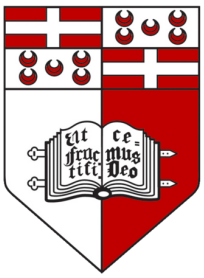
Face Photo-Sketch Recognition using Deeply-Learned and Engineered Features

Christian Galea

Supervised by Dr. Ing. Reuben A. Farrugia



Department of Communications & Computer Engineering

Faculty of Information & Communication Technology

University of Malta

January 2018

Submitted in partial fulfilment of the requirements for the degree of Ph.D. in ICT

L-Università ta' Malta

University of Malta Library – Electronic Thesis & Dissertations (ETD) Repository

The copyright of this thesis/dissertation belongs to the author. The author's rights in respect of this work are as defined by the Copyright Act (Chapter 415) of the Laws of Malta or as modified by any successive legislation.

Users may access this full-text thesis/dissertation and can make use of the information contained in accordance with the Copyright Act provided that the author must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the prior permission of the copyright holder.

Plagiarism Declaration

Plagiarism is defined as "the unacknowledged use, as one's own, of work of another person, whether or not such work has been published, and as may be further elaborated in Faculty or University guidelines" (University Assessment Regulations, 2009, Regulation 39 (b)(i), University of Malta).

I, the undersigned, declare that the dissertation submitted is my work, except where acknowledged and referenced.

I understand that the penalties for committing a breach of the regulations include loss of marks; cancellation of examination results; enforced suspension of studies; or expulsion from the degree programme.

Work submitted without this signed declaration will not be corrected, and will be given zero marks.

Student Name & Surname: Christian Galea *Signature:* 

Title of work submitted: Face Photo-Sketch Recognition using Deeply-Learned and Engineered Features

Date: 26th January 2018

Declaration of Authenticity

Student's I.D./Code: 258291M

Student Name & Surname: Christian Galea

Title of Dissertation: Face Photo-Sketch Recognition using Deeply-Learned and Engineered Features

I hereby declare that I am the legitimate author of this Dissertation and that it is my original work.

No portion of this work has been submitted in support of an application for another degree or qualification of this or any other university or institution of higher education.

I hold the University of Malta harmless against any third party claims with regard to copyright violation, breach of confidentiality, defamation and any other third party right infringement.

As a Ph.D. student, as per Regulation 49 of the Doctor of Philosophy Regulations, I accept that my thesis be made publicly available on the University of Malta Institutional Repository.

Signature: 

Date: 26th January 2018

Copyright Notice

1. Copyright in text of this dissertation rests with the Author. Copies (by any process) either in full, or of extracts may be made **only** in accordance with regulations held by the Library of the University of Malta. Details may be obtained from the Librarian. This page must form part of any such copies made. Further copies (by any process) made in accordance with such instructions may only be made with the permission (in writing) of the Author.

2. Ownership of the right over any original intellectual property which may be contained in or derived from this dissertation is vested in the University of Malta and may not be made available for use by third parties without the written permission of the University, which will prescribe the terms and conditions of any such agreement.

Abstract

Face sketches created from eyewitness descriptions of criminals have proven to be useful in assisting law enforcement agencies to apprehend perpetrators, particularly in cases lacking evidence. These sketches are typically disseminated to the public and law enforcement officers so that any persons recognising the suspect in the sketch may come forward with information leading to an arrest. However, this process is time consuming and not guaranteed to be successful. In this dissertation, an investigation of popular and state-of-the-art face photo-sketch synthesis and recognition methods which can identify perpetrators automatically is performed using an evaluation set-up that reflects real-world scenarios, through the use of challenging sketches and an extended gallery which simulates the extensive mug-shot galleries maintained by law enforcement agencies. The University of Malta Software-Generated Face Sketch (UoM-SGFS) database was also created to enable the design and evaluation of algorithms when using software-generated sketches, that are nowadays being used more often than hand-drawn sketches. This database is the largest software-generated face sketch database, one of the few containing multiple sketches per subject, and the only one containing sketches represented in colour. Several novel methods have also been designed and evaluated, namely: (i) the Eigenpatches (EP) approach which improves upon the performance of the popular Eigentransformation (ET) method by transforming photos into sketches or sketches into photos on a local level, (ii) the log-Gabor-MLBP-SROCC (LGMS) method that extracts modality-invariant features, (iii) the DEEP (face) Photo-Sketch System (DEEPS) framework that applies transfer learning to a state-of-the-art face recognition system based on a Deep Convolutional Neural Network (DCNN) with the aid of an extensive set of synthetic images created using a 3D morphable model, (iv) the use of multiple synthetic sketches during system deployment, and (v) the fusion of intra- and inter-modality methods which are shown to be capable of providing complementary information. The finalised system fuses LGMS with DEEPS to yield a system outperforming state-of-the-art methods for all types of sketches, including real-world forensic sketches. Moreover, the proposed approach is efficient in terms of both computation time and template size, thereby permitting its implementation in the real-world.

Acknowledgements

Several persons have provided me with continuous support throughout the duration of the project. In particular, I would like to thank Dr. Ing. Reuben A. Farrugia for his patience, time and dedication in providing me with invaluable guidance during the entire course of this thesis.

I would also like to thank the staff within the Department of Communications and Computer Engineering for their help provided during this project, in particular Ing. Maria Abela Scicluna who helped me use resources to speed up implementation and testing of the system developed. Heartfelt gratitude also goes towards Dr. Keith Bugeja and Dr. Alessio Magro for providing two NVIDIA Tesla K20c GPUs.

Last but certainly not least, heartfelt gratitude goes to my family for their constant support and encouragement provided during my studies.

This research was partially funded by the Malta Government Scholarship Scheme (MGSS), and done in collaboration with the Malta Police Force who are thanked for their assistance. NVIDIA Corporation also donated an NVIDIA Titan X Pascal GPU for use in this research.

Publications

Published papers

C. Galea and R. A. Farrugia, "Fusion of intra- and inter-modality algorithms for face-sketch recognition," in *Computer Analysis of Images and Patterns*, ser. Lecture Notes in Computer Science, G. Azzopardi and N. Petkov, Eds. Springer International Publishing, Sep. 2015, vol. 9257, pp. 700–711.

C. Galea and R. A. Farrugia, "Face Photo-Sketch Recognition using Local and Global Texture Descriptors," in *European Signal Processing Conference (EUSIPCO)*, Aug. 2016.

C. Galea and R. A. Farrugia, "A Large-Scale Software-Generated Face Composite Sketch Database," in *Int. Conf. Biometrics Special Interest Group (BIOSIG)*, Sep. 2016, pp. 1–5.

C. Galea and R. A. Farrugia, "Forensic Face Photo-Sketch Recognition using a Deep Learning-based Architecture," *IEEE Signal Processing Letters*, vol. 24, no. 11, pp. 1586–1590, Nov. 2017.

C. Galea and R. A. Farrugia, "Matching Software-Generated Sketches to Face Photos with a Very Deep CNN, Morphed Faces, and Transfer Learning," *IEEE Transactions on Information Forensics and Security*, accepted for publication.

Publications in other research fields

R. A. Farrugia, C. Galea, S. Zammit, and A. Muscat, "Objective Video Quality Metrics for HDTV Services: A Survey" in *Proceedings of IEEE EUROCON2013*, pp. 170–176, 2013.

C. Galea and R. A. Farrugia, "A No-Reference Video Quality Metric Using a Natural Video Statistical Model," in *Proceedings of IEEE EUROCON2015*, Sep. 2015.

R. A. Farrugia, C. Galea and C. Guillemot, "Super Resolution of Light Field Images using Linear Subspace Projection of Patch-Volumes," *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 7, pp. 1058-1071, Oct. 2017.

Links to these papers and other resources may be found at
`https://cnpgs.wordpress.com`.

Contents

List of Figures

List of Tables

List of Abbreviations

ANOVA Analysis of Variance.

BI Bayesian Inference.

CBR Component-Based Representation.

CCA Canonical Correlation Analysis.

CCTV Closed-Circuit Television.

CMC Cumulative Match Curve.

COTS Commercial Off-the-Shelf.

CPU Central Processing Unit.

cuDNN NVIDIA CUDA Deep Neural Network.

D-RS Direct Random Subspaces.

DCNN Deep Convolutional Neural Network.

DEEPS DEEP (face) Photo-Sketch System.

DEEPS-M DEEPS Multi-sketch.

DoG Difference of Gaussian.

E-HMM Embedded Hidden Markov Model (HMM).

E-HMMI Embedded HMM (E-HMM) Inversion.

EER Equal Error Rate.

EP Eigenpatches.

ET Eigentransformation.

FAR False Accept Rate.

FH Face Hallucination.

FR Face Recognition.

FRS Face Recognition System.

FSIM Feature SIMilarity index.

FSR Face Super-Resolution.

GPU Graphical Processing Unit.

HAOG Histogram of Averaged Orientation Gradients.

HFR Heterogeneous Face Recognition.

HIM Histogram of Image Gradients.

HMM Hidden Markov Model.

HOG Histogram of Orientation Gradients.

HVS Human Visual System.

IQA Image Quality Assessment.

IR Infra-Red.

K-NN K-Nearest Neighbour (NN).

KNDA Kernel-based Non-linear Discriminant Analysis.

LBP Local Binary Pattern.

LDA Linear Discriminant Analysis.

LFDA Local Feature-based Discriminant Analysis.

LGMS log-Gabor-MLBP-SROCC.

LLE Locally Linear Embedding.

MAP Maximum a Posteriori.

MLBP Multiscale Local Binary Pattern (LBP).

MRF Markov Random Field.

MSE Mean Square Error.

NIR Near Infra-Red (IR).

NN Neural Network.

NN Nearest Neighbour.

NS Normalised Sum.

P-RS Prototype Random Susbspaces.

PCA Principal Component Analysis.

ReLU Rectified Linear Unit.

RMSE Root Mean-Square Error.

ROC Receiver Operating Characteristics.

RS-LDA Random Sampling Linear Discriminant Analysis (LDA).

SGD Stochastic Gradient Descent.

SIFT Scale-invariant Feature Transform.

SL Subspace Learning.

SROCC Spearman Rank Order Correlation Coefficient.

SSIM Structural SIMilarity index.

SVM Support Vector Machine.

t-SNE t-Distributed Stochastic Neighbour Embedding.

TAR True Accept Rate.

TDE Triplet Distance Embedding.

TSE Triplet Similarity Embedding.

UIQI Universal Image Quality Index.

UoM-SGFS University of Malta Software-Generated Face-Sketch.

VIS visible band.

Chapter 1

Introduction

The term *biometric* originates from the Greek words *bios* and *metron*, meaning life and measurement, respectively. As a result, *biometrics* may be defined as the measurements of living human bodies, or more formally as the use of unique physical or behavioural traits able to distinguish one person from another. Applications range from security to forensics and even aid convenience by eliminating the need to remember passwords or possess identification cards. Biometric traits commonly used include fingerprints, hand geometry, iris, voice, and the face, each having different strengths and weaknesses which determine their suitability for different applications [1–5].

The face is a biometric trait that can be captured non-intrusively, at a distance, and without the consent of a person, which has lead to its wide adoption in law enforcement and surveillance and security applications [4,6–8]. Automated Face Recognition Systems (FRSs) have therefore been employed to allow faster, more reliable and more efficient identification of individuals.

There exist several types of FRSs, with traditional algorithms typically operating on photos captured with a digital camera. However, recent research has focused on FRSs which process face images spanning different image modalities. One important use of such algorithms is when no evidence is available at the scene of the crime, except for the account of an eyewitness. In such cases, a forensic sketch artist works with the eyewitness to create a face sketch which resembles as closely as possible the face of the perpetrator. This is then disseminated to media outlets and law enforcement officers so that anyone recognising the perpetrator can come

forward with information leading to an arrest. However, this process is slow, not guaranteed to be successful and does not utilise available resources. A system that automatically matches a sketch to a gallery of mug-shot photos maintained by law enforcement agencies would thus be of great benefit in such situations [9].

1.1 Background to Face Recognition

A FRS may be defined as a system that automatically processes face images with the aim of identifying persons [5], with applications spanning from surveillance and security to law enforcement. More specifically, FRSs compare a person's face image, typically captured in the visible band (VIS) (i.e. a photograph or frame extracted from a video, captured by a normal digital camera) to a set of VIS face photos in a database of similar quality to try and determine the person's identity [1,3,5,10,11], as shown in Figure 1.1. Since comparisons are being made between images of the same modality, such algorithms can be labelled as *homogeneous* FRSs. Although homogeneous Face Recognition (FR) includes any scenario where images of the same modality are being compared, it is usually the case that comparisons are being made between VIS images.

On the other hand, *heterogeneous* FRSs attempt to compare images spanning different modalities, e.g. a Near Infra-Red (IR) (NIR) image to a VIS image (NIR-VIS matching) and a thermal Infra-Red (IR) image to a VIS image (Thermal-VIS matching). Research in this type of systems has only emerged recently, to offer

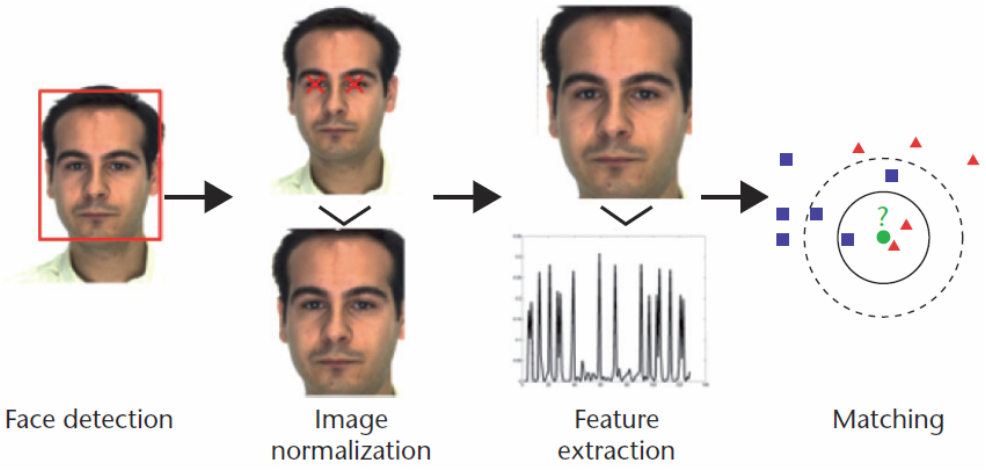


Figure 1.1: Main steps required in an automatic FRS [7]

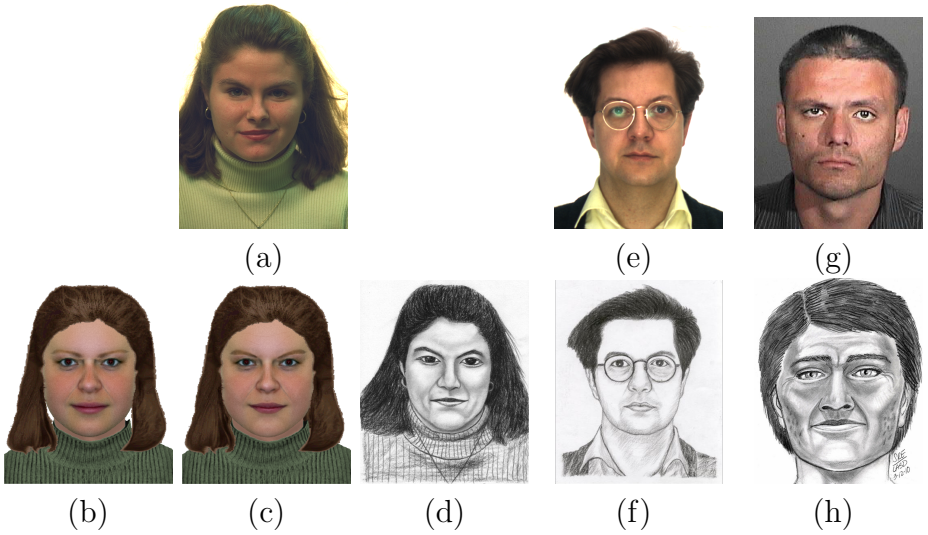


Figure 1.2: Examples of VIS photos and sketches used in this work: (a) Photo of a subject in the Color FERET database [13] and the corresponding viewed software-generated sketches in (b) Set A and (c) Set B of the UoM-SGFS database [14] and (d) the corresponding viewed hand-drawn sketch in the CUFSF database [15], (e) photo of a subject in the AR database [16] and (f) the corresponding viewed hand-drawn sketch in the CUFS database [17,18], (g) photo and (h) corresponding forensic hand-drawn sketch of a subject in the PRIP-HDC database [19].

solutions in numerous face recognition scenarios where the *query* image modality does not match the *gallery* modality, the latter containing face images with known identities. One example is when NIR images captured from Closed-Circuit Television (CCTV) cameras at night need to be compared to a gallery dataset containing VIS images. Indeed, although Heterogeneous Face Recognition (HFR) is defined as the matching of face images captured in any two modalities, typically the gallery dataset images contain VIS images [8,12].

Another HFR application involves the matching of photographs with sketches obtained from eyewitness descriptions of a criminal (forensic sketch-VIS matching). Sketches can either be drawn by artists, in which case they are called *hand-drawn sketches*, or generated using computer software such as IdentiKit [20], FACES [21] and EFIT-V [22], where they are called *software-generated composite sketches* [9]. Examples of such images are shown in Figure 1.2. In contrast to other HFR tasks, face photo-sketch recognition must not only contend with the *modality gap* but also other challenges that introduce several distortions and exaggerations in the texture and structure of sketches when compared to the corresponding photographs [10,19,23–25]:

- *Modality Gap*: photos and sketches bear inherent differences with respect to

their texture, since photos are captured in the real natural environment in VIS light whereas sketches are un-natural 'synthetic' images that are hand-drawn or computer-generated.

- *Memory Gap*: a sketch may bear incomplete information regarding the subject, arising from the fact that eyewitnesses generally cannot exactly recall the details of a suspect's face due to memory loss effects. In addition, the face shape might be exaggerated especially if the suspect has any distinguishing characteristics, and texture might be lost or replaced.

- *Communication Gap*: Mis-communication between eyewitness and sketch artist or difficulty in describing certain details may lead to inaccurate details in the sketch.

- *Other factors*: people tend to have difficulty in recognising and processing faces of people belonging to races other than their own (known as the "other-race effect", and affects both eyewitnesses and forensic artists), and the styles of hand-drawn sketches vary among different artists.

Due to the above challenges, sketches often do not resemble closely the corresponding VIS photographs [12]. Viewed sketches drawn by artists whilst viewing a subject or a photo of the subject have also been used in research as a stepping stone to forensic sketches since they are easier to obtain, but are typically quite accurate compared to the corresponding photographs since they mainly cater for the modality gap only, and hence do not reflect real-world conditions. However, these sketches typically do contain some deformations which still make their comparison with face photographs challenging (*viewed sketch*-VIS matching). Since forensic sketches are typically distributed to the media and shown to the public in the hope that any persons recognising the person in the sketch come forward with information leading to an arrest, the time taken to identify a suspect can take several weeks (if at all successful). Hence, automated algorithms capable of forensic sketch-VIS matching can increase both the speed and success rates of criminal apprehensions [7,9,19,26].

It can be argued that even VIS-VIS matching is not an easy task [1,7,8,27], due to similarity between faces (e.g. identical twins), ageing and other challenges such as different pose, lighting, and long distances between camera and subject as shown in Figure 1.3. Consequently, much research has been focused on these areas to

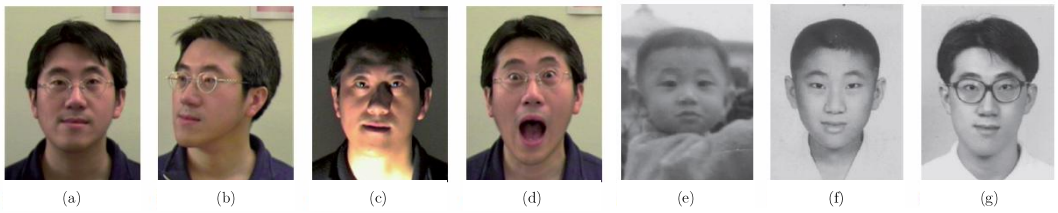


Figure 1.3: The four main challenges in an FRS: (a) image taken under ideal conditions with different (b) pose, (c) illumination, (d) expression and (e)-(g) ageing variations (subject in (a) is (e) 32, (f) 21 and (g) 15 years younger) [7].

improve FR when photos are not taken under ideal conditions (frontal pose, good lighting and close distance to camera, i.e. unconstrained FR). For example, Face Super-Resolution (FSR) is an area which is dedicated to solving the problem of capturing low-resolution face images (e.g. from CCTV cameras), where approaches attempt to approximate high-resolution images from low-resolution images [25] (whilst retaining important facial details that are vital for robust FR which may have been lost due to the low resolution). Due to the challenges involved in FR, it has been argued that FRSs do not replace humans, but rather augment their capabilities by first employing FRSs to retrieve the top K matches from a database containing thousands of subjects, where K is typically between 50 to 250 [6,7,12, 19,26], and then allowing humans to examine this small set of matches to identify the subject [7,28] as shown in Figure 1.4. This is especially applicable in cases involving challenging probe images and forensics applications where mistakes are

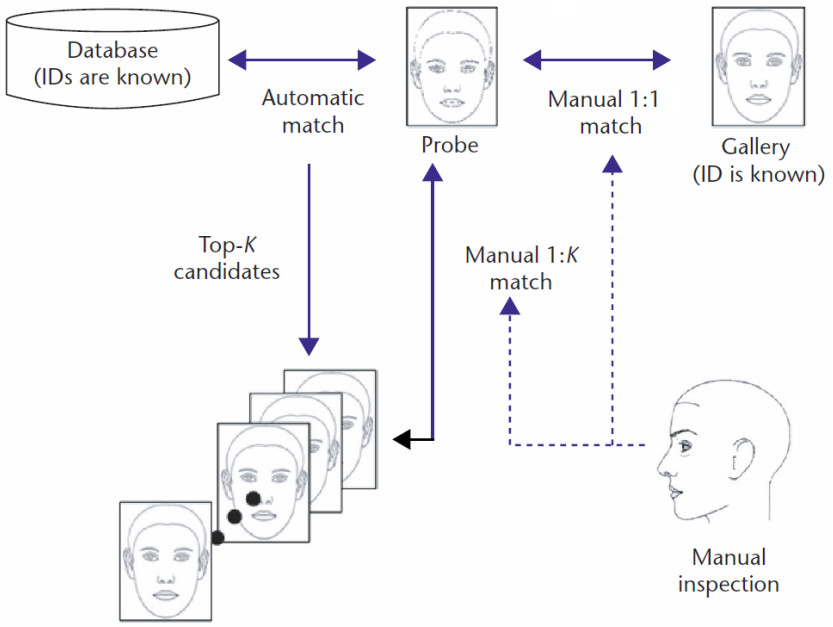


Figure 1.4: Forensic FR, a process that is not fully automatic but requires manual inspection of the top K retrieved matches from a large gallery of mug-shot photos [7].

critical, and consequently lead courts of justice to typically require some manual analysis to be involved when gathering evidence against a criminal.

1.2 Outline of the Dissertation

The rest of this dissertation is organised as follows: an overview of related work is given in Chapter 2 followed by a description of work done in this research in Chapters 3 and 4. The methodology used to analyse the performance of the proposed methods and leading algorithms proposed in literature is given in Chapter 5 followed by the corresponding results in Chapter 6. Concluding remarks are given in Chapter 7 followed by a discussion of proposed future research in Chapter 8. More in-depth analyses of the proposed methods are also provided, in Appendix A, while additional results are given in Appendix B. Images depicted in this dissertation are best viewed in colour and on a screen.

Chapter 2

Literature Review

There exist several types of Face Recognition Systems (FRSs), as introduced in Chapter 1. This chapter will give a review of the state-of-the-art work published to tackle the problem of Heterogeneous Face Recognition (HFR), with a focus on the area of face-sketch recognition.

An introduction to the main approaches adopted in developing HFR systems is given in Section 2.1 followed by a review of the two main types of algorithms, namely intra- and inter-modality algorithms, in Sections 2.2 and 2.3, respectively. An overview of other sketch-based recognition tasks and of deep learning are given in Section 2.4 and Section 2.5, respectively. A summary and concluding remarks are finally given in Section 2.6.

2.1 Background to Heterogeneous Face Recognition

Heterogeneous FRSs, which compare face images in different modalities, can be classified as either *inter*- or *intra*-modality techniques [6,10] as shown in Figure 2.1. *Inter-modality* techniques are practically FRSs designed specifically to compare face images in different modalities, by extracting features such as Scale-invariant Feature Transform (SIFT) [29,30] and Local Binary Pattern (LBP) [31] from the images to be compared and using the features themselves to determine the image similarities.

Intra-modality techniques attempt to reduce the modality gap by transforming one image into the domain of the other image. For example, in sketch-visible band (VIS) matching, a sketch (or photo) may be transformed to a photo (sketch) and the transformed sketch (photo), also known as a pseudo-photo (pseudo-sketch), is then compared to a database containing photos (sketches) using any FRS.

Face Hallucination (FH) techniques encompass methods where new face images are generated from other face images. One such group of techniques are known as Face Super-Resolution (FSR) methods, which generate high-resolution face images from low-resolution face images. FH also encompasses face photo-sketch synthesis (i.e. intra-modality methods) [25] since a probe image needs to be transformed to another image in both cases. The main advantage of such approaches is that any FRS may be used for person identification once an image is synthesised. This is especially useful when highly-performing FRSs which have been designed to cater for effects such as ageing and illumination in the target modality are available. However, intra-modality techniques are typically complex and require a significant amount of time to generate a synthesised image which then needs to be compared to images in the gallery by a FRS, further prolonging the time required to identify the person in the query image. As a result, the authors of [6,15] argue that intra-modality algorithms typically attempt to solve a more complex problem than the recognition task. Moreover, the performance of the FRS employed depends on the accuracy of the synthesised images. In fact, the synthesised images typically contain artefacts which inhibit the face recogniser's performance.

Recently, there has been a shift towards inter-modality approaches, which are practically specialised modal-insensitive FRSs. Inter-modality methods extract modality-invariant features from both photos and sketches such that inter-class separability is maximised while maintaining intra-class differences. Hence, the modality gap is reduced at the *classification stage*, whereas the modality gap is reduced at the *pre-processing stage* in intra-modality methods. Techniques to learn feature representations may also be implemented [6]. Despite being faster than intra-modality approaches, they are typically not designed to cater for challenges encountered in real-life scenarios such as ageing and illumination, i.e. they are designed for photos taken under ideal conditions (frontal pose, neutral expression, uniform lighting) and cannot take advantage of any FRSs designed to cater for these challenges in the required modalities.

Figure 2.1: Tree diagram for the different categories of HFR algorithms. Taxonomy for intra-modality methods based on that given in [25].

An overview of popular and state-of-the-art algorithms proposed in literature for each of these two types of categories will now be given in the following Sections.

2.2 Intra-modality approaches

Intra-modality approaches aim to transform two images occupying different modalities into a common modality, which is either one of the modalities of the original images or another modality entirely. A Face Recognition System (FRS) capable of matching the images in the transformed domain can then be used for recognition.

Face Hallucination techniques can be generally categorised as using either (i) the Subspace Learning (SL) framework, (ii) the Bayesian Inference (BI) framework, (iii) a combination of BI and SL, or (iv) Sparse Representation [25]. In the rest of this section, FH techniques will be discussed with an emphasis on literature tackling face-sketch synthesis. A more detailed review of FH techniques may be found in [25].

2.2.1 Subspace Learning framework

Subspace learning approaches aim to find a subspace embedded within a higher dimensionality subspace such that a projection matrix calculated by solving a standard eigenvalue decomposition problem from training data samples can then be used to project samples into the lower-dimensional subspace.

The Eigenfaces approach proposed in [37] was used in [17,33] to create the Eigen-transformation (ET) method, whereby photos are transformed to sketches by assuming that the weights representing a reconstructed photo are identical to those obtained if a reconstructed sketch were to be created. More specifically, in the traditional Eigenfaces approach [37], face photos are reconstructed by using a weighted summation of a set of eigenvectors $\mathbf{U}_p$ representing faces (consequently called Eigenfaces) as follows:

$$\vec{P_r} = \mathbf{U}_p \vec{b_p} \tag{2.1}$$

where $\vec{P}_r$ is a column vector representing the reconstructed face photo and $\vec{b}_p$ is a vector containing the projection coefficients in the eigenvector space. As demonstrated in [17], (2.1) may be rewritten as follows:

$$\vec{P}_r = \vec{\bar{P}} + \sum_{i=1}^{M} \vec{c}_p^{\{i\}} \vec{\Phi}^{\{i\}} \tag{2.2}$$

where $\vec{\bar{P}}$ is the mean face computed over M training face images, $\vec{\Phi}^{\{i\}} = \vec{P}^{\{i\}} - \vec{\bar{P}}$ is the centred face and $\vec{c}_p^{\{i\}}$ is a column vector of dimension M representing the contribution of the i^{th} training image $\vec{P}^{\{i\}}$ in the reconstruction of a test face image computed according to [17]. Hence, (2.2) shows that a reconstructed photo can be approximated to the original image using a weighted linear addition of the training images [17]. Since a photo and corresponding sketch should also be similar in terms of structure, (2.2) may be modified such that the training photos $\vec{P}^{\{i\}}$ are simply replaced by the corresponding training sketches $\vec{S}^{\{i\}}$, as follows:

$$\vec{S}_r = \vec{\bar{S}} + \sum_{i=1}^{M} \vec{c}_p^{\{i\}} \vec{\Psi}^{\{i\}} \tag{2.3}$$

where $\vec{S}_r$ is the reconstructed sketch, $\vec{\bar{S}}$ is the mean sketch, $\vec{\Psi}^{\{i\}} = \vec{S}^{\{i\}} - \vec{\bar{S}}$ and $\vec{S}^{\{i\}}$ is a column vector representing the i^{th} sketch. This is based on the hypothesis that if a photo contributes more weight to a reconstructed face photo, then the corresponding sketch will also contribute more weight to the reconstructed sketch [17].

It was shown that the Eigentransformation approach [17,33] outperforms the geometrical method (evaluating 26 measures of geometrical distances between fiducial points in a face) and using Eigenfaces as a face recogniser (performing recognition by treating a probing sketch as if it were a normal photo). Photo-to-sketch transformation was found to yield better performance than sketch-to-photo transformation since photos contain more detail and hence information is being compressed into a more compact representation in the former approach. Therefore, it is easier to convert a photo to a sketch, since in sketch-to-photo synthesis the more difficult operation of enlarging a compact representation to a full representation is being performed. Lastly, the performance of the proposed approach was shown to be superior to that of human beings in recognising sketches and shows that machines

can be used for automatic database searching using sketches as the probe images.

The authors of [38] also transform photos to pseudo-sketches using the Eigentransformation approach proposed in [17,33], but the algorithm is applied on both texture and shape separately to better satisfy the assumption that photos and sketches can be transformed using a linear representation of a training photo subset. Face photos and sketches are first normalised by representing them using a graph based on coordinates of fiducial points. The points of images in the training set are used to create a mean face shape which is then used such that face and sketch images are warped to this mean shape using affine interpolation via a set of triangles. Therefore, face images are aligned and texture and shape information can be extracted, followed by application of Eigentransformation on these two features. The generated texture is then warped to the generated sketch shape to yield the final synthesised pseudo-sketch. Recognition is performed using a Bayesian classifier using shape and texture vectors, trained using the shape/texture vectors obtained from real sketch and pseudo-sketches. The training set used to obtain the probabilistic subspace was different than that used to obtain the Eigentransformation coefficients. A total of 606 subjects from the CUFS database [39] were considered. It was shown that the proposed system outperformed Eigenfaces (Principal Component Analysis (PCA)) and the Elastic Bunch Graph Matching (EBGM) algorithm proposed in [40], and that the combination of texture and shape features improves recognition rate compared to when transformation and recognition is performed using (i) the two features separately and (ii) the whole cropped faces as was done in previous work. Lastly, the proposed system was also shown to outperform human sketch recognition using 100 photo/sketch pairs from the testing set and 30 candidates to rank the top 10 photos which best match a given sketch.

A local geometry preserving-based non-linear SL method is used in [34] to learn the mapping between photos and (viewed) hand-drawn sketches to enable transformation of photos to pseudo-sketches, inspired by the Locally Linear Embedding (LLE) manifold learning technique. The basic assumption is that small image patches in photos and sketches form manifolds having similar local geometry in the two image spaces such that a pseudo-sketch can be reconstructed by using K neighbours of each data point (i.e., patches) to compute a neighbour-preserving mapping between the original high-dimensional data and low-dimensional feature space [34]. This approach is fundamentally similar to ET except that synthesis is performed locally using a subset of the training images rather than all of them.

In fact, Equations (2.1) to (2.3) may be extended for LLE by evaluating them on each patch in an image, letting $M = K$ and calculating the weights $\vec{c}_p^{\{i,j\}}$ using the approach described by the authors of [34]. More specifically, Equation (2.3) showing the required synthesised sketch may be modified for LLE as follows:

$$\vec{S}_r^{\{j\}} = \vec{\bar{S}}^{\{j\}} + \sum_{i=1}^{K} \vec{c}_p^{\{i,j\}} \vec{\Psi}^{\{i,j\}} \quad \text{for } j = 1, 2, \ldots, n \ . \tag{2.4}$$

where n is the number of patches, $\mathbf{S}_r^{\{j\}}$ is the j^{th} patch of the synthesised sketch, $\vec{\Psi}^{\{i,j\}} = \vec{S}^{\{i,j\}} - \vec{\bar{S}}^{\{j\}}$, $\vec{S}^{\{i,j\}}$ is the i^{th} training sketch of patch j, $\vec{\bar{S}}^{\{j\}}$ is the j^{th} mean patch and $\vec{c}_p^{\{i,j\}}$ are the reconstruction weights for the j^{th} patch derived using the i^{th} training face image. However, since the non-linear relationship between photos and sketches is being approximated using a linear combination of neighbouring patches, it can be argued that this process is not truly non-linear [41]. Once the pseudo-sketches are synthesised, Kernel-based Non-linear Discriminant Analysis (KNDA) is used to match a probe sketch with the pseudo-sketches. This approach combines the non-linear kernel trick with Linear Discriminant Analysis (LDA) and caters for the fact that there exist complex non-linear variations in sketches and pseudo-sketches caused by man-made distortions and blurring artefacts. Similarity measurement is performed by finding the distance between the projections of the probe sketch and pseudo-sketches into the non-linear subspace produced by KNDA, using the magnitude of the difference of these projections. It was demonstrated that the proposed approach is able to create pseudo-sketches which are closer to the true sketches than PCA (Eigenfaces).

The authors of [42] argued that sketch-to-photo synthesis is a better approach than photo-to-sketch synthesis in practical applications, since transformation of photos to sketches involves loss of information due to the simpler nature of sketches. In addition, variations in illumination can cause artefacts when synthesising sketches using live footage. As a result, a sketch-to-photo synthesis approach is proposed, where eigenvectors are obtained using both sketches and photos in a training set (i.e. using the traditional PCA/Eigenfaces [37] approach on both types of images simultaneously rather than individually as is typically done). The resultant eigenfaces of a sketch and photo of a subject are correlated and look similar, as opposed to eigenfaces obtained by training on photos or sketches separately. Due to differences in intensity between the original photos and transformed photos, ad-

vanced correlation filters [43] are then used for recognition. Rank-1 retrieval rates of 100% were achieved on the CMU-PIE database [44] for two types of illumination (with/without ambient light), even when only a few eigenvectors ($\sim 30\%$) are used. However, sketches used in this experiment were obtained using a non-linear sketch function available in a photo editing program, and thus the sketches do not accurately represent neither hand-drawn sketches nor software-generated sketches, which contain inaccuracies even in the case of viewed sketches.

2.2.2 Bayesian Inference framework

Methods using the BI framework synthesise photos or sketches by exploiting probability models, such as in the approach by the authors of [18] where the relationships between local patches were modelled using Multiscale Markov Random Fields (MRFs) via the Maximum a Posteriori (MAP) rule. Image quilting is also utilised to reduce the blurring and blocking effects created when using the popular averaging technique to stitch overlapping patches. Random Sampling Linear Discriminant Analysis (LDA) (RS-LDA) was then used for recognition to attain better performance than Eigenfaces, the method proposed in [34], and Eigentransformation [17]. The method was also shown to be relatively robust to pose variations.

The work in [18] was extended in [45] to cater specifically for lighting and pose variations by using shape priors designed for specific facial regions, more robust metrics to find candidate patches, and consideration of intensity and gradient compatibility when matching neighbouring sketch patches. The effectiveness of the proposed approach was validated on the CUHK student database consisting of 188 photo-sketch pairs.

The authors of [46] proposed the Markov Weighted Fields (MWF) model, which uses a weighted MRF to model the relation between photo and sketch patches to enable the synthesis of new sketch patches (i.e. patches not available in the training data) using a linear combination of 10 candidate patches. A cascade decomposition method is also used for the large-scale optimisation required in the proposed approach. It was shown that the good quality images could be synthesised even under varying pose and illumination conditions as shown in Figure 2.2.

A non-linear photo-to-sketch synthesis approach was proposed by the authors of

[41,47], using the Embedded Hidden Markov Model (HMM) (E-HMM) Inversion (E-HMMI) algorithm [48,49] to learn the non-linear mapping for photo/sketch pairs. The E-HMM and related E-HMMI do not require many training samples and consist of five super-states in the proposed approach, one each for the forehead, eyes, nose, mouth, and chin. Multiple models are then obtained for the training images and used to generate several pseudo-sketches which are fused together using the selective ensemble machine learning technique to improve the generalisation ability of the proposed system and thus allow synthesis of more accurate face pseudo-sketches. Recognition between probe sketches and pseudo-sketches is then performed with Eigenfaces. Using photo/sketch pairs from the Chinese University of Hong Kong (CUHK) database and a leave-one-out train/test methodology, it was shown that the 95.24% recognition rate using this approach was higher than that obtained using (i) no transformation (i.e. comparing sketches and photos directly) (19.05%) and (ii) the technique proposed in [34] (described above) (71.43%). Using the Universal Image Quality Index (UIQI) metric as a measure of quality/similarity between the original sketches and corresponding synthesised sketches, it was also

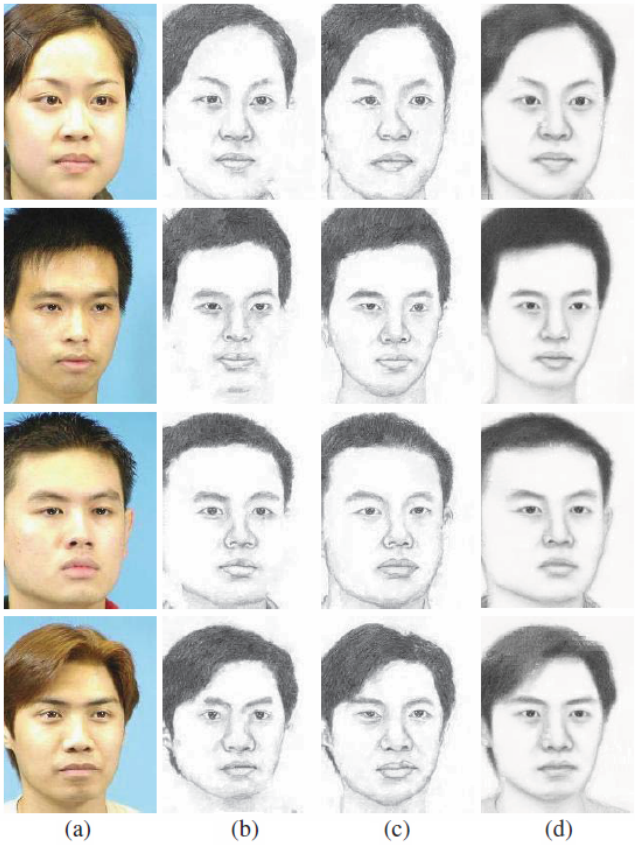


Figure 2.2: Examples of synthesised sketches with pose variation: (a) Original photo, (b) Images synthesised using approach in [18], (c) Images synthesised using approach in [45], (d) Images synthesised using approach in [46]

shown that the pseudo-sketches were more similar to the real sketches.

The recent method of [24] uses images in the Memory Gap Database (MGDB) that was also created for use in [24], containing sketches of 100 subjects that are constructed 1 and 24 hours after being viewed by an eyewitness, and also viewed and un-viewed sketches (where the latter are sketches that were viewed by an eyewitness and then immediately described to the forensic artist, thus modelling primarily the communication gap). Multi-task modelling using Gaussian Process Regression is then employed to model the memory and communication gaps and synthesise images that aim to reverse the effects of these gaps. The authors indicate that fidelity of the reconstructed sketches with respect to the corresponding photos can be improved and lead to improved matching performance.

2.2.3 Combination of Bayesian Inference framework and Subspace Learning framework

Approaches combining BI and SL have also been implemented, however most focus on face super-resolution. One approach that considers face-sketch synthesis is the statistical inference approach named Bayesian Tensor Inference proposed in [35], which first uses a method similar to that described by the authors of [34] to obtain an initial estimate and then uses the Bayesian MAP framework to estimate the high-frequency residual errors, which can be used to obtain more detailed images. Although the more complex problem of sketch-to-photo transformation was tackled with encouraging results, a large part of the database considered was used for training and only a limited set was used for testing with no indication of face matching rates.

