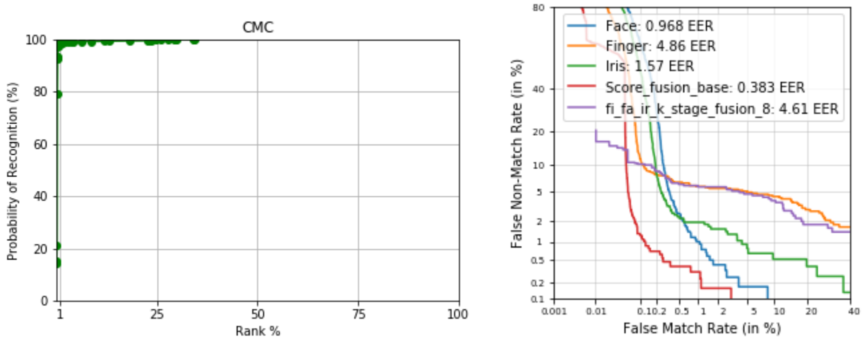
Figure B.38: K-stage fusion with iris-face-finger and threshold combination (3, 2, 3) with level-3 pre-selection denoting a Bad performance.



(a) CMC for finger-face-iris ordering and threshold combination (0, 2, 1)

(b) DET for finger-face-iris ordering and threshold combination (0, 2, 1)

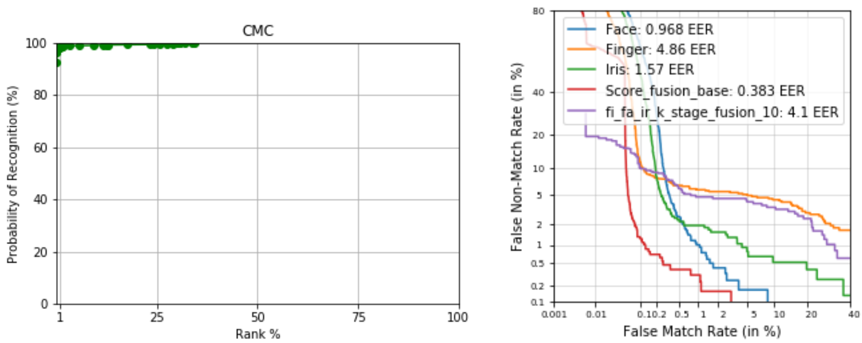
Figure B.39: K-stage fusion with finger-face-iris and threshold combination (0, 2, 1) denoting a Medium performance.



(a) CMC for finger-face-iris ordering and threshold combination (1, 2, 0)

(b) DET for finger-face-iris ordering and threshold combination (1, 2, 0)

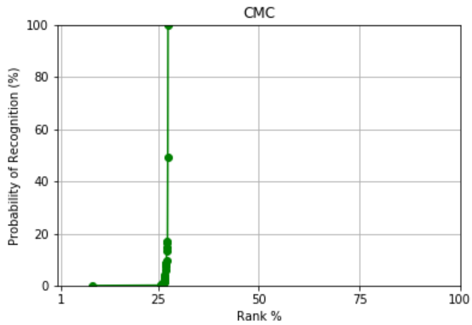
Figure B.40: K-stage fusion with finger-face-iris and threshold combination (1, 2, 0) denoting a Medium-good performance.



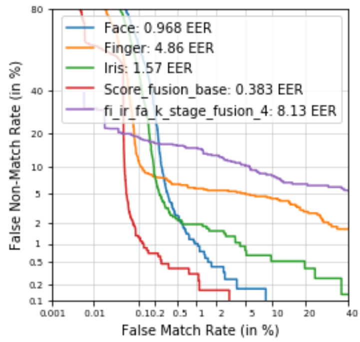
(a) CMC for finger-face-iris ordering and threshold combination (1, 3, 0)

(b) DET for finger-face-iris ordering and threshold combination (1, 3, 0)

Figure B.41: K-stage fusion with finger-face-iris and threshold combination (1, 3, 0) denoting Medium-good performance.

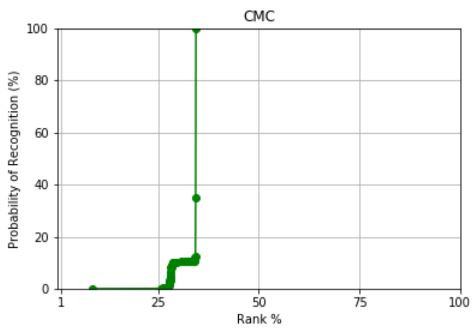


(a) CMC for finger-iris-face ordering and threshold combination (0, 3, 1)

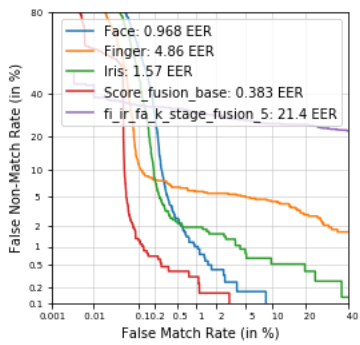


(b) DET for finger-iris-face ordering and threshold combination (0, 3, 1)

Figure B.42: K-stage fusion with finger-iris-face and threshold combination (0, 3, 1) denoting a Medium-bad performance.

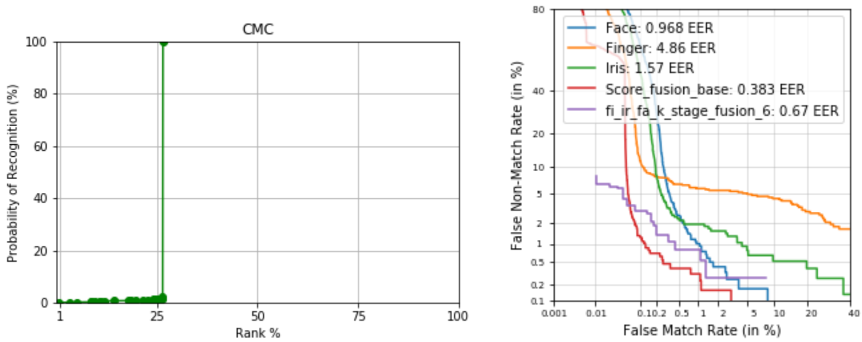


(a) CMC for finger-iris-face ordering and threshold combination (0, 3, 2)



(b) DET for finger-iris-face ordering and threshold combination (0, 3, 2)

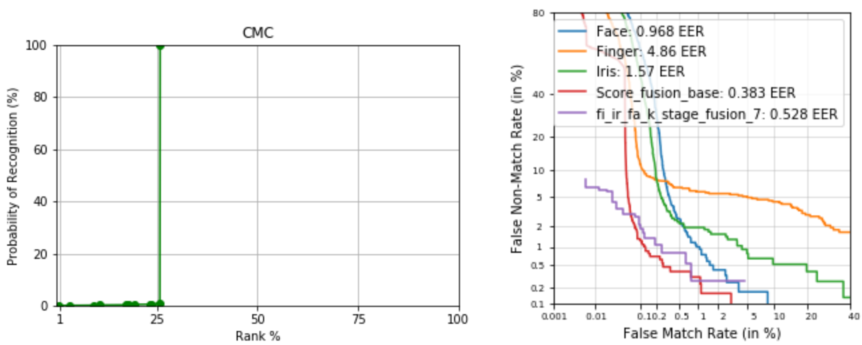
Figure B.43: K-stage fusion with finger-iris-face and threshold combination (0, 3, 2) denoting a Bad performance.



(a) CMC for finger-iris-face ordering and threshold combination (1, 0, 2)

(b) DET for finger-iris-face ordering and threshold combination (1, 0, 2)

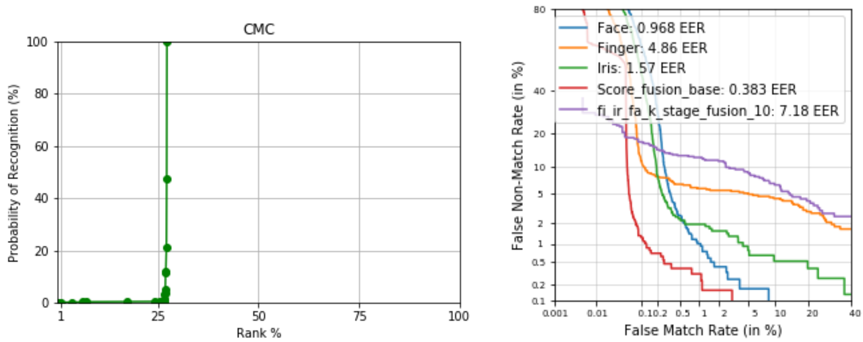
Figure B.44: K-stage fusion with finger-iris-face and threshold combination (1, 0, 2) denoting a Good performance.