2.2.4 Sparse Representation-based methods

Sparse-representation-based approaches decompose a signal into a combination of basis signals and choose those which represent the original signal in the most compact manner, with the aid of coupled dictionaries. Similar to algorithms using a combination of BI and SL above, most literature using sparse representation deals with the problem of face super-resolution.

An approach utilising sparse representation for face-sketch synthesis was proposed in [36], where sketches are synthesised from photos using a coupled dictionary on a local level. In other words, reconstruction is performed using overlapping patches whose overlapping areas are averaged in the final image under the assumption that corresponding face photo and sketch patch share a common sparse representation. The use of sparse representation follows the observation that sparseness is involved in human perception and human vision [36]. Compared with sketches synthesised by the approaches described in [18,34] on the CUHK dataset, it was shown that sketches generated with sparse representation yield sketches with acceptable quality with relatively low computation time.

The work in [36] was extended in [50] to reduce blocking artefacts by using a smoothness-constrained method across patches that is modelled as an energy minimisation function problem. To improve efficiency, a series of small-scale convex optimisation are utilised [50]. Using the 188 photo-sketch pairs in the CUHK student database, it was shown that the proposed method synthesises sketches from photos that are subjectively superior to those generated using the method proposed in [34] and comparable to the more time-consuming method proposed in [18].

The authors of [51] use sparse neighbour selection (SNS) to find close neighbours and create an initial estimate of pseudo-sketches or pseudo-photos followed by sparse-representation-based enhancement (SRE) for further quality improvement by constructing a coupled sparse representation model of the relationship between photo and sketch patches. Using the sparse representation classification (SRC) [52] algorithm for recognition, which was shown to outperform the Eigenfaces [37], Fisherfaces [53] and Locality Preserving Projection (LPP) [54] algorithms, the SNS-SRE algorithm outperformed the methods proposed in [34,55,56] in terms of recognition rates on a private database containing five sketches drawn by different artists for each of the 200 subjects considered.

Lastly, the recently proposed method in [57] obtains the sparse coefficient for each overlapping patch of an image and uses a greedy search algorithm to find the closest neighbouring patches in the training set. High frequency information/intensity of the test patch and the candidate photo patches are used to improve the selected neighbours. Bayesian inference using the Markov network is finally implemented to synthesise the final sketch image. For neighbour selection, all patches are used rather than those within a close local region with little increase in computation

time. The proposed algorithm was also shown to be robust to lighting, pose, alignment and background variations as well as to hair and accessories such as eye glasses and hairpins which may not be present in the training set images. The quality of synthesised sketches was also shown to be superior to those proposed in [18,34,46,51,58,59] using the Structural SIMilarity index (SSIM) [60] and Feature SIMilarity index (FSIM) [61] objective Image Quality Assessment (IQA) metrics, while rank-recognition performance was also comparable to the approaches proposed in [46,58] for sketches corresponding to subjects in the CUHK and AR databases and superior to all approaches considered on the XM2VTS database [57].

2.3 Inter-modality approaches

Inter-modality algorithms learn or extract modality-invariant features such that inter-class separability is maximised whilst maintaining intra-class differences [6, 62,63]. In essence, they could be described as specialised FRSs [7].

The authors of [64] presented one of the first feature-based approach, in which 128-D SIFT descriptors were computed at overlapping patches in photos and sketches. Two approaches were then considered for matching: (i) direct matching, in which the descriptors of all patches are combined for each photo and sketch, respectively, and then compared using Euclidean distance, and (ii) common representation matching, where the distance between the feature descriptors of each patch and patches of photo-sketch pairs in a dictionary are evaluated using probe sketches and training sketches, and gallery photos and training sketches, respectively. The distance between resultant vectors containing the common representations of probe sketches and gallery photos are then evaluated for person identification. This approach was implemented due to the concern that directly comparing SIFT descriptors obtained from photos and sketches would be unsuccessful due to the modality gap. As shown in Figure 2.3, the descriptors are in fact quite similar. In addition, SIFT descriptors were computed at two scales to yield a multi-scale representation. The fusion of the two matching approaches yielded the best performance at Rank-1 on the CUFS dataset of viewed hand-drawn sketches [39], followed by the direct matcher, the common-representation matcher, the MRF synthesis approach in [18] (ref. Section 2.2), the FaceVACS commercial FRS [65] and Eigentransformation [38], respectively. However, at higher ranks, the MRF algorithm in [18] outper-

formed the fused feature-based approach. Fusion of the feature-based approach with FaceVACS yielded further improvements in matching accuracy.

The approach proposed in [66] uses the Haar Transform on both sketches and photos to project the two types of images into a common modality. Diagonal details (HH decomposition) were used at scale 3. The corresponding negative images are then derived, followed by application of PCA for feature reduction. K-Nearest Neighbour (NN) (K-NN) and Support Vector Machine (SVM) were then considered separately for identification of a probe sketch from a gallery of photos. It was shown, using the Root Mean-Square Error (RMSE), that the modality difference was reduced with the transformed images. In addition, the proposed approach outperformed the Eigentransformation [17] and LLE [34] approaches, particularly at lower ranks. However, the train/test methodologies of the algorithms considered and the proposed approach were not the same; specifically, the authors of [66] trained the algorithms on a set of 100 photos whose corresponding sketches were used for testing, whereas the other two approaches contained different subjects in the training and test sets.

A feature-based approach designed to match forensic sketches was proposed in [9,67], where the SIFT and Multiscale LBP (MLBP) feature representations are employed due to the success achieved in previous works [64] and [68]. Each feature representation is computed on sliding windows with sizes of 16×16 and 32×32 (for both photos and sketches) and combined using the sum fusion rule, which is also used to combine MLBP and SIFT scores. To counteract the problems of high-dimensionality and the small sample size problem, an ensemble of linear discriminant classifiers termed Local Feature-based Discriminant Analysis (LFDA) is implemented as a training strategy by performing discriminant analysis on the

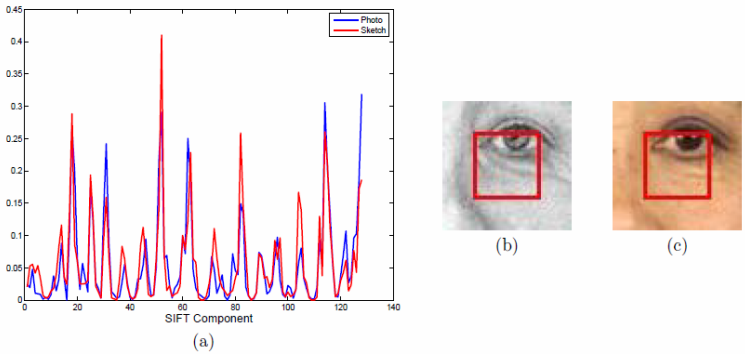


Figure 2.3: Similarity between sketch and photo patches of the same subject: (a) SIFT descriptors computed at the patch in the sketch shown in (b) and the photo shown in (c) [64]

concatenation of feature descriptor vectors derived from patches in the vertical direction. Evaluation was performed on the CUFS dataset [18,39], using 306 photo-sketch pairs for training and the remaining 300 pairs for testing, where it was shown that the proposed approach outperformed the SIFT-based algorithm in [64], the MRF algorithm in [18] and the FaceVACS Commercial Off-the-Shelf (COTS)-FRS [65]. It was also shown that the combination of MLBP and SIFT yields the best performance compared to the use of MLBP and SIFT individually and that LFDA improves recognition performance. A set of 159 hand-drawn forensic sketches and a gallery of 10,159 mug shot images were also considered, where classification of sketches was performed depending on their resemblance to subjects depicted in the mug shot photos. Specifically, 49 sketches were labelled as 'good' (i.e. only around 31% of the total number of sketches) while 110 were labelled as 'poor'. A significant difference in performance was reported for these two sets of sketches, highlighting the challenge of face recognition using forensic sketches. Training the proposed approach on the viewed sketches (due to the limited number of forensic sketches), it was shown that its Rank-50 retrieval rate of 32.65% outperforms the 8.16% retrieval rate obtained by FaceVACS. In addition, the performance of the proposed approach using poor sketches was roughly similar to the performance of FaceVACS when the latter used good sketches. Lastly, it was observed that several incorrect Rank-1 retrievals when using good sketches were due to the fact that the retrieved faces were visually more similar to the sketch than the true subject. This further emphasises the difficulty of matching forensic sketches, which often contain incomplete and inaccurate information [9].

Galoogahi and Sim [10] indicate that the modality gap between face photos and sketches is caused by differences of visual information. In addition, the largest modality gap is in visual information of fine face texture (low contrast areas of skin such as moles and shadows/light reflections). Moreover, the synthesised images of intra-modality methods greatly affect the success of the face recognition method employed. As a result, the authors of [10] argue not only that modality gap reduction in the feature extraction stage is preferable to modality transformation, but also indicate that feature extraction should be done using coarse texture (representing facial component boundaries with high contrast) since it has minimal effect on the modality gap. This is achieved by using a descriptor called Histogram of Averaged Orientation Gradients (HAOG). In this approach, orientation gradients are extracted on both fine and coarse textures but squared magnitudes are used to emphasize coarse textures and weaken fine face textures at the same time

since coarse textures are typically characterised by higher (stronger) magnitudes. The Chi-Square distance between a histogram describing a probe sketch and a histogram describing a photo from the gallery) is used for recognition. Compared with the methods proposed in [9,15,18] on the CUFS dataset [39], it was shown that the proposed method achieved the highest recognition rate, of 100%. It was also shown that the proposed approach provided better discrimination than the related Histogram of Orientation Gradients (HOG) descriptor. However, it is known that the CUFS database contains sketches which resemble very closely the original sketches, and therefore do not truly represent sketches obtained in real-world scenarios.

The Prototype Random Susbspaces (P-RS) approach proposed in [12] represents each image as a vector of kernel similarities to a set of prototypes. This is done by filtering each image with a Gaussian filter, a Difference of Gaussian (DoG) filter and a Centre-Surround Divisive Normalisation (CSDN) filter to (i) compensate for intensity variations and (ii) cater for differences in appearance between modalities. 128-D SIFT (HOG) and 236-D MLBP feature descriptors are then extracted from overlapping patches of these filtered images. Hence, there are six feature representations for each image. The cosine kernel is then used to compare the feature descriptors of a test image with those of the prototypes. Feature projection is performed using RS-LDA and final matching is performed using the cosine similarity measure. Scores from each of the filters are then added after min-max normalisation. It was shown that P-RS outperforms the FaceVACS COTS FRS on thermal Infra-Red (IR), NIR, viewed sketch and forensic sketch images. The algorithm was also shown to outperform the the Direct Random Subspaces (D-RS) approach proposed in [69] (which is similar to the proposed approach but computes similarities directly rather than using kernel similarities) on all the different modalities considered except on hand-drawn sketches.

Following the observation that most software-generated sketches are constructed using individual facial components (e.g. face shape, hair, eyes etc.), the authors of [6] proposed the Component-Based Representation (CBR) system where a component-based representation is employed to describe both face photographs and composite sketches by segmenting each image according to its components and then extracting MLBP descriptors from overlapping patches. Matching is first done patch-wise and then the scores from all patches are added to yield a component score. The scores from all components are finally summed to yield the final score. It was shown that the best matching performance, which outperformed FaceVACS, was achieved

using only the four most discriminative components (eyebrows, nose, hair and mouth). It was also observed that the quality of the software-generated sketches depends not only on the composite sketching software used but also on the user of the system due to the "other-race effect", where people of a certain race have difficulty recognising faces of people belong to other races [70,71]. Consequently, due to the cross-race bias, an Asian user would choose facial components which s/he is familiar with when sketching Caucasian subjects. In fact, since under 5% of the 123 subjects in the AR database are non-Asian, sketches created by a Caucasian user achieved significantly higher recognition rates than the sketches created by an Asian user (difference in rank-200 retrieval rates of about 20%). This also affects the performance of any face-sketch recognition system used.

The D-RS holistic-based algorithm in [12] and the CBR component-based algorithm in [6] were compared in [26] by utilising 75 mugshots with corresponding forensic hand-drawn and software-generated sketches, and a gallery populated with 10,000 mugshots from the Pinellas County Sheriff's Office (PCSO). The hand-drawn sketches used were created by sketch artists for real-world criminal investigations, while the composite sketches were obtained by first asking volunteers to view a mugshot of a suspect for one minute and then describing the facial features two days later to a FACES operator, who used a cognitive interview technique [72,73] to enhance the volunteer's memory of the mugshot observed. It was shown that the two approaches yield similar performance but CBR achieves the highest overall retrieval rates, and performance on composite sketches was superior to that attained on hand-drawn sketches when training was applied.

The work done by the authors of [26] was extended in [19], where a deployable software system called FaceSketchID is proposed. It uses modified versions of the D-RS and CBR algorithms proposed in [12] and [6], respectively, whose scores are combined to provide a single matching score. Modifications were done not only to tune performance but also to improve speed. The system was evaluated using viewed and forensic hand-drawn and software-generated composites, and hand-drawn sketches based on low-quality images captured from surveillance cameras. An extended gallery of 100,000 mugshots from PCSO is also utilised. It was shown that the databases from which the training images are acquired can significantly affect matching performance, D-RS outperforms CBR when the algorithms use the same training methodology, the fusion of CBR and D-RS provides the best performance which exceeds that of three COTS systems considered in this study,

and demographic filtering of subjects can improve performance considerably.

Approaches which compare two different types of images using coupled projections have also been proposed, with the most prominent methods including Canonical Correlation Analysis (CCA) [74,75], Coupled Locality Preserving Mappings (CLPM) [76], Simultaneous Discriminant Analysis (SDA) [32], Coupled Marginal Fisher Analysis (CMFA) [77], Coupled Spectral Regression (CSR) [78], two implementations of the Locality Constraint in Kernel Space (LCKS)-based coupled discriminant analysis method proposed in [63], and Maximum-Margin Coupled Mappings (MMCM) [79]. However, these algorithms have been implemented for the task of comparing high- and low-resolution photos, with the exceptions of the approaches in [63,78] where VIS-NIR comparison was also performed and the method proposed in [79] where ocular recognition was considered. In addition, multiple images per subject are required, thus making them unsuitable for the area of face-sketch recognition since only one sketch is typically created per criminal (even when multiple witnesses are present).

An algorithm using the concept of coupled projection was proposed by the authors of [15], where Coupled information-theoretic encoding was used to maximise the mutual information between features obtained from photos and sketches and therefore allow high correlation between codes obtained from photos and sketches of the same subject. In turn, this leads to a low inter-modality gap. Hence, in contrast to most inter-modality approaches, the modality gap is reduced at the feature extraction stage rather than the classification stage [15]. More specifically, photos and sketches are first aligned using affine transformation and then processed with a DoG filter to remove high- and low- frequency illumination variations. Next, the horizontal and vertical image gradients are computed and the neighbouring pixels of each pixel are sampled to form one vector per pixel. These vectors are encoded using a Coupled Information Theoretic Projection (CITP) tree, followed by construction of histograms on local regions of the resultant codes which are concatenated for the final Coupled Information-Theoretic Encoding (CITE) descriptor. These are used by a PCA-Linear Discriminant Analysis (LDA) classifier to compute the dissimilarity between a photo and sketch. A Randomised CITP Forest is utilised such that multiple trees are created, allowing the creation of multiple CITE descriptors which are fused using linear SVM. The proposed approach was shown to outperform several methods, including the MRF [18] and LFDA [9] approaches, when evaluated on the CUFS database and the CUFSF database that

was used for the first time in [15].

An analysis into the use of multiple sketches was performed in [80], by analysing any benefits that can be acquired through several types of fusion strategies. Both viewed hand-drawn and software-generated sketches were utilised. For the methods considered, it was shown that the use of multiple sketches at test-time can indeed be beneficial, with the best fusion techniques determined to be pixel-level and score-level fusion. However, no performance gain was achieved in the case of software-generated sketches. In addition, the quantity of subjects and number of sketches per subject used were limited.

A graphical representation-based method was proposed in [81] by using Markov networks to represent heterogeneous image patches separately, thus considering spatial information. The proposed approach was shown to yield promising performance in several HFR tasks including hand-drawn and software-generated sketches.

The authors of [82] used a genetic algorithm to train weights for features extracted locally using HOG [83] and Histogram of Image Gradients (HIM) [84] features, and applied transfer learning by first performing training with hand-drawn sketches and then using a subset of software-generated composite sketches to tune the parameters. It was shown that transfer learning was beneficial, and that the use of both HOG and HIM features yielded the best performance.

Finally, the authors of [85] propose a common encoding feature discriminant approach where intensity values are sampled into eight vectors for each pixel, which are then encoded using a feature transformation function that is learned from training data that is optimised to reduce the mean square error. The resultant coded image is then divided into patches of varying sizes, to enable multi-scale testing, from which histograms are extracted. The histograms from each patch and each scale are concatenated into one vector that is divided into slices on which PCA followed by LDA is applied. The result is projected using a method based on expectation maximisation with final comparison performed using the cosine distance. The method was shown to outperform several algorithms proposed in literature including LFDA [9] and CITE [15], and was also found to be effective for VIS-NIR matching. However, an extended gallery to mimic the extensive mug-shot galleries was not considered.

More information about inter-modality algorithms may also be found in [25,86].

2.4 Other sketch-based recognition tasks

Caricatures that are purposely exaggerated and distorted versions of the original face, as shown in Figure 2.4 can also be considered as sketches along with viewed, semi-forensic and forensic sketches. Since caricatures are not typically used in forensic investigations and largely depict simplified (but sometimes highly distorted) versions of a face image [87], they are not considered in this research. More information may be found in [12,23,87].

Some work has also been done on image retrieval using objects depicted in sketches. Since the task tends to involve comparison among different objects which typically exhibit greatly varying attributes (instead of recognition of the same object type as in the case of face photo-sketch recognition), object identification and retrieval using sketches can be considered a simpler task that is also less affected by memory and communication problems. Consequently, this line of research is also not investigated in this work. More information may be found in [88–91].

(a)

(b)

Figure 2.4: Two other types of sketches: (a) Sketch of an object and the corresponding valid matching photographs [88], (b) a caricature and the corresponding photograph of subject [23].

2.5 Deep Learning

Deep learning is a field that has attracted much attention in recent years due to the great performance that has been attained in several application domains, including traditional VIS-VIS FR. Since a deep learning-based system is proposed in this research, a brief background to deep learning will be given in Section 2.5.1 followed by an overview of state-of-the-art deep learning-based FRSs in Section 2.5.2.

2.5.1 Background

Neural Networks (NNs) have been used successfully in several machine learning tasks such as object recognition and natural language processing, and aim to emulate the vast networks of neurons that are used in biological brains to perform complex tasks. Indeed, as shown in Figure 2.5, the main component of NNs is the *artificial neuron* that outputs a value dependant upon the inputs and their importance, the latter controlled by *weights*. A *bias* term $\vec{b}$ is also included, as follows:

$$\text{output} = f(\vec{w} \cdot \vec{x} + \vec{b}) \tag{2.5}$$

where $\vec{w}$ and $\vec{x}$ are vectors whose components represent the weights and inputs, respectively, and $f(z)$ is an *activation function* where generally $z = w \cdot x + b$. A popular activation function is the sigmoid function:

$$f(z) = \frac{1}{1 + e^{-z}} \tag{2.6}$$

A network typically consists of several layers of artificial neurons, where each connection to a neuron has its own weight and every neuron has a bias value. Since neurons in later layers use the results of previous neurons, they generally make decisions at more abstract and complex levels than neurons in the initial layers [92–94]. The layout of a network is thus defined by the number of hidden layers and the amount of artificial neurons in each layer, which determine the number of weights and biases to be learned. Training a network to find optimal values

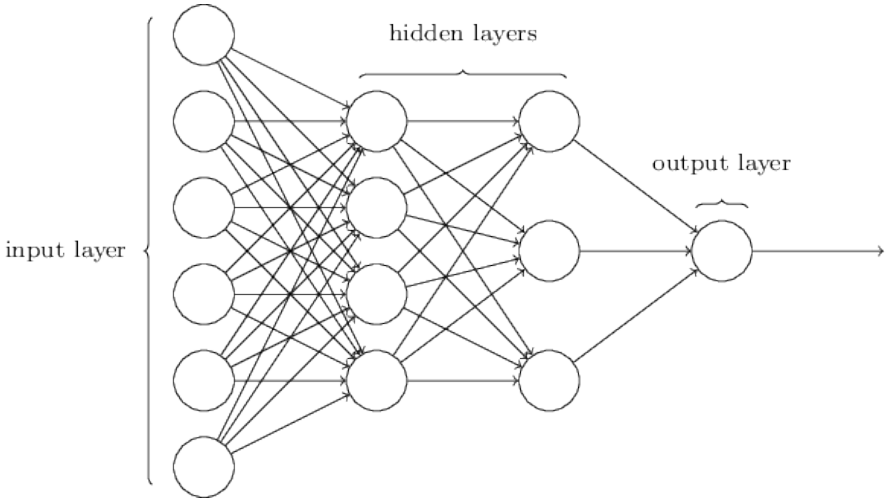


Figure 2.5: Example of a neural network [92]

for these parameters involves the minimisation of some *cost function*[1], such as the Mean Square Error (MSE) or *log-likelihood* function, via its gradient. This is typically done using *backpropagation* of gradients with *Stochastic Gradient Descent (SGD)*, where the objective function is minimised iteratively (i) over several *batches* that each contain a subset of the training data samples and (ii) over a number of *epochs*, where each epoch encompasses all the samples in the training set. Given a cost function C, a weight w_k and a bias b_j are updated as follows [92]:

$$w_k \rightarrow w'_k = w_k - \eta \frac{\partial C}{\partial w_k} \tag{2.7}$$

$$b_j \rightarrow b'_j = b_j - \eta \frac{\partial C}{\partial b_j} \tag{2.8}$$

where η is the *learning rate* controlling the amount by which the parameters are adjusted. Stochastic gradient descent has also been modified to use a *momentum coefficient* μ controlling the rate of change of the parameters [92]:

$$v_k \rightarrow v'_k = \mu v_k - \eta \frac{\partial C}{\partial w_k} \tag{2.9}$$

$$w_k \rightarrow w'_k = w_k + v'_k \tag{2.10}$$

[1]Also known as the *objective* or *loss* function

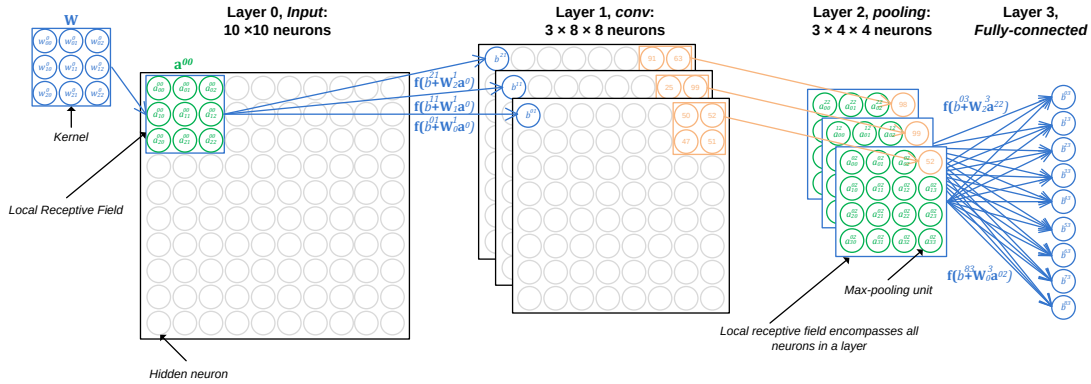


Figure 2.6: Example of a DCNN, showing the basic elements. Input layer contains 10 rows and 10 columns of neurons, corresponding to pixels in a 10×10 image. As shown in Equation (2.11), a matrix of weights $\mathbf{W}$, called a *kernel*, is applied to a group of neurons of size 3×3, called the *local receptive field*. This kernel is shared for all neurons. Three different kernels are applied, representing three features. A bias b is also shared among the neurons for each filter. Neurons in the first layer typically apply an *activation function* such as the sigmoid or ReLU functions. Next, groups of neurons are pooled to yield a single value; in this case, the max-pooling operator is applied on groups of 2×2 neurons, to yield the maximal value in each group. The final layer consists of all neurons in the second layer connected to all output neurons, thereby called a fully-connected layer. These typically correspond to class labels or a feature vector.

Deep networks are conceptually similar to neural networks, but contain more layers. These 'deeper' networks are able to learn more powerful functions and capture more complex information from the training data [92,95–99], but in practice they are also harder to train due to issues such as the 'vanishing gradient' and the related 'exploding gradient' problems, hyper-parameter selection, choice of network architecture, etc. [92,100]. Much research has been done to develop techniques capable of overcoming these issues, including weight initialisation methods and regularisation techniques such as *dropout*, where a random selection of neurons are disabled to effectively allow the network to learn the average effects of a large number of different networks [92,94,101], and the recent batch normalisation technique proposed in [102].

The Deep Convolutional Neural Network (DCNN), an example of which is shown in Figure 2.6, was designed to exploit spatial structure in images as demonstrated in the seminal work of LeCun *et al.* [103]. It is also able to reduce the substantial number of trainable parameters that arise from deep networks. In DCNNs, a neuron in a layer is connected to a small region of neurons in the previous layer called the *local receptive field*. The same weights and bias used for a region are also applied for the rest of the regions in the same layer. More specifically, the output for the

j, k-th hidden neuron for feature i at layer l is as follows [92]:

$$\text{output} = f\left(b^{i,l} + \sum_{n=0}^{N}\sum_{m=0}^{M} w^{l}_{n,m} a^{i,l-1}_{j+n,k+m}\right) \tag{2.11}$$

where N and M are the vertical and horizontal sizes, respectively, of the receptive field, and $a^{c,d}_{x,y}$ is the activation from the input neuron at location x, y for feature c at layer d. Hence, a layer utilising a 5×5 local receptive field will need just 25 weight values and one bias value. Consequently, the neurons detect the same type of feature but across different locations in the input image, which not only helps to reduce the number of parameters in a neural network but also allows for translation invariance. The operation in Equation (2.11) is often referred to as *convolution*, from which the name of the DCNN is derived, while the shared weights and bias are often referred to as defining a *filter* or *kernel*. Hence, an input image is convolved with the set of filters in a convolutional layer, which are learned over a number of batches and epochs such that they can accomplish the required task[2] as successfully as possible. It should be noted that the convolutional layers are linear, requiring other types of layers to enable non-linearity. Examples include the use of activation functions such as the Rectified Linear Unit (ReLU)[3] having the form:

$$f(z) = \max\{0, z\} \tag{2.12}$$

Pooling layers which output a single value for a group of neurons are also used to simplify the information output from a convolutional layer. Commonly used functions include max-pooling and mean-pooling where the output is simply the maximum or average of the activations considered, respectively. Additionally, *fully-connected* layers consist of neurons which are connected to all of the neurons in the previous layer (similar to traditional neural networks) and are essentially special cases of the convolutional layer [93,104]. The final layer in any network includes neurons which output a results such as a class label, typically using a function such

[2]Tasks performed by a deep network can include (i) regression where another image is output (e.g. a network trained for de-blurring will output a sharpened version of a blurred image presented to the network), (ii) a feature descriptor or (iii) the class of the input image (e.g. the type of object present in the image)

[3]Rectified linear activation functions often enable higher performance than other functions; while it is believed that this may be due to the avoidance of saturation in ReLUs, the superior performance is still not well understood[92]

as *softmax* defined as follows [92]:

$$f(z^L) = \frac{e^{z_j^L}}{\sum_k e^{z_k^L}} \tag{2.13}$$

where z_j^L is the j-th output neuron in the last layer L. Hence, the output of a neuron is practically normalised by the sum of the outputs of all neurons in the layer. Hence, the outputs of the final layer can be thought of as corresponding to a probability distribution, since all values are positive and always sum to one [92].

When training a network, the final layer is followed by an objective function as mentioned previously, which is typically removed prior to network deployment (when the net is used for evaluation after the network has been trained). Consequently, a DCNN will ultimately learn a function obtained through a series of linear and non-linear layers [93].

Deep networks are often slow to train on Central Processing Units (CPUs), which are not optimised to exploit the parallelisation of the vast number of operations entailed by these networks. On the other hand, Graphical Processing Units (GPUs) allow much parallelisation and are well-suited to handle matrix operations that can be used in DCNNs to improve efficiency, especially when equipped with libraries designed to optimise functions that are used for deep learning such as the NVIDIA CUDA Deep Neural Network (cuDNN) library [105]. Significant speed-ups are thus obtained, typically amounting to approximately 10 times the rate achieved by CPUs [104] (depending on the CPUs and GPUs used).

A more detailed review of neural networks and deep learning may be found in [92,93,106].

2.5.2 Deep Learning-based Face Recognition

One of the earliest and most popular deep-learning approaches is the AlexNet DCNN architecture [94] that was trained for the task of object classification. Several superior approaches based on deep-learning have since been introduced, along with new methods to improve the performance of a network. Of particular interest in this paper are face recognition methods such as Facebook's DeepFace [107] that was extended in [108], the further extension of DeepFace in the DeepID series [109–112], Google's FaceNet [113], the end-to-end face verification systems in [114,115], and VGG-Face [93], which have provided important observations such as the superior performance that is generally obtained by using more layers [112], the benefit of a high amount of training data (especially for 'deeper' networks having more trainable parameters) [93,113], the use of multiple DCNNs [109], and a "triplet-based" objective function which aims at decreasing the distance between features of the same subject and increasing the distance between features of different subjects [93,113]. Deep learning-based face recognition methods are typically reported to achieve over 90% accuracy on popular datasets such as Labelled Faces in the Wild (LFW) [116], alluding that the problem of face recognition is largely solved. However, the recent MegaFace challenge involving a large number of subjects in the gallery (1 million face images of 690,572 subjects) has shown that even the top-performing FaceNet method, which attained a near-perfect accuracy on the LFW dataset, exhibited a significant loss in performance to approximately 75% accuracy [117]. Indeed, the saturation of performance on popular datasets such as LFW and YTF has led to the creation of the IANUS Jarpa Benchmark A (IJB-A) dataset and its extended version, Janus Challenging Set 2 (JANUS CS2) which pose greater challenges to face recognition algorithms [114,118]. The recent study in [119] also demonstrated the significant challenge of a large gallery by using 80 million face images, where it was further shown that an approach utilising a DCNN (based on the one described in [120]) yielded the best performance among the methods considered whilst requiring low computation times. Thus, while it is evident that the problem of face recognition is far from solved, the results achieved by deep learning-based algorithms is encouraging. More information regarding deep-learning-based FRSs may be found in [93,109–114].

To the best of the authors' knowledge, few works have considered the use of deep learning concepts for face photo-sketch recognition. Notable methods include the

approach in [121], where an autoencoder and deep belief networks were bootstrapped to learn a feature representation of normal VIS face photos and were then fine-tuned using transfer learning for face photo-sketch recognition. However, the system is shallow and does not exploit the spatial relationships inherently present in images, which are important for face recognition [6]. The approach in [122] uses a convolutional neural network for photo-to-sketch synthesis, although the network is quite shallow (six layers deep) and was implemented using the easy photo-sketch pairs in the CUHK student dataset; since these sketches bear a very close resemblance to the original photographs, essentially only the modality gap is being tackled while ignoring the memory and communication gaps. The method in [123] also uses a convolutional neural network for several HFR tasks, which is trained primarily using face photos and then fine-tuned with a few subjects having images in both of the domains being considered. A shared projection matrix is also learned. The proposed approach attained similar performance to the method in [121] on the PRIP-VSGC database [6,19], and higher performance on the EPRIP database [121,124]. Lastly, the recent intra-modality method described in [125] uses a branched convolutional neural network in which a photo and global prior are used as input to three shared layers. The output is then passed through two separate networks each having a further three layers, with one network focusing on predictions of structures and the other network focusing on textures. Similar to the approach in [57], the network is thus relatively shallow and was also implemented using the easy CUFS dataset [17,18].

2.6 Summary

Numerous algorithms have been proposed in literature for face photo-sketch recognition, which can be classified as either intra-modality or inter-modality approaches. In the former approach, photos and sketches are transformed into a common modality to facilitate recognition by a face recogniser designed to work in the target modality, while the aim of inter-modality approaches is to extract or learn modality-invariant features and thus create specialised FRSs for photo-sketch recognition. A summary of notable methods is shown in Table 2.1.

Intra-modality approaches are desirable in theory but have been criticised for attempting to solve a more complex problem than recognition itself, and most approaches are in fact complex and thus require significant computation time and memory requirements [6,12,25]. Consequently, recent research has focused on inter-modality approaches, which tend to outperform intra-modality methods. However, several approaches use hand-crafted descriptors such as LBP and SIFT which were not specifically designed for face-sketch recognition, and hence may not be the optimal choice [10]. However, the learning of feature descriptors is constrained by the limited amount of available data, namely the number of subjects and quantity of sketch images per subject.

Whilst good results have been attained by both types of algorithms, most approaches (especially intra-modality algorithms) have been trained and tested on sketches which closely resemble the original photo. For example, even the hair component is very well matched in terms of shape and texture in the sketches of the popular CUFS database [15,126]. In such cases, the problem is simplified to the extent that the reported matching performance is not highly representative of real-world performance where forensic sketches obtained from eyewitness descriptions of criminals are utilised. In fact, it has been shown that a good COTS FRS can achieve performance comparable to leading face-sketch recognition algorithms on the CUFS database [6,12]. Moreover, numerous methods fail to utilise an extended photo gallery to simulate the extensive mug-shot galleries maintained by law enforcement agencies. Consequently, some face-sketch recognition algorithms have reportedly achieved a perfect 100% recognition rate at low ranks. As shown by the performances achieved by leading methods in Chapter 6, this rate is hard to achieve when using a realistic implementation set-up consisting of challenging

sketches and an extended gallery as found in real-world criminal investigations.

Lastly, a few intra- and inter-modality methods using deep learning have been proposed. However, these methods typically employ relatively shallow networks, are trained predominantly using face photos and only a limited amount of photo-sketch pairs, and are evaluated on easy databases without the use of an extended gallery.

Table 2.1: Summary of primary methods in literature. DB = database, Ext. = Extended, subjs. = subjects, X2Y = transformation of X to Y, where X,Y∈ {Photo (P), Sketch (S)}, SR = Sparse Representation, DL = Deep-Learning, 'v' = viewed sketches, 'u' = un-viewed sketches, 'f' = forensic sketches, 'hdc' = hand-drawn composite sketches, 'sgc' = software-generated composite sketches. Approximate values are provided when results are only shown graphically.

Type	Method	Photo/Sketch DB(s)		Ext. Gallery DB(s)		Performance (%)		
		DB Name(s)	# Train/Test subjs.	DB Name(s)	# subjs.	Synth. Mode	Rank-1	Rank-10
Intra (SL)	ET [17,33]	CUHK [17] (v-hdc)	88/100	–	0	P2S S2P	71.0 57.0	96.0 88.0
Intra (SL)	Tang & Wang [38]	CUFS [18,39] (v-hdc)	153/300	–	0	P2S	81.3	97.0
Intra (SL)	LLE [34]	CUFS [18,39] (v-hdc)	306/300	–	0	P2S	87.7	99.0
Intra (BI)	MRF [18]	CUFS [18,39] (v-hdc)	153/300	–	0	P2S S2P	93.3 96.3	99.0 99.7
Intra (BI)	MWF [46]	CUHK [18] (v-hdc)	88/100	–	0	P2S	≃88.0	≃100.0
Intra (BI)	E-HMM [41,47]	CUHK [18] (v-hdc)	605/1[a]	–	0	P2S	95.2	N/A
Intra (BI)	Ouyang *et al.* [24]	MGDB [24] (v,u-hdc)	68/32	–	0	S2P	86–99	N/A
		MGDB [24] (v,u-hdc); PRIP-HDC[b] [19] (f-hdc)	68/49	(Private)[c]	10,030	S2S	38.0	42.0
		MGDB [24] (v,u-hdc); (Private)[b] (u-sgc)	68/51	(Private)[c]	10,030	S2S	14.0	18.0
Intra (SR)	SNS-SRE [51]	CUFS [18,39] (v-hdc)	310/296[a]	–	0	P2S	93.1	N/A
		CUFS [18,39] (v-hdc)	310/296[a]	–	0	S2P	96.5	N/A
		VIPSL (Private) (v-hdc)	100/100	–	0	P2S	99.0	N/A
		VIPSL (Private) (v-hdc)	100/100	–	0	S2P	89.0	N/A
Inter	LFDA [9,67]	CUFS [18,39] (v-hdc)	306/300	–	0	–	99.5	N/A
		CUFS [18,39] (v-hdc); PRIP-HDC[b] [19] (f-hdc)	306/159	MSP (Private)	10k	–	≃0.0–16.3[d]	≃4.0–22.5[d]
	HAOG [10]	CUFS [18,39] (v-hdc)	0/606	–	0	–	100.0	100.0
	P-RS [12]	CUFS [18,39] (v-hdc)	404/202	PCSO (Private)	10k	–	74.6	≃97.0
		PRIP-HDC[b] [19] (f-hdc)	106/53	PCSO (Private)	10k	–	≃1.0	≃6.0
	D-RS [12,69]	CUFS [18,39] (v-hdc)	404/202	PCSO (Private)	10k	–	96.4	≃99.0
		PRIP-HDC[b] [19] (f-hdc)	106/53	PCSO (Private)	10k	–	≃4.0	≃13.0
	CBR [6]	PRIP-VSGC[e] [6,127] (v-sgc)	0/123	PCSO (Private)	10k	–	10.6	≃35.0
				MEDS-II [128]	1193	–	12.2	≃42.0
	FaceSketchID [127]	CUFS [18,39], CUFSF [15,129] (v-hdc); PRIP-HDC[b] [19] (f-hdc)	1800+212/53	PCSO (Private)	100k	–	≃5.0	≃12.0
		CUFS [18,39], CUFSF [15,129] (v-hdc); PRIP-SGC (Private) (u-sgc)	1800/75	PCSO (Private)	100k	–	≃4.0	≃10.0
		CUFS [18,39], CUFSF [15,129] (v-hdc); PRIP-VSGC[b] [6,127] (v-sgc)	1677/123	PCSO (Private)	10k	–	≃11.0–24.0	≃30.0–38.0
	CITE+CITP [15]	CUFS [18,39] (v-hdc)	306/300	–	0	–	99.9	N/A
	Gong *et al.* [85]	CUFSF [15,129] (v-hdc)	500/694	–	0	–	93.6	N/A
Inter (DL)	Mittal *et al.* [121]	CMU-PIE [44]; PRIP-VSGC[b] [6,127] (v-sgc)	30k+48/75	–	0	–	≃8.0	52.0±2.4
		CMU-PIE [44]; EPRIP [121,124] (v-sgc)	30k+48/75	–	0	–	≃5.0	60.2±2.9
	Saxena & Verbeek [123]	CASIA Webface [130]; PRIP-VSGC [6,127] (v-sgc)	500k+48/75	–	0	–	N/A	51.5±4.0
		CASIA Webface [130]; EPRIP [121,124] (v-sgc)	500k+48/75	–	0	–	≃12.0	65.6±3.7

[a] A leave-one-out strategy is employed for some databases considered

[b] Only a subset of sketches/photos are publicly available

[c] Mugshots downloaded from `mugshots.com`, but list of file names has not been made publicly available

[d] Lower and upper ranges are accuracies reported for 'poor' and 'good' sketches, respectively

[e] Sketches are created using the IdentiKit software program and FACES software program; authors of [6] focus on FACES sketches which are unavailable

Chapter 3

Proposed Methods

Several face photo-sketch recognition methods have been designed in this project, spanning both intra- and inter-modality methods. The sequence of the descriptions hereunder will follow chronological order, starting with the proposals of an intra-modality approach and the fusion of intra- and inter-modality approaches in Section 3.1 that were also described in a paper published at the *International Conference on Computer Analysis of Images and Patterns (CAIP2015)*, [62]. This is followed by the description in Section 3.2 of a state-of-the-art inter-modality method that was published in a paper presented at the *European Signal Processing Conference (EUSIPCO2016)*, [131], and finally the description of a novel method utilising a DCNN in Section 3.3 which is described in a letter published in the *IEEE Signal Processing Letters* publication, [132], and in a journal paper accepted for publication in the *IEEE Transactions on Information Forensics and Security*, [133]. In-depth analyses of several components of these proposed systems, including parameter selection, are provided in Appendix A.

Due to the limited public availability of software-generated composite sketches, the University of Malta Software-Generated Face-Sketch (UoM-SGFS) database has also been created during the course of this project. Details of the UoM-SGFS database have been published in a paper presented at the *Biometrics Signal Processing Conference (BIOSIG2016)*, [14]. A description of the database will be given in Chapter 4.

3.1 Eigenpatches and fusion of intra- and inter-modality algorithms

The first algorithm proposed in this work is the adaptation of the Eigenpatches (EP) intra-modality method used for face super-resolution as described in [134] to operate for face photo-sketch synthesis. EP is based on the ET approach described in Section 2.2.1 on Page 10 and operates on local patches instead of the global face image, thus allowing higher detail to be synthesised in local regions. In this work, EP was implemented for face photo-sketch synthesis to determine if synthesis at a local level would be superior to global face synthesis for photo-sketch recognition. In addition, the effect of fusing ET and EP is evaluated to determine if these methods provide complementary information.

EP may be applied for sketch (or photo) synthesis by synthesising each patch (instead of the whole image) and can be obtained by learning the optimal linear combination of patches found in the same local area of images in the training set. Formally, starting from the ET equation which performs synthesis on the whole image:

$$\vec{S}_r = \vec{\bar{S}} + \sum_{i=1}^{M} \vec{c}_p^{\{i\}} \vec{\Psi}^{\{i\}} \tag{3.1}$$

where $\vec{S}_r$ is the reconstructed sketch, $\vec{\bar{S}}$ is the mean sketch, $\vec{\Psi}^{\{i\}} = \vec{S}^{\{i\}} - \vec{\bar{S}}$, $\vec{S}^{\{i\}}$ is a column vector representing the i^{th} sketch, and $\vec{c}_p^{\{i\}}$ is a column vector of dimension M representing the contribution of the i^{th} training photo image $\vec{P}^{\{i\}}$ in the reconstruction of a test face image computed according to [17]. Then, Equation (3.1) is modified such that it is applied locally for each of the n patches in an image:

$$\vec{S}_r^{\{j\}} = \vec{\bar{S}}^{\{j\}} + \sum_{i=1}^{M} \vec{c}_p^{\{i,j\}} \vec{\Psi}^{\{i,j\}} \quad \text{for } j = 1, 2, \ldots, n \ . \tag{3.2}$$

where $\vec{S}_r^{\{j\}}$ is the j^{th} patch of the synthesised sketch, $\vec{\Psi}^{\{i,j\}} = \vec{S}^{\{i,j\}} - \vec{\bar{S}}^{\{j\}}$, $\vec{S}^{\{i,j\}}$ is the i^{th} training sketch of patch j, $\vec{\bar{S}}^{\{j\}}$ is the j^{th} mean patch and $\vec{c}_p^{\{i,j\}}$ are the reconstruction weights for the j^{th} patch derived using the i^{th} training face image. To the best of the author's knowledge, Eigenpatches has thus far not been used

(a) **(b)** **(c)** **(d)** **(e)** **(f)**

Figure 3.1: Synthesised images of one subject in the Color FERET/CUFSF datasets: (a) Original photo, (b) Original sketch, (c) Eigentransformation P2S, (d) Eigentransformation S2P, (e) Eigenpatches P2S, (f) Eigenpatches S2P, where P2S=photo-to-sketch synthesis and S2P=sketch-to-photo synthesis.

for face photo-sketch synthesis. Similar to other approaches involving patch-based operations, the overlap of patches with their neighbours is arbitrarily fixed at half the patch size and a simple averaging operation is done on overlapping areas to combine patches together. The patch size is set to 128×128 since it provided the best results, as detailed in [62] and Appendix A.1. Examples of images synthesised with Eigenpatches are shown in Figure 3.1.

For both methods, the role of sketches and photos is simply interchanged for pseudo-photo synthesis. Although any face recogniser can then be used to match the pseudo-sketches (photos) with the original sketches (photos), PCA (Eigenfaces method) [37] is implemented due to its widespread use in the literature for intra-modality methods. Hence, photo and sketch subspaces are learned and the face images to be compared are projected into these subspaces, similar to the approach in [17].

The fusion of intra- and inter-modality algorithms is also considered by fusing ET, EP and HAOG as shown in Figure 3.2. Fusion is performed at the matching score level, by first normalising the scores output from face recogniser using min-max normalisation as shown in Equation (5.1) on Page 61. The sum-of-scores method is then used to fuse the normalised scores together:

$$\vec{F} = \sum_{k=1}^{L} \vec{\hat{s}}_k \tag{3.3}$$

where L represents the number of intra- and inter- modality methods considered, $\vec{\hat{s}}_k$ is the score of k^{th} face recognition method, and $\vec{F}$ is the final similarity score

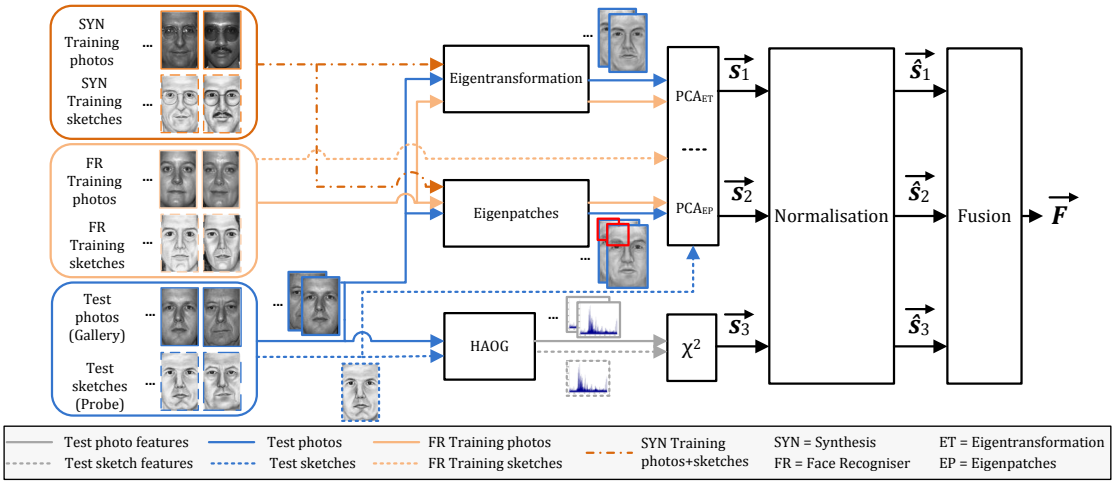


Figure 3.2: System flow diagram of the proposed fusion of intra- and inter-modality methods, for photo-to-sketch synthesis. Photos and sketches simply switch roles for sketch-to-photo synthesis.

between a sketch (pseudo-photo) and the N pseudo-sketches (photos) [135]. Sum-of-scores fusion and min-max normalisation were chosen since they have been shown to provide some of the best results for fusion of multi-biometric systems [2,80,135].

It is shown in Chapter 6 that Eigenpatches outperforms Eigentransformation at all ranks (especially at lower ranks) for viewed hand-drawn sketches, viewed software-generated sketches, and for real-world forensic sketches. This shows that the reconstruction of local regions can provide images that are able to discriminate between persons more reliably due to the utilisation of local features, while Eigentransformation encodes global spatial information that enables it to perform better at higher ranks when faces are being compared at a more global level. In fact, the fusion of Eigentransformation and Eigenpatches was shown in [62] to yield gains in performance, indicating that the local-based Eigenpatches approach and the holistic Eigentransformation approach provide complementary information. Further gains in performance were achieved when fusing Eigentransformation and Eigenpatches with the HAOG inter-modality method, showing the benefit in combining intra- and inter-modality algorithms. This is also shown to hold for viewed hand-drawn and software-generated sketches, and for real-world forensic sketches in Chapter 6.

3.2 LGMS inter-modality approach

The third proposed approach is an inter-modality method called log-Gabor-MLBP-SROCC (LGMS) [131]. The system flow diagram of the proposed LGMS method is shown in Figure 3.3. First, all photos and sketches are aligned such that the eyes and mouth are in the same position for all images as detailed in Section 5.1, which are then filtered with 32 log-Gabor filters to yield 32 images for each sketch and each photo. Gabor filters are able to represent signals localised in both time/space and frequency [136] and have been used in a vast number of applications [137]. Their use is motivated by the observation that these filters can model the Human Visual System (HVS) [136,137] and have yielded good performance within their application domains. However, log-Gabor filters were proposed in [136] to better model natural images, to remove the DC component, and to reduce the number of filter banks required [61,137]. They are less commonly used in literature than Gabor filters and to the best of the authors' knowledge have thus far not been used for face photo-sketch recognition. MLBP descriptors from overlapping patches of the images derived in the filtering stage are then extracted and concatenated. Although both MLBP and log-Gabor filters can be said to extract texture information, MLBP is able to characterise the *type* of texture present within local areas. Hence, log-Gabor filtering extracts texture information at a *global* level, while MLBP extracts

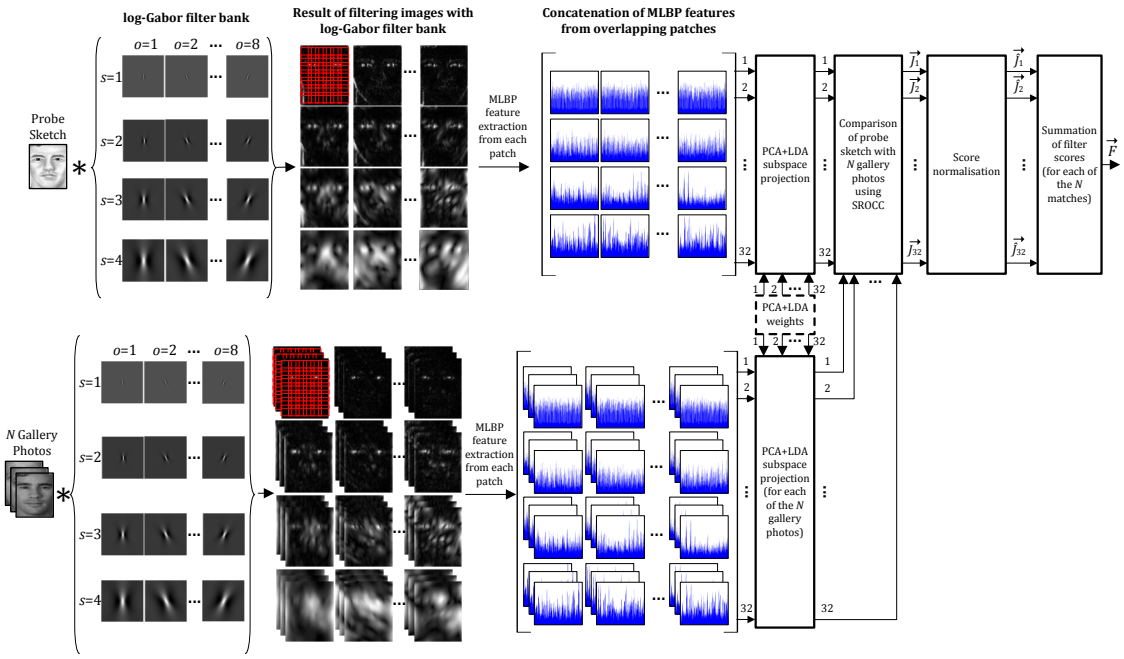


Figure 3.3: System flow diagram of the proposed LGMS approach for $O = 8$ and $S = 4$. Dotted block represents data obtained from the training stage.

local texture information. Following discriminant analysis on the concatenated MLBP vectors using PCA followed by LDA, an approach proven to be beneficial for face recognition [12], the Spearman Rank Order Correlation Coefficient (SROCC) between the resultant descriptors of the sketches and photos to be compared is then found and used as a similarity measure, for each of the 32 filters. Whilst not often used for FR, it is shown in Appendix A.2 that SROCC outperforms popular comparison metrics. Scores are finally normalised and summed to yield the final matching score. More details will now be given hereunder.

3.2.1 Image Filtering

Each geometrically normalised photo and sketch is filtered with a bank of log-Gabor filters that are selective in terms of frequency and orientation. The 2D log-Gabor function defined using the Gaussian spreading function is as follows [61,137]:

$$LG_{o,s}(f,\theta) = \exp\left(-\frac{\ln^2(f/f_0)}{2\ln(\kappa_\beta)}\right)\exp\left(-\frac{(\theta-\theta_0)^2}{2\sigma_\theta^2}\right) \tag{3.4}$$

where $o = 1, 2, \ldots, O$ and $s = 1, 2, \ldots, S$ are the orientation and scale of the filter, respectively, $\kappa_\beta = 0.55$ is related to the filter bandwidth, $\sigma_\theta = 0.3272$ is the angular bandwidth, f_0 is the centre frequency, and θ_0 is the centre orientation. Setting $O = 8$ and $S = 4$, a total of $G = 8 \times 4 = 32$ filters are defined. These values were chosen such that a good balance between performance and computational complexity is achieved. In addition, it was ensured that a Nyquist rate of at least n times the standard deviation apart from f_0 (on a logarithmic scale) was maintained to avoid aliasing, using the following condition [138]:

$$f_0 < \pi\kappa_\beta^n \tag{3.5}$$

where the authors of [138] state that $n = 3$ is typically sufficient; hence, n was also set to a value of 3 in this work.

3.2.2 Feature Extraction

After the filtered images have been obtained, they are divided into $p \times p$ patches with an overlap of $p/2$ both vertically and horizontally, where $p = 32$. Overlapping patches are able to consider the relationship among neighbouring regions and thus encode spatial information that is useful for recognition. For an image of size 200×250 as used in this work, 154 patches are obtained from which LBP features are then extracted. LBP is a gray-scale invariant texture operator which thresholds local circularly symmetric neighbourhoods into a binary pattern. More specifically, the operator uses the gray values of P equally spaced pixels on a circle of radius R, and assigns a binomial factor to each pixel [31]. In the proposed approach, 'uniform' patterns at $P = 8$ sampling locations with radius $R = 1$, similar to the implementation in [12] and as described in [31] are used

$$LBP_{P,R} = \begin{cases} \sum_{p=0}^{P-1} s(g_p - g_c)2^p, & \text{if } U(LBP_{P,R}) \leq 2 \\ P + 1, & \text{otherwise} \end{cases} \tag{3.6}$$

where g_c is the gray value of the centre pixel of the neighbourhood g_p ($p = 0, \ldots, P-1$), and $U(LBP_{P,R})$ is defined as follows:

$$U(LBP_{P,R}) = |s(g_{P-1} - g_c) - s(g_0 - g_c)| + \sum_{p=1}^{P-1} |s(g_p - g_c) - s(g_{p-1} - g_c)| \tag{3.7}$$

such that patterns having more than two binary transitions are assigned the common code $P + 1$. Thus, as shown in Figure 3.4, there are eight possible patterns for each of pattern codes 1–7 (since rotation invariance is not used). The cases when all eight sampling locations are 0 or 1 yield an additional two patterns, while the remaining patterns are all assigned the same 'uniform' code. Hence, the total

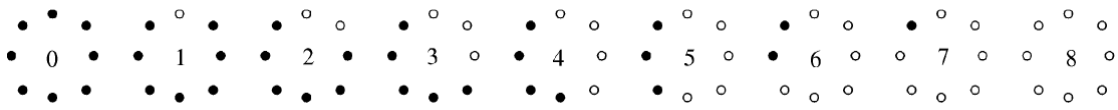


Figure 3.4: Binary patterns that can occur in a circularly symmetric neighbour set of $P = 8$ pixels with radius $R = 1$, using 'uniform' patterns and without rotation invariance. Black and white circles correspond to bit values g_p of 0 and 1, respectively [31].

number of values that can be assigned to a neighbourhood is 59, which are then used to construct a histogram such that 59D vectors are obtained for each patch.

The MLBP extension is also used, which concatenates LBP descriptors computed with radii $R = \{1, 3, 5, 7\}$ to yield 236D vectors. These parameters have been chosen since they are similar to those used in [12] and yielded the best performance when evaluated on viewed hand-drawn sketches.

Whilst SIFT or HOG descriptors computed on patches are often also used with M/LBP in approaches proposed in literature to extract shape information, their use in this work is not highly beneficial. This is because such descriptors represent the frequency of occurrence of orientations. However, the images used for feature extraction in the proposed system contain the responses at only one specific orientation. Consequently, descriptors of this type do not offer much useful information since shape information is already being implicitly considered.

3.2.3 Sketch-Photo Matching

The MLBP features from each patch are concatenated into one $154 \times 236 = 36,344$D vector. Discriminant analysis is then performed for each of the G 36,344D feature vectors by first applying PCA followed by LDA on a training set of images [53], an approach shown to be beneficial in face recognition [12]. The 'PhD Toolbox'[1] was utilised for this task [139,140]. The number of eigenvectors used is the upper bound $c-1$, where c is the number of classes (subjects) as described in [53]. After the projection matrix has been obtained, the mean-subtracted features of each sketch and the gallery photos are projected onto the subspace. Comparison between sketches and photos is performed by measuring the Spearman Rank Order Correlation Coefficient (SROCC) between the resultant vectors of a sketch and a gallery photo to yield a similarity score for each filter. More specifically, SROCC measures the strength and direction of the monotonic relationship between two ranked vectors [141] such that scores lie within the range $[-1, 1]$, where a value of 0 indicates no correlation and 1 and -1 represent maximum positive correlation (where the values of both vectors tend to increase) and negative correlation (the values of one vector increase while the values of the other vector decrease), respectively. SROCC

[1] Available at: `http://luks.fe.uni-lj.si/sl/osebje/vitomir/face_tools/PhDface/index.html`

is computed as follows:

$$\rho = \frac{\sum_i (R(x_i) - \overline{R(x)})(R(y_i) - \overline{R(y)})}{\sqrt{\sum_i (R(x_i) - \overline{R(x)})^2 \sum_i (R(y_i) - \overline{R(y)})^2}} \tag{3.8}$$

while Equation (3.9) hereunder can be used when the data does not contain tied ranks:

$$\rho = 1 - \frac{6 \sum_i (R(x_i) - R(y_i))^2}{n(n^2 - 1)} \tag{3.9}$$

where ρ is the correlation result between two vectors x and y of equal length, $R(x)$ and $R(y)$ are the ranks of x and y, respectively, $\overline{R(x)}$ and $\overline{R(y)}$ are the mean values of $R(x)$ and $R(y)$, respectively, and n is the number of variables in x and y. [141,142].

Given a gallery containing N subjects, there are G vectors $\vec{J_k} \in \mathbb{R}^N, k = 1, 2, \ldots, G$ containing the scores for each comparison between a probe sketch and a gallery photo to yield $G \times N$ scores for each probe sketch. Min-max normalisation is applied using Equation (5.1) to obtain $\vec{\hat{J}_k} \in \mathbb{R}^N, k = 1, 2, \ldots, G$. These G vectors, representing the matching scores of the G filters, are finally fused using the sum-of-scores method as shown in Equation (3.3), similar to the approach described in Section 3.1 for the fusion of intra- and inter-modality algorithms. This yields a vector $\vec{F} \in \mathbb{R}^N$ containing N LGMS scores for each probe sketch, representing the similarity between the sketch and all the photos in the gallery.

3.3 DEEPS and DEEPS-M

The majority of inter-modality algorithms use hand-crafted features such as the SIFT and MLBP, and have attained good performance [12,14,131]. However, it is unlikely that such features are optimal since they were not designed for inter-modality face recognition [15], and it would therefore be desirable to design and use potentially superior feature descriptors that are better adapted for the task of face photo-sketch recognition. This can be performed with the aid of *deep learning*, which has become a hot research topic owing to its great success in several application domains including traditional VIS-VIS face recognition and image super-resolution [93,107,113,143], as discussed in Section 2.5. However, there has been limited work in using deep learning for face photo-sketch recognition. This was thus investigated, with the following contributions:

- It is shown that simply applying a deep network, even one trained for unconstrained face recognition that achieved state-of-the-art performance, typically results in poor recognition rates when matching sketches with photographs (ref. performance of VGG-Face in Chapter 6).
- While transfer learning is beneficial, it is shown that performance still lags behind leading methods since traditional data augmentation techniques are insufficient to avoid the problem of having only one sketch per subject (ref. Appendix A.4.1).
- A 3D morphable model is used to automatically create a set of synthetic images to artificially expand the number of images per subject. Applying transfer learning to a deep network using these images yields a framework that outperforms state-of-the-art algorithms (ref. Chapter 6).
- Since forensic sketches typically contain several inaccuracies, the synthetic sketches can bear a better liking to the matching photo than the original sketch (ref. Section 3.3.4). In fact, performance is improved when multiple sketches for each subject are used for comparison with the gallery photos (ref. Section 6.4).
- The fusion of the proposed architecture with the LGMS algorithm is shown to yield further improved performance on both viewed and forensic sketches (ref. Chapter 6).

As shown in Figure 3.5, the proposed framework consists of a Deep Convolutional Neural Network (DCNN) together with a triplet embedding that optimises the features for verification, and a data augmentation approach to circumvent the lack of multiple images per subject. The framework is demonstrated to outperform leading methods, with the resultant method denoted the DEEP (face) Photo-Sketch System (DEEPS). A description of the components of this system will now be provided.

3.3.1 Deep Convolutional Neural Network

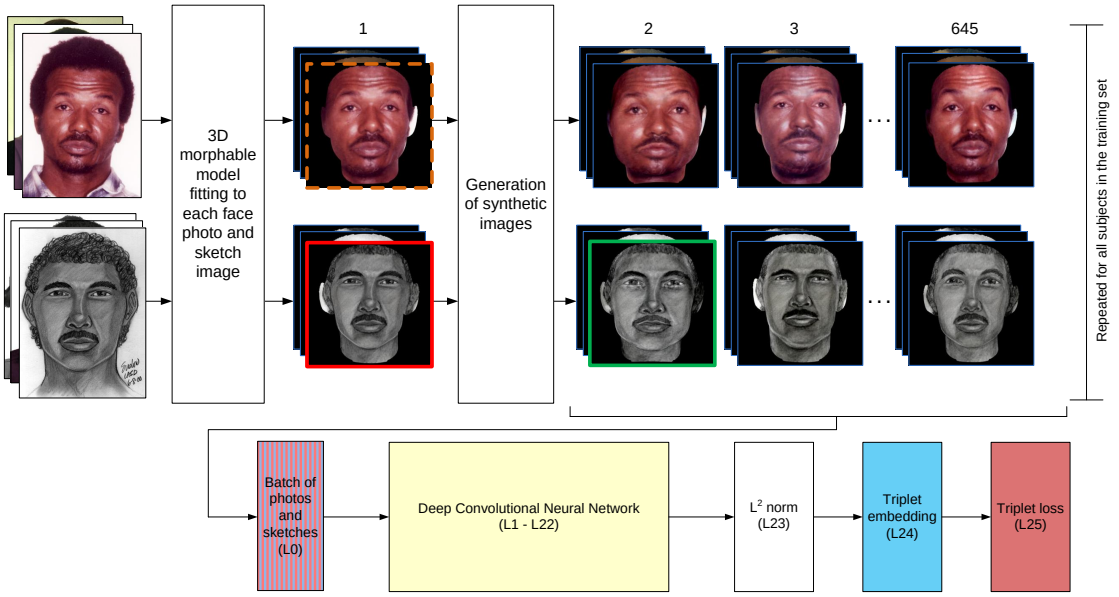


Figure 3.5: The proposed architecture, where synthetic images are created and used to train the DCNN in [93] via transfer learning. The first and second rows contain original and synthesised photos and sketches, respectively, of a subject in the PRIP-HDC forensic sketch database [19]. Column '1' contains images fitted with a 3D morphable model, and '2' to '645' are synthesised versions of '1'. The synthetic sketch of variation '2' (represented with a green border) has a more rounded appearance than the original sketch (red border) and bears a subjectively better similarity to the corresponding original photo (dashed orange border). Detailed architecture of the deep network is shown in Table 3.1.

Although there exist several deep learning architectures, the DCNN is the one most suited for image processing tasks since it is able to exploit the spatial relationships within images. Moreover, the proposed LGMS method exhibits several similarities to the basic layout of a DCNN, including image filtering, feature extraction, and feature dimensionality reduction. Due to the good performance attained by LGMS as discussed in Chapter 6, it was determined that a DCNN would be the most suitable deep architecture for DEEPS.

Table 3.1: Network architecture when training for verification using the triplet loss, based on the VGG-Face descriptor network [93]. An entry $\{\text{conv\_}R : [H \times W \times F]\}$ represents a convolutional layer having an $R \times R$ receptive field size, an input of dimension $H \times W$, and F filters. All convolutional layers except L24 use the ReLU function [94].

L0	input: $[224 \times 224 \times 3]$	L13	conv_3: $[28 \times 28 \times 512]$
L1	conv_3: $[224 \times 224 \times 64]$	L14	mpool_2: $[14 \times 14 \times 512]$
L2	conv_3: $[224 \times 224 \times 64]$	L15	conv_3: $[14 \times 14 \times 512]$
L3	mpool_2: $[112 \times 112 \times 64]$	L16	conv_3: $[14 \times 14 \times 512]$
L4	conv_3: $[112 \times 112 \times 128]$	L17	conv_3: $[14 \times 14 \times 512]$
L5	conv_3: $[112 \times 112 \times 128]$	L18	mpool_2: $[7 \times 7 \times 512]$
L6	mpool_2: $[56 \times 56 \times 128]$	L19	conv_7: $[1 \times 1 \times 4096]$
L7	conv_3: $[56 \times 56 \times 256]$	L20	dropout_1: $[1 \times 1 \times 4096]$
L8	conv_3: $[56 \times 56 \times 256]$	L21	conv_1: $[1 \times 1 \times 4096]$
L9	conv_3: $[56 \times 56 \times 256]$	L22	dropout_1: $[1 \times 1 \times 4096]$
L10	mpool_2: $[28 \times 28 \times 256]$	L23	norm_1: $[1 \times 1 \times 4096]$
L11	conv_3: $[28 \times 28 \times 512]$	L24	conv_1: $[1 \times 1 \times 1024]$
L12	conv_3: $[28 \times 28 \times 512]$	L25	Triplet loss function

Since performance of neural networks tends to increase with more layers and filters [93], it would be ideal to design a wide and deep network for face photo-sketch recognition. However, a significant amount of training data is required, which also leads to long training times (often on the order of weeks). As a result, researchers often apply the process of *transfer learning*, where a pre-trained network's parameters are fine-tuned with a training set that contains samples from the database on which the network will be employed. Hence, the use of a pre-trained network enables faster convergence, decreases the probability of finding poor local minima, and leverages the regularisation effect that enables better generalisation [144,145].

This work also benefits from transfer learning by starting with the VGG-Face FRS DCNN described in [93] and shown in Table 3.1, that was trained with 2.6M face images of 2,622 subjects. This network was chosen for several reasons: first, it was designed specifically for recognition of faces as done in this work, and the face photos used for training also form one of the modalities used in the proposed approach. Secondly, the VGG-Face network was shown to be among the leading FRSs for unconstrained face recognition [93]. Hence, the network provides a better basis on which fine-tuning can be performed when compared to a network trained for other tasks.

A similar implementation methodology to that used for the VGG-Face network is employed, whereby all the parameters of the DCNN are first trained for the task of classification using the softmax log-loss objective function (tuning layers L1-L21),

after which training for verification using the triplet-loss is performed for the triplet embedding layer only (i.e. learning L24) while freezing the parameters of the rest of the network (i.e. setting their learning rate to zero). Stochastic Gradient Descent (SGD) with momentum is used to train the network in each case. However, the last fully-connected layer of the VGG-Face network (which maps the D-dimensional feature descriptor to a number of classes corresponding to the number of distinct identities in the training set) must be re-initialised since the network was trained using different subjects than the ones considered in this work. Following the procedure in [93], biases are set to 0 and weights are randomly sampled from a Gaussian distribution with zero mean and 10^{-2} standard deviation. Given that this layer is re-initialised, unlike the rest of the network, the learning rate of its parameters is set to ten times the global learning rate of 10^{-3}. The low learning rate was chosen to limit the rate of change of the parameters and thus enable better convergence, since the parameters should not require great adjustments given that they are already pre-trained. On the other hand, the learning rate of the last layer must be higher since it is randomly initialised without any prior training whatsoever.

Photos and the corresponding sketches are used for each subject in the training set, allowing the network to learn the relationship between the two modalities. In other words, the aim of the network is to learn modality-invariant parameters such that it may correctly classify both photos *and* sketches.

After the network is trained for classification, the last two layers (the last fully-connected layer and the softmax log-loss layer) are replaced with three layers: (i) a layer that normalises the output feature vector to unit length (L23), followed by (ii) a fully-connected layer (L24) consisting of D inputs and L outputs, $L \ll D$, and (iii) a triplet-loss layer (L25). Layer L24 performs dimensionality reduction and outputs a vector that is suitable for verification, which should yield vectors whose distance with respect to other vectors is small for input face images of the same subject, and large for different subjects. L25 computes the error of the objective function to determine how the parameters must be adjusted using SGD.

The input to the network is a patch of size 224×224 that is randomly cropped from a face image and flipped with 50% probability, with the mean of the images in the training set subtracted to ascertain stability of the learning algorithm [93]. At test time, a process similar to that employed for the original VGG-Face network as described in [93] is performed, namely the dropout layers are removed and images

are scaled to three sizes ($256 \times 256, 384 \times 384$ and 512×512) to enable multi-scale testing. Feature vectors are then computed for ten patches (the four corners, the centre, and their horizontal flips), extracted at each scale. The final descriptor is the average of the resultant 30 L-D feature vectors that are obtained for each probe (sketch) image and gallery (photo) image. As elaborated in Section 3.3.3, the feature vector of a probe image is then compared with those of all the gallery images using either Euclidean distance [93] or vector dot products [113]. Lastly, the results when tuning several parameters of the proposed system are shown and discussed in Appendix A.3.

3.3.2 Data Augmentation

A drawback of deep learning methods is the requirement of a large amount of data for robust learning, to reduce effects such as over-fitting and to learn more effective functions [93,94]. Extensive datasets containing not only a high number of unique classes but also numerous examples for each class have been created for tasks such as object and face recognition, and thus allow researchers to train and test their algorithms well. For example, the ImageNet database [146] contains a training set having 1.2 million images of 1000 categories. Such high numbers are possible due to the sheer availability of images on the Internet, such that search engines can be used to automatically retrieve images of interest. In the case of face recognition, databases such as the one used to train the VGG-Face network [93] are typically created by using celebrities as the individual subjects, many photos of whom are often captured and thus allow a database to be quite easily populated with multiple images per subject. This approach cannot be undertaken in the case of face sketches due to their limited availability as a result of privacy protection issues in the case of real-world forensic sketches, and the time consuming nature of sketch creation in the case of publicly available viewed sketch datasets. This means that the number of subjects represented with a sketch image is quite limited, even when combining all available databases. Moreover, sketch databases typically contain only one sketch per subject, and the face photo datasets used to construct sketch databases often contain a limited number of photos per subject also. However, object and face databases contain hundreds of examples for each unique entity which exhibit several variations. In the case of face recognition, these variations span factors such as expression and pose, and allow a network to

be robust to intra-class differences. Consequently, a deep network trained using just two images per subject (a sketch and a photo) would find it hard to reliably distinguish them from different identities and at the same time learn intra-class similarities [147]. Even algorithms making use of vast amounts of data for training have found data augmentation techniques beneficial for system performance [93, 94,147]. However, it is shown in Appendix A.4 that even traditional augmentation methods yield limited performance improvement when using one photo and one sketch (the original images).

To circumvent this problem, the use of a 3D face morphable model[2] [148] (along with the approach in [149] to fit the model to face images using edge features[3]) is proposed to enable the automatic generation of synthetic face photos and sketches, with the additional benefit of normalising off-pose faces to be frontally aligned with no rotation (which is particularly useful in the case of photos). Changes to a face image include: (i) local changes to the individual facial features[4], and more global changes in terms of (ii) age (older or younger), (iii) gender (more female or more male), (iv) height (taller or shorter) and (v) weight (fatter or thinner). Of course, there is a virtually infinite number of ways in which a face image can be altered. In this work, 644 images are created for each face image. These include five random adjustments to the four facial components individually (yielding 20 images), and 624 adjustments to the age, gender, height and weight, both individually (i.e. changing one attribute at a time) and also when multiple attributes are changed simultaneously. The original image is also used, for a total of 645 photos and 645 sketches per subject[5]. Some examples of face photos and face sketches created with this approach are shown in Figure 3.5.

The proposed system therefore allows any face database to be expanded with an arbitrarily large number of images. This is especially the case for sketches, of which there is typically only one per subject in both publicly available datasets and in real-life. The increased number of samples will be shown in Appendix A.4 to be greatly beneficial for training a deep network.

[2]Available at: `http://faces.cs.unibas.ch/bfm/main.php`
[3]Available at: `https://github.com/waps101/3DMM_edges`
[4]Facial features encompass the eyes, nose, mouth and the rest of the face
[5]The exact parameters may be found at: `http://wp.me/P6CDe8-7D`

3.3.3 Triplet-loss embedding scheme

The triplet-loss objective function has been used in several state-of-the-art systems to train the network for the final application of a FRS, namely identification via verification. Given a triplet $\{a, p, n\}$, the aim is to reduce the distance (or increase similarity) between an image a of a subject called the *anchor* and another image p of the same subject known as the *positive* example, while increasing the distance (minimising the similarity) between the anchor and an example n from a different subject called the *negative* example. Two main methods have been proposed in literature: *Triplet Distance Embedding (TDE)* based on Euclidean distance [93,113], and *Triplet Similarity Embedding (TSE)* based on vector dot product similarities [114]. For both methods, an input face image l_i, $i \in \{a, p, n\}$ yields an output $\phi(l_i) \in \mathbb{R}^D$ that is projected to a $L \ll D$ dimensional space to obtain $\vec{x}_i \in \mathbb{R}^L$:

$$\vec{x}_i = W \frac{\phi(l_i)}{\|\phi(l_i)\|_2} \tag{3.10}$$

where $W \in \mathbb{R}^{L \times D}$ is the projection matrix that is learned, with L and D set to 1024 and 4096, respectively, as done for the VGG-Face descriptor network [93]. The last fully-connected layer now corresponds to the weight matrix W, corresponding to L24 in Table 3.1. Biases and their learning rate are set to zero along with the parameters of the rest of the network, such that the training process only affects W.

The condition that is imposed for TDE is:

$$\|\vec{x}_a - \vec{x}_p\|_2^2 + \alpha < \|\vec{x}_a - \vec{x}_n\|_2^2 \tag{3.11}$$

where $\alpha \geq 0$ is a fixed scalar denoting a margin between the distances comparing the anchor and positive, and the anchor and negative. The objective function to be minimised is thus:

$$E(W) = \sum_{(a,p,n) \in T} \max\{C, \|\vec{x}_a - \vec{x}_p\|_2^2 + \alpha - \|\vec{x}_a - \vec{x}_n\|_2^2\} \tag{3.12}$$

where T is a set consisting of training triplets, and the $C = 0$ is the minimum allowable error. For SGD as used in this work, the gradient of this function with respect to the inputs must be calculated to determine how the weights must be

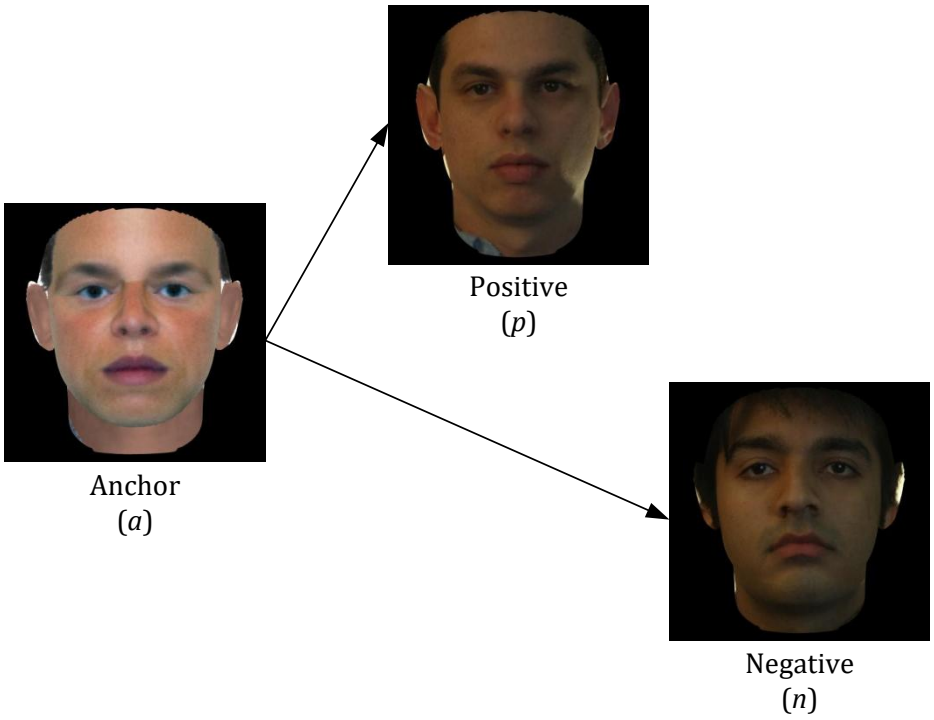


Figure 3.6: The aim of using the triplet loss when training the network is to minimise the distance (or maximise the similarity) between the anchor and positive examples that depict the same subject, and to maximise the distance (or minimise the similarity) between the anchor and negative examples that have a different identity.

adjusted in order to reduce the error in backpropagation, as follows [150]:

$$\frac{\partial E(W)}{\partial \vec{x}_i} = \begin{cases} 2(\vec{x}_n - \vec{x}_p), & \text{if } i = a \text{ and } E(W) > C \\ 2(\vec{x}_p - \vec{x}_a), & \text{if } i = p \text{ and } E(W) > C \\ 2(\vec{x}_a - \vec{x}_n), & \text{if } i = n \text{ and } E(W) > C \\ 0, & \text{otherwise} \end{cases}$$

The TSE approach [114] maximises similarities between the anchor and positive, while minimising similarities between the anchor and negative with the following condition:

$$\vec{x'}_a \cdot \vec{x}_n + \alpha < \vec{x'}_a \cdot \vec{x}_p \tag{3.13}$$

where $\vec{x'}$ denotes the transpose of $\vec{x}$. The function to be minimised is thus:

$$E(W) = \sum_{(a,p,n)\in T} \max\{C, \vec{x'}_a \cdot \vec{x}_n - \vec{x'}_a \cdot \vec{x}_p + \alpha\} \tag{3.14}$$

and the gradients to be backpropagated are as follows:

$$\frac{\partial E(W)}{\partial \vec{x}_i} = \begin{cases} \vec{x}_n - \vec{x}_p, & \text{if } i = a \text{ and } E(W) > C \\ -\vec{x}_a, & \text{if } i = p \text{ and } E(W) > C \\ \vec{x}_a, & \text{if } i = n \text{ and } E(W) > C \\ 0, & \text{otherwise} \end{cases}$$

Following recommendations in literature [93,113], all positive pairs (a, p) are selected and n is chosen at random among those images which violate the triplet loss margin. Triplets are chosen while the network is being trained.

3.3.4 Multiple synthetic sketches for testing

While the synthetic images aid learning by permitting the DCNN to be flexible for disparities in the facial attributes of photos and sketches, it will be shown in Section 6.4 that they may also be useful during system deployment when determining the identity of a subject in a sketch (where the synthetic test sketches are created using the same procedure outlined in Section 3.3.2). This leverages the observation that adjustment of facial attributes may yield a sketch which counteracts the distortions and exaggerations that are typically present within sketches when compared to the corresponding photos. A real-world example of such a case is depicted in Figure 3.5. Hence, the features of a subset of 199 synthetic sketches and the original are computed for each subject. Since a sketch is compared to the gallery photos, it is represented with 200 distance measures for each subject in the gallery. These distances are fused using: (i) the median of the top 49 matches and the match to the original sketch, and (ii) the best match among 9 sketches and the original. Two distance values are thus obtained for each comparison of a sketch with a photo, that are combined using sum-of-scores fusion and min-max normalisation as described by Equations (3.3) and (5.1), respectively, which are

reportedly among the best fusion approaches [2,80,135]. This method is denoted DEEPS Multi-sketch (DEEPS-M) and is only applied on forensic sketches, since viewed sketches generally bear an already close resemblance to the original photo and any alteration to the facial attributes will likely reduce similarity.

Work in literature considering the use of multiple sketches at test time is quite limited. Most relevant to this paper is the work done in [51,80], but the number of subjects and sketches used were both limited since the latter were manually created by employing several artists or software operators, making the process costly and time-consuming. These problems are critical, especially in the time-sensitive nature of real-world criminal investigations. This is in contrast to the proposed approach which generates images automatically and hence allows more examples to be created for both training and testing.

3.3.5 System fusion

Empirically, it was observed that there are several cases where LGMS performs noticeably worse than DEEPS, and vice versa, on both viewed and forensic sketches as shown in Figure 6.4 on Page 84. Fusion of methods based on engineered features and learned features has also been shown to be capable of enabling performance improvement in other domains such as video indexing [151]. Hence, the two approaches are combined to determine if these methods can benefit from complementary information, using also sum-of-scores fusion and min-max normalisation [135] similar to the fusion of intra- and inter-modality algorithms described in Section 3.1 on Pages 38 to 39.

3.4 Summary

Four primary methods have been proposed in this research, spanning both intra- and inter-modality techniques. The first method is the extension of the Eigen-transformation (ET) intra-modality approach to operate on local patches, yielding Eigenpatches (EP). Fusion of intra- and inter-modality methods is also considered, and is shown in Chapter 6 to be capable of providing improved performance. This work was also described in a paper presented at the International Conference on Computer Analysis of Images and Patterns (CAIP2015), [62].