(a) CMC for finger-iris-face ordering and threshold combination (1, 0, 3)

(b) DET for finger-iris-face ordering and threshold combination (1, 0, 3)

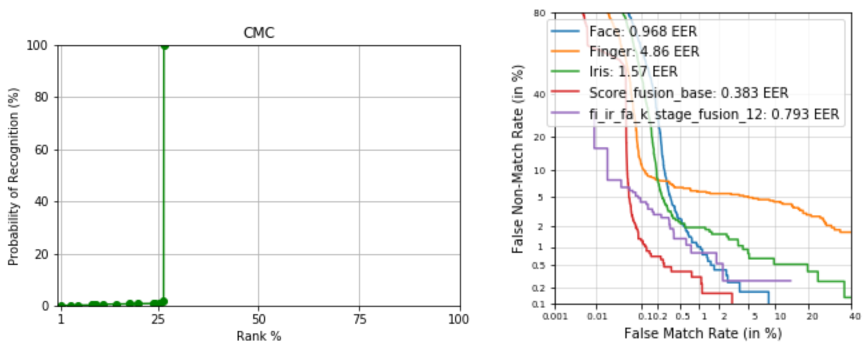
Figure B.45: K-stage fusion with finger-iris-face and threshold combination (1, 0, 3) denoting a Good performance.



(a) CMC for finger-iris-face ordering and threshold combination (1, 3, 0)

(b) DET for finger-iris-face ordering and threshold combination (1, 3, 0)

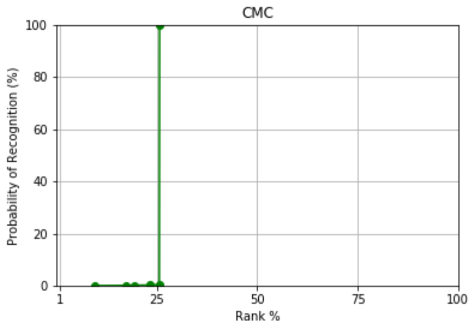
Figure B.46: K-stage fusion with finger-iris-face and threshold combination (1, 3, 0) denoting a Medium performance.

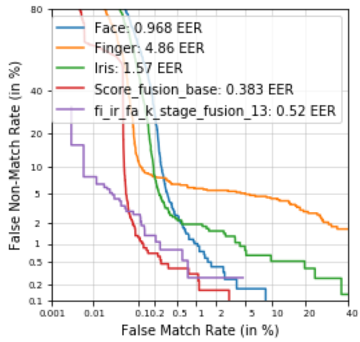


(a) CMC for finger-iris-face ordering and threshold combination (2, 0, 1)

(b) DET for finger-iris-face ordering and threshold combination (2, 0, 1)

Figure B.47: K-stage fusion with finger-iris-face and threshold combination (2, 0, 1) denoting a Good performance.

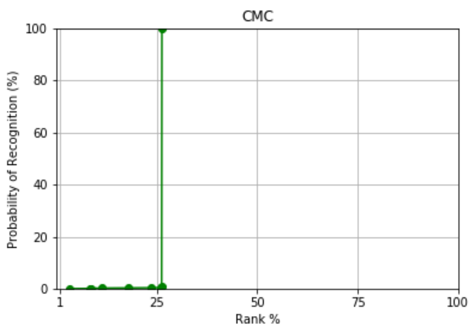


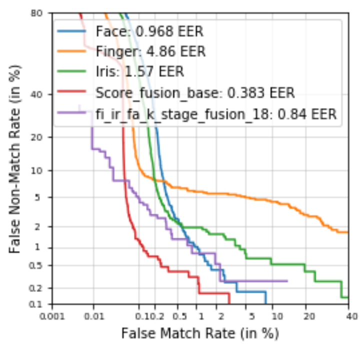


(a) CMC for finger-iris-face ordering and threshold combination (2, 0, 3)

(b) DET for finger-iris-face ordering and threshold combination (2, 0, 3)

Figure B.48: K-stage fusion with finger-iris-face and threshold combination (2, 0, 3) denoting a Good performance.

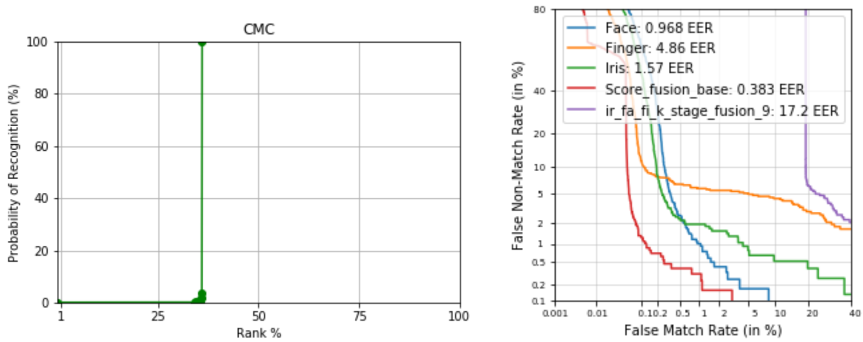




(a) CMC for finger-iris-face ordering and threshold combination (3, 0, 1)

(b) DET for finger-iris-face ordering and threshold combination (3, 0, 1)

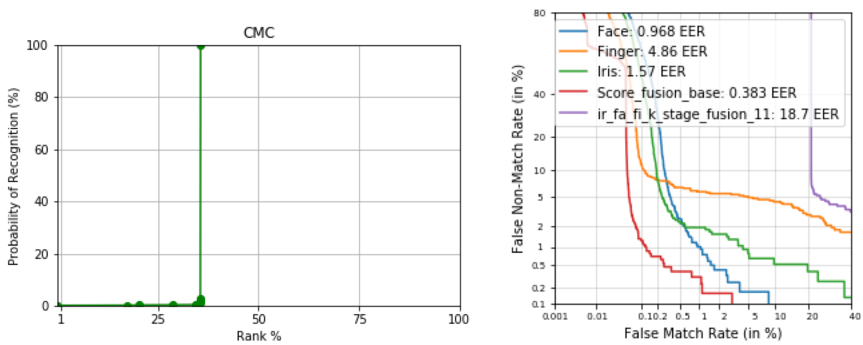
Figure B.49: K-stage fusion with finger-iris-face and threshold combination (3, 0, 1) denoting a Good performance.



(a) CMC for iris-face-finger ordering and threshold combination (1, 2, 3)

(b) DET for iris-face-finger ordering and threshold combination (1, 2, 3)

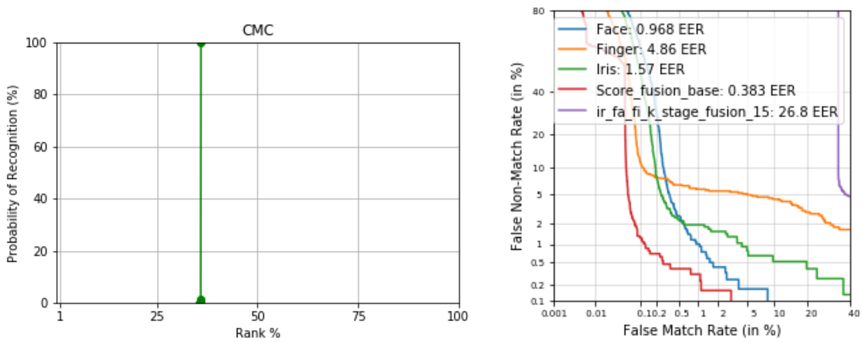
Figure B.50: K-stage fusion with iris-face-finger and threshold combination (1, 2, 3) denoting Bad performance.



(a) CMC for iris-face-finger ordering and threshold combination (1, 3, 2)

(b) DET for iris-face-finger ordering and threshold combination (1, 3, 2)

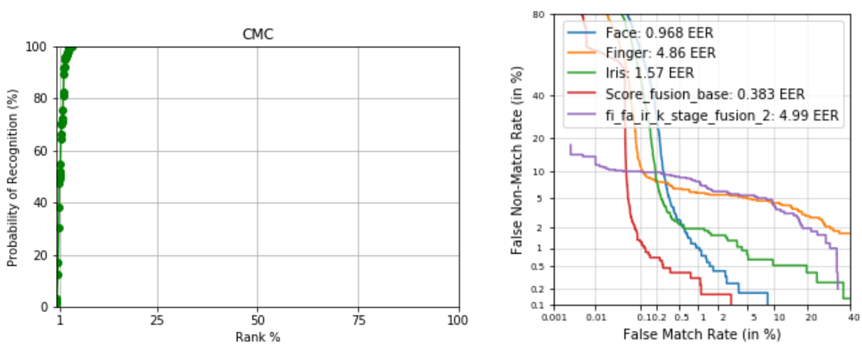
Figure B.51: K-stage fusion with iris-face-finger and threshold combination (1, 3, 2) denoting a Bad performance.



(a) CMC for iris-face-finger ordering and threshold combination (2, 1, 3)

(b) DET for iris-face-finger ordering and threshold combination (2, 1, 3)

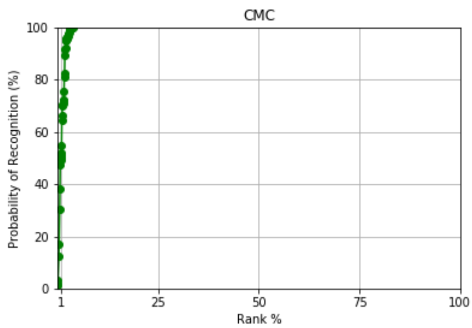
Figure B.52: K-stage fusion with iris-face-finger and threshold combination (2, 1, 3) denoting a Bad performance.

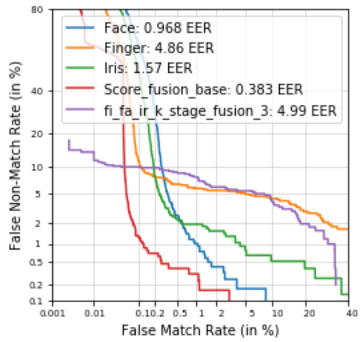


(a) CMC for finger-face-iris ordering and threshold combination (0, 2, 1) with level-3

(b) DET for finger-face-iris ordering and threshold combination (0, 2, 1) with level-3

Figure B.53: K-stage fusion with finger-face-iris and threshold combination (0, 2, 1) with level-3 pre-selection denoting a Medium-good performance.

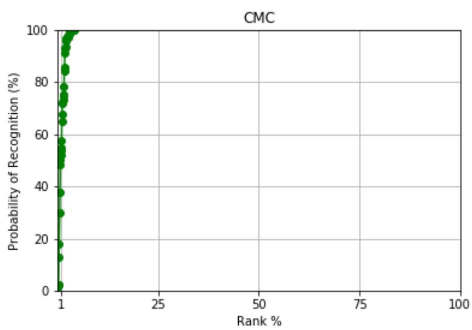


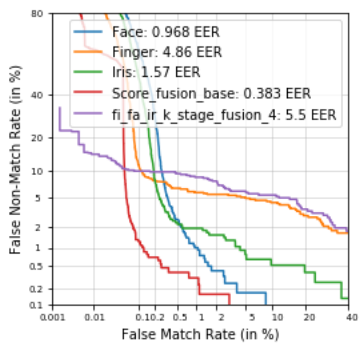


(a) CMC for finger-face-iris ordering and threshold combination (0, 2, 3) with level-3

(b) DET for finger-face-iris ordering and threshold combination (0, 2, 3) with level-3

Figure B.54: K-stage fusion with finger-face-iris and threshold combination (0, 2, 3) with level-3 pre-selection denoting a Medium-good performance.

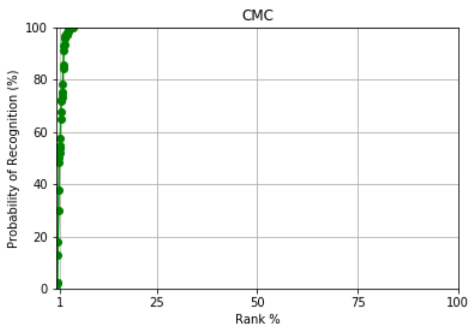


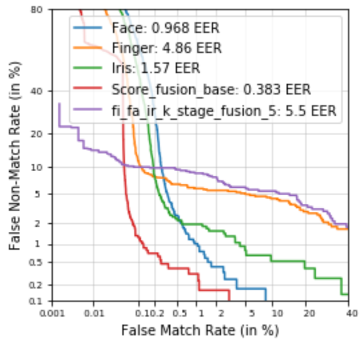


(a) CMC for finger-face-iris ordering and threshold combination (0, 3, 1) with level-3

(b) DET for finger-face-iris ordering and threshold combination (0, 3, 1) with level-3

Figure B.55: K-stage fusion with finger-face-iris and threshold combination (0, 3, 1) with level-3 pre-selection denoting a Medium performance.

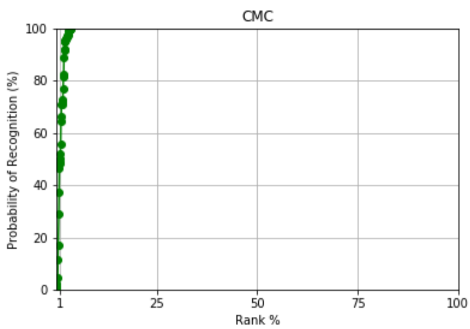


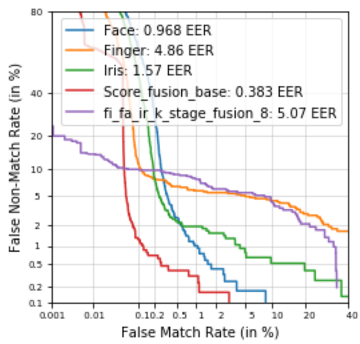


(a) CMC for finger-face-iris ordering and threshold combination (0, 3, 2) with level-3

(b) DET for finger-face-iris ordering and threshold combination (0, 3, 2) with level-3

Figure B.56: K-stage fusion with finger-face-iris and threshold combination (0, 3, 2) with lelvel-3 pre-selection denoting Medium performance.

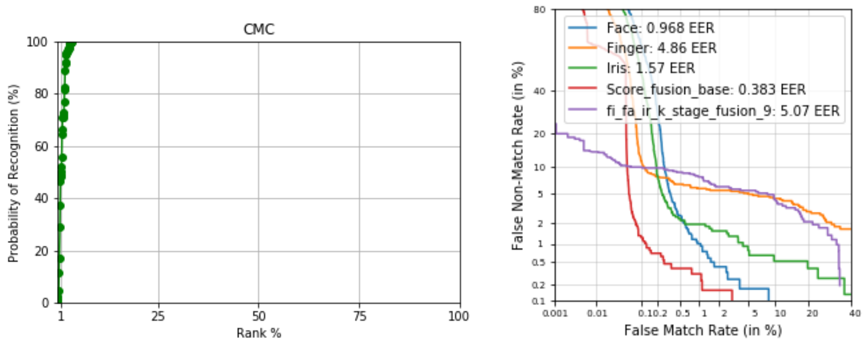




(a) CMC for finger-face-iris ordering and threshold combination (1, 2, 0) with level-3

(b) DET for finger-face-iris ordering and threshold combination (1, 2, 0) with level-3

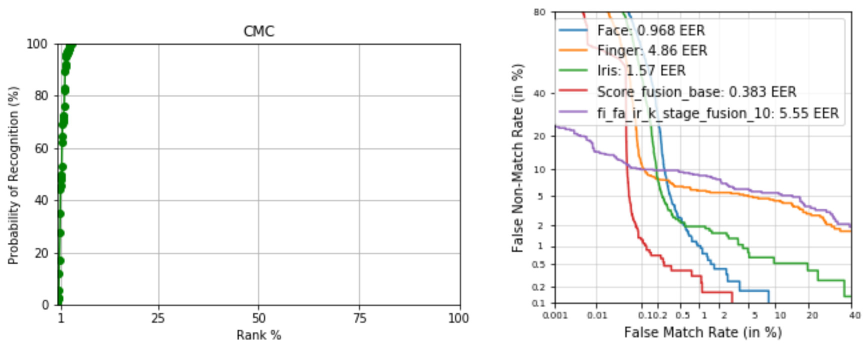
Figure B.57: K-stage fusion with finger-face-iris and threshold combination (1, 2, 0) with level-3 pre-selection denoting a Medium performance.