An inter-modality method called log-Gabor-MLBP-SROCC (LGMS) was then created, which was described in a paper presented at the International European Signal Processing Conference (EUSIPCO), [131]. LGMS uses engineered features together with subspace projection and comparison using correlation of the resultant feature vectors. This method is shown in Chapter 6 to be able to outperform state-of-the-art methods proposed in literature.

Finally, the DEEP (face) Photo-Sketch System (DEEPS) was created, and is the first method to use a very deep convolutional neural network for face photo-sketch recognition. To circumvent the single sketch image per subject problem, which inhibits a network from learning robust features, a 3D morphable model was used to enable variation of facial attributes and automatically create a new large set of synthetic images. Due to the distortions and exaggerations found in sketch images when compared to the original face photo, these variations also allow a network to be flexible in the presence of these artefacts. Indeed, the variations may yield synthetic sketches that counteract some of these distortions to make them more similar to the corresponding photo than the original sketch. In fact, it will also be shown in Chapter 6 that they can be beneficial even during system deployment. This work was also described in a letter published in the IEEE Signal Processing Letters publication, [132], and in a paper accepted for publication in the IEEE Transactions on Information Forensics and Security, [133].

Chapter 4

UoM-SGFS Database

The University of Malta Software-Generated Face-Sketch (UoM-SGFS) database[1] was constructed in this research to circumvent the limited public availability of software-generated sketches, which were created using the EFIT-V software program [22] that is used by numerous law enforcement agencies. Presented in [14], the database initially contained 600 sketches of 300 subjects in the Color FERET database [13,152] and was later doubled in size to contain 1200 sketches of 600 subjects. The UoM-SGFS database is thus:

- One of the largest face sketch databases
- The largest face sketch database containing software-generated sketches
- The only database containing all sketches represented in full-colour
- The only database containing sketches created with EFIT-V [22]
- One of the few databases containing more than one sketch image per subject

As a result, this database is highly useful to researchers working in the field of face photo-sketch recognition. In fact, 13 requests to use this database have been received at the time of writing this dissertation.

The motivation to create the database will first be given, followed by an overview of EFIT-V and a more detailed description of this new database.

[1] UoM-SGFS database and additional data available at: `http://wp.me/P6CDe8-4q`, `http://goo.gl/KYeQxt`

4.1 Motivation to create the database

Despite the more prevalent use of software-generated sketches in the real-world (compared to hand-drawn sketches) [6], there exist few such sketches which are publicly available. Indeed, only two databases containing software-generated sketches are publicly available. The first is the PRIP Viewed Software-Generated Composite (PRIP-VSGC) dataset [6,19] containing viewed software-generated composite sketches created using Identi-Kit. Although other sketches were created by another user and by another software program, these have not been made publicly available. As a result, this database only contains one sketch for each of the 123 subjects considered. The Extended PRIP (EPRIP) database [121,124] contains 123 sketches of the same subjects, created by an Indian software operator.

Other databases have been used in literature, but these have remained private. As a result, only the PRIP-VSGC and the EPRIP databases contain software-generated composite sketches that can be used by researchers. However, the number of sketches available is quite low for robust algorithm evaluation and the sketches in the PRIP-VSGC often look highly dissimilar to the corresponding photos. Indeed, the creators of the database themselves use the sketches generated by a different operator and a different software program for algorithm evaluation, due to better representational accuracy [6]. Moreover, colour information which has been found to be capable of improving face recognition performance [153] cannot be exploited since all sketches are grayscale. Lastly, there is no publicly available database containing sketches with the EFIT-V software [22] that is used by several law enforcement agencies around the world. The UoM-SGFS database was therefore created to counteract these issues, primarily the lack of publicly available (i) software-generated composite sketches, (ii) sketches represented in full-colour, and (iii) sketches created with the popular EFIT-V software.

4.2 EFIT-V overview

Most facial composite systems involve the selection and spatial configuration of individual facial features, thus relying on the witness' ability to recall and describe these features. However, human recognition and synthesis of faces is a global process relying on the interaction of all features in the face [154]. Thus, EFIT-V

uses a global face model which allows witnesses to recognise perpetrators' faces rather than recall them. This is done by presenting the witness with a set of faces, who retains or rejects them depending on their similarity to the face of the criminal. A genetic algorithm learns the witness's choices and presents a new set of faces through a process of mutations. The procedure is repeated again until the witness is satisfied that a better likeness cannot be generated. Adjustments such as translation, scaling, and rotation may also be performed for each facial component.

EFIT-V allows features other than the facial components to be depicted, including jewellery and glasses. These may not only help the eyewitness during the composite generation process but also allow the creation of sketches which can depict discriminative and thus vital information about the perpetrators. Image editing software is directly supported in EFIT-V, to allow fine-tuning of details [155].

The older EFIT software has proven to be effective in generating a good likeness to suspects, and was shown to be able to outperform not only other software systems but also hand-drawn sketches [155]. As a result, the enhanced EFIT-V is nowadays used by several law enforcement agencies in more than 30 countries around the world including New Scotland Yard, New South Wales Police in Australia and the Malta Police Force. More information may be found in [22,154].

4.3 Database details

The face photographs for the UoM-SGFS database were obtained from the Color FERET database [13,152], chosen since it contains a large number of good quality frontal images and a variety of different age ranges and ethnicities. Most subjects selected are White to minimise the 'other-race effect', since the software operator creating the sketches was also White. However, subjects belonging to different races to that of the operator were still included to represent real-life conditions where both law enforcement officers and eyewitnesses may belong to a different race than the perpetrator. In addition, the EFIT-V operator was trained by a qualified forensic scientist from a local police force so as to ensure that practices adopted in real-life were adopted in the creation of the new database. Moreover, photos in the Color FERET database often contain non-uniform lighting and slight head rotations, making the recognition task more challenging but also more realistic.

The UoM-SGFS dataset contains two viewed sketches for each of the 600 subjects considered and is partitioned into two sets. *Set A* contains those sketches created using EFIT-V, where the number of steps performed in the program was minimised to avoid producing composites that are overly similar to the original photo. The sketches in Set A were then lightly altered with the Corel PaintShop Pro X7 image editing program [156] to fine-tune details that cannot be easily modified with EFIT-V (as done in real-life), yielding *Set B*. Hence, sketches in Set B are generally closer in appearance to the corresponding face photos.

EFIT-V also allows the depiction of shoulders, which can indicate the type of clothes that the perpetrator was wearing and the physique (e.g. fat, muscular, etc.). While the type of clothing is important, more emphasis was given to the physique of the subject since it provides more salient information. In addition, any accessories such as jewellery and hats are generally slightly different to those shown in the original photograph and sometimes omitted in the UoM-SGFS database sketches, to mimic memory losses of eyewitnesses. Lastly, to the best of the authors' knowledge, all sketches currently available are represented in grayscale. However, the sketches in the new database are all colourful. Since colour has been shown to yield improved performance for face recognition [153], its use may now also be considered for photo-sketch recognition to potentially improve performance. Some examples of sketches in the UoM-SGFS database may be found in Figure 4.1.

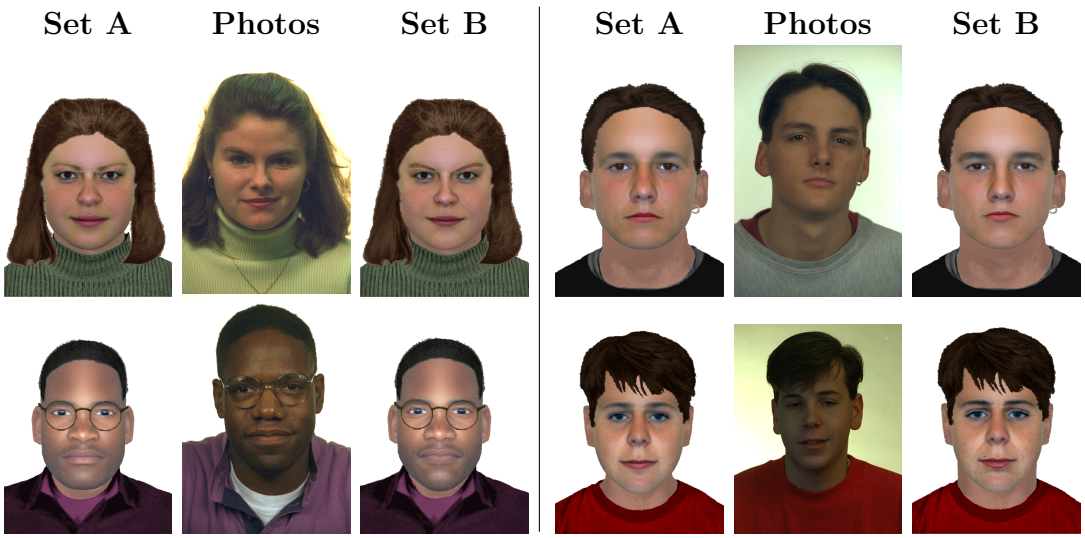


Figure 4.1: Photos of four subjects in the Color FERET database [13,152] and the corresponding sketches from the two sets of the UoM-SGFS database

Chapter 5

Implementation & Evaluation Methodology

The algorithms designed in this project will now be outlined and their performance compared to popular and state-of-the-art algorithms. The architecture used to implement and evaluate face photo-sketch recognition algorithms will be given in Section 5.1 followed by a description of the databases used in Section 5.2. An overview of the algorithms proposed in literature that have been chosen as a basis for comparison is given in Section 5.3 while an overview of the methods proposed in this research that will be evaluated is given in Section 5.4. The protocol used to evaluate the algorithms will be given in Section 5.5 followed by presentation and discussion of results in Chapter 6.

5.1 Framework used for evaluation

The system flow diagram of the framework used to evaluate face-sketch recognition algorithms is shown in Figure 5.1 and follows the approach presented in [18]. A given dataset of face photo-sketch pairs is first partitioned into three sets, such that one is used to train the intra-modality algorithm, one is used to train the face recogniser or inter-modality algorithm, and another is used to test the system. When evaluating intra-modality methods, training of the face recogniser/inter-modality algorithm is done by transforming the photos in the set used to train

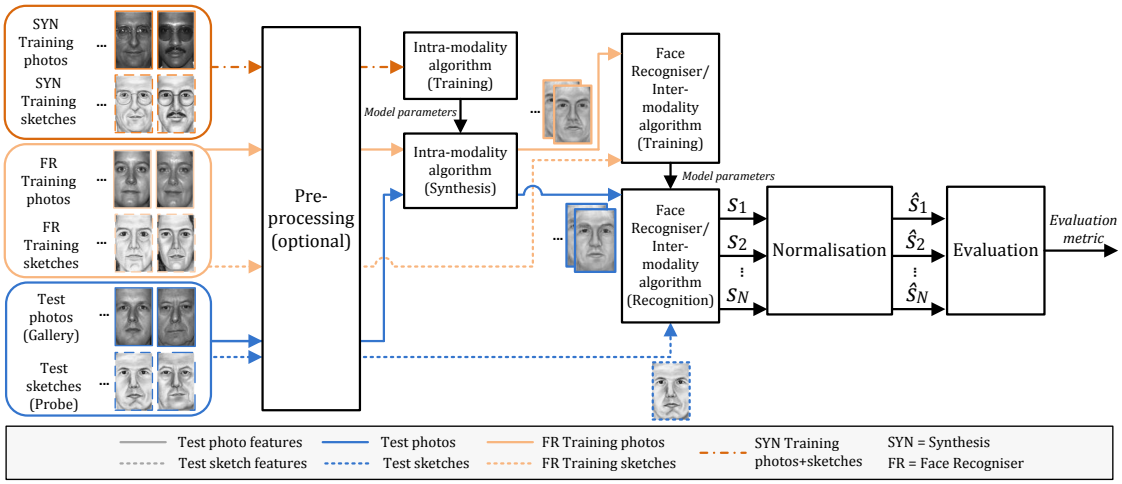
the face recogniser to pseudo-sketches and used together with the sketches of the same set when photo-to-sketch synthesis is being performed. Next, the photos in the test set are appended with an additional set of face photos which do not have a corresponding sketch image, to more closely represent the large mug-shot galleries used by law enforcement agencies (ref. Section 5.2.4). All photos are then transformed into sketches using the chosen intra-modality method, and each probe sketch is compared to the synthesised pseudo-sketches using a face recogniser or an inter-modality method. Photo-to-sketch synthesis is preferred since this mode achieves better performance than sketch-to-photo synthesis [62]. A total of N scores are thus generated for each query sketch, representing the similarity or distance between the sketch and the gallery of photos and are normalised using min-max normalisation:

$$\hat{\vec{s}} = \frac{(\vec{s} - min(\vec{s}))}{(max(\vec{s}) - min(\vec{s}))} \tag{5.1}$$

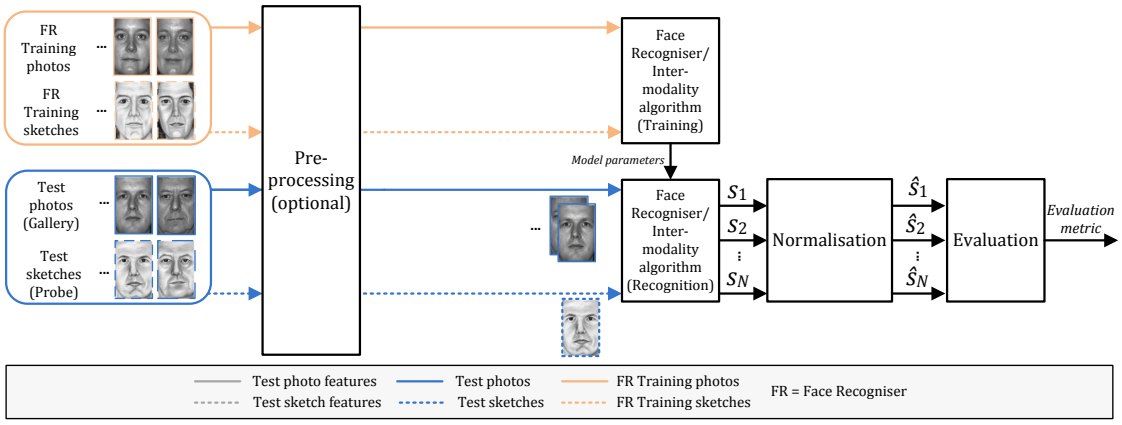
where $\vec{s} = [s_1, s_2, \ldots, s_N]$ is a vector of dimension N containing the scores for matching a probe with the N gallery images, $\hat{\vec{s}}$ contains the corresponding normalised scores, and $min(\vec{x})$ and $max(\vec{x})$ represent the minimum and maximum values of $\vec{x}$, respectively [135]. The scores can finally be used for evaluation as elaborated in Section 5.5.

When no intra-modality methods are used (i.e. for evaluation of only a face recogniser or inter-modality method), then the same approach is used but no synthesis is performed. Also, the images to be compared are the original query sketches with all VIS photos in the gallery. Therefore, training of the face recogniser is performed using only the photos and sketches in the set used to train the face recogniser.

For each algorithm, images are first geometrically normalised by rotating them such that the angle between the two eye centres is zero degrees and scaled such that the distance between eye centres and between the eyes and mouth is the same for all images. The coordinates of the eyes and mouths used for this task were those provided in the databases considered, while some were also labelled manually. Then, all images are tightly cropped to a height of 250 pixels and a width of 200 pixels such that the amount of hair visible in the image is minimised, as shown in Figure 5.1. Although it has been argued that hair should be retained to maximise the already limited amount of information present in sketch images



(a)



(b)

Figure 5.1: System flow diagram of the framework used for the evaluation of any algorithms considered, where N represents the number of gallery images: (a) Framework for intra-modality methods, where the process for photo-to-sketch synthesis is shown. For sketch-to-photo synthesis, the roles of sketches and photos simply change roles; (b) Framework for face recognisers and inter-modality methods, which is the same as that for intra-modality methods but excluding any synthesis-related processing.

[6], hair is not a reliable source of information since it is often difficult to be precisely described by an eyewitness and is a feature that can be easily changed in appearance. As a result, the hair depicted in a sketch may be very different from the hair in the corresponding mug-shot photo and may actually inhibit the performance of a recogniser in practice. Other pre-processing steps may optionally be done depending on the intra- and inter-algorithms considered.

5.2 Databases used

There are two main types of databases, namely face photo databases containing VIS face photos of subjects and those containing face sketches of subjects in a photo database. The databases considered in this research will now be described[1].

5.2.1 Viewed Hand-drawn face sketch databases

Since there are a greater number of viewed hand-drawn sketches than real-world hand-drawn forensic sketches, the former are primarily used for algorithm evaluation. Nevertheless, forensic sketches are also used, as described in Section 5.2.3.

The first hand-drawn sketch database used is the popular CUFS database[2] [17, 18,39], consisting of 606 sketches of the corresponding frontal face photos in the AR [16], XM2VTS [157] and CUHK student databases. However, these sketches have been criticised in literature as often being too similar to the corresponding photos [15,126]. Therefore, another 946 sketches are derived from the CUFSF database[3] [15,129], with the corresponding face photos obtained from the Color FERET database[4] [13,152]. Sketches in the CUFSF database were created such that they contain distortions and shape exaggerations, and are therefore more similar to sketches used by law enforcement agencies [15]. Each subject in these databases is represented using only one sketch and the corresponding photograph, i.e. there are 1552 subjects each having one photo-sketch pair.

Examples of the viewed hand-drawn sketches used in this work are depicted in Figure 1.2 on Page 3.

5.2.2 Viewed Software-generated face sketch database

The UoM-SGFS database as described in Chapter 4 is employed to evaluate the algorithms considered on software-generated composite sketches, with the corre-

[1]The detailed protocol may be found at `https://wp.me/P6CDe8-8B`
[2]Available at: `http://mmlab.ie.cuhk.edu.hk/archive/facesketch.html`
[3]Available at: `http://mmlab.ie.cuhk.edu.hk/archive/cufsf/`
[4]Available at: `http://www.nist.gov/itl/iad/ig/colorferet.cfm`

sponding photos obtained from the Color FERET database [13,152]. In total, the database contains 1200 sketches of 600 subjects. The sketches are split into two sets, which are evaluated independently (i.e. the sketches in Set A and Set B are not mixed, such that one sketch per subject is available for each algorithm).

Examples of the viewed software-generated composite sketches used in this work are depicted in Figure 4.1 on Page 59.

5.2.3 Forensic face sketch database

The PRIP-HDC dataset [19] is used to perform evaluation with real-world images, containing hand-drawn forensic sketches of 47 subjects created from eye-witness accounts in real-world investigations. The mug-shot photos were only available after the suspects in the sketches were identified [9,12]. All subjects were only used for testing by employing the same models that were trained on the viewed hand-drawn sketches, due to the dataset's small size. Also for this reason, traditional performance measures may yield inaccurate results and therefore analysis is performed using the ranks at which algorithms retrieve each subject directly.

5.2.4 Face photo databases

The photo gallery set is extended with 1521 subjects to mimic the mug-shot galleries maintained by law-enforcement agencies. These include 509 subjects from the MEDS-II database [128], 476 subjects from the FRGC v2.0 database[5], 337 subjects from the Multi-PIE database [158], and 199 subjects from the FEI database[6].

5.3 Baseline algorithms

The intra- and inter-modality algorithms proposed in literature that are chosen as a basis for comparison are some of the most popular and best-performing systems currently available, as follows:

[5]Available at: `http://www.nist.gov/itl/iad/ig/frgc.cfm`
[6]Available at: `http://fei.edu.br/~cet/facedatabase.html`

- **Face recognisers**
  - The traditional Eigenfaces (PCA) FRS [37] is used as a baseline for the intra-modality methods, which also use PCA as a face recogniser.
  - VGG-Face [93] is a state-of-the-art FRS that achieved 97-99% accuracy for unconstrained face recognition. Hence, the default network[7] is used as the benchmark for all methods, to determine if approaches designed specifically for face photo-sketch recognition can exceed the performance of a leading FRS. The VGG-Face network performance also serves as a baseline performance measure given that the proposed DEEPS and DEEPS-M systems are initialised with the parameters of this network.
- **Intra-modality algorithms with Eigenfaces (PCA) [37] as the FRS**
  - Eigentransformation (ET) [17,33], chosen since it is one of the simplest and most popular approaches for pseudo-photo/sketch synthesis
  - Locally Linear Embedding (LLE)-based approach in [34,159] chosen since, in contrast to ET, it operates on local regions and is reported to achieve good performance
- **Inter-modality methods**
  - Histogram of Averaged Orientation Gradients (HAOG) [10], chosen because it was reported to achieve 100% recognition rate even at low ranks on the CUFS database
  - CBR algorithm [6], which was designed to operate on software-generated sketches
  - D-RS algorithm [12,69], chosen since it performs even better than the more recent P-RS algorithm [12] on hand-drawn sketches and outperformed the FaceVACS COTS FRS which is considered to be one of the best commercial face matchers especially in HFR scenarios [12].
  - The fusion of the CBR and D-RS algorithms (CBR + D-RS), yielding the FaceSketchID system [19], is an approach that was shown to outperform both CBR and D-RS on viewed hand-drawn and software-generated sketches, and forensic hand-drawn sketches. CBR + D-RS is considered a state-of-the-art face photo-sketch recognition system.

[7] Available at: `http://www.robots.ox.ac.uk/~vgg/software/vgg_face`

In the case of the intra-modality methods, only the results of photo-to-sketch synthesis are reported due to better performance than sketch-to-photo synthesis, as demonstrated in Appendix A.1 and [62]. Also, only one sketch and one photo for each subject are used during testing of all the above methods (the original non-synthetic images). Lastly, since virtually no intra- or inter-modality methods proposed in literature were readily available, all algorithms considered have been implemented from scratch. To determine whether the re-implemented algorithms are faithful replicas of the originals, the same evaluation protocol as described in the papers was employed and results were verified to be virtually identical. In cases where the exact protocol could not be recreated, paper authors were contacted to provide the exact parameters of the algorithms to ensure a good implementation. Further discussion is also provided in Section 6.7.

5.4 Algorithms proposed in this work

The methods proposed in literature as described in Section 5.3 are compared to the work proposed in this research as detailed in Chapter 3, primarily: (i) EP [62], (ii) LGMS [131], and (iii) DEEPS and DEEPS-M. Fusion of these methods is also considered, via combination of both intra- and inter-modality methods and combination of multiple inter-modality methods (i.e. LGMS and DEEPS/DEEPS-M, following observations in literature that methods based on deeply learned features can still benefit from methods based on engineered features [151]). Hence, the performance of the following fused methods are also considered, using min-max normalisation and sum-of-scores fusion [135]: (i) ET+EP+HAOG as presented in [62], (ii) LGMS+EP as presented in [131], (iii) DEEPS/DEEPS-M+EP, (iv) LGMS+DEEPS/DEEPS-M, and (v) LGMS+DEEPS/DEEPS-M+EP. The last method represents the culmination of all proposed work in this research, namely the combination of all methods and the observation that intra- and inter-modality methods can provide useful complementary information.

5.5 Evaluation methodology

Due to the various database sizes, different evaluation protocols will be applied for the viewed hand-drawn sketches, for the viewed software-generated composite sketches and for the forensic sketches. Details will now be given hereunder.

5.5.1 Performance metrics

Since sketches are typically used when heinous crimes are committed, a large amount of attention is dedicated by investigators in solving the crime. Therefore, the top K best matching subjects are often given equal importance to the best match [6,12] in criminal investigations. K typically lies in the range $[50, 200]$ [6,7,9,12]. As a result, the main algorithm evaluation metrics are the Rank-retrieval rates, which indicate how many subjects were correctly identified at Rank-X. For example, a Rank-50 rate of 60% indicates that the algorithm correctly identified 60% of subjects as being within the top 50 best matches. Rank retrieval rates are displayed graphically using Cumulative Match Curves (CMCs). False Accept Rates (FARs) at fixed True Accept Rates (TARs)[8] along with the corresponding Receiver Operating Characteristics (ROC) curves are also provided. Lastly, the error rate at the ROC operating point where the number of false positives and false negatives are equal, known as the Equal Error Rate (EER), is also reported. The ROC curve and the EER are often used to measure the performance of FRSs [118,160].

It should be noted that all algorithms use the same train/test sets; hence, the reported standard deviations are a measure of the consistency in performance of each algorithm.

5.5.2 Database partitioning

An overview of the number of subjects used for training and testing the algorithms considered will now be provided. Since photo-to-sketch synthesis is employed, it

[8]FARs represent how many subjects are accepted as matches when they actually do not match the query image. This is done by setting a threshold on the similarity score (or distance measure) and determining the number of non-matching subjects having a higher score (smaller distance) than the threshold. The correct matches are used to calculate the TARs.

should be noted that the face recogniser used for the intra-modality methods is trained using both the sketches of the subjects in the training set and the corresponding synthesised sketches, similar to the approach in [18].

Viewed hand-drawn sketches

The 1552 photo-sketch pairs are partitioned into three sets as described in Section 5.1. In this project, 800 subjects are selected to train the inter-modality methods and face recognisers, 350 subjects are used to train the intra-modality methods, and the remaining remaining 402 subjects are used for testing. The sets are disjoint, such that each subject is used in only one of the three sets. Since the training and test sets are constructed by random selection of subjects, each algorithm is evaluated on three different training and test sets and the average and standard deviations of results recorded.

Viewed software-generated composite sketches

The UoM-SGFS database created in this work is partitioned such that 450 subjects are used for training while the remaining 150 subjects are used for testing. To avoid using too few subjects to train the intra modality algorithms and inter-modality methods or face recognisers, the same training set is used for both types of methods. Each algorithm is evaluated on five different train/test set-splits and the average and standard deviations of results recorded.

Forensic sketches

All 47 hand-drawn forensic sketches in the PRIP-HDC database [19] are used for testing only due to the low quantity of sketches available. The same models obtained after training the algorithms on the viewed hand-drawn sketches as discussed above are then employed on the forensic sketches.

Extended gallery

The gallery is extended by the face photos of 1521 subjects as described in Section 5.2.4, for all types of sketches considered. As a result, the test gallery set contains (i) $N = 402 + 1521 = 1923$ subjects when using the viewed hand-drawn sketches, (ii) $N = 150 + 1521 = 1671$ subjects when using the software-generated sketches, and (iii) $N = 47 + 1521 = 1568$ subjects when using the forensic sketches.

5.6 Summary

A robust evaluation methodology is employed, where several popular and state-of-the-art face photo-sketch synthesis and recognition methods are employed on both types of sketches that are used by law enforcement agencies, namely hand-drawn and software-generated sketches. Moreover, most viewed sketches were selected such that they contain several distortions and exaggerations which mimic real-world sketches. Furthermore, real-world forensic sketches are also used for evaluation. The gallery of photos with which sketches are compared is also populated with additional subjects to mimic the extensive mug-shot galleries maintained by law enforcement agencies. Thus, it is ensured that the task of determining the identity of subjects by comparing their sketch images with a gallery of photos corresponds to challenges faced in the real-world.

Chapter 6

Experimental Results

The experimental protocol outlined in Chapter 5 is used to obtain the results of the algorithms considered, which will be presented in Sections 6.1 to 6.4 when using viewed hand-drawn sketches, viewed software-generated sketches and real-world forensic sketches. Analyses of the proposed DEEPS network visualisation, and computation time and feature sizes of the algorithms considered are given in Section 6.5 and Section 6.6, respectively. Concluding remarks are finally provided in Section 6.8. Results of further in-depth analyses of the proposed methods are also provided in Appendix A while additional results are provided in Appendix B. The best result for each performance metric considered will be highlighted in boldface in the tables presented in this chapter and in the appendices.

6.1 Hand-drawn sketches

As shown in Table 6.1 and Figure 6.1, the intra-modality methods are superior to the FRSs including PCA which is used as the face recogniser for this type of methods, indicating that they are able to successfully reduce the modality gap. The higher performance of the proposed EP method compared to that attained by ET also shows that performing synthesis at a local level instead at a global level is beneficial. However, the intra-modality methods still typically lag behind the performance of the inter-modality approaches. The only exception is CBR, whose poor performance might be a result of being designed to operate on software-generated sketches. However, its fusion with D-RS as done in [19] to create

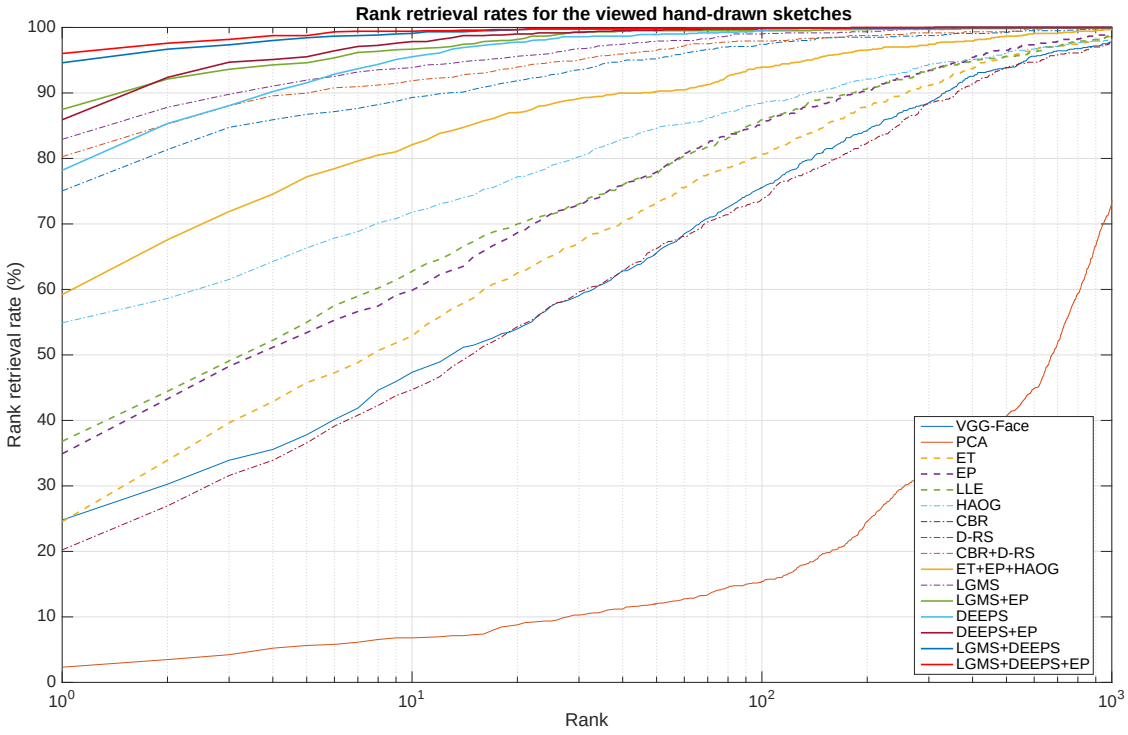
the state-of-the-art FaceSketchID system improves performance such that it is the second-best performing algorithm. The LGMS method proposed in this research outperforms all algorithms described in literature across every performance metric used. LGMS's matching rates also exhibit some of the lowest standard deviation values of all methods, showing that the proposed method performs consistently well on different train/test sets in contrast to algorithms such as ET and D-RS at lower ranks. This is especially impressive considering that challenging sketches from the CUFSF database, which contain several deformations and exaggerations that make the identification task more challenging, were used. In fact, this can be demonstrated by the decreased performance of the algorithms considered when compared to that reported in literature. For example, the HAOG algorithm that was reported to achieve a Rank-1 rate of 100% on the easier CUFS database only managed a rate of 54.9% with an extended gallery and additional sketches from the CUFSF database. Moreover, LGMS outperforms D-RS despite using LDA as the feature projection algorithm, which is theoretically inferior to RS-LDA as used in D-RS. However, as shown in Section 6.6, RS-LDA yields larger features that require more storage space than the use of LDA. Of course, this also holds for the method fusing CBR and D-RS, which uses significantly larger features compared to those of LGMS but only approaches similar performance at higher ranks.

DEEPS also achieves higher performance than all methods, including LGMS, except below Rank-5. However, since law enforcement agencies would still manually examine a few tens or hundreds of top matches, the performance at such low ranks is arguably less important than other ranks. Comparison of DEEPS with VGG-Face, which formed the basis of DEEPS, indicates that the proposed artificial expansion of the training set and application of transfer learning are beneficial given the significantly improved performance of DEEPS over the VGG-Face network. This is also shown to be true for the software-generated sketches in Appendix A.4.

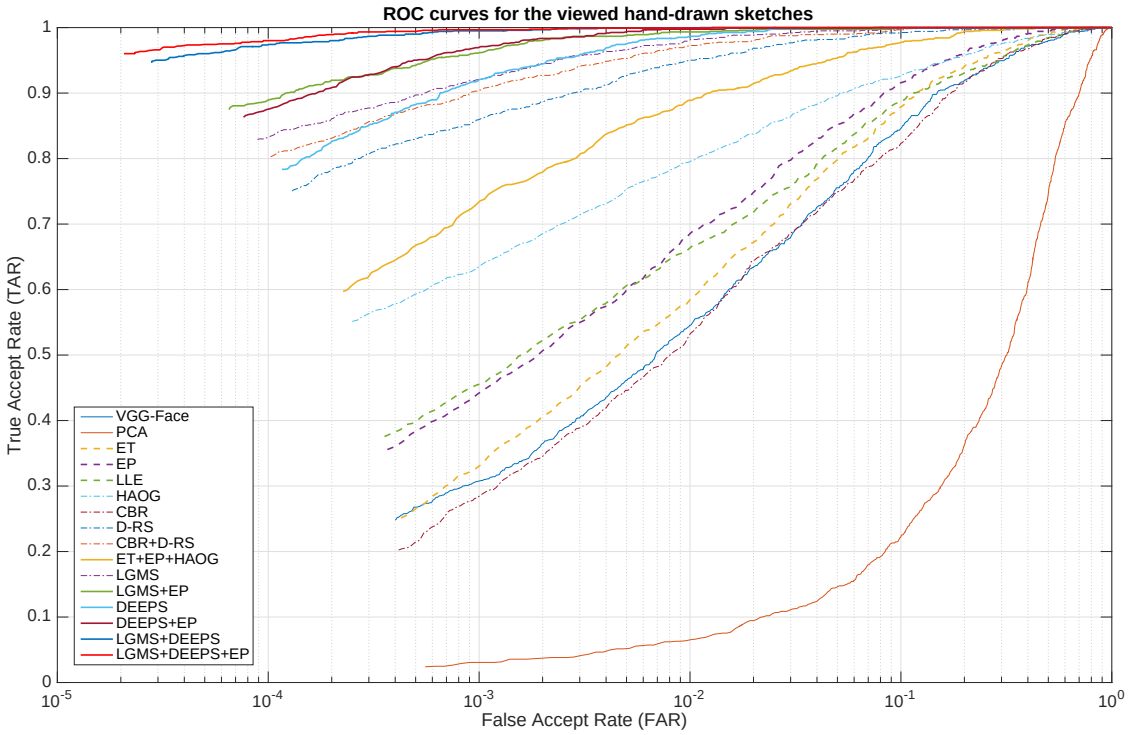
Fusion of intra- and inter-modality methods also enabled improved performance compared to the individual algorithm performance, for the fusion of HAOG with ET and EP (as published in [62]), the fusion of LGMS with EP (as published in [131], and fusion of DEEPS with EP.

Empirically, it was also observed that there are several cases where LGMS performs noticeably worse than DEEPS, and vice versa. As shown in Figure 6.4, this phenomenon also holds true for the forensic sketches. In fact, the combination of the

two methods using min-max normalisation and sum-of-scores fusion [135] yields improved performance, demonstrating that the two approaches provide complementary information. The resultant fused system comprehensively outperforms all other methods and is able to reduce error rates by 75.7% and 80.7% compared to DEEPS and LGMS, respectively. Finally, fusion of DEEPS and LGMS with EP improves performance further still, correctly retrieving over 96% of subjects at only Rank-1 and yielding the lowest EER of just 0.17%.



(a)



(b)

Figure 6.1: Results for algorithms considered when evaluated using the viewed hand-drawn sketches, averaged over 3 set splits: (a) CMC curves, (b) ROC curves

Table 6.1: Means and standard deviations over 3 train/test-set splits for algorithms evaluated on viewed hand-drawn sketches. Methods proposed in this research are shown in italics.

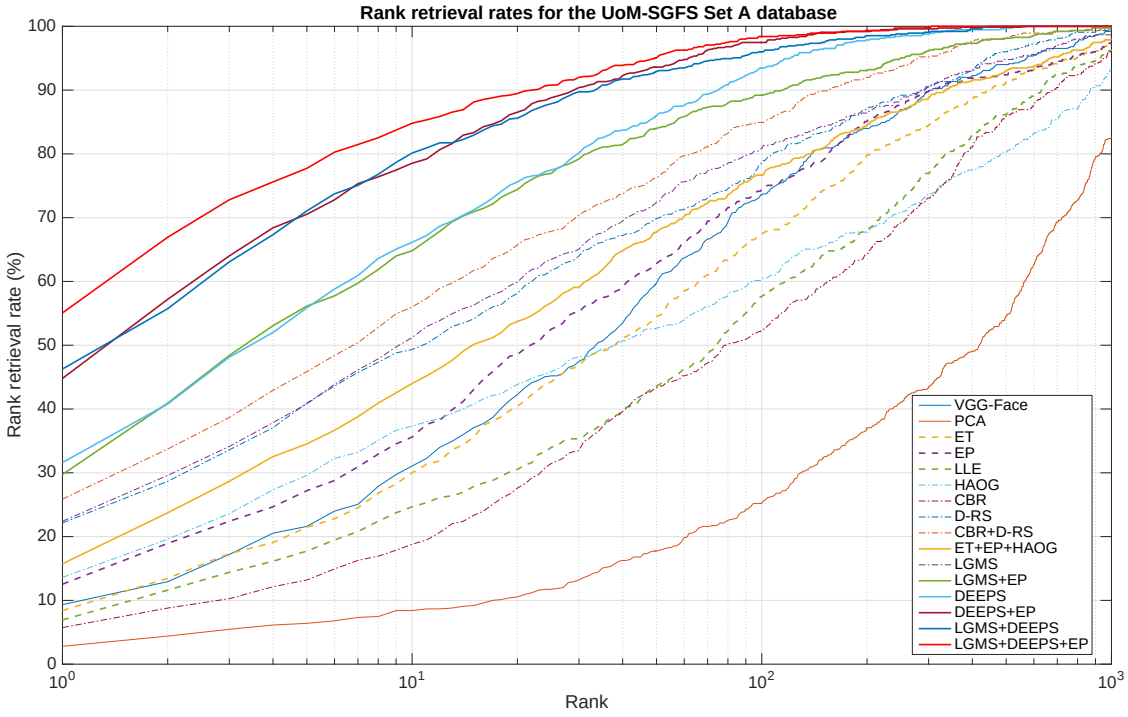
#	Method	Matching Rate (%) at Rank-X					TAR@FAR=0.1%	TAR@FAR=1.0%	EER (%)
		X=1	X=10	X=50	X=100	X=150			
1	VGG-Face [93]	24.79±1.87	47.35±1.37	65.59±1.23	75.54±0.76	81.09±1.51	30.76±2.35	54.56±2.23	12.51±0.53
2	PCA [37]	2.32±0.80	6.80±0.94	12.02±0.29	15.42±1.08	19.65±1.00	3.07±0.38	6.47±0.90	39.40±1.19
3	ET (+PCA) [17]	24.54±2.61	52.90±1.80	73.38±1.74	80.68±1.04	84.74±0.80	33.17±1.83	58.71±1.08	10.89±0.52
4	LLE (+PCA) [159]	36.82±1.97	62.77±2.52	77.53±0.52	85.90±1.01	89.22±0.76	46.19±1.25	66.42±1.14	10.49±0.75
5	HAOG [10]	54.89±2.17	71.81±0.80	84.58±0.50	88.47±0.76	90.46±0.29	63.60±2.09	80.43±0.87	7.68±0.31
6	CBR [6]	20.23±2.36	44.69±1.41	66.25±0.76	73.80±1.50	78.94±0.80	28.28±2.61	52.74±0.86	13.42±0.36
7	D-RS [12,69]	75.04±1.37	89.30±1.29	95.19±0.63	97.43±0.57	98.42±0.52	85.82±1.88	94.94±0.76	2.58±0.13
8	D-RS+CBR [19]	80.27±0.63	91.87±0.14	96.52±0.90	97.93±0.63	98.42±0.76	90.55±0.90	97.26±1.14	1.88±0.36
9	*EP* (+PCA)	34.91±3.31	59.87±1.75	77.94±0.63	85.32±0.66	88.39±0.29	44.20±1.46	68.41±1.32	9.06±0.16
10	*ET + EP + HAOG*	59.20±3.46	82.09±0.86	90.22±0.14	93.95±0.52	95.61±0.38	73.22±2.14	88.89±0.29	4.75±0.02
11	*LGMS*	82.92±1.25	93.86±0.38	97.93±0.63	99.00±0.00	99.17±0.14	92.04±0.90	98.01±0.25	1.35±0.25
12	*LGMS + EP*	87.48±2.26	96.68±0.29	99.50±0.25	99.50±0.25	99.67±0.14	96.19±0.38	99.34±0.29	0.82±0.15
13	*DEEPS*	78.19±0.52	95.52±1.49	98.92±0.76	99.50±0.25	**99.83±0.29**	91.96±0.72	98.67±0.52	1.07±0.25
14	*DEEPS + EP*	85.90±1.66	97.84±0.52	99.59±0.14	**99.83±0.14**	**99.92±0.14**	97.01±0.66	**99.75±0.43**	0.63±0.14
15	*LGMS + DEEPS*	94.61±0.80	99.09±0.14	**99.83±0.14**	**99.92±0.14**	**99.92±0.14**	**99.50±0.25**	**99.92±0.14**	0.26±0.04
16	*LGMS + DEEPS + EP*	**96.02±0.86**	**99.42±0.29**	**99.83±0.14**	**99.92±0.14**	**99.92±0.14**	**99.67±0.38**	**99.92±0.14**	**0.17±0.12**

6.2 Software-generated composite sketches

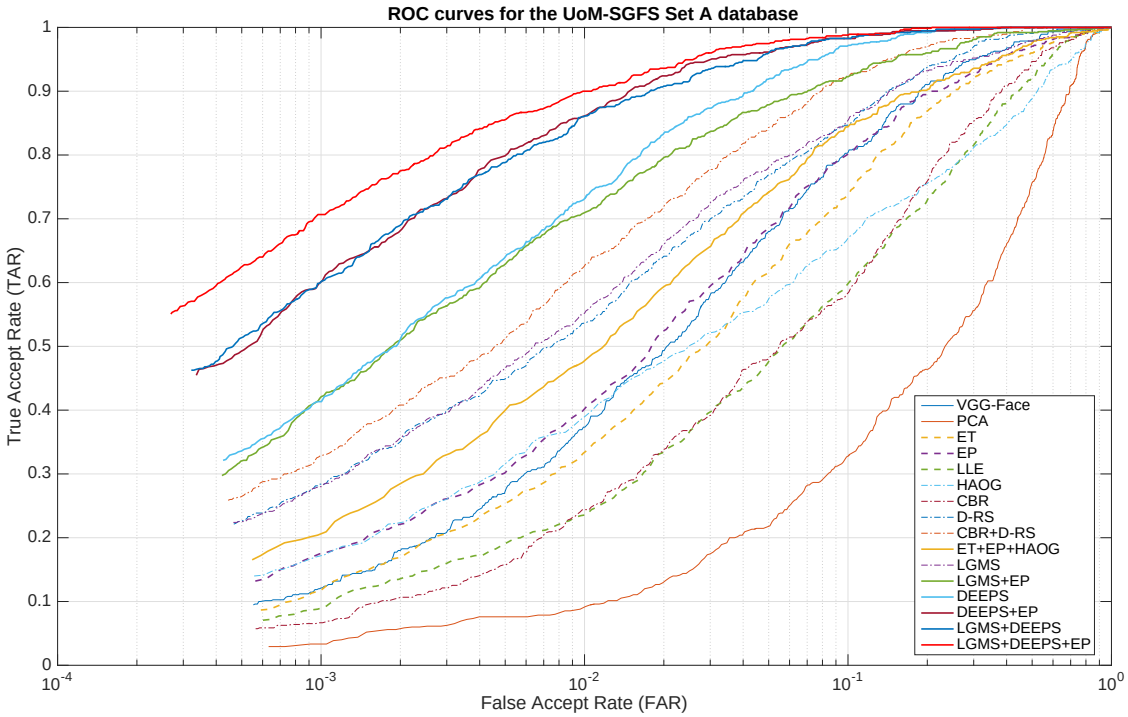
The results of the algorithms considered are shown in Table 6.2 and Figure 6.2 when evaluated on the UoM-SGFS Set A database and in Table 6.3 and Figure 6.3 when evaluated on the UoM-SGFS Set B database.