(a) CMC for finger-face-iris ordering and threshold combination (1, 2, 3) with level-3

(b) DET for finger-face-iris ordering and threshold combination (1, 2, 3) with level-3

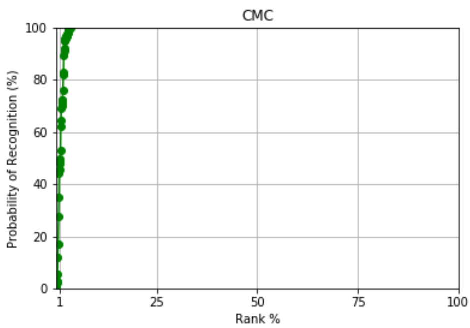
Figure B.58: K-stage fusion with finger-face-iris and threshold combination (1, 2, 3) with level-3 pre-selection denoting a Medium performance.



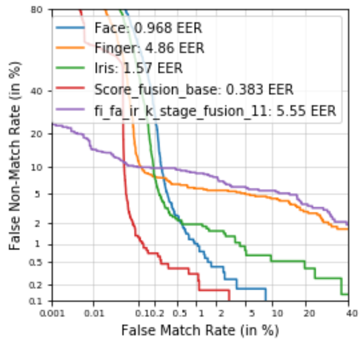
(a) CMC for finger-face-iris ordering and threshold combination (1, 3, 0) with level-3

(b) DET for finger-face-iris ordering and threshold combination (1, 3, 0) with level-3

Figure B.59: K-stage fusion with finger-face-iris and threshold combination (1, 3, 0) with level-3 pre-selection denoting a Medium performance.

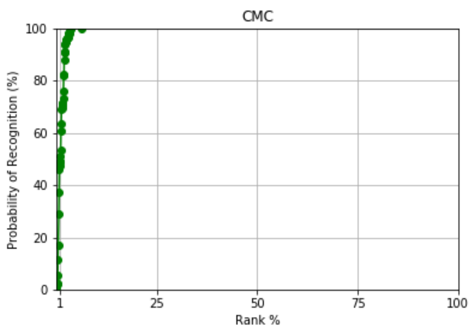


(a) CMC for finger-face-iris ordering and threshold combination (1, 3, 2) with level-3

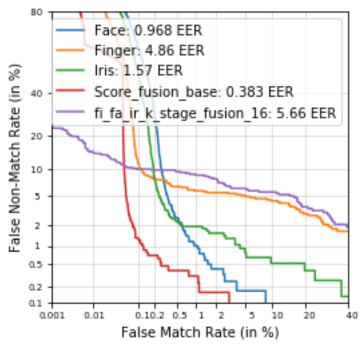


(b) DET for finger-face-iris ordering and threshold combination (1, 3, 2) with level-3

Figure B.60: K-stage fusion with finger-face-iris and threshold combination (1, 3, 2) with level-3 pre-selection denoting a Medium performance.

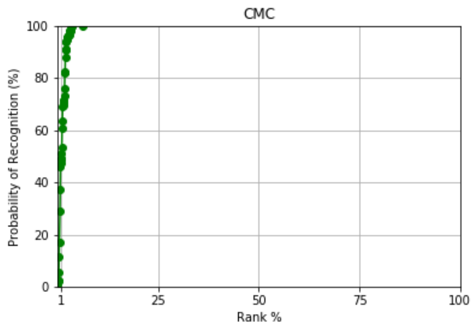


(a) CMC for finger-face-iris ordering and threshold combination (2, 3, 0) with level-3

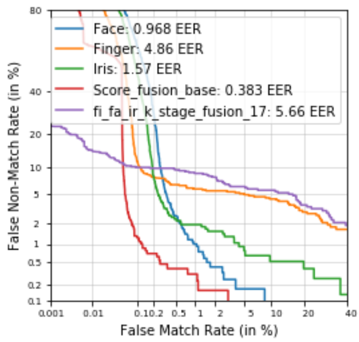


(b) DET for finger-face-iris ordering and threshold combination (2, 3, 0) with level-3

Figure B.61: K-stage fusion with finger-face-iris and threshold combination (2, 3, 0) with level-3 pre-selection denoting a Medium performance.

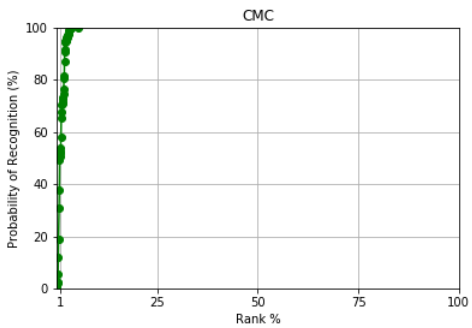


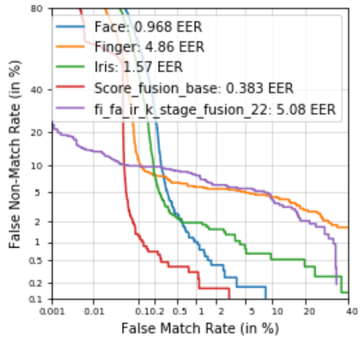
(a) CMC for finger-face-iris ordering and threshold combination (2, 3, 1) with level-3



(b) DET for finger-face-iris ordering and threshold combination (2, 3, 1) with level-3

Figure B.62: K-stage fusion with finger-face-iris and threshold combination (2, 3, 1) with level-3 pre-selection dentoting a Medium performance.



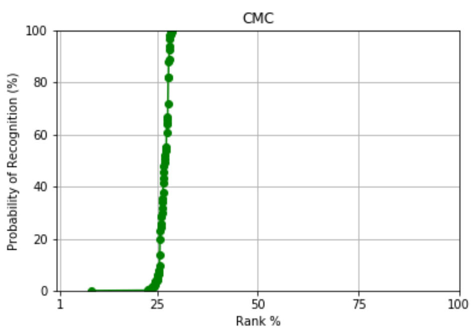


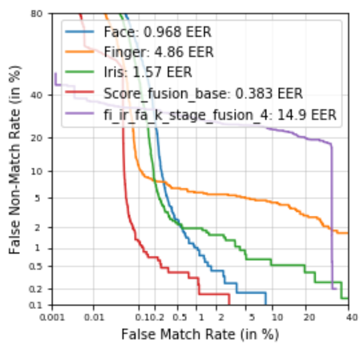
(a) CMC for finger-face-iris ordering and threshold combination (3, 2, 0) with level-3

(b) DET for finger-face-iris ordering and threshold combination (3, 2, 0) with level-3

Figure B.63: K-stage fusion with finger-face-iris and threshold combination (3, 2, 0) with a level-3 pre-selection denoting a Medium performance.

The figure show the similar tendencies. The curvature of the DET for the k-stage configuration show a steady decrease with a more dramatic drop at its end points. The curve is comparative to the baselines, however, still with a relatively higher EER score at approximately 5%. The CMC have a very low 1. rank probability and a very low maximum rank under 10% rank with a dramatic increase starting from rank 1 % (see figures B.53, B.54, B.55, B.56, B.57, B.58, B.59, B.60, B.61, B.62, B.63).

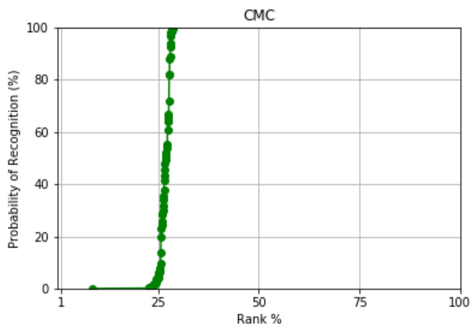




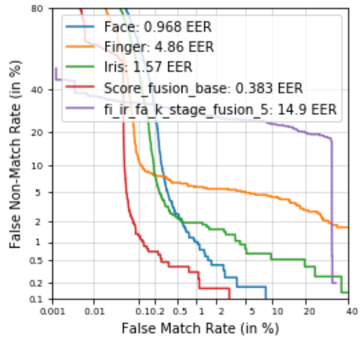
(a) CMC for finger-iris-face and threshold combination (0, 3, 1) with level-3

(b) DET for finger-iris-face and threshold combination (0, 3, 1) with level-3

Figure B.64: K-stage fusion with finger-iris-face and threshold combination (0, 3, 1) with level-3 pre-selection denoting a Bad performance.

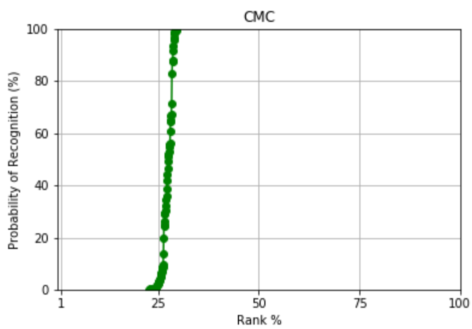


(a) CMC for finger-iris-face and threshold combination (0, 3, 2) with level-3

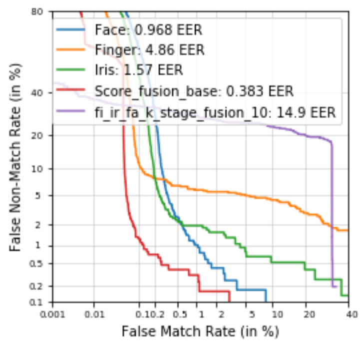


(b) DET for finger-iris-face and threshold combination (0, 3, 2) with level-3

Figure B.65: K-stage fusion with finger-iris-face and threshold combination (0, 3, 2) with level-3 pre-selection denoting a Bad performance.



(a) CMC for finger-iris-face and threshold combination (1, 3, 0) with level-3

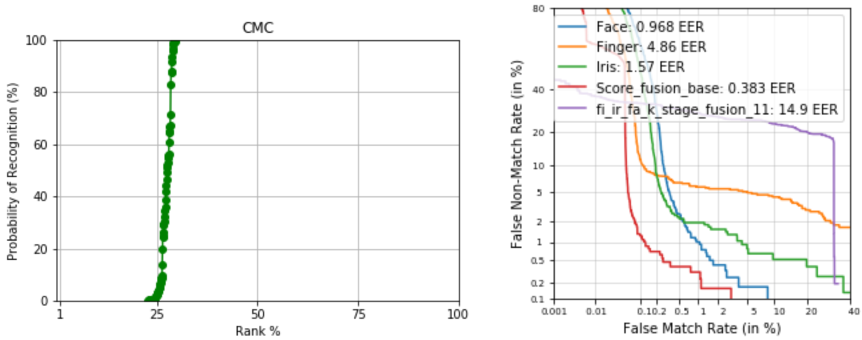


(b) DET for finger-iris-face and threshold combination (1, 3, 0) with level-3

Figure B.66: K-stage fusion with finger-iris-face and threshold combination (1, 3, 0) with level-3 pre-selection denoting a Bad performance.

The figures show the similar tendencies. The curvature of the DET for the k-stage configuration show a steady decrease with a dramatic drop at a comparatively higher level. The curve is comparative to the baselines, however, still with a relatively higher EER score. The CMC have a very low 1. rank probability which is kept until around 25% rank whereafter a steep increase till 100% probability with a rank at around 26 % (see figures B.64, B.65, B.66, B.67, B.68.

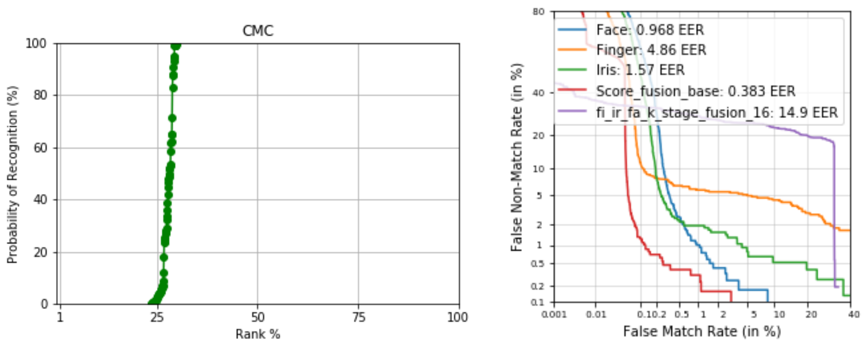
The figure show the similar tendencies. The curvature of the DET for the k-stage configuration show similar behaviour to the baseline DET curves with a



(a) CMC for finger-iris-face and threshold combination (1, 3, 2) with level-3

(b) DET for finger-iris-face and threshold combination (1, 3, 2) with level-3

Figure B.67: K-stage fusion with finger-iris-face and threshold combination (1, 3, 2) with level-3 pre-selection denoting a Bad performance.



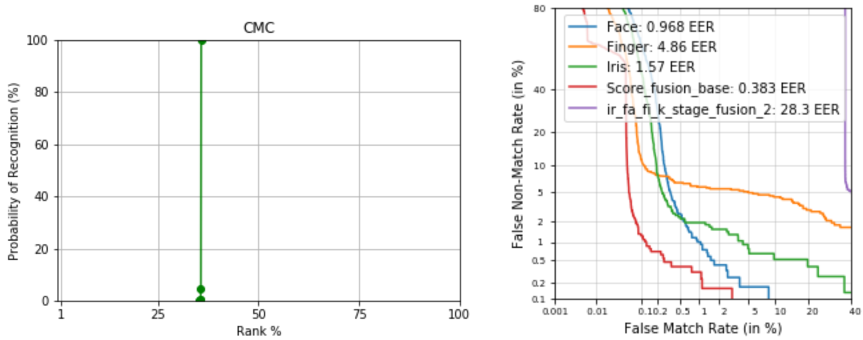
(a) CMC for finger-iris-face and threshold combination (2, 3, 0) with level-3

(b) DET for finger-iris-face and threshold combination (2, 3, 0) with level-3

Figure B.68: K-stage fusion with finger-iris-face and threshold combination (2, 3, 0) with level-3 pre-selection denoting a Bad performance.

sudden dramatic increase at a certain point. The curve is comparative to the baselines, however, still with a relatively higher EER score. The CMC have a very low 1. rank probability which is kept until around 25% rank whereafter a steep increase till 100% probability with a rank at around 26 % (see figure B.69).

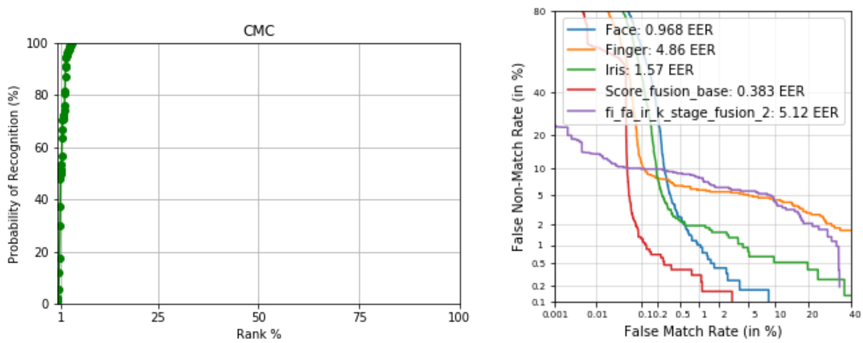
The figures show the similar tendencies.The curvature of the DET for the k-stage configuration show a steady decrease. The curve is comparative to the baselines, however, still with a relatively higher EER score at around 5%. The



(a) CMC for iris-face-finger and threshold combination (2,2,2)

(b) DET for iris-face-finger and threshold combination (2,2,2)

Figure B.69: K-stage fusion with iris-face-finger and threshold combination (2,2,2) with a Bad performance.



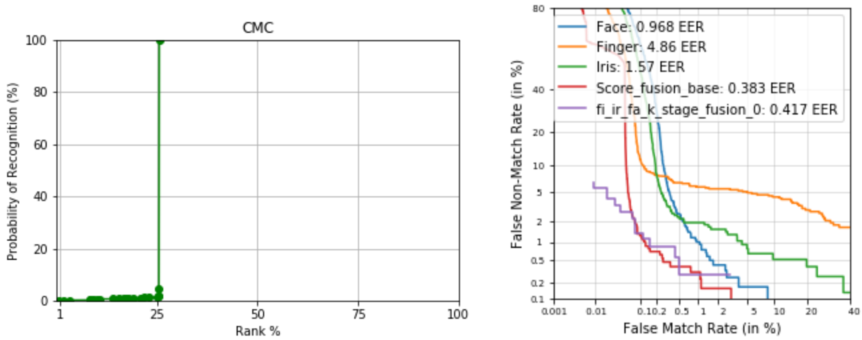
(a) CMC for finger-face-iris and threshold combination (2,2,2)

(b) DET for finger-face-iris and threshold combination (2,2,2)

Figure B.70: K-stage fusion with finger-face-iris and threshold combination (2,2,2) with a Medium performance.