Although algorithms achieve higher performance when operating on the Set B sketches than the harder Set A sketches, trends in performance for both sets are similar to those observed in the case of the hand-drawn sketches in Section 6.1, namely the intra-modality methods typically achieve superior performance to the FRSs but are inferior to the inter-modality methods. The proposed EP method also improves upon the performance of ET, while fusion of intra- and inter-modality methods is also beneficial. However, LGMS is now inferior to D-RS+CBR, the latter likely benefiting from the complementary information provided by the CBR algorithm that was designed to operate on software-generated sketches. Nevertheless, LGMS requires a similar amount of time to match its features, which also require significantly less storage space than D-RS+CBR as elaborated in Section 6.6.

DEEPS comprehensively outperforms all methods considered on both sets of the UoM-SGFS database, and its fusion with LGMS and EP further improves performance to once again yield the best performing system. In fact, the performance on the Set A sketches is superior to the state-of-the-art D-RS+CBR system when operating on the easier Set B sketches. DEEPS and its fused variants are also the only systems that exceed rank retrieval rates of 95% by Rank-150 on the Set A sketches, where 96-99% of subjects are correctly identified.

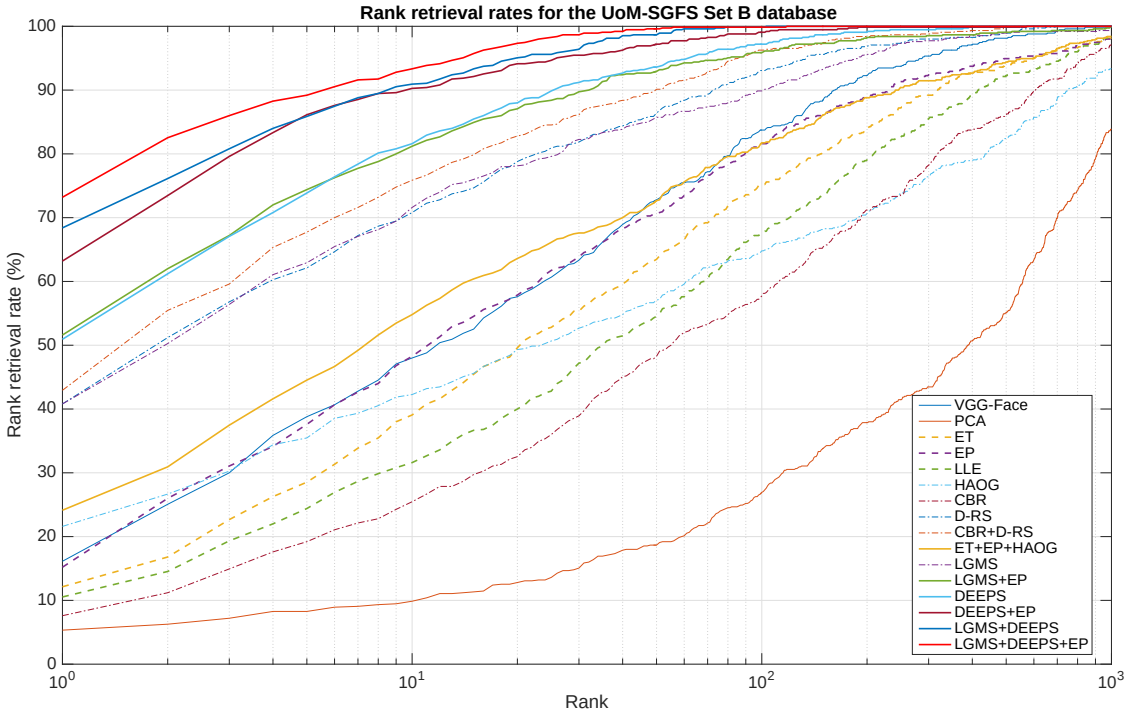


(a)

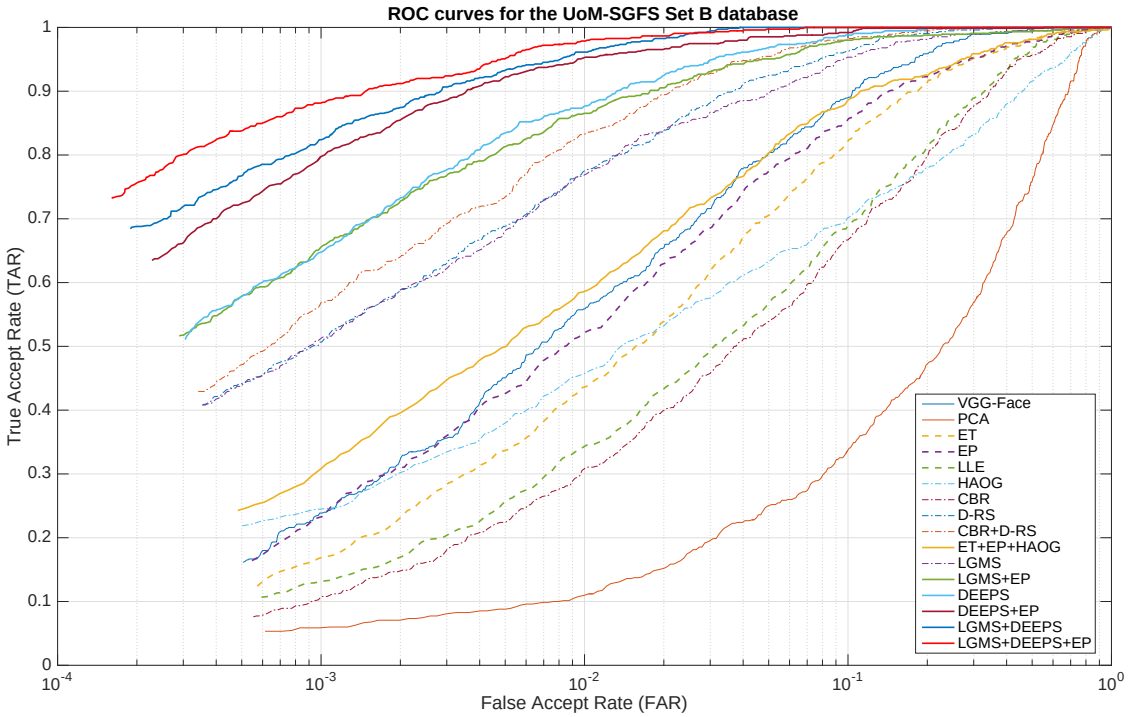


(b)

Figure 6.2: Results for algorithms considered when evaluated using the UoM-SGFS Set A software-generated sketches, averaged over 5 set splits: (a) CMC curves, (b) ROC curves



(a)



(b)

Figure 6.3: Results for algorithms considered when evaluated using the UoM-SGFS Set B software-generated sketches, averaged over 5 set splits: (a) CMC curves, (b) ROC curves

Table 6.2: Means and standard deviations over 5 train/test-set splits for algorithms evaluated on UoM-SGFS Set A sketches.

#	Method	Matching Rate (%) at Rank-X					TAR@FAR=0.1%	TAR@FAR=1.0%	EER (%)
		X=1	X=10	X=50	X=100	X=150			
1	VGG-Face [93]	9.33±2.45	31.07±3.73	59.73±2.52	73.60±3.58	80.80±2.72	11.87±2.47	37.33±4.40	14.18±1.05
2	PCA [37]	2.80±1.19	8.40±2.03	17.73±3.22	25.20±3.35	32.40±3.11	3.33±1.56	9.33±1.94	37.13±0.64
3	ET (+PCA) [17]	8.40±2.14	30.00±3.62	54.53±5.82	67.47±2.28	74.27±3.29	11.73±2.81	34.13±4.33	16.13±2.06
4	LLE (+PCA) [159]	6.93±1.92	24.67±2.98	43.60±2.34	57.60±3.79	64.13±3.57	8.93±1.80	24.13±2.64	23.68±1.00
5	HAOG [10]	13.60±2.29	37.33±1.94	52.67±2.62	60.27±1.67	65.47±1.85	16.93±2.29	39.07±3.73	23.12±1.78
6	CBR [6]	5.73±2.09	18.80±1.28	43.33±1.94	52.40±2.14	59.33±2.49	6.67±2.45	24.27±2.65	21.62±1.29
7	D-RS [12,69]	22.13±1.45	49.33±4.24	69.87±2.18	78.67±2.26	83.33±2.54	28.53±2.38	53.60±3.35	12.30±0.93
8	D-RS + CBR [19]	25.87±4.43	56.00±3.80	76.27±3.90	84.93±1.92	89.87±1.28	32.27±3.25	62.67±3.89	8.84±0.82
9	*EP* (+PCA)	12.53±2.08	35.60±2.19	62.80±2.88	74.40±3.61	79.07±4.13	17.20±1.66	40.00±1.70	14.49±1.29
10	*ET + EP + HAOG*	15.73±1.53	44.00±2.05	67.73±3.52	76.67±2.54	80.93±2.14	20.93±3.15	47.87±3.44	12.73±1.84
11	*LGMS*	22.40±5.05	51.20±3.63	72.40±3.48	80.80±3.18	84.27±2.03	27.73±6.01	55.33±4.59	12.32±1.20
12	*LGMS + EP*	29.73±2.65	64.80±5.61	84.00±2.79	89.20±2.18	92.13±1.73	42.13±3.93	70.80±2.18	8.52±0.58
13	*DEEPS*	31.60±1.12	66.13±2.47	86.00±1.25	93.47±1.85	96.40±1.21	41.87±3.11	73.47±2.08	6.26±0.60
14	*DEEPS + EP*	44.80±1.73	78.53±2.23	93.60±2.14	**97.47±1.37**	**99.07±0.76**	59.87±2.47	86.40±1.67	4.16±0.95
15	*LGMS + DEEPS*	46.27±3.22	80.13±2.42	92.93±1.38	96.00±0.94	97.47±1.10	60.27±4.10	86.13±3.44	4.60±0.68
16	*LGMS + DEEPS + EP*	**55.07±3.67**	**84.80±2.23**	**95.07±1.80**	**98.40±1.53**	**98.93±1.21**	**70.53±2.28**	**89.73±1.92**	**3.42±0.54**

Table 6.3: Means and standard deviations over 5 train/test-set splits for algorithms evaluated on UoM-SGFS Set B sketches.

#	Method	Matching Rate (%) at Rank-X					TAR@FAR=0.1%	TAR@FAR=1.0%	EER (%)
		X=1	X=10	X=50	X=100	X=150			
1	VGG-Face [93]	16.13±2.72	48.00±3.30	72.80±2.84	83.73±2.24	88.53±3.14	23.60±2.73	56.00±4.83	10.24±1.13
2	PCA [37]	5.33±2.11	9.87±0.99	18.67±3.43	26.93±3.08	33.87±1.85	5.87±1.85	11.07±1.98	36.05±1.08
3	ET (+PCA) [17]	12.13±1.10	39.07±4.73	63.47±3.18	75.07±3.55	80.27±2.39	16.80±1.37	43.20±6.04	13.04±1.14
4	LLE (+PCA) [159]	10.53±1.59	31.60±1.01	54.53±2.02	67.60±3.42	73.47±3.87	13.07±0.76	34.40±2.34	19.13±1.42
5	HAOG [10]	21.60±3.67	42.27±2.89	57.07±3.39	64.80±2.23	68.27±2.52	24.40±3.90	45.47±1.37	21.49±1.35
6	CBR [6]	7.60±1.98	25.47±2.38	48.27±1.30	57.73±2.03	65.47±2.80	10.67±3.13	30.53±0.99	20.11±1.21
7	D-RS [12,69]	40.80±1.45	70.80±1.85	86.40±0.89	93.07±1.12	95.60±0.76	50.93±1.38	77.20±1.97	6.42±0.51
8	D-RS + CBR [19]	42.93±1.38	75.87±1.59	90.13±1.45	96.27±1.21	97.47±1.73	56.80±0.87	83.07±1.80	4.35±1.11
9	*EP* (+PCA)	15.20±1.28	48.27±3.04	70.67±2.49	81.60±2.97	86.00±2.54	23.33±1.33	52.27±2.14	12.04±1.11
10	*ET + EP + HAOG*	24.13±3.25	54.80±3.84	72.67±5.14	81.73±2.97	86.13±1.73	30.67±3.68	58.93±4.28	10.29±1.32
11	*LGMS*	40.80±4.79	71.60±3.35	85.73±1.53	89.87±2.08	93.33±1.49	50.80±5.24	76.93±2.24	7.32±1.03
12	*LGMS + EP*	51.60±3.22	81.20±1.73	92.93±1.61	95.87±0.56	97.33±0.47	65.60±1.74	86.40±1.98	4.96±0.52
13	*DEEPS*	50.93±3.61	81.60±2.52	93.60±1.21	97.20±1.45	98.67±0.47	65.07±4.54	87.73±1.67	3.99±0.74
14	*DEEPS + EP*	63.20±2.56	90.27±2.24	97.20±0.99	99.07±0.37	**99.47±0.56**	79.47±2.38	95.33±1.05	2.65±0.47
15	*LGMS + DEEPS*	68.40±2.43	90.93±1.53	98.67±1.05	**99.87±0.30**	**100.00±0.00**	82.27±2.39	96.40±1.67	1.86±0.46
16	*LGMS + DEEPS + EP*	**73.20±2.96**	**93.33±1.41**	**99.60±0.37**	**99.87±0.30**	**100.00±0.00**	**88.27±0.76**	**98.00±0.67**	**1.49±0.29**

6.3 Evaluation on PRIP-VSGC and EPRIP datasets

Since implementations for the recent deep learning-based methods in [121,123] are unavailable, the proposed DEEPS+LGMS methods are evaluated with the same protocol employed to obtain the results shown in these papers. Specifically, two databases that each contain viewed software-generated composite sketches of the 123 subjects in the AR database [16] are used: (i) the PRIP Viewed Software-Generated Composite (PRIP-VSGC) dataset [6,127] having sketches created by an Asian operator using Identi-Kit, and (ii) the Extended PRIP (EPRIP) database [121,124] containing the sketches of the same subjects, created by an Indian software operator using the FACES software. The same terminology as used in literature will be employed to refer to these two databases, namely *IdentiKit (As)* and *FACES (In)* for the PRIP-VSGC and EPRIP databases, respectively. Only the Rank-10 retrieval rates will be reported since they are the only results reported numerically (shown using exact values) in the papers. In addition, since the databases contain viewed sketches only, DEEPS-M is not evaluated for the same reasons stated in Section 3.3.4 (i.e. viewed sketches tend to bear a generally good resemblance to the original photos and any amendments to the facial features as done in DEEPS-M will likely reduce similarity). 49 subjects are reserved for training while the remaining 74 subjects are used for testing, with no extended gallery. Five cross-validation folds are performed. However, the training subjects are not used to re-train DEEPS due to the limited quantity available, and to determine the robustness of the proposed approach on images that are different to those used during training. The results are provided in Table 6.4.

As shown, the proposed DEEPS framework outperforms the methods in [121,123] on both datasets, particularly in the case of the EPRIP databases containing sketches created using the FACES software by an Indian user. This is despite DEEPS not being re-trained on the new databases. This indicates that the proposed approach yielded a network that did not suffer from over-fitting and that generalises quite well to different types of sketches which were unseen by the system during training. In the case of the LGMS method, its performance is relatively poor on the PRIP-VSGC database and its fusion with DEEPS yields improved performance compared to itself but inferior performance compared to DEEPS. LGMS also performs more poorly than DEEPS on the EPRIP database, but is notice-

Table 6.4: Rank-10 retrieval rates (%) computed on the PRIP-VSGC and EPRIP databases, corresponding to IdentiKit (As) and FACES (In) results, respectively, in [121] and [123].

Method	Matching Rate at Rank-10 (%)	
	IdentiKit (As)	FACES (In)
Mittal *et al.* [121]	52.0 ± 2.4	60.2 ± 2.9
Saxena and Verbeek [123]	51.5 ± 4.0	65.6 ± 3.7
LGMS	26.1 ± 1.5	67.2 ± 2.2
DEEPS	$\mathbf{55.2 \pm 1.2}$	83.7 ± 2.4
DEEPS+LGMS	51.7 ± 2.2	$\mathbf{91.7 \pm 2.6}$

ably superior to [121] and marginally superior to [123]. The fusion of LGMS with DEEPS on this dataset yields substantially improved performance, correctly retrieving 91.7% of subjects at Rank-10 compared to 65.6% of subjects retrieved by the method in [123]. It should also be noted that the standard deviations of LGMS, DEEPS, and DEEPS+LGMS are smaller than those of the methods in [121,123], thereby indicating that their performance is more consistent. As can be observed, all methods appear to be challenged on the PRIP-VSGC database whose sketches were generated using IdentiKit. One of the main reasons may be that the sketches themselves generally bear a subjectively poor resemblance to the original photographs; indeed, even the creators of the database noticed significantly degraded performance when using these sketches and did not use them extensively in their experiments [6]:

> *"A comparison between the two component sketch kits FACES and IdentiKit shows that FACES provides a more accurate depiction for the different faces in the AR database. Thus, in the following experiments, we focus on the recognition of composite sketches generated using only FACES".*

6.4 Forensic sketches

All subjects were only used for testing by employing the same models that were trained on the viewed hand-drawn sketches, due to the forensic sketch dataset's small size. Also for this reason, traditional performance measures may yield inaccurate results, particularly for the computation of the True Accept Rates and False Accept Rates. Indeed, the results of some metrics are identical for multiple algorithms, leading to null standard deviation values. However, the relatively small number of sketches facilities the analysis of the ranks at which each subject in the dataset is retrieved by the algorithms considered. These are provided and discussed in Appendix B.1, where the query forensic sketch and corresponding gallery photo are shown to facilitate the determination of any challenges faced by an algorithm when comparing the two images. To summarise the performance of the methods considered, the mean values of the ranks at which the true identity of the 47 subjects is retrieved are computed instead of the rank retrieval rates and TARs at given FARs, which are shown in Table 6.5.

From the results given in Table 6.5, some differences with respect to the viewed hand-drawn sketches can be noticed. In particular, although EP also improves upon the performance obtained by ET, both intra-modality methods perform worse than the PCA method which is also used as the face recogniser for the two methods. In addition, LLE only improves performance marginally. This poor performance is likely due to the significant differences between forensic sketches and the corresponding photos, since these intra-modality methods primarily aim to reduce the modality gap only. Hence, the synthetic pseudo-sketches obtained are still rather dissimilar to the corresponding sketches, and the face recogniser is unable to cater for these differences.

The inter-modality algorithms, which are designed to use features that are robust for both the sketch and photo domains, perform noticeably better. However, it should also be noted that the VGG-Face algorithm also performs relatively well. Moreover, in contrast to the viewed sketches, CBR performs significantly better than D-RS such that their fusion yields better performance than that of CBR alone. Indeed, this algorithm yields the best performance of algorithms proposed in literature.

Comparing DEEPS to LGMS in terms of differences in the ranks at which the

identity of the subjects are retrieved as shown in Figure 6.4, it is evident that both methods perform relatively well on the forensic sketches and achieve mean rank retrieval values of 325.02 and 398.27, respectively. This demonstrates that, overall, DEEPS is able to successfully retrieve subjects at smaller ranks than LGMS. That said, both LGMS and DEEPS lag behind the performance of D-RS+CBR. However, as elaborated hereunder, fusion of LGMS and DEEPS yields a system that outperforms all methods considered, including D-RS+CBR. Furthermore, the performance of LGMS can be improved through its fusion with EP, despite the poor performance of the latter on forensic sketches. Fusion of EP with HAOG and ET also improves performance with respect to the performance of the individual methods, further demonstrating that fusion of intra- and inter-modality algorithms as investigated in this research can benefit from complementary information.

As shown in Figure 6.5, DEEPS-M is generally able to retrieve subjects at better ranks than DEEPS, lowering the mean rank retrieval values to 312.11. This indicates that the proposed use of multiple sketches can indeed be beneficial during deployment. It is likely that performance can be improved with the use of a more flexible morphable model that allows better variation of the facial features, which are also able to reflect more closely the distortions and exaggerations that are typically found in forensic sketches.

As observed in the case of viewed sketches, LGMS and DEEPS can provide complementary information to yield improved performance. This also holds for the forensic sketches, yielding a mean retrieval rank of 272.58 which is lower than either algorithm. Fusion of the superior DEEPS-M with LGMS also leads to smaller ranks than either approach as shown in Figure 6.7, indicating that the two methods also provide complementary information for forensic sketches. Indeed, the average rank value is reduced by 13.9% and 32.5% compared to DEEPS-M and LGMS, respectively, to 268.82. This demonstrates that the fusion is overall substantially beneficial. Moreover, as depicted in Figure 6.6, instances where subjects are retrieved at large ranks can be simply a consequence of the top matches being more similar to a probe sketch than the true corresponding photo, and not due to a failure of the algorithm. Finally, fusion of EP with DEEPS+LGMS and DEEPS-M+LGMS fails to improve performance on the forensic sketches, due to the poor performance of EP as discussed above.

Table 6.5: Average values over all ranks of the 47 subjects in the PRIP-HDC [19] forensic sketch database, for the algorithms considered after averaging over 3 set splits.

#	Algorithm	Average rank
1	VGG-Face [93]	430.16
2	PCA [37]	485.24
3	ET (+PCA) [17]	660.99
4	LLE (+PCA) [34]	481.84
5	HAOG [10]	576.21
6	CBR [6]	351.28
7	D-RS [12,69]	482.11
8	CBR + D-RS [12]	312.00
9	*EP* (+PCA)	611.83
10	*ET + EP + HAOG*	570.43
11	*LGMS*	398.27
12	*LGMS + EP*	383.70
13	*DEEPS*	325.02
14	*DEEPS + EP*	376.27
15	*DEEPS + LGMS*	**272.58**
16	*DEEPS-M*	312.11
17	*DEEPS-M + LGMS*	**268.82**
18	*DEEPS + LGMS + EP*	289.46
19	*DEEPS-M + LGMS + EP*	286.11

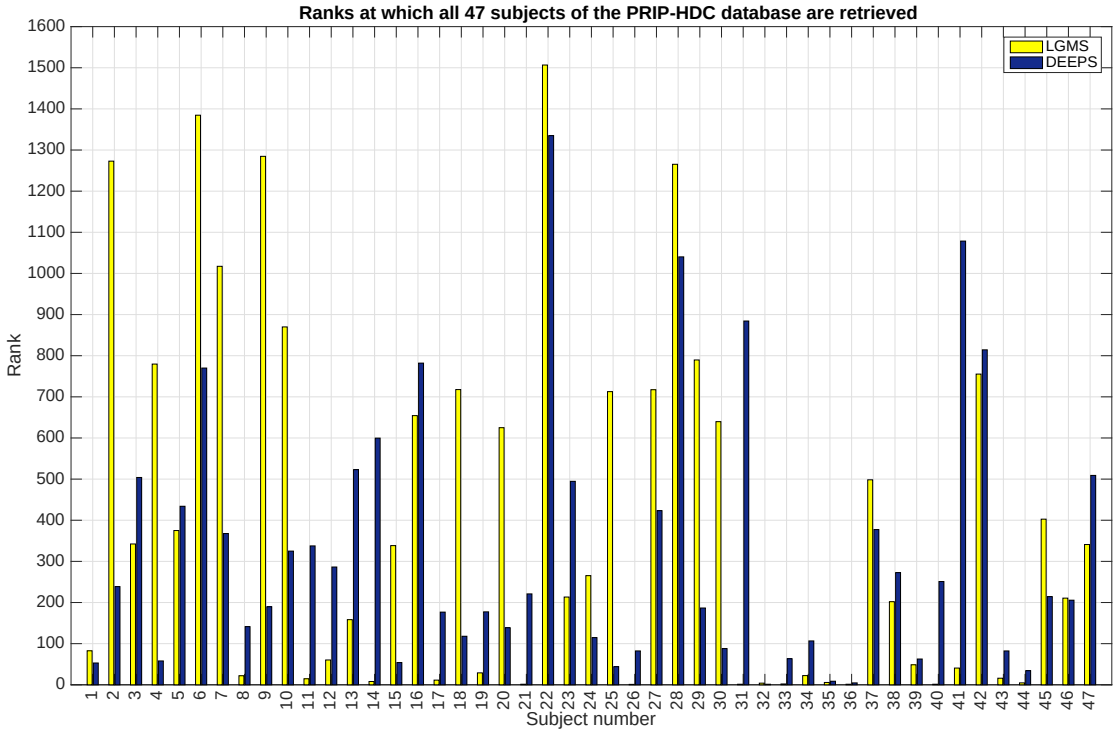


Figure 6.4: Ranks of all 47 subjects in the PRIP-HDC database [19] for LGMS and DEEPS. Smaller values are desired.

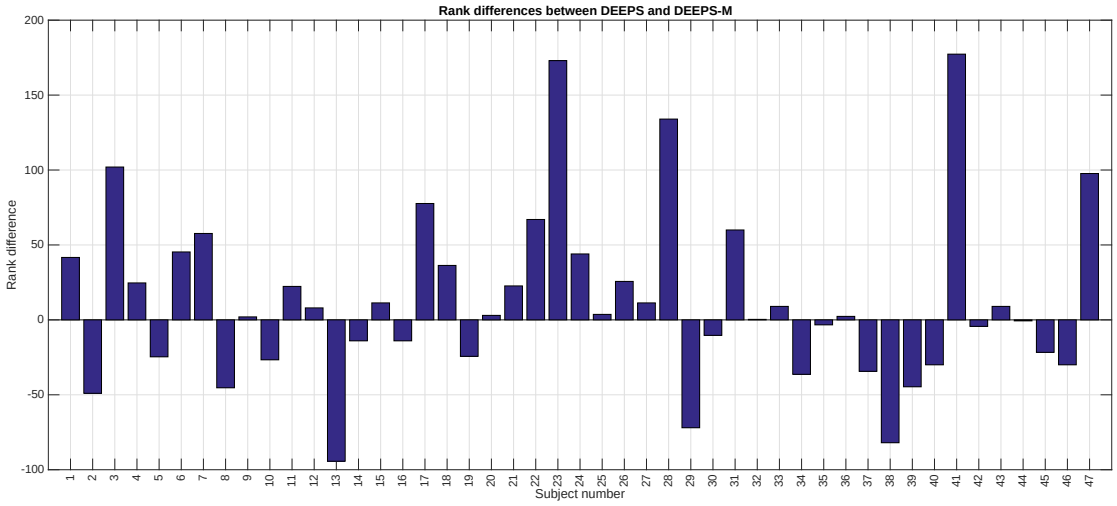


Figure 6.5: Rank differences for all 47 subjects in the PRIP-HDC database [19] when comparing DEEPS with DEEPS-M (positive values indicate DEEPS-M rank improvements).

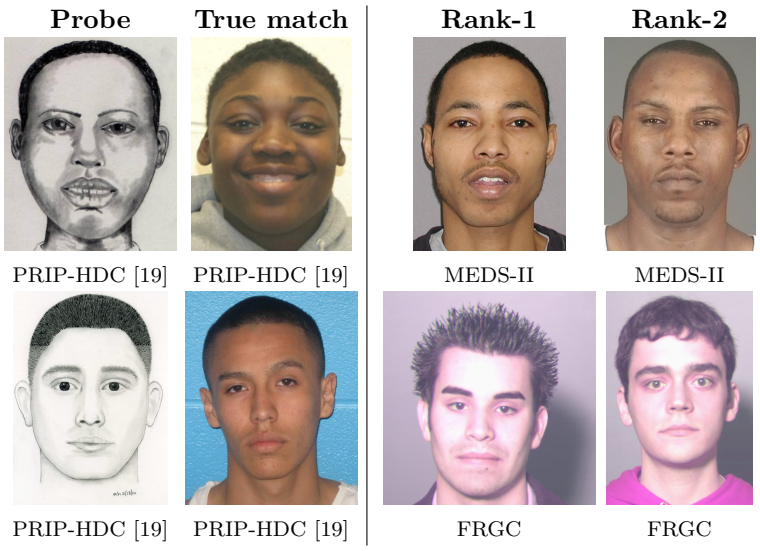


Figure 6.6: Examples where best matches retrieved by LGMS + DEEPS-M bear a better liking to probe than the true match. Subject in first row is retrieved at rank 213, while subject in second row is retrieved at Rank 293.

Method	Rank
D-RS+CBR [19]	499.00
LGMS	82.67
DEEPS	53.00
DEEPS-M	11.33
LGMS+DEEPS	38.33
LGMS+DEEPS-M	22.33

Method	Rank
D-RS+CBR [19]	880.00
LGMS	342.33
DEEPS	504.00
DEEPS-M	402.00
LGMS+DEEPS	329.33
LGMS+DEEPS-M	275.33

Method	Rank
D-RS+CBR [19]	321.00
LGMS	14.67
DEEPS	337.67
DEEPS-M	315.33
LGMS+DEEPS	33.00
LGMS+DEEPS-M	27.67

Method	Rank
D-RS+CBR [19]	1037.00
LGMS	717.67
DEEPS	118.00
DEEPS-M	81.67
LGMS+DEEPS	207.00
LGMS+DEEPS-M	154.00

Method	Rank
D-RS+CBR [19]	112.00
LGMS	712.67
DEEPS	44.00
DEEPS-M	40.33
LGMS+DEEPS	171.00
LGMS+DEEPS-M	160.00

Figure 6.7: Examples of ranks (averaged over three set splits) at which the correct photo is retrieved given a query forensic sketch. Images available in the PRIP-HDC database [19]. Ranks for all subjects in the database are given in Appendix B.1.

6.5 DEEPS/DEEPS-M network visualisation

To understand what contributes to the performance of DEEPS and the related DEEPS-M method, two approaches are considered: (i) the approach in [161], which approximates the inverse of the output representation for each layer (i.e. inferring the input from the output representation), and (ii) the t-Distributed Stochastic Neighbour Embedding (t-SNE) approach [162] which performs non-linear dimensionality reduction of the 1024-D features derived from photos and sketches to 2-D, an approach that is well-suited to visualise high-dimensional datasets. More information about these methods and additional results are given in Appendix A.5.

In terms of the output at each layer by the method in [161] as shown in Figure 6.8, it can be observed that most layers retain the structural information of face images and is invariant not only to limited variations in the facial components, but also to translation and larger differences in the facial components. Moreover, the network appears capable of inferring color information given a grayscale sketch image, for both light-skinned and dark-skinned subjects.

The visualisations produced by the t-SNE approach, as shown in Figure 6.9, indicate that the distribution of face photos and the corresponding face sketches are quite similar, showing that the network handles photos and sketches in a similar manner. In other words, the network appears to have successfully learned modality-invariant features, such that a single network is capable of handling two modalities. Moreover, it can also be observed that certain classes of images are clustered together despite no supervision being provided to the network to enable this behaviour, e.g. by race, gender, or finer characteristics such as facial hair. In this manner, the network is implicitly performing demographic filtering.

The above observations indicate that the learned network is robust to the significant cross-domain gap between photos and sketches, despite using a single network to handle two modalities.

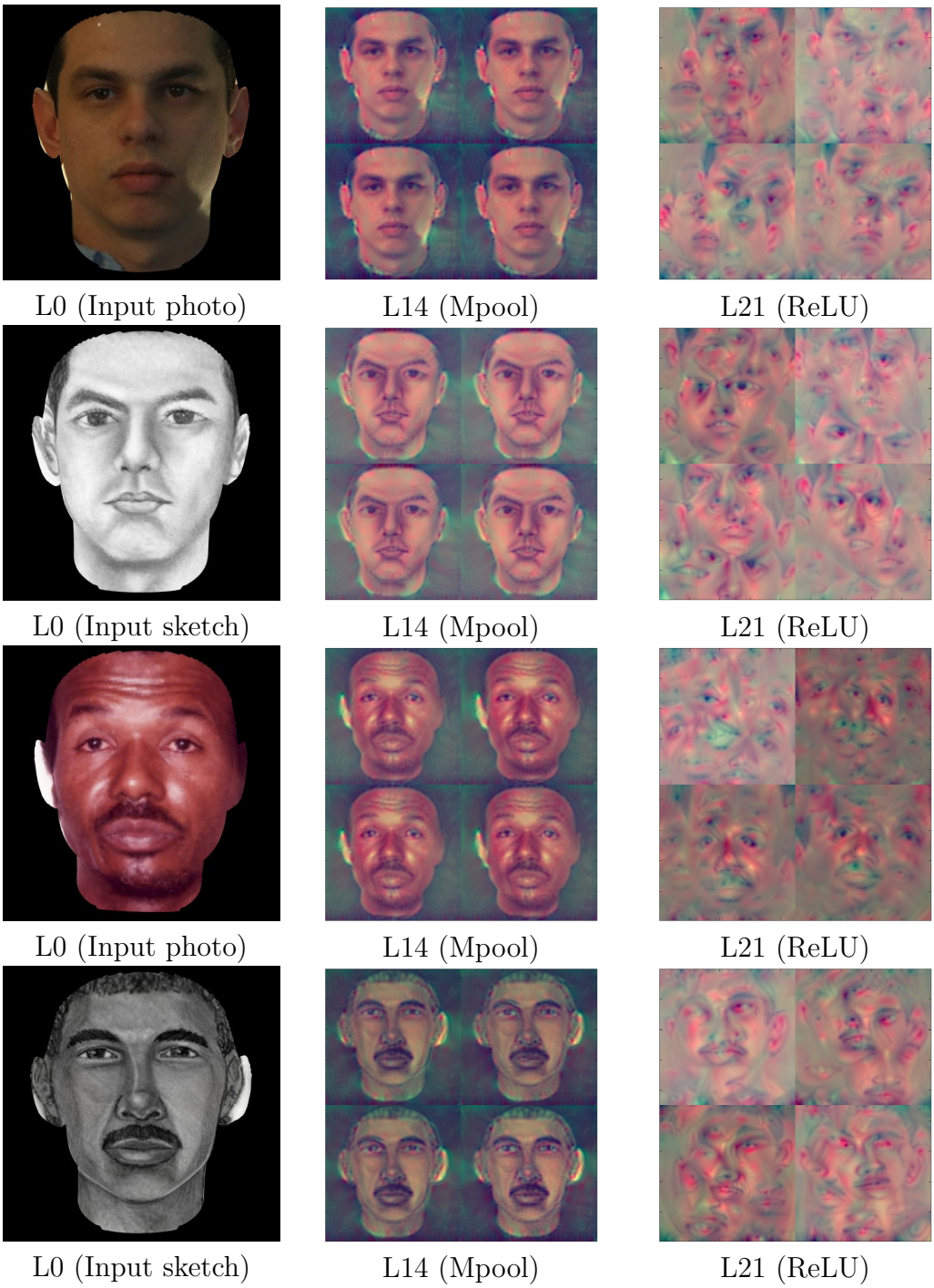


Figure 6.8: Examples of network visualisation, using the method in [161]: first row: photo of a subject in the Color FERET database [13,152], second row: corresponding hand-drawn sketch in the CUFSF database [15,129], third row: photo of a subject in the PRIP-HDC database [19], fourth row: corresponding forensic hand-drawn sketch in the PRIP-HDC database [19]

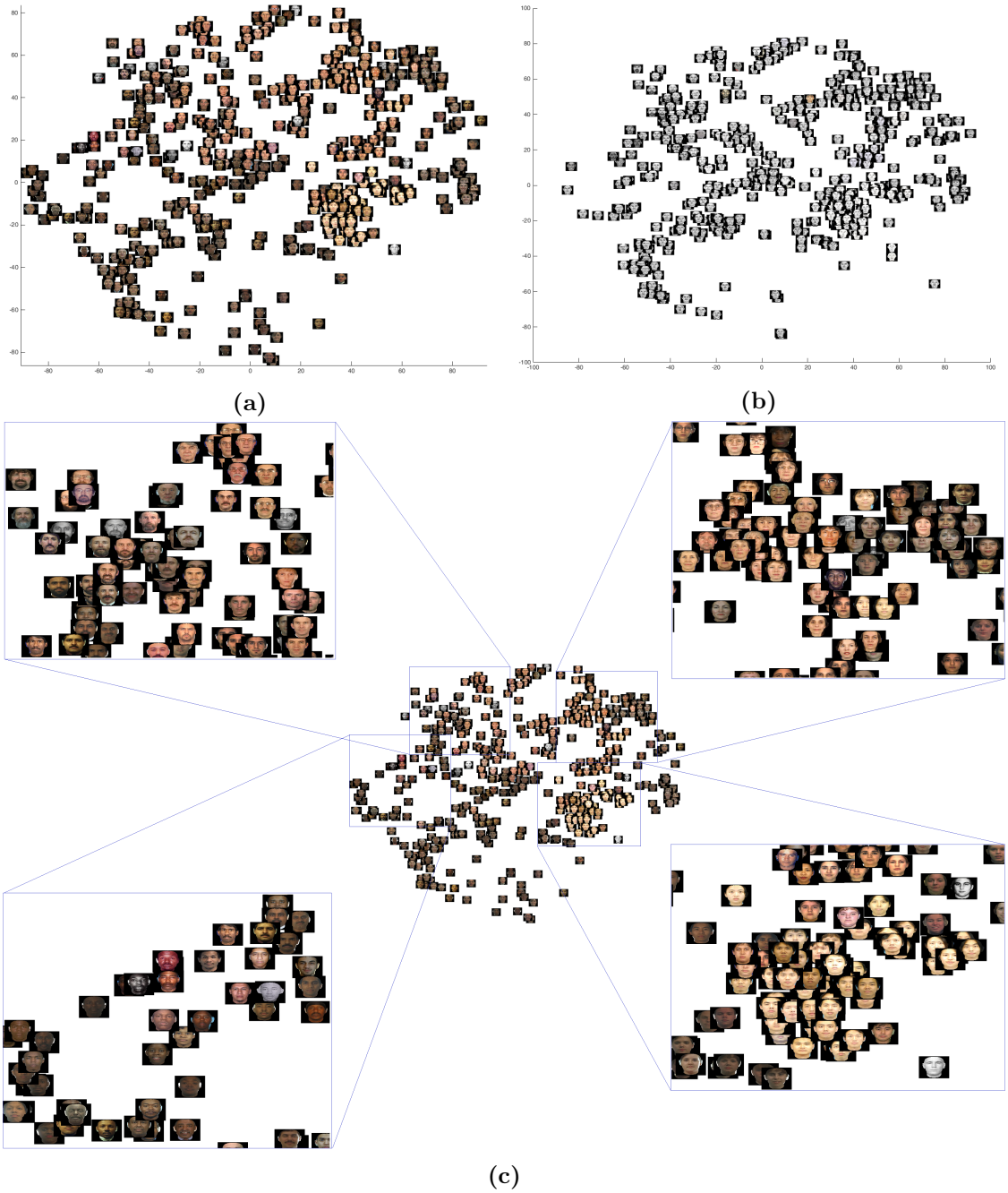


Figure 6.9: Visualisation of network using the t-SNE method [162]: (a) Photos and (b) corresponding hand-drawn sketches, (c) zoomed areas of photos visualisation. Categories appear to have been automatically arranged by semantic similarity despite not being enforced during training, such as men with beards (top left), females (top right), Asian subjects (lower right), and dark-skinned subjects (lower left). Similar observations can also be made for the sketches visualisation. Images best viewed on a screen.