CMC have a very low 1. rank probability which is kept until around 5% rank whereafter a steep increase till 100% probability with a rank at around 6 % (see figure B.70).

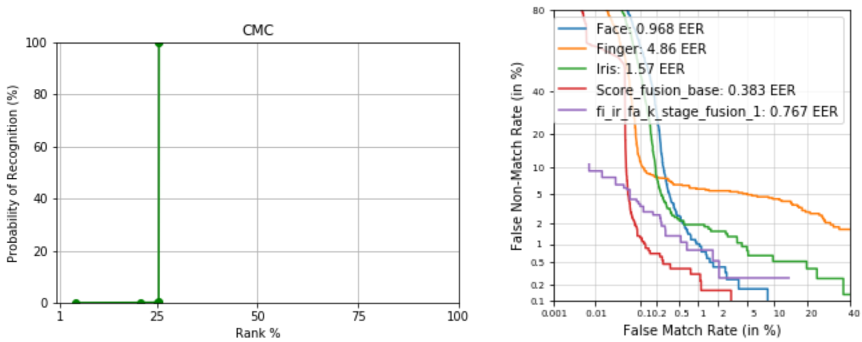
The curvature of the DET for the k-stage configuration show a steady decrease. The curve is comparative to the baselines, however, still with a relatively lower EER score. The CMC have a very low 1. rank probability which is kept until around 25% rank whereafter a steep increase till 100% probability with a rank at around 26 % (see figure B.71).



(a) CMC for finger-iris-face and threshold combination (0,0,0) with level-3

(b) DET for finger-iris-face and threshold combination (0,0,0) with level-3

Figure B.71: K-stage fusion with finger-iris-face and threshold combination (0,0,0) with level-3 pre-selection denoting a Good performance.

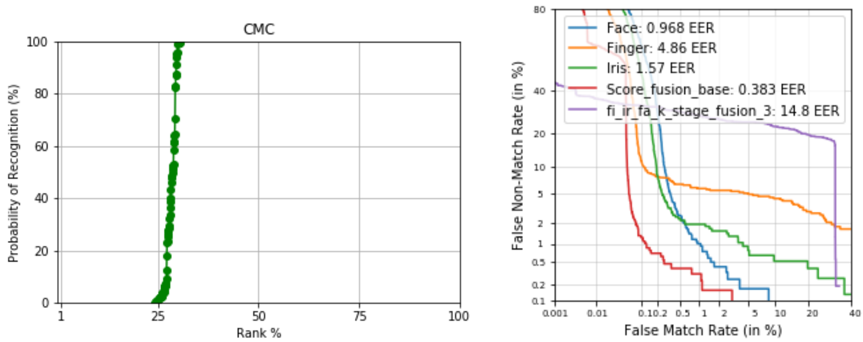


(a) CMC for finger-iris-face and pre-selection sizes finger=25%, iris=25% and final selection with face

(b) DET for finger-iris-face and pre-selection sizes finger=25%, iris=25%, and final selection with face

Figure B.72: K-stage fusion with finger-iris-face and pre-selection sizes finger=10%, iris=10%, and final selection with face. This k-stage system configurations has a 0.767 EER. Of interest is the basic score fusion (red graph) and the k-stage fusion (purple graph) where the comparative performance shows that the k-stage configuration is slightly worse.

The figure show the similar tendencies. The curvature of the DET for the k-stage configuration show a steady decrease. The curve is comparative to the baselines, however, still with a relatively higher EER score. The CMC have a very low 1. rank probability which is kept low until around 25% rank whereafter a steep increase till 100% probability with a rank at around 26 % (see figure B.72)).

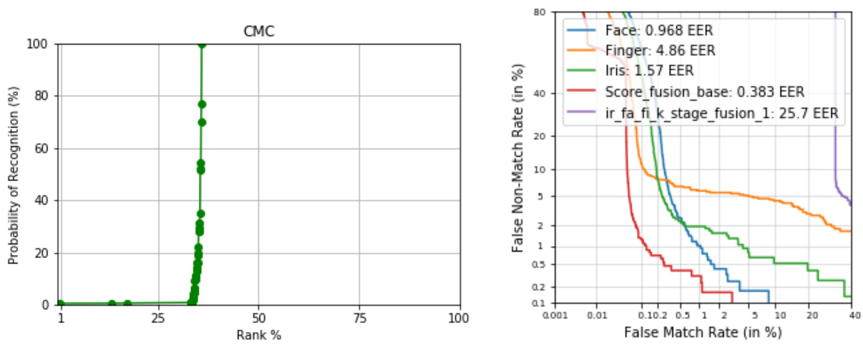


(a) CMC for finger-iris-face and threshold combination (3,3,3) with level-3

(b) DET for finger-iris-face and threshold combination (3,3,3) with level-3

Figure B.73: K-stage fusion with finger-iris-face and threshold combination (3,3,3) with level-3 pre-selection denoting a Bad performance.

The curvature of the DET for the k-stage configuration show a steady decrease with a sudden drop at a relatively higher point. The curve is comparative to the baselines, however, still with a relatively higher EER score. The CMC have a very low 1. rank probability which is kept until around 5% rank whereafter a steep increase till 100% probability with a rank at around 6 % (see figure B.73).

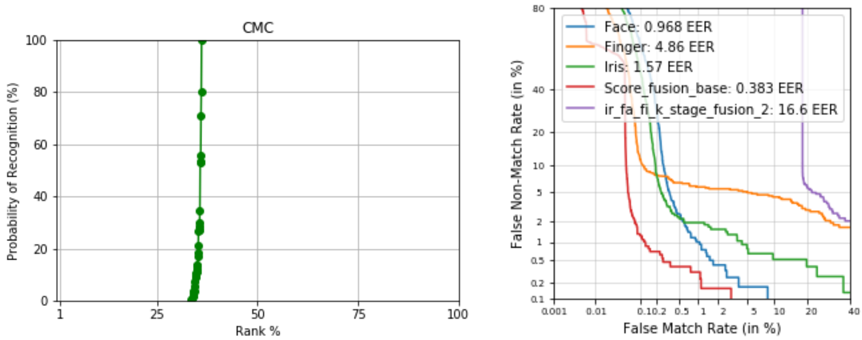


(a) CMC for iris-face-finger and threshold combination (1,1,1) with level-3

(b) DET for iris-face-finger and threshold combination (1,1,1) with level-3

Figure B.74: K-stage fusion with iris-face-finger and threshold combination (1,1,1) with level-3 pre-selection denoting a Bad performance.

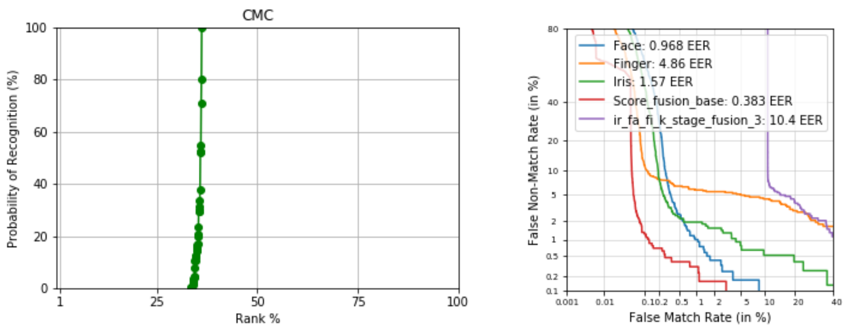
The figure show the similar tendencies.The curvature of the DET for the k-stage



(a) CMC for iris-face-finger and threshold combination (2,2,2) with level-3

(b) DET for iris-face-finger and threshold combination (2,2,2) with level-3

Figure B.75: K-stage fusion with iris-face-finger and threshold combination (2,2,2) with level-3 pre-selection denoting a Bad performance.



(a) CMC for iris-face-finger and threshold combination (3,3,3) with level-3

(b) CMC for iris-face-finger and threshold combination (3,3,3) with level-3

Figure B.76: K-stage fusion with iris-face-finger and threshold combination (3,3,3) denoting a Bad performance.

configuration show similar behavior to the baselines with a sudden increase at a certain point. The curve is comparative to the baselines, however, still with a relatively higher EER score. The CMC have a very low 1. rank probability which is kept until around 30% rank whereafter a steep increase till 100% probability with a rank at around 32 % (see figure B.74, B.75, B.76).