6.6 Computation time and feature sizes

The time required to compute and compare the features of the algorithms considered will now be discussed. However, practical implementations of face photo-sketch recognition systems would likely pre-compute the features of all subjects in the law enforcement agencies' database and store them as *biometric templates*, similar to the process done for other biometric traits [4,163]. Then, the image of a new subject is simply enrolled by computing its features and storing them along with the other templates. Hence, the storage space requirements are arguably more important than computation time.

All algorithms were implemented in MATLAB R2015b on a computer running an Intel Core i7-4770k processor, 32GB RAM, and Windows 10. In the case of the methods based on DCNNs (i.e. VGG-Face [93], DEEPS and DEEPS-M), the MatConvNet toolbox [104] was used in conjunction with an NVIDIA Titan X Pascal GPU provided by NVIDIA Corporation to enable faster feature computation times. Feature matching is still performed by the Intel Core i7-4770k CPU.

It should be noted that the values for methods that are fused are simply the addition of the feature dimensionality and computation times for each individual method. Of course, computation time can be reduced by performing the computations of each method in parallel where possible. Furthermore, the individual algorithms are not optimised for speed and are also serially implemented; although several algorithms can be made faster by performing computations in parallel provided that they can be handled by the computation system employed (in terms of both processing power and memory required), the worst case scenario where methods are computed serially is assumed for the rest of this discussion.

The computation times and feature dimensionalities are given in Table 6.6. It is evident that the proposed LGMS and DEEPS methods are quite efficient in terms of the template size, which is one of the most important parameters for systems implemented in the real-world. Another important aspect is the time to match templates, where DEEPS is the fastest algorithm considered while LGMS requires the most amount of time. However, the fused variants of these methods yield overall significantly smaller templates and a similar overall matching time compared to the state-of-the-art D-RS+CBR system. Although the intra-modality

methods are also quite efficient[1], their performance is quite poor when compared to the inter-modality methods as discussed previously. The matching times for the proposed DEEPS-M method are comparable to other methods, but utilises more storage space than DEEPS since a subject is represented using 200 sketches. Hence, a system implemented in the real-world could utilise the DEEPS system instead of DEEPS-M if there exist tight storage limitations, given that DEEPS and its fusion with LGMS still perform well.

[1]The intra-modality methods use PCA as the face recogniser, which retains 99% of the eigenvalue variance; thus, the feature sizes are variable and the results shown in Table 6.6 are average values

Table 6.6: Mean computation times and feature dimensionality of all algorithms considered, using images of size 200×250. The model size, corresponding to the number of parameters that would need to be used during testing, is also shown when applicable. Feature dimensionality for methods using PCA is an average figure. 'Feature computation' time encompasses time to produce an image for intra-modality methods and features extracted (e.g. MLBP) for the inter-modality methods; time to filter images is also included (where applicable). It is also assumed that computations are done serially. N represents the number of training samples used, which is set to 575 for both intra- and inter-modality algorithms. $T = 200$ is the number of sketches used during system deployment for the DEEPS-M method, and $B = 30$ is the number of 'bags' used for RS-LDA in D-RS and D-RS+CBR.

#	Algorithm	No. of parameters	Feature dimensionality	Execution time (seconds)	
				Feature computation	Matching
1	VGG-Face [93]	134.3M	4096	0.3657	0.0799
2	PCA [37]	14.3M	$\simeq 285$	0.0094	0.0048
3	ET (+PCA) [17]	43.7M + 29.8M[a] = 73.5M	$\simeq 595$	0.0579 + 0.0094 = 0.0673	0.0048
4	LLE (+PCA) [34]	211.0M + 42.8M[a] = 253.8M	$\simeq 850$	3.9587 + 0.0094 = 3.9681	0.0048
5	HAOG [10]	18	32,879	0.4840	0.4611
6	D-RS [12,69]	270.0M	$(N-1) \times B \times 6 = 103,320$	2.8688	0.0023
7	CBR [6]	–	48,144	1.2075	3.7963
8	EP (+PCA) [62]	153.1M + 31.1M[a] = 184.2M	$\simeq 620$	0.9895 + 0.0094 = 0.9989	0.0048
9	LGMS	670.3M	$(N-1) \times 32 = 18,368$	24.5525	5.3243
10	DEEPS	138.5M	1024	0.3461	0.4667×10^{-4}
11	DEEPS-M	138.5M	$T \times 1024 = 204,800$	$T \times 0.3461 = 69.22$	$T \times (0.4667 \times 10^{-4}) = 0.0093$

[a]Figure refers to number of parameters associated with PCA

6.7 Evaluation of method re-implementation integrity

All methods proposed in literature and used for evaluation in this chapter, except for PCA and the VGG-Face descriptor, have been implemented from scratch in this research since their code is unavailable. To determine whether the re-implementations are a faithful representation to the original methods, the same evaluation protocols were implemented and the results obtained were compared to those published in literature. However, as shown in Table 6.7, most methods use one or more databases which are either unavailable or whose protocol (i.e. the list of image names used) is not provided. Consequently, an exact comparison cannot be performed. Hence, it was ensured that the re-implementations were done by closely following the published method descriptions and contacting paper authors when any missing information was encountered.

In the case of the ET, LLE and HAOG algorithms, the images used for their evaluation in literature are publicly available. Hence, the performance obtained in this research could be verified to be almost identical to that reported in literature[2].

For the other algorithms, an analysis into observations which could indicate whether the re-implementations are faithful representations of the original were performed in lieu of exact results. For example, in the case of D-RS, the implementation in [12] used a publicly available photo-sketch dataset but also used a private extended gallery. Since the use of an extended gallery typically makes the identification task more difficult (thus reducing performance[3].), then it should at least be expected that the use of the photo-sketch dataset by itself would enable better performance than that reported in [12]. This was indeed the case, indicating that the algorithm was implemented successfully.

For the deep learning-based methods in [123] and [124], re-implementations were not possible since the algorithms use a vast amount of images to train deep networks; hence, there are numerous random factors which cannot be easily replicated (e.g. initialisation of network (which is done randomly and can affect convergence),

[2]Small deviations in results are to be expected, due to minor variations in image alignment (e.g. due to different faducial coordinates), and details such as the programming language used and variable types (which are often unspecified).

[3]More details may be found in [6,9], and Appendix B.3

any data augmentation techniques involving random decisions such as image flips, etc.). However, the EPRIP database and one set in the PRIP-VSGC database that were used in both methods are publicly available. Hence, the proposed methods were evaluated using the same protocol to enable a direct comparison with the results reported in [123] and [124].

Table 6.7: Summary of results reported in literature, based on Table 2.1. DB = database, Ext. = Extended, subjs. = subjects, X2Y = transformation of X to Y, where X,Y∈ {Photo (P), Sketch (S)}, SR = Sparse Representation, DL = Deep-Learning, 'v' = viewed sketches, 'u' = un-viewed sketches, 'f' = forensic sketches, 'hdc' = hand-drawn composite sketches, 'sgc' = software-generated composite sketches. Approximate values are provided when results are only shown graphically.

Type	Method	Photo/Sketch DB(s)		Ext. Gallery DB(s)		Synth. Mode	Performance (%)	
		DB Name(s)	# Train/Test subjs.	DB Name(s)	# subjs.		Rank-1	Rank-10
FRS	PCA [17,33,37]	CUHK [17] (v-hdc)	88/100	–	0	–	31.0	67.0
	VGG-Face [93]	VGG Face DB [93]; LFW [116]	982,803/13,233	–	0	–	99.0	N/A
		VGG Face DB [93]; YTF [164]	982,803/3,425[a]	–	0	–	97.3	N/A
Intra (SL)	ET [17,33]	CUHK [17] (v-hdc)	88/100	–	0	P2S	71.0	96.0
						S2P	57.0	88.0
Intra (SL)	LLE [34]	CUFS [18,39] (v-hdc)	306/300	–	0	P2S	87.7	99.0
Inter	HAOG [10]	CUFS [18,39] (v-hdc)	0/606	–	0	–	100.0	100.0
	D-RS [12,69]	CUFS [18,39] (v-hdc)	404/202	PCSO (Private)	10k	–	96.4	≃99.0
		PRIP-HDC[b] [19] (f-hdc)	106/53	PCSO (Private)	10k	–	≃4.0	≃13.0
	CBR [6]	PRIP-VSGC[c] [6,127] (v-sgc)	0/123	PCSO (Private)	10k	–	10.6	≃35.0
				MEDS-II [128]	1193	–	12.2	≃42.0
	FaceSketchID [127]	CUFS [18,39], CUFSF [15,129] (v-hdc); PRIP-HDC[b] [19] (f-hdc)	1800+212/53	PCSO (Private)	100k	–	≃5.0	≃12.0
		CUFS [18,39], CUFSF [15,129] (v-hdc); PRIP-SGC (Private) (u-sgc)	1800/75	PCSO (Private)	100k	–	≃4.0	≃10.0
		CUFS [18,39], CUFSF [15,129] (v-hdc); PRIP-VSGC[b] [6,127] (v-sgc)	1677/123	PCSO (Private)	10k	–	≃11.0–24.0	≃30.0–38.0
DL	Mittal *et al.* [121]	CMU-PIE [44]; PRIP-VSGC[b] [6,127] (v-sgc)	30k+48/75	–	0	–	≃8.0	52.0±2.4
		CMU-PIE [44]; EPRIP [121,124] (v-sgc)	30k+48/75	–	0	–	≃5.0	60.2±2.9
	Saxena & Verbeek [123]	CASIA Webface [130]; PRIP-VSGC [6,127] (v-sgc)	500k+48/75	–	0	–	N/A	51.5±4.0
		CASIA Webface [130]; EPRIP [121,124] (v-sgc)	500k+48/75	–	0	–	≃12.0	65.6±3.7

[a]The YTF database evaluation involves the comparison of subjects appearing in videos; thus, the quoted figure refers to the number of videos

[b]Only a subset of sketch-photo pairs are publicly available

[c]Sketches are created using the IdentiKit software program and FACES software program; authors of [6] focus on FACES sketches which are unavailable

6.8 Summary

The main methods proposed in literature were discussed in Chapter 2, and some of the most popular and best-performing methods were evaluated. The methods proposed in this research as outlined in Chapter 3 were also evaluated. It was ensured that the identification task was as realistic as possible by using challenging hand-drawn sketches in the CUFSF database and software-generated sketches in the UoM-SGFS database created in this research, along with an extended gallery to simulate the extensive mug-shot galleries maintained by law enforcement agencies. Real-world forensic hand-drawn sketches were also utilised.

It was demonstrated that the proposed EP intra-modality method can improve upon the performance of the related ET method, highlighting the importance of using local patch-based processing. However, while the performance of intra-modality methods is generally superior to traditional FRSs, they lag behind the performance of inter-modality methods. Moreover, the synthesised images often contain artefacts which affect a face recogniser's performance especially when photos are synthesised from sketches, since an image of higher complexity than the source is being constructed. Whilst good quality images have been synthesised by methods reported in literature, these were typically evaluated on databases containing sketches that are very similar to the corresponding photos and typically fail when more challenging images as found in real-life are used instead.

LGMS was shown to outperform leading algorithms on viewed hand-drawn sketches and is only slightly inferior on software-generated sketches and forensic sketches. The proposed transfer learning of a DCNN using a training set enlarged with the aid of a 3D morphable model, which is the first work to use such an approach, was shown to be greatly beneficial. In fact, the resultant DEEPS method generally outperformed LGMS and indeed all algorithms proposed in literature. The performances of both LGMS and DEEPS are remarkable considering the lower time required to generate and compute features, and the lower burden on storage requirements compared to the leading methods proposed in literature. Moreover, no forensic sketches were used for training due to their low quantity. Therefore, the use of more real-world forensic sketches could potentially yield improved performance (if and when they become available to researchers).

Fusion of intra- and inter-modality algorithms, which has been inadequately studied

in literature, was also shown to be beneficial in several instances. The use of multiple sketches during test-time, which has also been rarely explored in literature, was shown to be beneficial for real-world forensic sketches. The proposed approach is also the first to generate these sketches automatically, enabling a timely and cost effective real-world solution.

Ultimately, the best-performing approaches on the viewed sketches were the fusion of LGMS with DEEPS or DEEPS-M and of LGMS, DEEPS/DEEPS-M, and EP. However, the performance of EP and the other intra-modality methods was relatively poor on the real-world forensic sketches, and its fusion with inter-modality methods generally hindered performance. As a result, the final proposed methods are the fusion of LGMS with DEEPS, and of LGMS with DEEPS-M. While the latter approach yielded superior performance, it also requires more computation time and, more importantly, storage space. Hence, the choice to determine which system is to be used in a practical scenario depends primarily on the storage space constraints of the particular application. Even with the technically inferior DEEPS, performance is still sufficiently robust with considerably lower computation time and storage requirements than not only LGMS+DEEPS-M, but also current state-of-the-art methods proposed in literature.

Chapter 7

Conclusions

The problem of identifying subjects based on sketches obtained from eyewitness descriptions of criminals has been tackled in this dissertation. Algorithms have been studied and created for both of the two main approaches for face photo-sketch recognition, namely intra- and inter-modality methods.

The initial Eigenpatches (EP) approach described in [62] improved performance over Eigentransformation (ET) on which it was based, while the benefits of fusing intra- and inter-modality methods were also demonstrated. However, intra-modality methods tend to be complex and computationally intensive while attempting to solve a more complex task than the problem of recognition itself, so that focus was then shifted towards inter-modality approaches. To this end, the log-Gabor-MLBP-SROCC (LGMS) approach was created and was shown to outperform some of the most popular and state-of-the-art methods proposed in literature despite using a relatively simple training algorithm [131].

Although most law enforcement agencies now use software-generated sketches rather than hand-drawn sketches [6], publicly available software-generated sketch databases are limited. Hence, the University of Malta Software-Generated Face-Sketch (UoM-SGFS) database was created, initially consisting of 600 sketches of 300 subjects as described in [14] and later doubled in size to contain 1200 sketches of 600 subjects. This database is the largest software-generated face sketch database and indeed one of the largest face sketch databases available, the only one consisting of sketches which are all represented in colour, the only one containing sketches created with the popular EFIT-V software, and one of the few to contain multiple sketches per

subject. The database was used to enable evaluation of algorithms operating on software-generated sketches, along with hand-drawn sketches. In addition, even the viewed sketches used in this work contained several distortions and exaggerations similar to those found in real-life, while an extended gallery to simulate the extensive mug-shot galleries maintained by law enforcement agencies was also employed. Real-world forensic sketches were also used for evaluation, thus ensuring that the proposed algorithms and state-of-the-art algorithms proposed in literature were evaluated under realistic conditions.

In this research, a novel Deep Convolutional Neural Network (DCNN)-based framework called DEEP (face) Photo-Sketch System (DEEPS) was also proposed and was shown to outperform all methods considered on both hand-drawn and software-generated sketches, by applying transfer learning to a leading DCNN-based FRS using a large set of synthetic images that were created automatically with the aid of a 3D morphable model. DEEPS was also extended to use multiple synthetic sketches during system deployment to yield the DEEPS Multi-sketch (DEEPS-M) system, which was shown to further improve performance. The fusion of LGMS and DEEPS or DEEPS-M were shown to yield the best performance of all algorithms, even on real-world forensic sketches. In addition, these methods are quite efficient in terms of both computation time and storage space requirements, particularly when compared to existing methods; hence, their implementation in the real-world by law enforcement agencies is largely feasible.

The proposed algorithms serve as a solid basis for any future work to be carried out, including the use of even more photos and sketches to aid training (particularly in the case of forensic sketches), and the application of the proposed methods to other heterogeneous face recognition tasks.

Chapter 8

Future Work

While the proposed framework has achieved good performance on viewed, forensic, hand-drawn and software-generated sketches, some aspects could be improved. Proposals for future improvements will thus be given hereunder, in Section 8.1. Suggestions will also be given in Section 8.2 for alternative approaches that can be undertaken to tackle the problem of face photo-sketch recognition.

8.1 Extensions of proposed work

Despite its great success, the current iteration of DEEPS has a few drawbacks that could be improved in the future. One of the most prominent is the susceptibility of the 3D morphable model to create synthetic images containing artefacts when the facial components are varied using large values. In addition, some off-pose faces are also handled incorrectly. Hence, the use of more advanced face fitting and morphable models such as the approach in [165] that is based on deep learning and which has attained promising performance, or commercial systems as employed in [166], could be used to render images of higher quality that can also aid both training and testing performance. Moreover, a more advanced morphable model would also allow greater flexibility in the adjustment of the facial components and facial expression to enable the network to be even more resilient to such variations that can occur in facial images. Multiple pose and lighting variations could also be considered to allow matching of sketches with photographs obtained in the wild from surveillance cameras, smartphone cameras, etc. Empirically, it was also

observed that DEEPS can be susceptible to colour intensity. While it is not entirely undesirable, since it enables implicit racial filtering, pre-processing of the photos and sketches could be performed to normalise image illumination, in addition to mean subtraction removal. A more detailed study into the use of multiple synthetic sketches during system deployment could also be carried out, primarily in terms of the fusion techniques to be used.

Different deep neural network architectures may also prove to be superior to the network used in this paper, along with new types of layers such as batch normalisation that has enabled better convergences [102]. The performance of intermediate features of the network employed could also be explored in the future, along with an analysis of the effect on performance when the output feature dimensionality is reduced.

With regards to the LGMS method, the log-Gabor filters are all given equal importance in the present work. However, it may be the case that certain filters are more important than others. Hence, an analysis into the usefulness of each of the log-Gabor filters could be performed to learn weights that can potentially improve performance. An alternative training method could also be explored, since the PCA+LDA approach employed in the current work is computationally inexpensive but is inferior to other approaches such as RS-LDA.

The combination of multiple datasets for algorithm training could also be explored, namely the use of both hand-drawn and software-generated sketches as done in [80] and the amalgamation of the two sets of the UoM-SGFS database. The use of the approach proposed in [24] could also be considered to synthesise sketches that contain less artefacts arising from the memory and communication gaps, which can then be used for training and testing face recognisers and inter-modality methods.

Although the UoM-SGFS database is the largest software-generated sketch database available, it could also be further enlarged to contain all subjects in the Color FERET database on which it is based. This would allow larger training and testing sets and thus potentially better performance and more robust results.

Finally, a larger extended gallery could also be employed to simulate the extensive mug-shot databases maintained by law enforcement agencies more accurately.

8.2 Alternative approaches

Different approaches to the implementation of deep learning could be considered. Specifically, instead of using a single network to handle both photos and sketches, two networks can be trained for each domain separately, similar to the approach in [167]. Hence, one network and the resultant features would be adapted for face recognition using photos, whereas the other network would be suited for face recognition using sketches. During system deployment, photos are then input to the network trained using face photos, whereas sketches are input to the network trained on face sketches, to yield two sets of features that need to be compared to determine if the photo and sketch belong to the same subject. There are numerous ways in which this can be done, including training another network that uses the two features to determine their similarity, or using subspace projection techniques such as the classical Canonical Correlation Analysis (CCA) method [168] where the two features are projected into a common subspace to facilitate comparison.

Algorithms can also be made more robust through the use of a larger number of sketches, particularly real-world forensic sketches. The limited number of such sketches (primarily due to privacy protection issues) at present are insufficient for robust training and thus impose limitations on the performance that can be attained. Unfortunately, this is also not directly within the control of researchers, who must depend on law enforcement agencies to provide and allow the use of such sketches. Particularly lacking at the time of writing this dissertation are software-generated forensic sketches. Their availability would aid researchers in developing algorithms capable of performing well when using software-generated sketches in real-life scenarios.

More use of ancillary information could also be made. The authors of [23,87,169, 170] extracted several attributes describing low-level characteristics such as the general shapes and sizes of facial components, which were used as a means of filtering the subjects together with the traditional demographic filtering, which were shown to improve performance. While the focus has been on attributes pertaining directly to the face, use of external information such as shoulder shape could also be considered as a means to identify the stature of the subject (e.g. weight, height, etc.).

Given the success of DCNNs, their use for intra-modality recognition could also be

investigated. Moreover, following the observations in this work that fusion of intra- and inter-modality methods can be beneficial despite the relatively low performance of intra-modality methods, the fusion of a good-performing intra-modality method with the proposed DCNN system may yield substantial gains in performance.

Lastly, the algorithms considered in this research could also be applied for other HFR tasks such as VIS-NIR matching, which also have important applications in surveillance and security.

8.3 Summary

While the proposed methods have been shown to exceed the performance of state-of-the-art methods, certain aspects could be improved to further increase performance. The novel DCNN-based method proposed in this work also serves as proof that such methods can not only be applied for the task of face photo-sketch recognition, but can also yield good performance; hence, it serves as a good starting point from where future research on the use of deep learning for face photo-sketch recognition can resume in further detail. Apart from improvements to the algorithms themselves, face photo-sketch databases could also be enlarged to enable better training and testing of such algorithms, particularly in the case of real-world forensic sketches. Investigations into other HFR domains could also be explored to determine if the proposed work also attains good performance on these important applications.

Appendix A

Dissection of Proposed Methods

An analysis of the various components of the proposed algorithms, namely EP, LGMS, and DEEPS and DEEPS-M, will now be given. The evaluation protocol is based on that outlined in Chapter 5 unless otherwise stated.

A.1 Eigenpatches parameter tuning

The rank retrieval rates using different patch sizes are given in Figure A.1. Patch sizes were set to be equal to $2^p \times 2^p$ for $p = 3, 4, \ldots, 7$. Evaluation was performed using the sketches of 842 subjects in the CUFSF database [15,129] and the corresponding face photos in the Color FERET database [13,152], where 211 subjects are used to train EP, 211 subjects are used to train the face recogniser, and the remaining 420 subjects are used for testing. Results were also reported in [62].

Firstly, it can be observed that photo-to-sketch (P2S) synthesis is superior to sketch-to-photo (S2P) synthesis for all patch sizes at all ranks. This is likely due to the fact that in transforming photos to sketches, a large amount of information is being compressed into a smaller representation since photos tend to be more complex and therefore contain more information. Transforming a photo to a sketch is a more stable operation than attempting to expand a small representation into a larger one, as in the case of S2P synthesis [17]. As can also be observed from Figure 3.1 on Page 38, it is evident that S2P synthesis is unable to recover a face photo adequately. From empirical observations, low patch sizes resulted in high

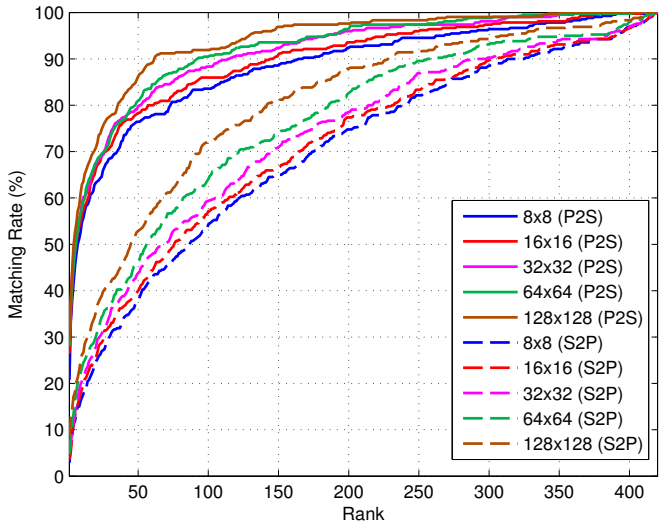


Figure A.1: Recognition rate for varying patch sizes of the EP algorithm. S2P = sketch-to-photo synthesis, P2S = photo-to-sketch synthesis.

amounts of noise which decreased with larger patch sizes, leading to superior performance with increasing patch size. In addition, patches of size 128×128 pixels roughly overlap with the individual facial components. Hence, synthesis is virtually being done for each component and is thus a contributor to increased quality of the synthesised images. Based on these observations, Eigenpatches with a patch size of 128×128 pixels is used for the experiments conducted in this research.

A.2 LGMS ablation study

From the results of method 2 in Table A.1, the effectiveness of log-Gabor filtering is clear given that the intensity features alone are able to provide good performance, as demonstrated by the results of method 1. Nonetheless, the use of MLBP yields noticeable improvements. However, the poor results when using MLBP on the unfiltered images show the benefit in using both the local and global texture feature descriptors. Also, the importance in using adequate distance or similarity measures is highlighted by the significantly higher performance obtained using SROCC compared to the Euclidean distance and Cosine similarity that are often used as feature or histogram comparison measures. The ensemble of the log-Gabor image filtering, feature extraction using MLBP, and SROCC yields the best performance, indicating that these components complement each other well for the task of face photo-sketch recognition.

Table A.1: Means and standard deviations over five random train/test set-splits when using hand-drawn sketches, without demographic filtering. Training set contains 300 subjects while testing set contains 952 subjects. An extended gallery containing 1522 subjects is also used. LG = log-Gabor.

#	Method	Matching Rate (%) at Rank-X					TAR@FAR=0.1%	TAR@FAR=1.0%	EER (%)
		X=1	X=10	X=50	X=100	X=150			
1	MLBP & SROCC	39.12±1.22	61.37±0.81	76.89±2.06	83.76±1.59	87.39±1.27	50.53±1.27	71.01±1.24	9.49±0.98
2	LG & SROCC	71.43±0.78	85.92±0.68	92.65±0.67	95.11±0.20	96.20±0.20	82.94±0.48	92.63±0.35	3.69±0.29
3	LG & MLBP & Euclidean	52.84±1.14	67.42±1.47	79.87±0.91	84.66±0.94	87.35±0.72	62.25±1.54	78.30±1.12	7.58±0.39
4	LG & MLBP & Cosine	73.61±1.62	90.53±1.00	96.68±0.48	98.17±0.35	**98.74±0.32**	87.56±1.29	95.55±0.77	2.49±0.44
5	*LG & MLBP & SROCC (LGMS)*	**81.37±0.42**	**93.72±0.39**	**97.46±0.37**	**98.49±0.34**	**98.89±0.24**	**92.79±0.39**	**97.94±0.46**	**1.55±0.27**

A.3 DEEPS parameter tuning

The DCNN used in the DEEPS/DEEPS-M framework contains several parameters that can be adjusted, most of which were set to the same values used to train the VGG-Face network [93] that forms the basis of the DCNN as elaborated in Section 3.3.1. However, the triplet distance margin (ref. Section 3.3.3) and batch size when performing triplet embedding were not provided in [93]. Hence, these parameters were tuned on a validation set extracted from the training set. Utilising the hand-drawn sketches due to the higher number of subjects available, the training set is reduced from 800 subjects to 600 subjects and the remaining 200 subjects are used for validation.

As shown in Table A.2, the effect of the parameter α which determines the distance margin between positive and negative triplets is minor. Hence, the mid-point value of $\alpha = 1.00$ was chosen. With regards to the batch size, it is evident in Table A.3 that the smallest batch size of 250 images generally yields poorer performance than larger batch sizes. However, differences between the results of the other batch sizes are relatively minimal. Hence, a batch size of 500 exemplars was chosen due to faster training times compared to the other batch sizes. The above triplet margin and batch size values were thus used to compute the results in Chapter 6.

Table A.2: Means and standard deviations over two random train/test set-splits when using hand-drawn sketches, without demographic filtering. Training set contains 600 subjects while validation set contains 200 subjects. An extended gallery containing 1521 subjects is also used.

#	α	Matching Rate (%) at Rank-X					TAR@FAR=0.1%	TAR@FAR=1.0%	EER (%)
		X=1	X=10	X=50	X=100	X=150			
1	0.00	75.75±2.47	93.75±0.35	98.25±0.35	99.75±0.35	99.75±0.35	88.50±0.71	97.25±0.35	1.55±0.08
2	0.25	75.50±3.54	93.50±0.71	98.50±0.00	99.50±0.00	99.75±0.35	89.00±1.41	97.25±0.35	1.56±0.03
3	0.50	75.25±3.89	93.25±1.06	98.50±0.00	99.50±0.00	100.00±0.00	88.75±0.35	97.75±0.35	1.52±0.03
4	0.75	75.75±2.47	93.50±0.71	98.75±0.35	99.25±0.35	99.75±0.35	88.75±1.06	97.75±0.35	1.48±0.05
5	1.00	75.50±4.24	92.75±1.06	98.50±0.00	99.75±0.35	99.75±0.35	89.50±0.71	97.75±0.35	1.48±0.03
6	1.25	76.25±2.47	93.75±0.35	99.00±0.00	99.75±0.35	99.75±0.35	88.50±0.71	97.25±0.35	1.45±0.07
7	1.50	76.75±1.77	93.75±1.06	99.00±0.71	99.75±0.35	99.75±0.35	88.50±1.41	98.50±0.00	1.50±0.01
8	1.75	75.25±2.47	93.25±0.35	98.75±0.35	99.50±0.00	99.75±0.35	87.25±1.06	97.75±0.35	1.52±0.03
9	2.00	75.50±2.83	93.00±0.00	98.75±0.35	99.25±0.35	99.75±0.35	89.75±0.35	97.50±0.71	1.46±0.05

Table A.3: Means and standard deviations over two random train/test set-splits when using hand-drawn sketches, without demographic filtering. Training set contains 600 subjects while validation set contains 200 subjects. An extended gallery containing 1521 subjects is also used.

#	Batch size	Matching Rate (%) at Rank-X					TAR@FAR=0.1%	TAR@FAR=1.0%	EER (%)
		X=1	X=10	X=50	X=100	X=150			
1	250	69.50±1.41	91.25±1.06	**98.25±0.35**	**99.25±0.35**	**99.50±0.71**	83.50±1.41	96.50±1.41	1.71±0.29
2	500	**75.50±4.24**	**92.75±1.06**	**98.50±0.00**	**99.75±0.35**	**99.75±0.35**	**89.50±0.71**	**97.75±0.35**	**1.48±0.03**
3	750	**75.75±2.47**	**93.00±0.71**	**98.75±0.35**	**99.50±0.00**	**99.75±0.35**	**88.50±0.71**	**97.50±0.00**	**1.50±0.00**
4	1000	**74.50±3.54**	**92.75±0.35**	**98.50±0.00**	**99.50±0.00**	**99.75±0.35**	**87.50±1.41**	**97.75±0.35**	**1.48±0.02**

A.4 DEEPS components

An analysis into the performance of each component of the proposed DEEPS framework is now given, namely the benefit of transfer learning, data augmentation, triplet embedding and the triplet embedding scheme used. Experiments are performed on the UoM-SGFS Set A database and the results of several set-ups as described in Table A.4 are shown in Table A.5.

A.4.1 Transfer learning & data augmentation

The set-up denoted by S2 considers the fine-tuning of the VGG-Face network parameters and the layer employing TDE using only one photo and one sketch per subject (the original images), which yields noticeable improvements compared to the basic VGG-Face network since it is now trained with both modalities. However, performance is only on par with the best-performing algorithm, namely D-RS+CBR, results of which are shown in Section 6.2. The use of the synthesised images created with the 3D morphable model for training as performed in S3 yields substantial improvements across all performance measures, enabling the retrieval of 6-13% more subjects and a reduction in the error rate by 33%. Since images used to train S2 were randomly cropped and flipped, which are two of the most common approaches employed to augment data, its inferior performance compared to S3 indicates that the traditional data augmentation techniques are insufficient for the task of face photo-sketch recognition. Compared to the performance of the original VGG-Face method, transfer learning and the proposed data augmentation increase rank retrieval rates by 15-35% and reduce the error rate by 55.9%.

Table A.4: Overview of different set-ups used to generate the results in Table A.5. All configurations apply transfer learning to the basic VGG-Face network [93].

Config.	Description
S1	No triplet embedding, using all images (original + synthetic)
S2	Using TDE for triplet loss and one image per subject (original)
S3	Using TDE for triplet loss and all images (original + synthetic)
S4	Using TSE for triplet loss and all images (original + synthetic)

A.4.2 Triplet embedding

Employing a triplet embedding scheme is also beneficial, as observed in comparing the performance of configuration S1 (which did not employ this scheme) with S3 and S4 which both consist of S1 concatenated with a layer employing TDE and TSE, respectively. More specifically, the additional layer that was tuned for verification (and which also serves as a means of dimensionality reduction) enables the retrieval of approximately 10% more subjects at most ranks, and reduces the error rate by 37%.

A.4.3 Triplet embedding scheme

Comparing the results of configurations S3 and S4 indicates that the two approaches generally yield relatively similar performance. However, TDE tends to consistently outperform TSE. Configuration S3 (which uses TDE) is therefore selected as the final proposed architecture.

A.4.4 Facial adjustments

The proposed data augmentation strategy varies several attributes of a face image in order to generate new synthetic images that are used for model training. An ablative analysis of the effect on performance when omitting individual changes is thus performed, to determine if any set of changes is more important than others. Configuration S3 is used as the benchmark, since the same network set-up is used and employs all images. As shown in Table A.6, the omission of any group of changes leads to reduced performance, with the lack of weight and height variations leading to the largest losses. This is likely a result of these variations affecting the roundness of a face, which can easily be represented inaccurately in a sketch image. Hence, the inclusion of images exhibiting these variations allow a network to be more robust to such commonly encountered differences. The least performance loss is observed in the case of age variations. This is likely a result of sketches being generated while viewing the photos, such that the age is represented quite accurately. However, law enforcement agencies may only have old photos of a suspect, yielding significant differences between the age of the subject in the photo

and the sketch. Hence, the effect of this attribute would likely be more prevalent in real-world applications.

A.4.5 Fusion

The ranks of LGMS and DEEPS-M are compared to the system fusing both methods in Figure A.2 to determine the effects of fusion. Although fusion sometimes leads to performance loss compared to one of the methods, it usually yields improvement when compared to the other method. The fused system therefore performs a trade-off between the benefits and weaknesses of both systems. Overall, the ranks of most subjects improve considerably over those of both LGMS and DEEPS-M as shown in Table 6.5 on Page 84 and thus LGMS+DEEPS-M is better generalised than either method alone. Similar observations can also be made for the fusion of LGMS and DEEPS.

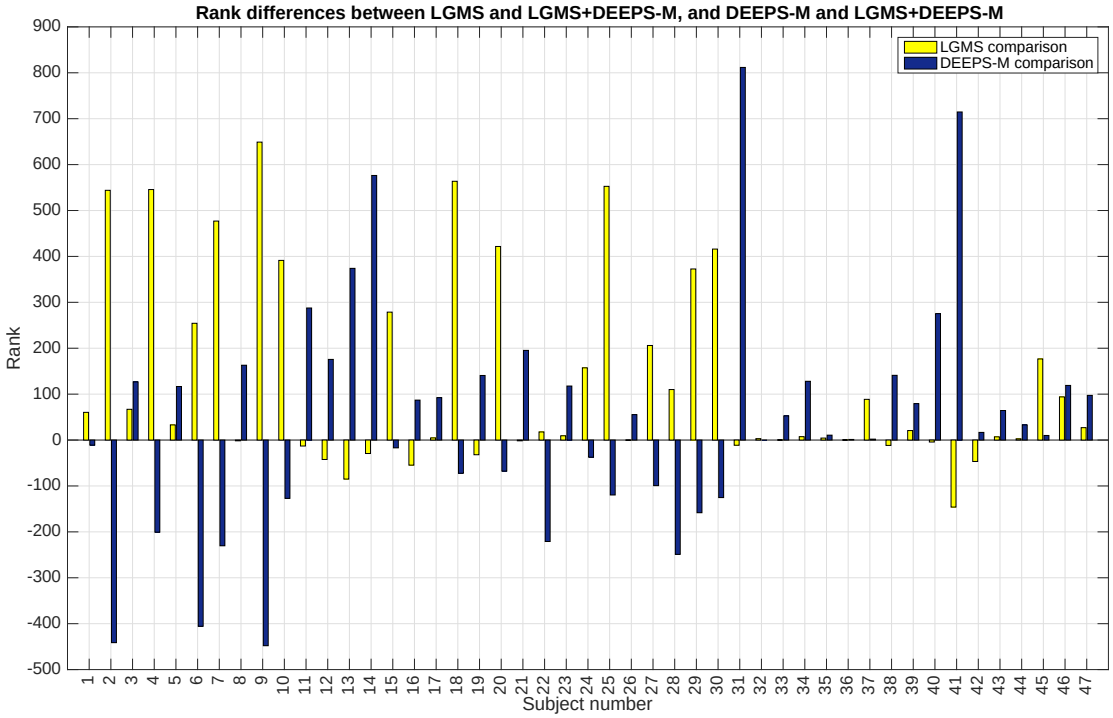


Figure A.2: Rank differences between all 47 subjects in the PRIP-HDC database [19] when comparing LGMS and DEEPS-M with LGMS+DEEPS-M (positive values indicate LGMS+DEEPS-M rank improvements).

Table A.5: Means and standard deviations over 5 train/test set-splits for variations of DEEPS when evaluated on UoM-SGFS Set A software-generated sketches. Configuration descriptions are given in Table A.4.

Algorithm	Matching Rate (%) at Rank-X					TAR@FAR=0.1%	TAR@FAR=1.0%	EER (%)
	X=1	X=10	X=50	X=100	X=150			
VGG-Face	9.33±2.45	31.07±3.73	59.73±2.52	73.60±3.58	80.80±2.72	11.87±2.47	37.33±4.40	14.18±1.05
S1	22.80±2.88	49.33±3.65	73.07±2.69	82.53±1.85	87.47±2.08	29.60±5.53	54.40±3.90	9.99±1.29
S2	20.93±2.14	52.93±3.96	76.40±3.25	85.47±3.25	90.40±3.79	27.47±2.72	61.07±4.68	9.35±1.34
S3	**31.60±1.12**	**66.13±2.47**	**86.00±1.25**	**93.47±1.85**	**96.40±1.21**	**41.87±3.11**	**73.47±2.08**	**6.26±0.60**
S4	29.73±3.35	62.67±2.26	**86.00±2.00**	92.40±2.85	**96.27±2.19**	39.60±2.56	70.93±1.86	**6.39±0.88**

Table A.6: Means and standard deviations over 5 train/test set-splits when evaluated on UoM-SGFS Set A software-generated sketches and omitting individual attribute variations (i.e. age, gender, weight, and height) for photo and sketch generation.

Description	Matching Rate (%) at Rank-N					TAR@FAR=0.1%	TAR@FAR=1.0%	EER (%)
	N=1	N=10	N=50	N=100	N=150			
All variations (Config. S3)	**31.60±1.12**	**66.13±2.47**	**86.00±1.25**	**93.47±1.85**	**96.40±1.21**	**41.87±3.11**	**73.47±2.08**	**6.26±0.60**
No age variations	26.93±2.85	60.53±3.78	85.47±2.28	91.47±1.45	95.73±1.38	35.73±4.10	69.73±3.08	6.78±0.97
No gender variations	27.07±1.86	59.47±4.31	85.20±3.38	90.67±1.33	94.67±1.49	35.87±2.18	68.27±5.28	6.93±0.39
No height variations	24.67±3.50	55.87±6.30	80.80±4.51	90.00±2.45	93.73±1.80	31.47±4.33	65.20±7.92	7.24±0.61
No weight variations	26.27±3.39	58.13±8.74	81.47±3.44	89.73±2.39	93.33±2.75	35.20±5.91	65.47±6.17	7.62±1.65

A.5 DEEPS/DEEPS-M network visualisation

As mentioned in Section 6.5, two methods are utilised to understand what contributes to the performance of DEEPS and the related DEEPS-M method: the first is the approach in [161], which is used to visualise the network at each layer. More specifically, this method depicts what information is retained by the network by approximating the inverse of the output representation to find the image whose representation best matches the one given. This is performed by modelling the representation output from a network as a function $\Phi(\mathbf{x})$ of the image $\mathbf{x}$ and then computing the inverse Φ^{-1}, formulated as a regularised regression problem that essentially aims to minimise the Euclidean distance between the representation to be inverted and the representation of the image that is found during the inversion optimisation process. The solution (i.e. the image to be found) is initialised with random noise and gradient descent is employed to minimise the objective function. Since representations should mitigate irrelevant differences in images (e.g. illumination, face pose, viewpoint etc.), the network function should not be uniquely invertible. Therefore, several reconstructions can be obtained which are virtually indistinguishable from the network's viewpoint. In this dissertation, four images are obtained from each layer to cater for this observation.

The results when using the network trained using photo and hand-drawn sketch pairs are shown in Figures A.4 to A.6 when inputting a photo and in Figures A.7 to A.9 when inputting a sketch. The results when also inputting a hand-drawn forensic sketch and the corresponding photo are shown in Figures A.13 to A.15 and Figures A.10 to A.12, respectively. The results when using the network trained using photo and software-generated sketch pairs are shown in Figures A.16 to A.18 when inputting a photo and in Figures A.19 to A.21 when inputting a sketch. The original input images are depicted in Figure A.3.

As shown in the figures below, most layers retain the structural information of the face images, although some differences in the facial components are evident at deeper layers. This indicates that the network is invariant to slight differences in the facial components, as desired. An interesting phenomenon also occurs in the case of the sketches, where the network appears to attempt colour inference from layer L8 onwards. This is true for subjects with white skin tones and also darker skin tones. At the final layers (L19-L24), the network also becomes highly robust

to translation and larger differences in the facial components.

The second network visualisation approach is via the t-Distributed Stochastic Neighbour Embedding (t-SNE) method [162], which performs non-linear dimensionality reduction of features (i.e. the 1024-D DEEPS features of the test images) to 2-D, such that similar objects are located nearby while dissimilar objects are farther apart. Results are shown in Figure A.22 when using hand-drawn sketches and the corresponding photos, and in Figure A.23 when using software-generated sketches and the corresponding photos. Due to the limited number of test images, the results are inconclusive in the case of the software-generated sketches. However, the greater quantities of hand-drawn sketches and corresponding photos allow several observations to be made (as also discussed in Section 6.5). First, the distributions of face photos and the corresponding face sketches are quite similar, showing that the network handles images from the two modalities in a similar manner. Hence, the network appears to have successfully learned modality-invariant features. Moreover, it can also be observed that certain classes of images are clustered together despite no supervision being provided to the network to enable this behaviour, e.g. by race, gender, or finer characteristics such as facial hair. In this manner, the network is implicitly computing a form of demographic filtering.

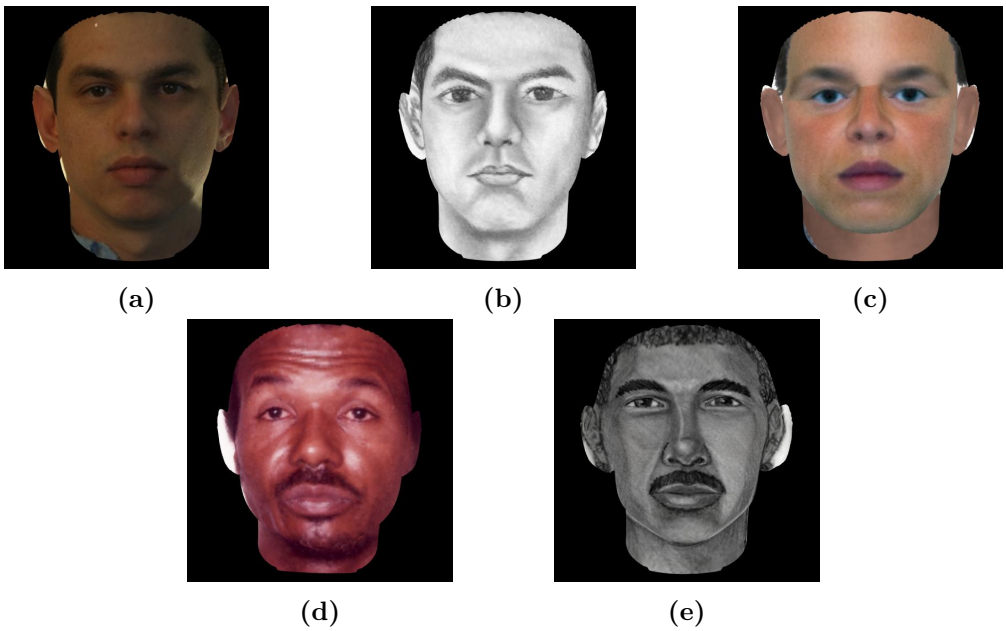


Figure A.3: Images used as input to networks: (a) Photo of a subject in the Color FERET database [13,152], (b) corresponding hand-drawn sketch in the CUFSF database [15,129], (c) corresponding software-generated sketch in the UoM-SGFS Set A database, (d) Photo of a subject in the PRIP-HDC database [19], (e) corresponding forensic hand-drawn sketch in the PRIP-HDC database [19]

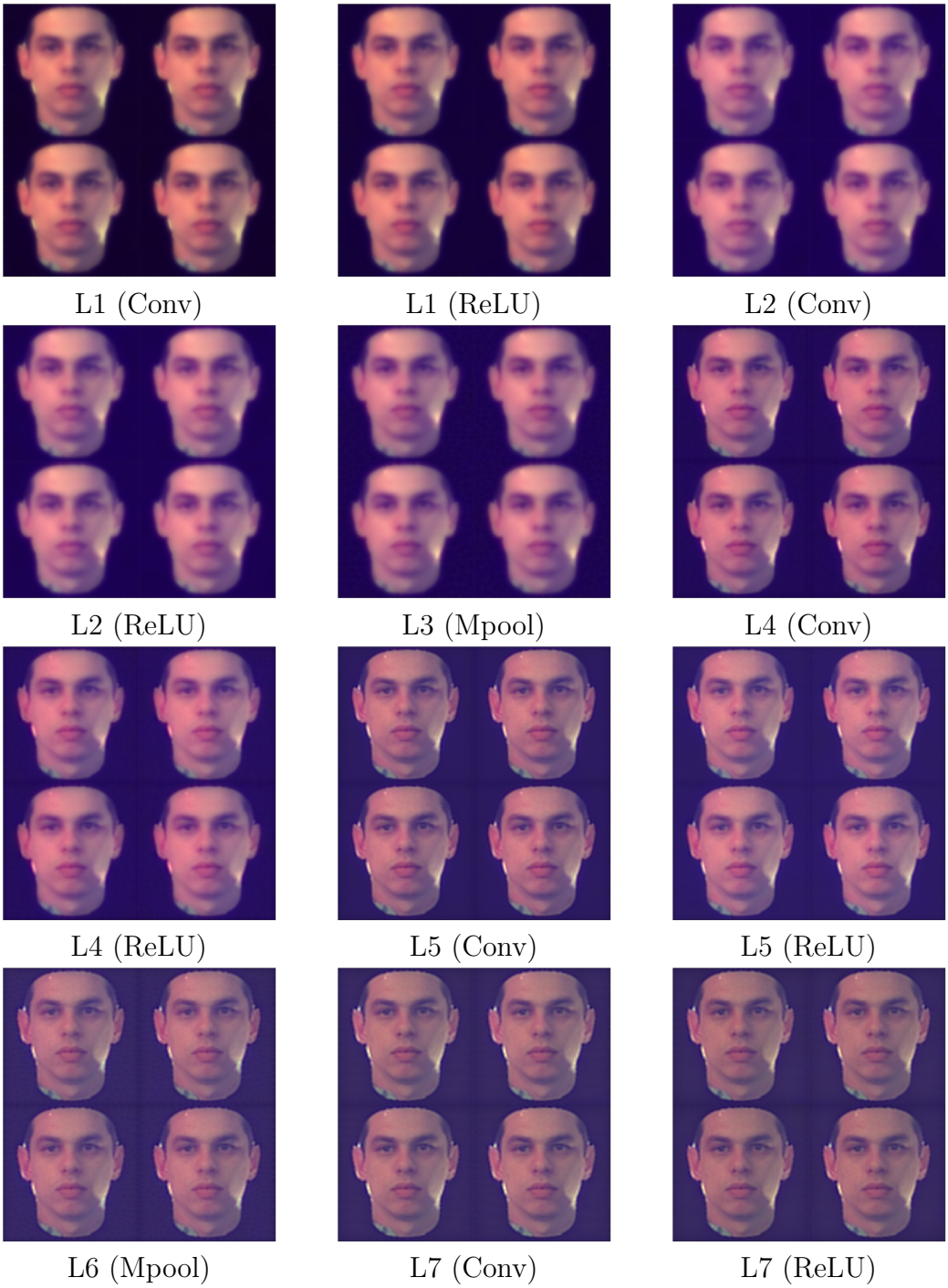


Figure A.4: Visualisation of network (shown in Table 3.1) when using photo/hand-drawn sketch pairs for training and a photo as input.

Figure A.5: Visualisation of network (shown in Table 3.1) when using photo/hand-drawn sketch pairs for training and a photo as input.

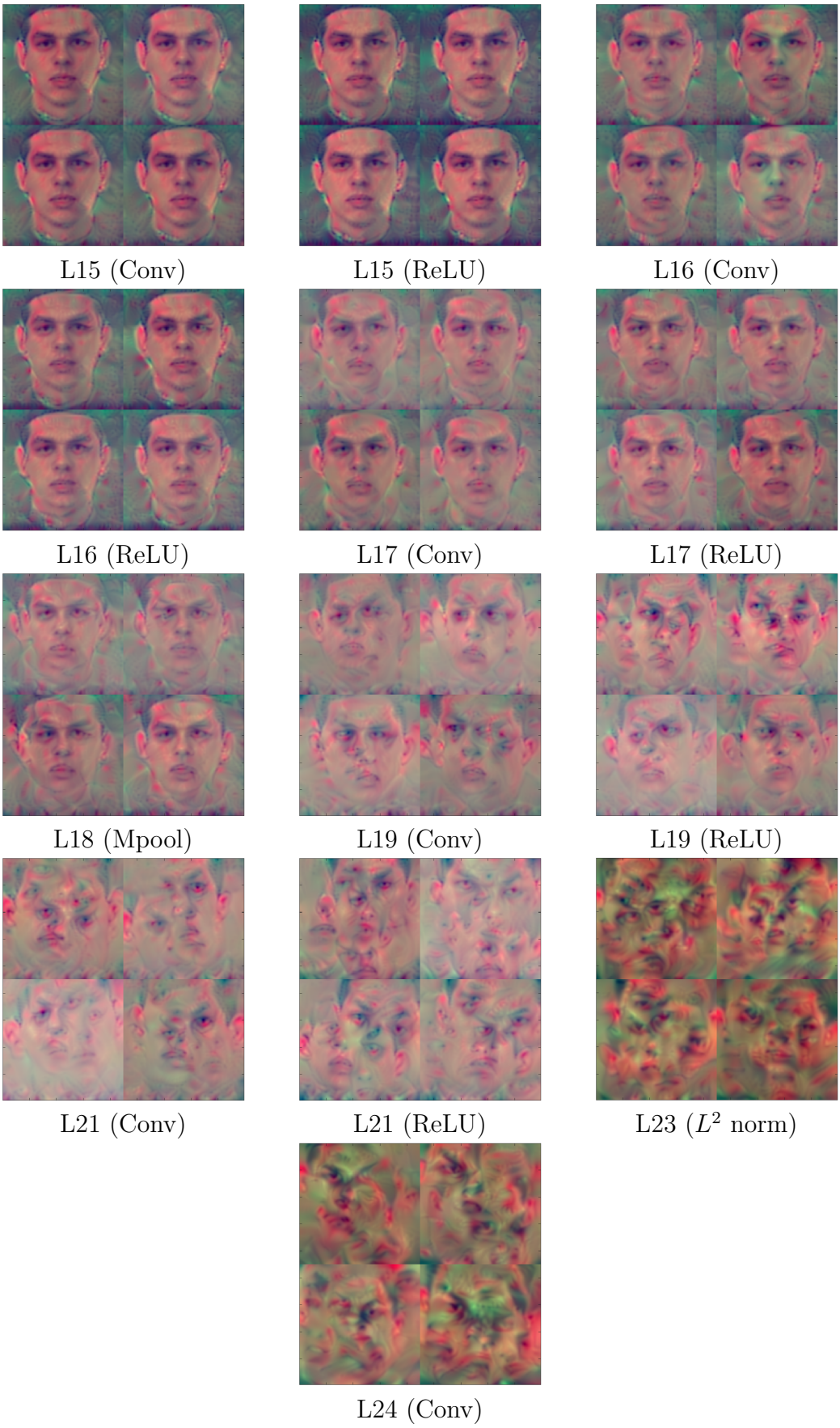


Figure A.6: Visualisation of network (shown in Table 3.1) when using photo/hand-drawn sketch pairs for training and a photo as input.

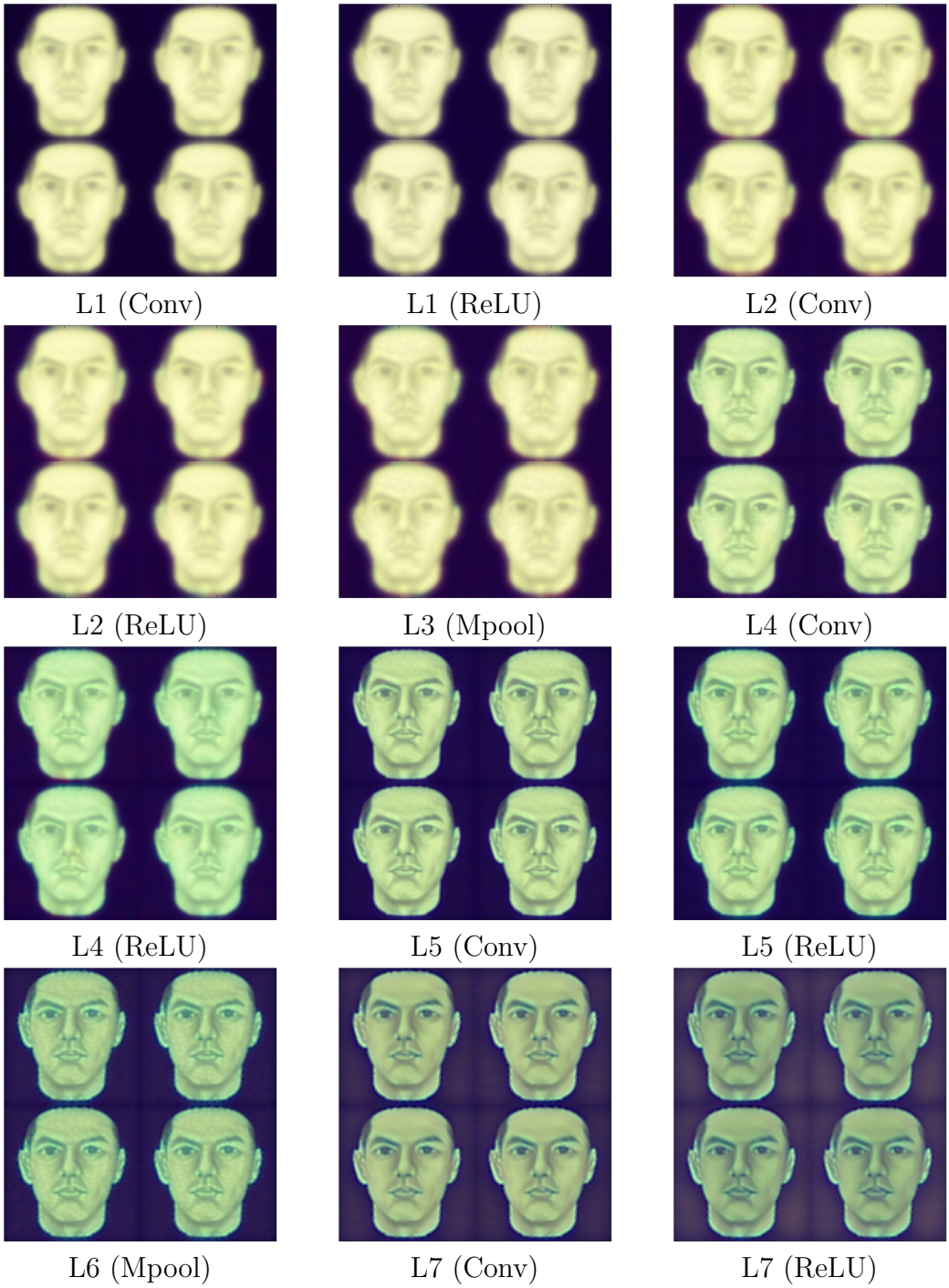


Figure A.7: Visualisation of network (shown in Table 3.1) when using photo/hand-drawn sketch pairs for training and a sketch as input.

Figure A.8: Visualisation of network (shown in Table 3.1) when using photo/hand-drawn sketch pairs for training and a sketch as input.

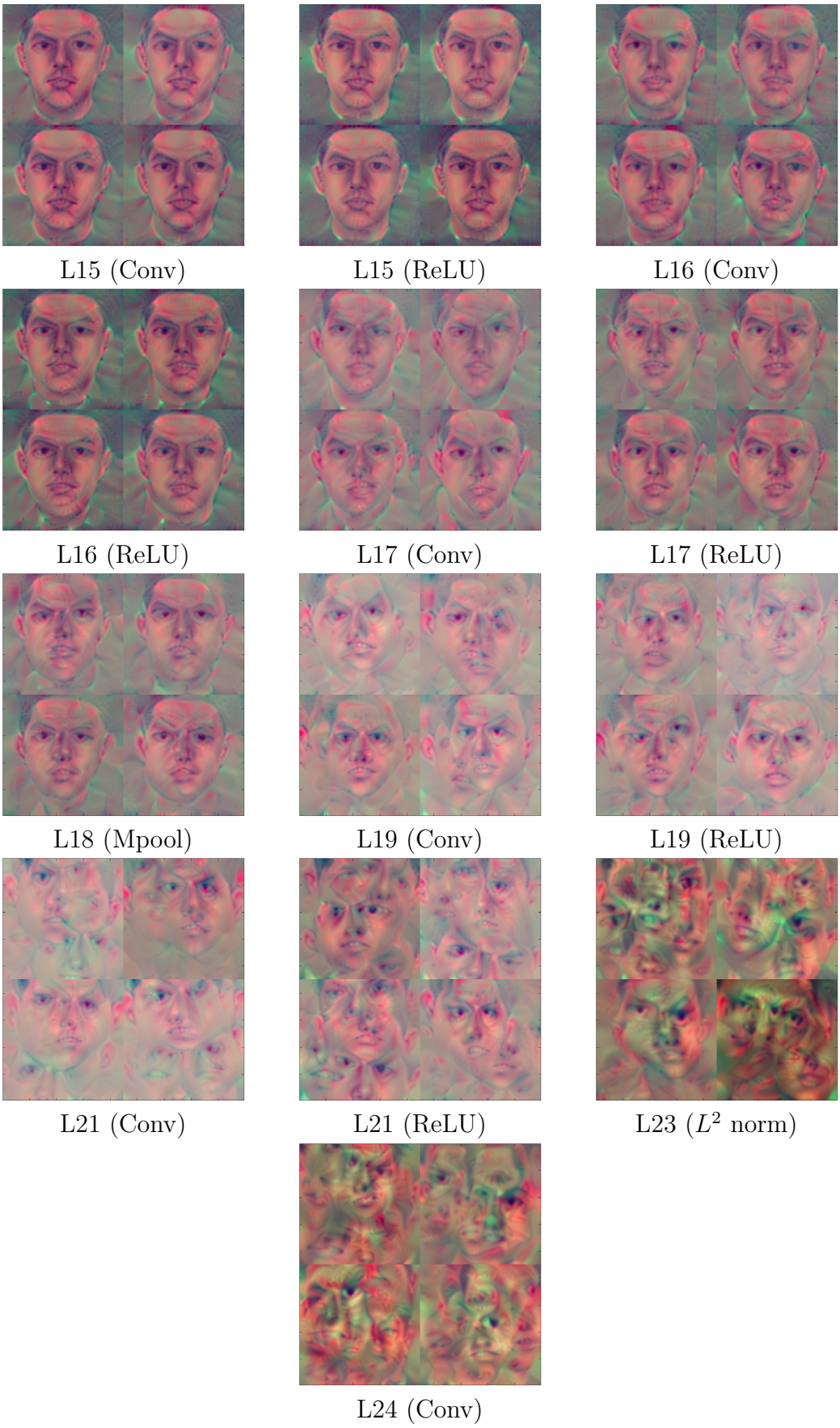


Figure A.9: Visualisation of network (shown in Table 3.1) when using photo/hand-drawn sketch pairs for training and a sketch as input.

Figure A.10: Visualisation of network (shown in Table 3.1) when using photo/hand-drawn sketch pairs for training and a photo as input.

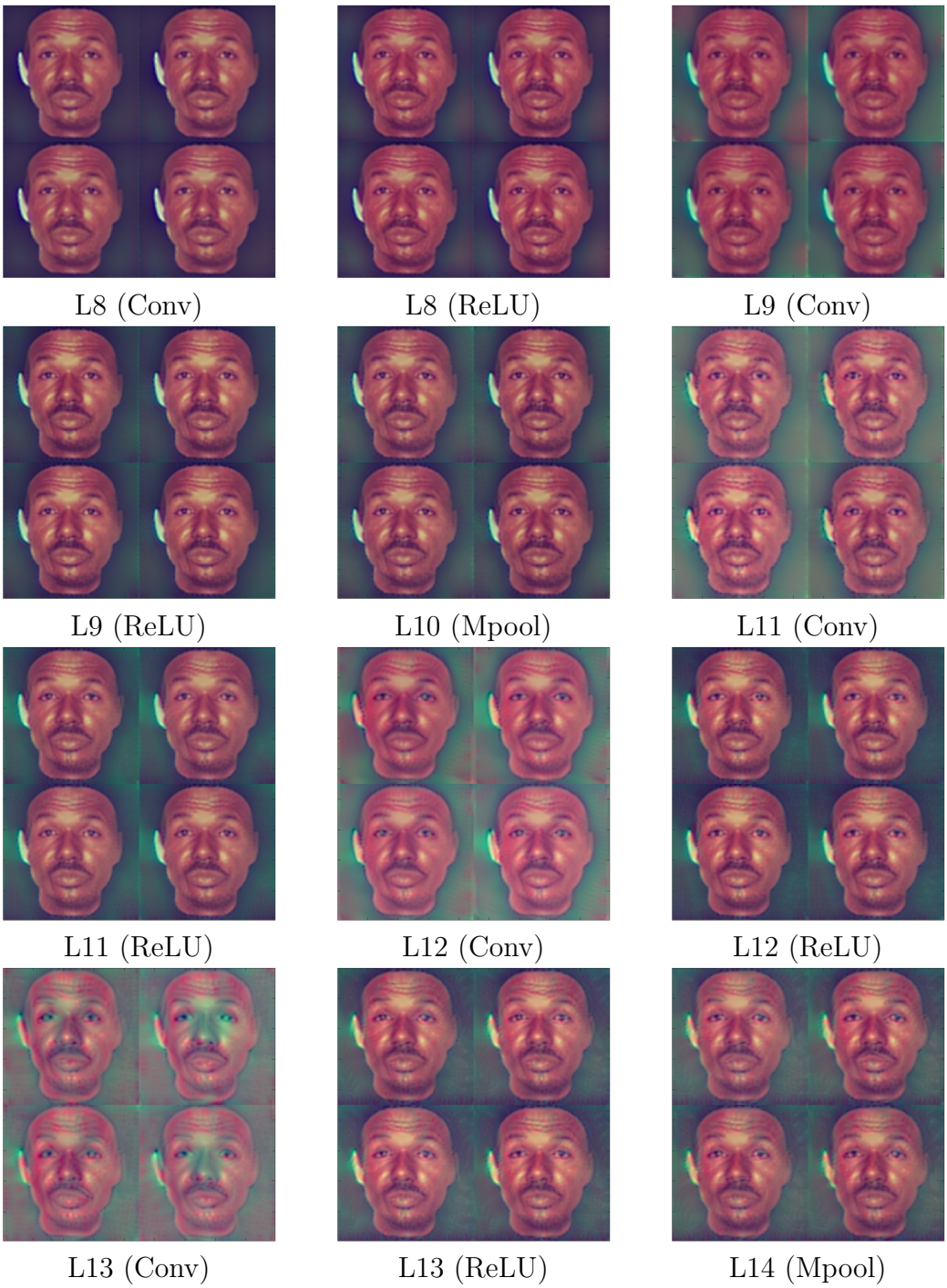


Figure A.11: Visualisation of network (shown in Table 3.1) when using photo/hand-drawn sketch pairs for training and a photo as input.

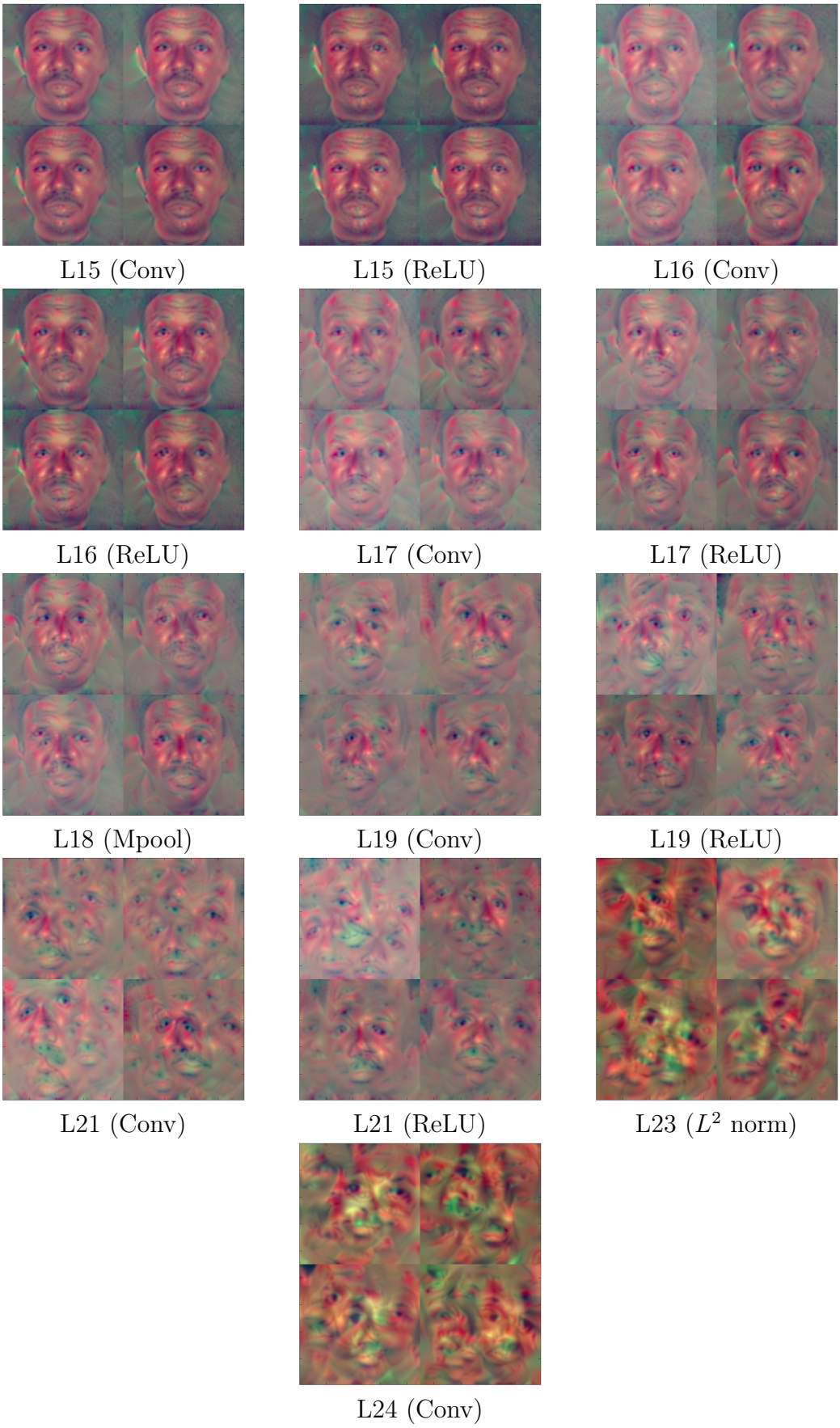


Figure A.12: Visualisation of network (shown in Table 3.1) when using photo/hand-drawn sketch pairs for training and a photo as input.

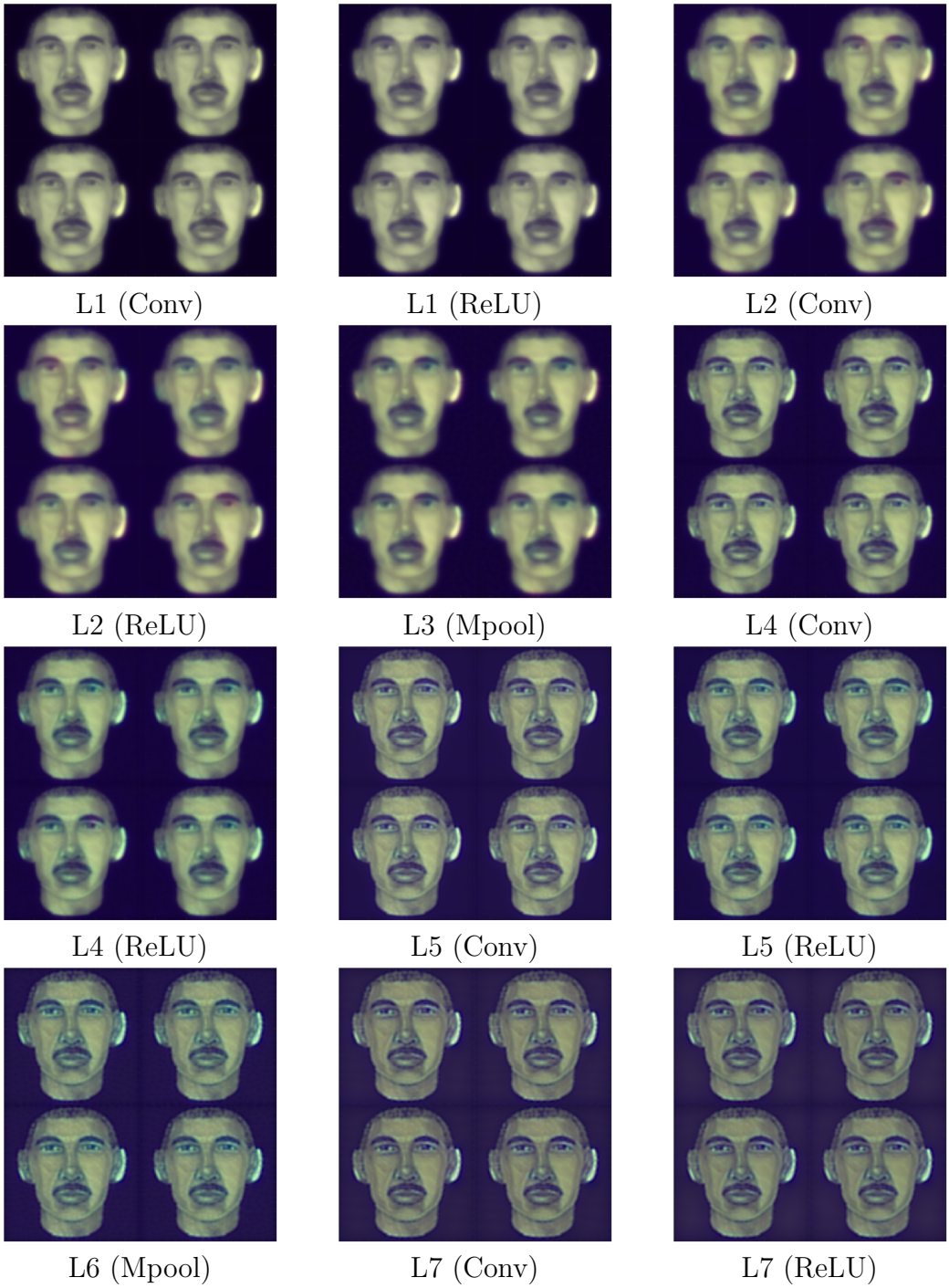


Figure A.13: Visualisation of network (shown in Table 3.1) when using photo/hand-drawn sketch pairs for training and a sketch as input.

Figure A.14: Visualisation of network (shown in Table 3.1) when using photo/hand-drawn sketch pairs for training and a sketch as input.

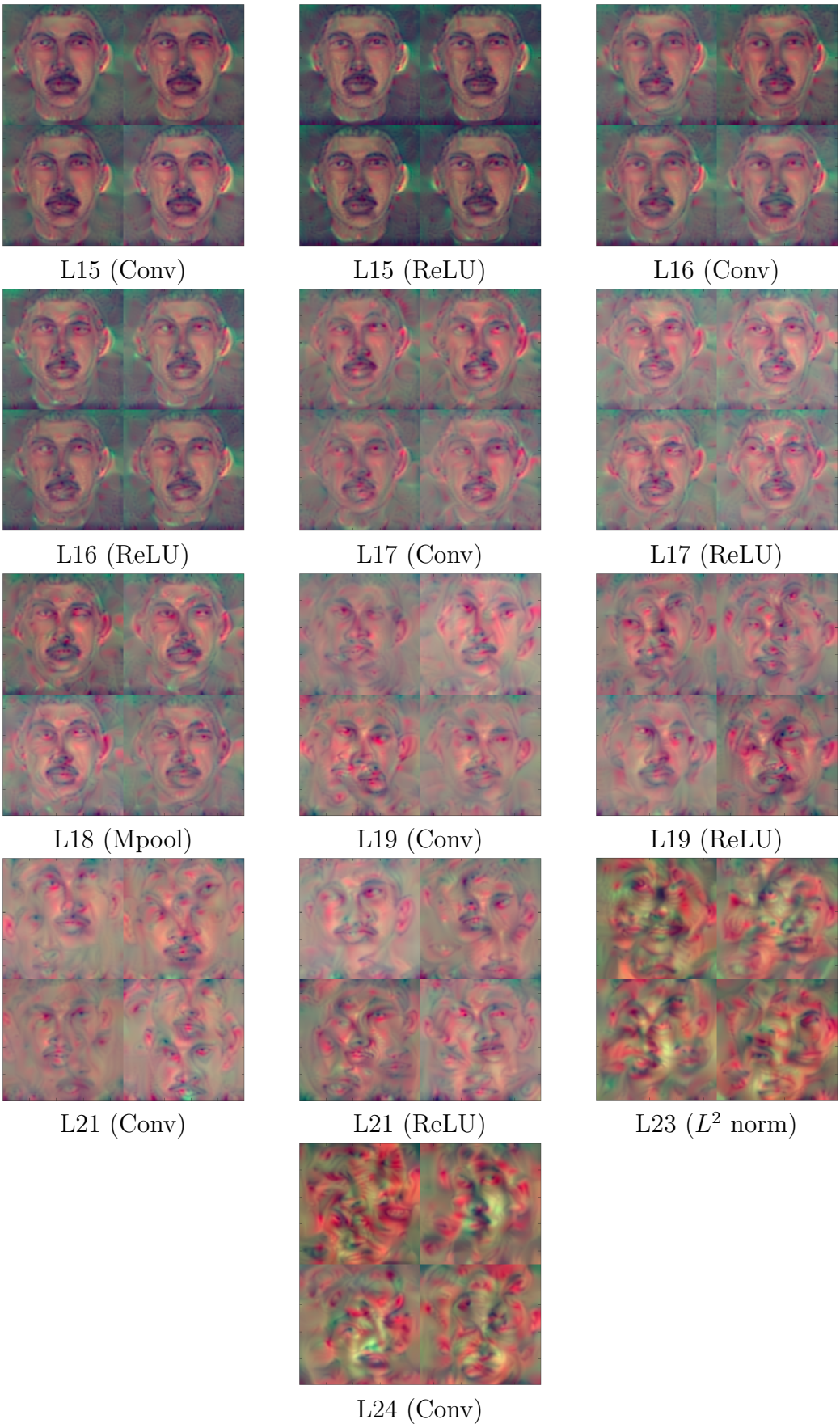


Figure A.15: Visualisation of network (shown in Table 3.1) when using photo/hand-drawn sketch pairs for training and a sketch as input.

Figure A.16: Visualisation of network (shown in Table 3.1) when using photo/software-generated sketch pairs for training and a photo as input.

Figure A.17: Visualisation of network (shown in Table 3.1) when using photo/software-generated sketch pairs for training and a photo as input.

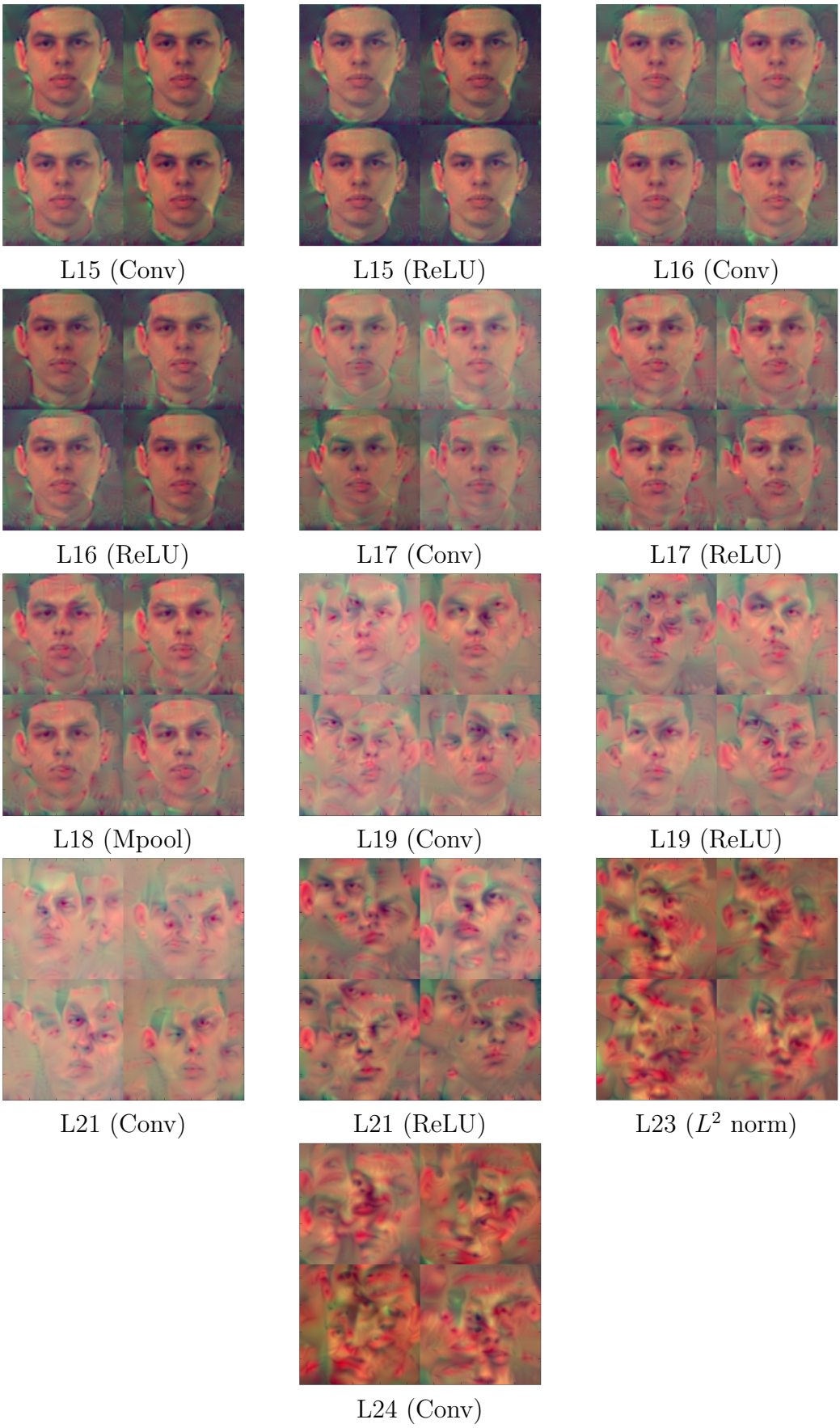


Figure A.18: Visualisation of network (shown in Table 3.1) when using photo/software-generated sketch pairs for training and a photo as input.