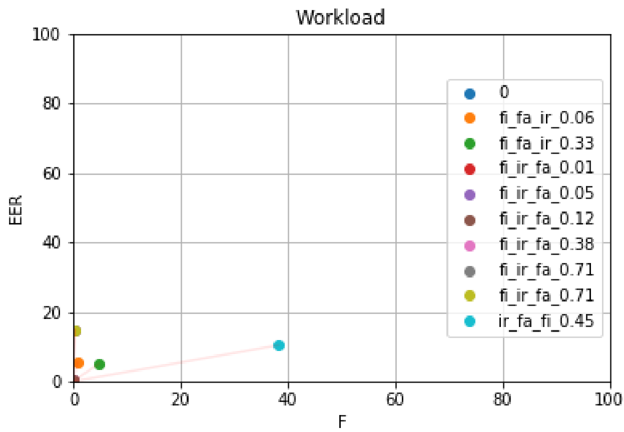


Figure B.77: Full scale Workload reduction presentation using EER as accuracy.

B.3 Workload

This section includes some other iterations of the workload illustrations in the first k-stage experiment (see figure B.77).

Appendix C

Second Experiment results

This section some more of the best/most interesting results of the second k-stage experiment with the synthetic datasets for DET (see figures C.1, C.2, C.3) and its associated CMC (see figure C.4, C.5).

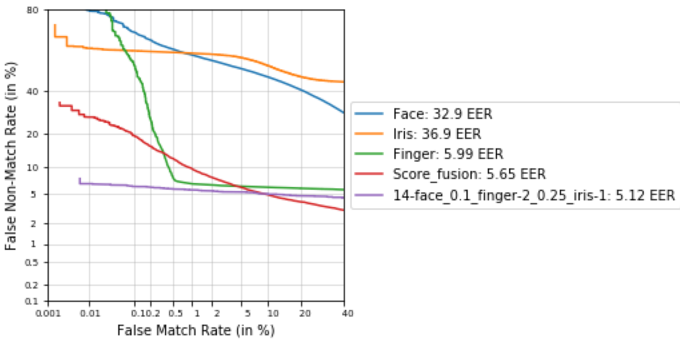


Figure C.1: This configuration have 10% pre-selection with the face modality on the 1.level. It has 25% pre-selection with the finger modality on 2.level. Final selection was done with the iris modality. It can be seen that this configuration denote the lowest False-non rate at false-match rate 0.01 % and that this configuration denote a 5.12 EER.

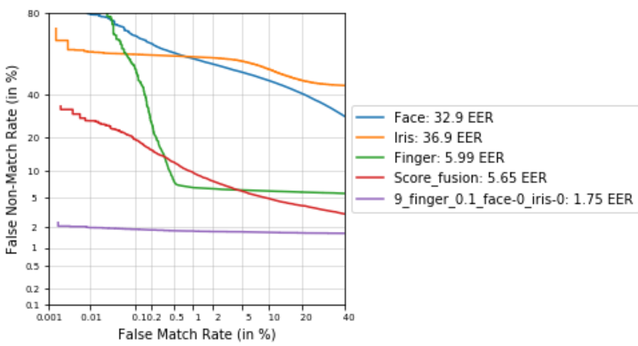


Figure C.2: This configuration have 10% pre-selection with the finger modality on the 1.level. It has final selection was done with the face modality. It can be seen that this configuration denote the lowest False-non rate at false-match rate 0.01 % and that this configuration denote a 1.75 EER.

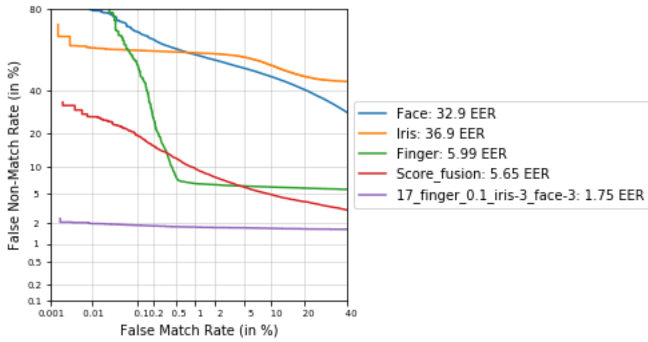


Figure C.3: This configuration have 10% pre-selection with the finger modality on the 1.level. It has final selection was done with the iris modality. It can be seen that this configuration denote the lowest False-non rate at false-match rate 0.01 % and that this configuration denote a 1.75 EER.

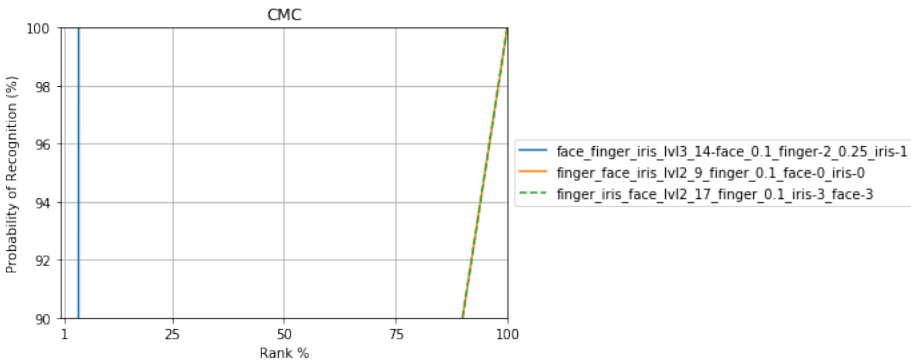


Figure C.4: The full-scale version of a summary illustration of the CMC curves of the second experiment associated to the configurations respective DET curves.

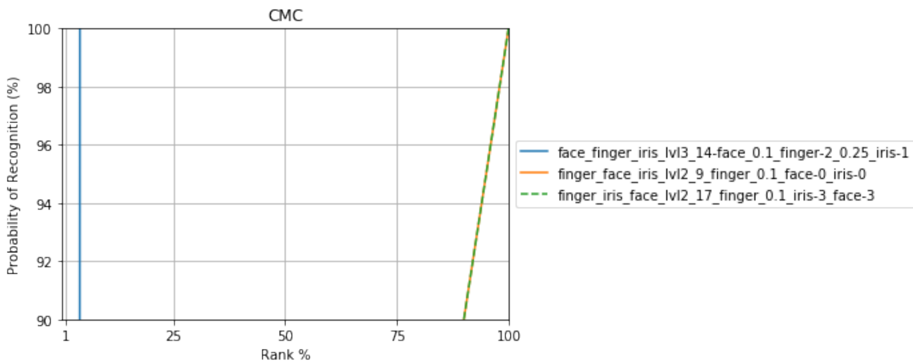


Figure C.5: The zoomed version of a summary illustration of the CMC curves of the second experiment associated to the configurations respective DET curves.

Bibliography

[Aad] Aadhaar, india's biometric id system, gets its day in court. https://spectrum.ieee.org/computing/software/aadhaar-indias-biometric-id-system-gets-its-day-in-court.

[ARJ06] K. Nandakumar A.A. Ross and A.K. Jain. *Handbook of Multibiometrics (International Series on Biometrics)*. Springer-Verlag, Berlin, Heidelberg, 2006.

[BO15] M. Agu B. Ogbuokiri. An enhanced authentication system using face and fingerprint technologies. 17:2278–661, 12 2015.

[BRB+14] S. Billeb, C. Rathgeb, M. Buschbeck, H. Reininger, and K. Kasper. Efficient two-stage speaker identification based on universal background models. pages 1–6, Sept 2014.

[Bus06] C. Busch. Information technology — biometrics tutorial. *ISO/IEC*, 2006.

[Cap04] R. Cappelli. "sfinge: an approach to synthetic fingerprint generation". http://biolab.csr.unibo.it/research.asp?organize=Activities&select=&selObj=12&pathSubj=111%7C%7C12&, 2004.

[C.B16] C.Busch. Nfiq 2.0 - nist fingerprint image quality. *NIST*, 2016.

[CFM10] R. Cappelli, M. Ferrara, and D. Maltoni. Minutia cylinder-code: A new representation and matching technique for fingerprint recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(12):2128–2141, Dec 2010.

[Dau00] J. Daugman. *Biometric Decision Landscapes*. Number UCAM-CL-TR-482. University of Cambridge, Computer Laboratory, 2000.

[Dau04] J. Daugman. How iris recognition works. *IEEE Transactions on Circuits and Systems for Video Technology*, 14(1):21–30, Jan 2004.

[DH17] L. M. Dinca and G. P. Hancke. The fall of one, the rise of many: A survey on multi-biometric fusion methods. *IEEE Access*, 5:6247–6289, 2017.

[Dp11] inc Open Source Eddition Digital persona. Fingerjetfx ose - fingerprint feature extractor. `https://github.com/FingerJetFXOSE/FingerJetFXOSE`, 2011.

[DR13] B. DeCann and A. Ross. Relating roc and cmc curves via the biometric menagerie. pages 1–8, Sept 2013.

[DRB17] P. Drozdowski, C. Rathgeb, and C. Busch. Multi-iris indexing and retrieval: Fusion strategies for bloom filter-based search structures. pages 46–53, Oct 2017.

[DRB18a] P. Drozdowski, C. Rathgeb, and C. Busch. Bloom filter-based search structures for indexing and retrieving iris-codes. *IET Biometrics*, 7(3):260–268, 2018.

[DRB18b] P. Drozdowski, C. Rathgeb, and C. Busch. Bloom filter-based search structures for indexing and retrieving iris-codes. *IET Biometrics*, 7(3):260–268, 2018.

[DRBK17] N. Damer, C. I. Rhaibani, A. Braun, and A. Kuijper. Trust the biometrie mainstream: Multi-biometric fusion and score coherence. pages 2191–2195, Aug 2017.

[Dro17] P. Drozdowski. Sic-gen: A synthetic iris-code generator. `https://github.com/dasec/sic-gen`, 2017.

[DTK+13] A. Drosou, A. Tsimpiris, D. Kugiumtzis, N. Porfyriou, D. Ioannidis, and D. Tzovaras. Dimensionality reduction for enhanced 3d face recognition. pages 1–8, July 2013.

[fBC05] Center for Biometrics and Security Research (CBSR). Casia. `http://www.cbsr.ia.ac.cn/`, 2005.

[GDR98] A. Martin M. Przybocki G. Doddington, W. Liggett and D. Reynolds. Sheep, goats, lambs and wolves: A statistical analysis of speaker performance in the nist 1998 speaker recognition evaluation. *Proc. Int'l Conf. Spoken Language Processing*, 1998.

[GRC09] J. E. Gentile, N. Ratha, and J. Connell. An efficient, two-stage iris recognition system. pages 1–5, Sept 2009.

[HHF+10] M. He, S.J. Horng, P. Fan, R.S. Run, R.J. Chen, J.-L. Lai, M. K. Khan, and K. O. Sentosa. Performance evaluation of score level fusion in multimodal biometric systems. *Pattern Recognition*, 43(5):1789 – 1800, 2010.

[IBMN18] G. Iovane, C. Bisogni, L. De Maio, and M. Nappi. An encryption approach using information fusion techniques involving prime numbers and face biometrics. *IEEE Transactions on Sustainable Computing*, pages 1–1, 2018.

[IEE18] Ieee-747 double precision. *IEEE-747*, 2018.

[IK17] C. Bhagvati I. Kavati, M. V. N. K. Prasad. Search space reduction in biometric databases: A review. *IGI-Global*, 2017.

[iso06] Information technology - biometric application programming interface, 2006.

[ISO11] International standard performance testing and reporting., 2011.

[J.F18a] R.Vera-Rodriquez D. Camacho J.Fierrez, A. Morales. Multiple classifiers in biometrics. part 1: Fundamentals and review. *Information Fusion*, 44:57 – 64, 2018.

[J.F18b] R.Vera-Rodriquez D. Camacho J.Fierrez, A. Morales. Multiple classifiers in biometrics. part 2: Trends and challenges. *Information Fusion*, 44:103 – 112, 2018.

[KD18] N. Kvitha-Devi. Authentication using multimodal biometric features. *International Journal of Computer Science and Mobile Computing*, 7(1):1–8, 2018.

[lfw07] Labeled faces in the wild home. `http://vis-www.cs.umass.edu/lfw/`, 2007.

[MEA12] C. Charrier M. El-Abed. Evaluation of biometric systems. new trends and developments in biometrics. *Hal-archives*, 1:149 – 169, 2012.

[nis] Nist. `https://www.nist.gov/itl/iad/image-group/resources/biometric-special-databases-and-software#Fingerprint`.

[NMR+17] A. Nautsch, D. Meuwly, D. Ramos, J. Lindh, and C. Busch. Making likelihood ratios digestible for cross-application performance assessment. *IEEE Signal Processing Letters*, 24(10):1552–1556, Oct 2017.

[OA17] O.Ayodele and A. Adegbenjo. Current practices in information fusion for multimodal biometrics. *American Journal of Engineering Research (AJER)*, 6(4), 2017.

[oB15] DISI University of Bologna. Biometric system laboratory. `http://biolab.csr.unibo.it/research.asp?organize=Activities&select=&selObj=12&pathSubj=111\delimiter"026B30D\delimiter"026B30D12&`, 2015.

[OP15] A. Zisseman O.M Pakhai, A. Vedaldi. Deep face recognition. *British Machine Vision Conference 2015*, 1:41.1–41.12, 01 2015.

[Oth11] N. Othman. A biometric reference system for iris osiris version 4.1. *Biosecure*, 2011.

[Pro13] H. Proencça. Iris biometrics: Indexing and retrieving heavily degraded data. *IEEE Transactions on Information Forensics and Security*, 8(12):1975–1985, Dec 2013.

[PRSB15] A. Pflug, C. Rathgeb, U. Scherhag, and C. Busch. Binarization of spectral histogram models: An application to efficient biometric identification. pages 501–506, June 2015.

[RBBB15] C. Rathgeb, F. Breitinger, H. Baier, and C. Busch. Towards bloom filter-based indexing of iris biometric data. pages 422–429, May 2015.

[RD18] S. Dey R-Dwivedi. A novel hybrid score level and decision level fusion scheme for cancelable multi-biometric verification. *CoRR*, abs/1805.10433, 2018.

[RG05] Arun. Ross and R. Govindarajan. Feature level fusion of hand and face biometrics. 5779, 03 2005.

[RGR18] D. Ramos and J. Gonzalez-Rodriguez. Cross-entropy analysis of the information in forensic speaker recognition. 07 2018.

[RJ03] A. Ross and A. Jain. Information fusion in biometrics. *Pattern Recognition Letters*, 24(13):2115 – 2125, 2003. Audio- and Video-based Biometric Person Authentication (AVBPA 2001).

[Ros07] A. Ross. An introduction to multibiometrics. pages 20–24, Sept 2007.

[Ros10] A. Ross. Iris recognition: The path forward. *Computer*, 43(2):30–35, Feb 2010.

[RR15] L. Van Gool R. Rothe, R. Timofte. Imdb-wiki – 500k+ face images with age and gender labels. `https://data.vision.ee.ethz.ch/cvl/rrothe/imdb-wiki/`, 2015.

[San] D. Sandberg. `https://github.com/davidsandberg/facenet`.

[S.B18] S.Barde. Multimodal biometrics: Most appropriate for person identification. *i-manager's Journal on Pattern Recognition*, 2018.

[SKP15] F. Schroff, D. Kalenichenko, and J. Philbin. Facenet: A unified embedding for face recognition and clustering. *CoRR*, abs/1503.03832, 2015.

[SUM+05] R. Snelick, U. Uludag, A. Mink, M. Indovina, and A. Jain. Large-scale evaluation of multimodal biometric authentication using state-of-the-art systems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(3):450–455, March 2005.

[Sus16] N. Susyanto. Semiparametric copula models for biometric score level fusion. *UvA-DARE*, 05 2016.

[tse] The superior electoral tribunal's biometric authentication server. `http://revistapesquisa.fapesp.br/en/2014/06/18/large-scale-fingerprint-recognition-systems/`.

[YE13] R.Adjoudj Y. Elmir, Z. Elberrichi. A hierarchical fusion strategy based multimodal biometric system. *ACIT*, 12 2013.

[ZI96] Y. Zuev and S. Ivanon. The voting as a way to increase the decision reliability, in foundations of information/ decision fusion with applications to engineering problems. *IJITKM*, 1996.