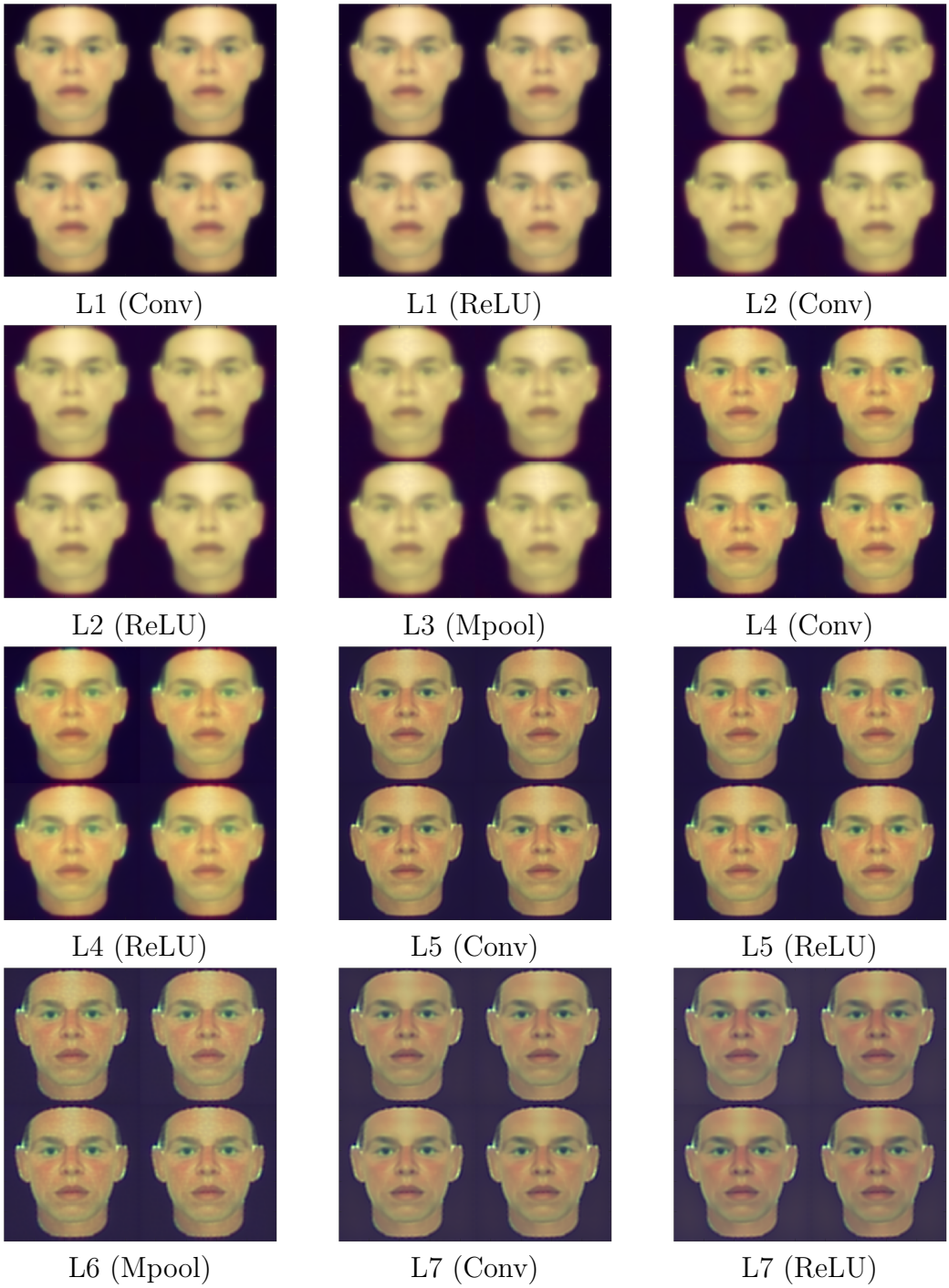


Figure A.19: Visualisation of network (shown in Table 3.1) when using photo/software-generated sketch pairs for training and a sketch as input.

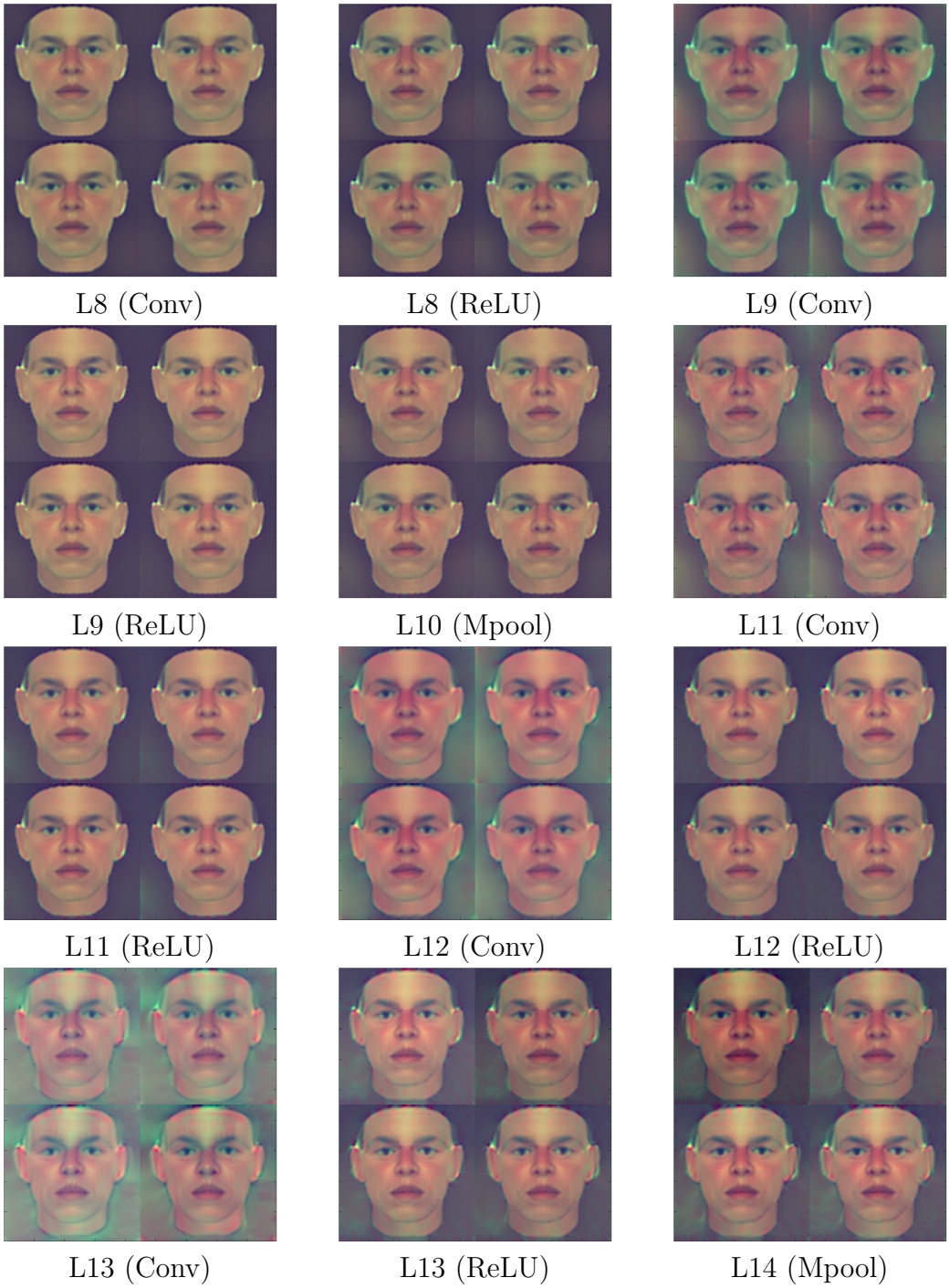


Figure A.20: Visualisation of network (shown in Table 3.1) when using photo/software-generated sketch pairs for training and a sketch as input.

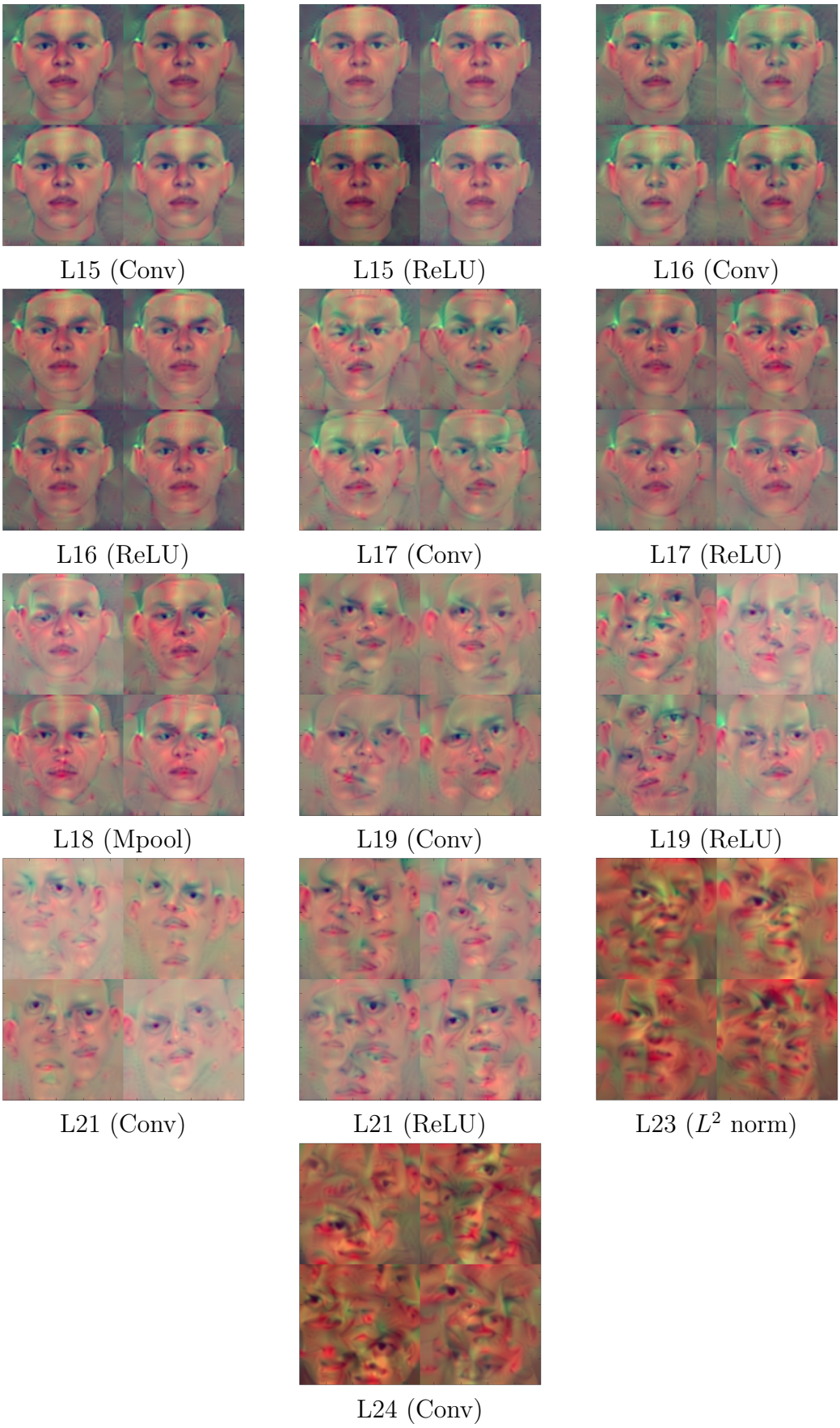


Figure A.21: Visualisation of network (shown in Table 3.1) when using photo/software-generated sketch pairs for training and a sketch as input.

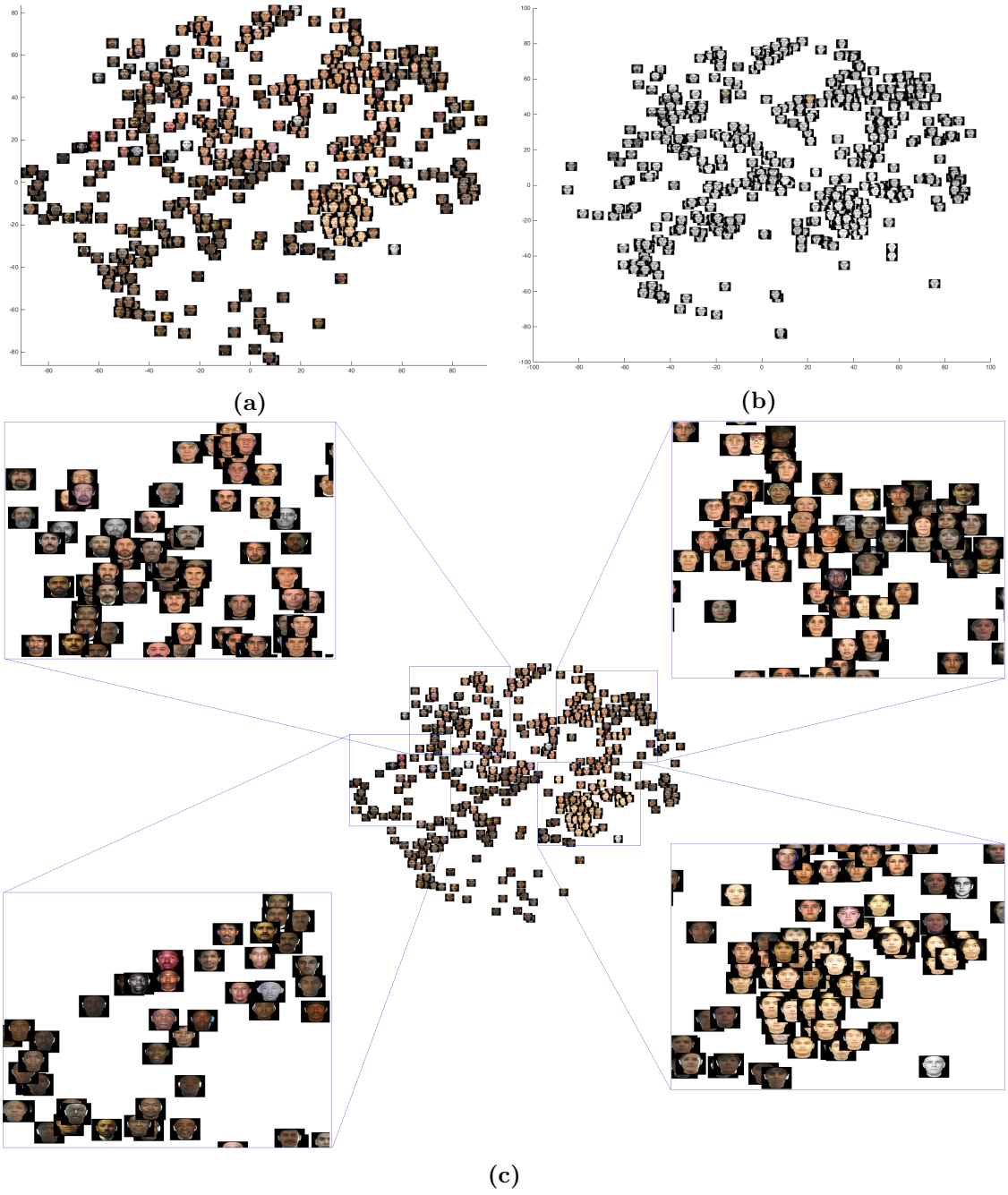


Figure A.22: Visualisation of network using the t-SNE method [162] on 449 sketch-photo pairs: (a) Photos and (b) corresponding hand-drawn sketches, (c) zoomed areas of photos visualisation. Categories appear to have been automatically arranged by semantic similarity despite not being enforced during training, such as men with beards (top left), females (top right), Asian subjects (lower right), and dark-skinned subjects (lower left). Similar observations can also be made for the sketches visualisation. Images best viewed on a screen.

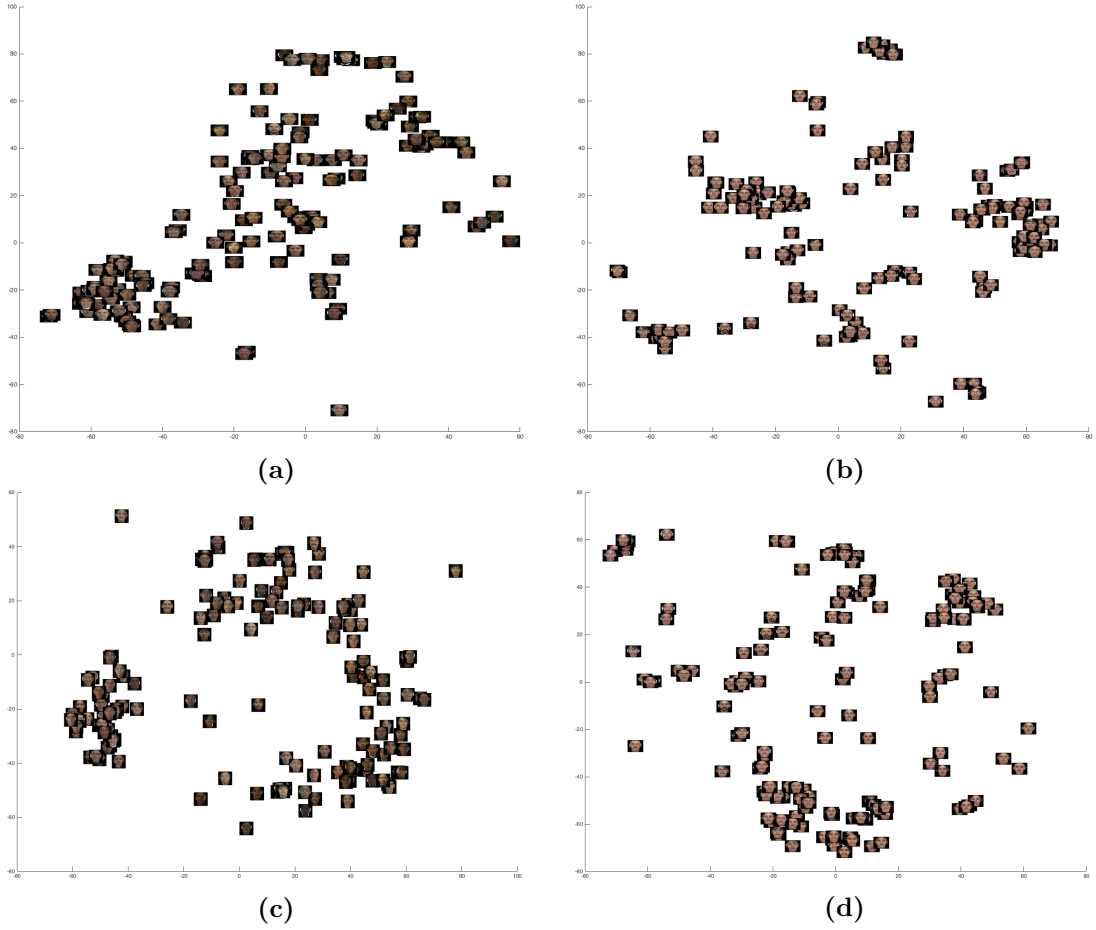


Figure A.23: Visualisation of network using the t-SNE method [162] on 150 photo-sketch pairs: (a) Photos and (b) corresponding UoM-SGFS Set A sketches, (c) photos and (d) corresponding UoM-SGFS Set B sketches. Images best viewed on a screen.

Appendix B

Additional Results

Additional results for the algorithms considered in Sections 5.3 and 5.4 are given in this Appendix, and serve as an extension of the results provided in Chapter 6.

B.1 Rank retrieval rates for forensic sketches

The exact rank values for some of the main algorithms considered are given for all 47 real-world forensic sketches of the PRIP-HDC database [19] in Figures B.2 to B.8. The query sketch and true matching photo are also depicted to provide an insight into the challenges that need to be overcome by the methods considered when matching the two images. The ranks of the proposed methods are also summarised in Figure B.1.

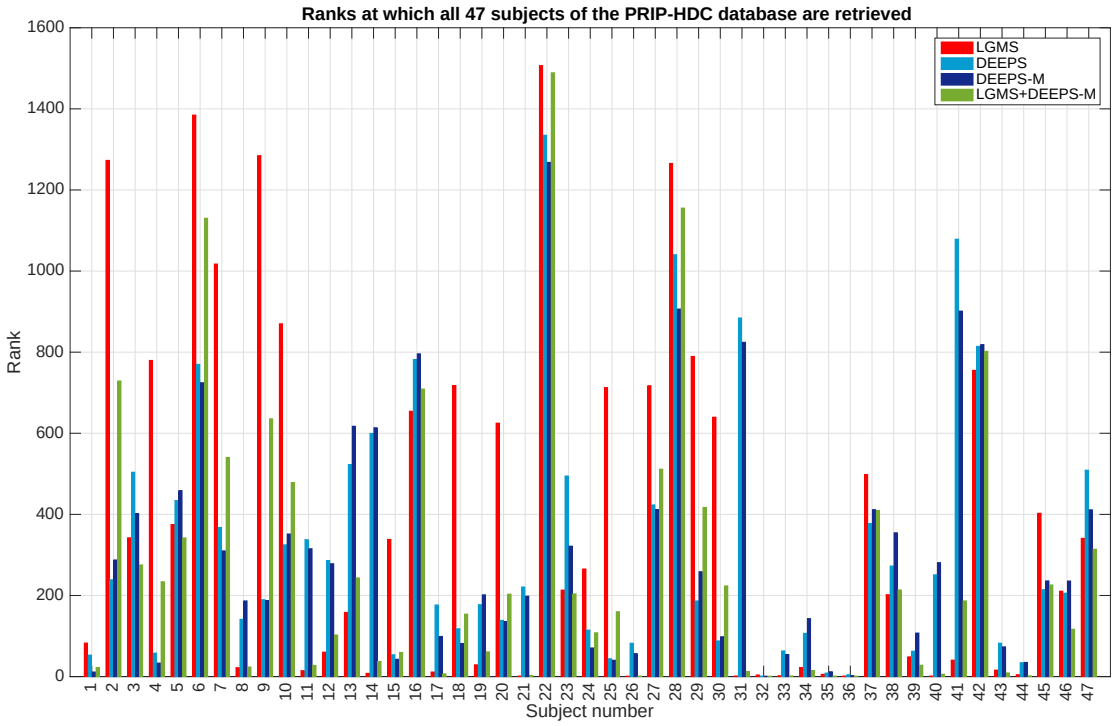
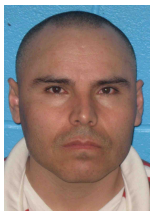


Figure B.1: Ranks of all 47 subjects in the PRIP-HDC database [19] for the methods considered. Smaller values are desired.

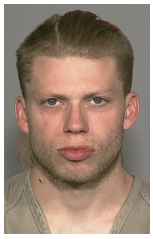



Method	Rank
D-RS+CBR [19]	499.00
LGMS	82.67
DEEPS	53.00
DEEPS-M	11.33
LGMS+DEEPS	38.33
LGMS+DEEPS-M	22.33

Subject 1 (MISC_001)

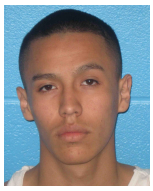

Method	Rank
D-RS+CBR [19]	1538.67
LGMS	1273.00
DEEPS	238.67
DEEPS-M	287.67
LGMS+DEEPS	690.67
LGMS+DEEPS-M	729.00

Subject 2 (MISC_002)

Method	Rank
D-RS+CBR [19]	880.00
LGMS	342.33
DEEPS	504.00
DEEPS-M	402.00
LGMS+DEEPS	329.33
LGMS+DEEPS-M	275.33

Subject 3 (MISC_003)

Method	Rank
D-RS+CBR [19]	144.33
LGMS	779.67
DEEPS	58.00
DEEPS-M	33.33
LGMS+DEEPS	260.33
LGMS+DEEPS-M	234.00

Subject 4 (MISC_004)


Method	Rank
D-RS+CBR [19]	218.00
LGMS	375.00
DEEPS	434.00
DEEPS-M	458.67
LGMS+DEEPS	323.33
LGMS+DEEPS-M	342.00

Subject 5 (MISC_005)


Method	Rank
D-RS+CBR [19]	283.33
LGMS	1384.67
DEEPS	770.00
DEEPS-M	724.67
LGMS+DEEPS	1147.00
LGMS+DEEPS-M	1130.33

Subject 6 (MISC_006)

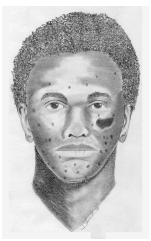
Figure B.2: Ranks (averaged over three set splits) at which the correct photo is retrieved given a query forensic sketch. Images available in the PRIP-HDC database [19].

Method	Rank
D-RS+CBR [19]	28.00
LGMS	1017.33
DEEPS	367.67
DEEPS-M	310.00
LGMS+DEEPS	573.67
LGMS+DEEPS-M	540.33

Subject 7 (MISC_007)

Method	Rank
D-RS+CBR [19]	54.00
LGMS	22.00
DEEPS	141.33
DEEPS-M	186.67
LGMS+DEEPS	20.67
LGMS+DEEPS-M	23.67

Subject 8 (MISC_008)



Method	Rank
D-RS+CBR [19]	10.00
LGMS	1284.67
DEEPS	189.67
DEEPS-M	187.67
LGMS+DEEPS	625.00
LGMS+DEEPS-M	635.67

Subject 9 (MISC_009)

Method	Rank
D-RS+CBR [19]	738.33
LGMS	870.00
DEEPS	325.00
DEEPS-M	351.67
LGMS+DEEPS	456.00
LGMS+DEEPS-M	478.67

Subject 10 (MISC_010)

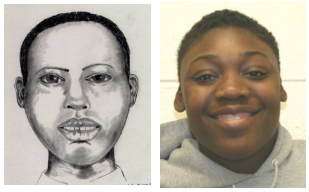
Method	Rank
D-RS+CBR [19]	321.00
LGMS	14.67
DEEPS	337.67
DEEPS-M	315.33
LGMS+DEEPS	33.00
LGMS+DEEPS-M	27.67

Subject 11 (MISC_011)

Method	Rank
D-RS+CBR [19]	360.33
LGMS	60.33
DEEPS	286.33
DEEPS-M	278.33
LGMS+DEEPS	110.67
LGMS+DEEPS-M	102.67

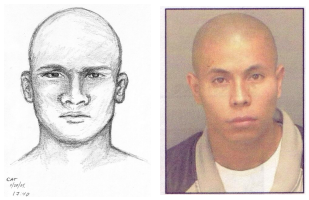
Subject 12 (MISC_012)

Figure B.3: Ranks (averaged over three set splits) at which the correct photo is retrieved given a query forensic sketch. Images available in the PRIP-HDC database [19].



Method	Rank
D-RS+CBR [19]	397.67
LGMS	158.33
DEEPS	523.00
DEEPS-M	617.33
LGMS+DEEPS	202.33
LGMS+DEEPS-M	243.33

Subject 13 (MISC_013)



Method	Rank
D-RS+CBR [19]	126.33
LGMS	8.00
DEEPS	599.67
DEEPS-M	613.67
LGMS+DEEPS	34.33
LGMS+DEEPS-M	37.33

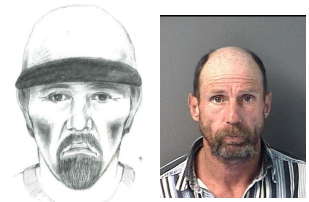
Subject 14 (MISC_014)

Method	Rank
D-RS+CBR [19]	115.67
LGMS	338.33
DEEPS	54.00
DEEPS-M	42.67
LGMS+DEEPS	67.00
LGMS+DEEPS-M	59.67

Subject 15 (MISC_015)

Method	Rank
D-RS+CBR [19]	1458.33
LGMS	654.33
DEEPS	782.00
DEEPS-M	796.00
LGMS+DEEPS	698.33
LGMS+DEEPS-M	709.00

Subject 16 (MISC_016)



Method	Rank
D-RS+CBR [19]	168.67
LGMS	11.33
DEEPS	176.67
DEEPS-M	99.00
LGMS+DEEPS	10.00
LGMS+DEEPS-M	6.67

Subject 17 (MISC_017)

Method	Rank
D-RS+CBR [19]	1037.00
LGMS	717.67
DEEPS	118.00
DEEPS-M	81.67
LGMS+DEEPS	207.00
LGMS+DEEPS-M	154.00

Subject 18 (MISC_018)

Figure B.4: Ranks (averaged over three set splits) at which the correct photo is retrieved given a query forensic sketch. Images available in the PRIP-HDC database [19].


Method	Rank
D-RS+CBR [19]	40.00
LGMS	29.00
DEEPS	177.33
DEEPS-M	201.67
LGMS+DEEPS	56.33
LGMS+DEEPS-M	61.00

Subject 19 (MISC_019)


Method	Rank
D-RS+CBR [19]	24.67
LGMS	625.00
DEEPS	138.67
DEEPS-M	135.67
LGMS+DEEPS	211.00
LGMS+DEEPS-M	203.33

Subject 20 (MISC_020)


Method	Rank
D-RS+CBR [19]	63.33
LGMS	1.33
DEEPS	221.00
DEEPS-M	198.33
LGMS+DEEPS	3.33
LGMS+DEEPS-M	3.00

Subject 21 (MISC_021)


Method	Rank
D-RS+CBR [19]	246.67
LGMS	1506.67
DEEPS	1335.00
DEEPS-M	1268.00
LGMS+DEEPS	1511.33
LGMS+DEEPS-M	1489.00

Subject 22 (MISC_022)


Method	Rank
D-RS+CBR [19]	286.00
LGMS	213.33
DEEPS	494.67
DEEPS-M	321.67
LGMS+DEEPS	280.67
LGMS+DEEPS-M	204.00

Subject 23 (MISC_023)


Method	Rank
D-RS+CBR [19]	329.33
LGMS	265.33
DEEPS	114.67
DEEPS-M	70.67
LGMS+DEEPS	129.67
LGMS+DEEPS-M	108.00

Subject 24 (MISC_024)

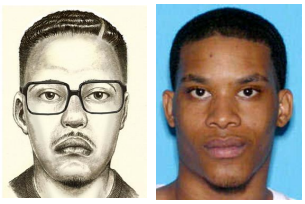
Figure B.5: Ranks (averaged over three set splits) at which the correct photo is retrieved given a query forensic sketch. Images available in the PRIP-HDC database [19].

Method	Rank
D-RS+CBR [19]	112.00
LGMS	712.67
DEEPS	44.00
DEEPS-M	40.33
LGMS+DEEPS	171.00
LGMS+DEEPS-M	160.00

Subject 25 (MISC_026)

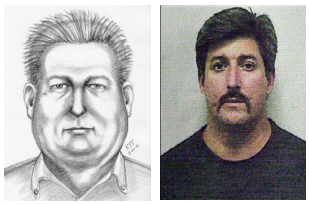
Method	Rank
D-RS+CBR [19]	34.67
LGMS	1.00
DEEPS	82.33
DEEPS-M	56.67
LGMS+DEEPS	1.33
LGMS+DEEPS-M	1.33

Subject 26 (MI_001)



Method	Rank
D-RS+CBR [19]	226.33
LGMS	717.33
DEEPS	423.33
DEEPS-M	412.00
LGMS+DEEPS	514.00
LGMS+DEEPS-M	511.33

Subject 27 (MSP_001)



Method	Rank
D-RS+CBR [19]	681.00
LGMS	1265.33
DEEPS	1040.33
DEEPS-M	906.33
LGMS+DEEPS	1237.33
LGMS+DEEPS-M	1155.33

Subject 28 (WEB_001)

Method	Rank
D-RS+CBR [19]	942.33
LGMS	789.67
DEEPS	186.67
DEEPS-M	258.67
LGMS+DEEPS	339.33
LGMS+DEEPS-M	417.00

Subject 29 (WEB_002)

Method	Rank
D-RS+CBR [19]	172.33
LGMS	639.67
DEEPS	88.00
DEEPS-M	98.33
LGMS+DEEPS	203.67
LGMS+DEEPS-M	223.67

Subject 30 (WEB_003)

Figure B.6: Ranks (averaged over three set splits) at which the correct photo is retrieved given a query forensic sketch. Images available in the PRIP-HDC database [19].

Method	Rank
D-RS+CBR [19]	1.67
LGMS	1.33
DEEPS	884.33
DEEPS-M	824.33
LGMS+DEEPS	14.00
LGMS+DEEPS-M	12.67

Subject 31 (WEB_004)

Method	Rank
D-RS+CBR [19]	1.33
LGMS	4.00
DEEPS	1.33
DEEPS-M	1.00
LGMS+DEEPS	1.00
LGMS+DEEPS-M	1.00

Subject 32 (WEB_005)

Method	Rank
D-RS+CBR [19]	4.00
LGMS	2.00
DEEPS	63.33
DEEPS-M	54.33
LGMS+DEEPS	1.67
LGMS+DEEPS-M	1.67

Subject 33 (WEB_006)

Method	Rank
D-RS+CBR [19]	55.00
LGMS	22.33
DEEPS	106.67
DEEPS-M	143.00
LGMS+DEEPS	16.67
LGMS+DEEPS-M	15.00

Subject 34 (WEB_007)

Method	Rank
D-RS+CBR [19]	19.00
LGMS	5.67
DEEPS	8.67
DEEPS-M	12.00
LGMS+DEEPS	1.00
LGMS+DEEPS-M	1.33

Subject 35 (WEB_008)

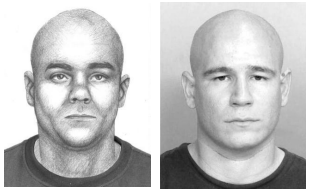
Method	Rank
D-RS+CBR [19]	1.00
LGMS	1.33
DEEPS	4.67
DEEPS-M	2.33
LGMS+DEEPS	1.00
LGMS+DEEPS-M	1.00

Subject 36 (WEB_009)

Figure B.7: Ranks (averaged over three set splits) at which the correct photo is retrieved given a query forensic sketch. Images available in the PRIP-HDC database [19].

Method	Rank
D-RS+CBR [19]	97.00
LGMS	498.33
DEEPS	377.33
DEEPS-M	411.67
LGMS+DEEPS	375.67
LGMS+DEEPS-M	409.67

Subject 37 (WEB_010)

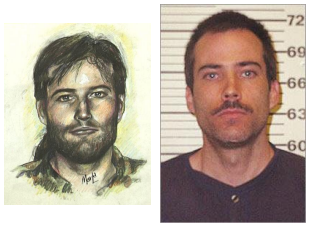


Method	Rank
D-RS+CBR [19]	60.67
LGMS	202.00
DEEPS	272.67
DEEPS-M	354.67
LGMS+DEEPS	173.33
LGMS+DEEPS-M	213.67

Subject 38 (WEB_011)

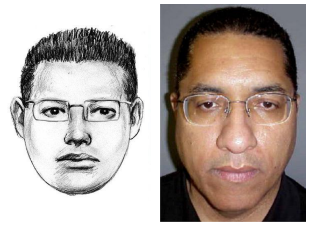
Method	Rank
D-RS+CBR [19]	41.67
LGMS	48.67
DEEPS	62.67
DEEPS-M	107.33
LGMS+DEEPS	20.00
LGMS+DEEPS-M	28.00

Subject 39 (WEB_012)



Method	Rank
D-RS+CBR [19]	1.33
LGMS	1.33
DEEPS	251.00
DEEPS-M	281.00
LGMS+DEEPS	5.00
LGMS+DEEPS-M	5.67

Subject 40 (WEB_013)



Method	Rank
D-RS+CBR [19]	568.00
LGMS	40.67
DEEPS	1078.67
DEEPS-M	901.33
LGMS+DEEPS	243.33
LGMS+DEEPS-M	186.67

Subject 41 (WEB_014)

Method	Rank
D-RS+CBR [19]	538.33
LGMS	755.33
DEEPS	814.33
DEEPS-M	818.67
LGMS+DEEPS	801.00
LGMS+DEEPS-M	802.00

Subject 42 (WEB_015)

Figure B.8: Ranks (averaged over three set splits) at which the correct photo is retrieved given a query forensic sketch. Images available in the PRIP-HDC database [19].

Method	Rank
D-RS+CBR [19]	10.33
LGMS	16.00
DEEPS	82.33
DEEPS-M	73.33
LGMS+DEEPS	9.67
LGMS+DEEPS-M	9.00

Subject 43 (WEB_016)

Method	Rank
D-RS+CBR [19]	4.00
LGMS	4.67
DEEPS	34.33
DEEPS-M	35.00
LGMS+DEEPS	2.33
LGMS+DEEPS-M	2.00

Subject 44 (WEB_017)

Method	Rank
D-RS+CBR [19]	227.00
LGMS	402.67
DEEPS	214.33
DEEPS-M	236.00
LGMS+DEEPS	206.33
LGMS+DEEPS-M	226.00

Subject 45 (WEB_018)

Method	Rank
D-RS+CBR [19]	31.67
LGMS	210.67
DEEPS	205.67
DEEPS-M	235.67
LGMS+DEEPS	104.67
LGMS+DEEPS-M	116.67

Subject 46 (WEB_019)

Method	Rank
D-RS+CBR [19]	1357.33
LGMS	341.00
DEEPS	509.00
DEEPS-M	411.33
LGMS+DEEPS	348.67
LGMS+DEEPS-M	314.00

Subject 47 (WEB_020)

Figure B.9: Ranks (averaged over three set splits) at which the correct photo is retrieved given a query forensic sketch. Images available in the PRIP-HDC database [19].

B.2 Demographic filtering

The use of demographic information of the subject in a probe sketch can improve retrieval rates since the gallery is effectively being filtered by discarding any subjects having differing characteristics to the subject at interest [26]. The effect of demographic filtering is thus investigated, using gender and race/ethnicity. The demographic statistics of the databases considered are shown in Tables B.1 and B.2 for the viewed hand-drawn and software-generated sketches, respectively, with labels based on those used in the Color FERET dataset. Filtering is not performed for the forensic sketches since the demographic data is unavailable.

The performance of the algorithms considered are shown in Tables B.3 to B.5 and Tables B.6 to B.11 for the viewed hand-drawn and software-generated sketches, respectively. It is evident that demographic filtering improves performance for all algorithms, with racial filtering typically being more beneficial than gender filtering. Using both gender and race to filter the gallery yields the best performance. Performance could potentially be further improved using age range, which was shown in [26] to be superior to gender and race/ethnicity. However, this information was not available for all subjects and could not be investigated.

Table B.1: Demographic statistics of the viewed hand-drawn sketches. AA = African-American, PS=statistics of face photo-sketch databases, EG=statistics of extended gallery, Combi=statistics of PS and EG combined.

Gender				Race			
	PS	EG	Combi		PS	EG	Combi
Male	59.15%	70.41%	64.73%	White	61.40%	57.79%	59.62%
Female	40.85%	29.59%	35.27%	Oriental	22.74%	14.73%	18.78%
				Black/AA	5.03%	15.65%	10.28%
				Hispanic	4.38%	4.67%	4.52%
				Asian	4.06%	6.11%	5.08%
				Other	2.38%	1.05%	1.72%

Table B.2: Demographic statistics of the viewed software-generated sketches. AA = African-American, PS=statistics of face photo-sketch database, EG=statistics of extended gallery, Combi=statistics of PS and EG combined.

Gender				Race			
	PS	EG	Combi		PS	EG	Combi
Male	53.00%	70.41%	65.49%	White	59.33%	57.79%	58.23%
Female	47.00%	29.59%	34.51%	Oriental	18.17%	14.73%	15.70%
				Black/AA	8.17%	15.65%	13.53%
				Hispanic	7.17%	4.67%	5.37%
				Asian	5.50%	6.11%	5.94%
				Other	1.67%	1.05%	1.23%

Table B.3: Means and standard deviations over 3 train/test-set splits for algorithms evaluated on viewed hand-drawn sketches, with gender demographic filtering. Methods proposed in this research are shown in italics.

#	Method	Matching Rate (%) at Rank-X					TAR@FAR=0.1%	TAR@FAR=1.0%	EER (%)
		X=1	X=10	X=50	X=100	X=150			
1	VGG-Face [93]	30.43±1.93	53.90±0.87	74.46±1.12	83.83±0.50	88.23±0.38	34.74±2.54	62.35±1.23	9.19±0.68
2	PCA [37]	3.40±0.72	8.29±0.52	15.75±1.04	23.71±2.01	31.34±2.60	3.57±0.52	8.96±0.43	26.25±0.32
3	ET (+PCA) [17]	31.34±1.39	62.19±2.04	81.43±1.18	86.57±0.66	90.88±0.63	39.39±1.90	66.50±1.58	8.07±0.31
4	LLE (+PCA) [159]	42.79±3.53	70.40±1.32	85.24±1.62	90.80±0.50	93.20±0.38	51.16±2.07	71.64±1.63	8.37±0.19
5	HAOG [10]	58.29±1.62	76.95±1.01	89.05±0.50	91.79±0.43	93.62±0.38	68.24±1.04	84.25±1.01	6.11±0.37
6	CBR [6]	25.12±2.49	54.39±1.37	75.95±1.62	84.08±0.50	87.81±0.25	34.08±2.93	63.10±0.80	9.70±0.25
7	D-RS [12,69]	78.77±1.80	92.04±1.14	97.18±0.29	98.51±0.50	98.84±0.52	88.47±1.37	96.43±0.38	1.92±0.15
8	D-RS+CBR [19]	83.91±0.63	94.03±1.32	97.84±0.63	98.76±0.43	99.00±0.43	92.37±0.63	97.84±0.76	1.47±0.28
9	*EP* (+PCA)	41.46±2.49	68.33±1.01	84.99±0.52	90.71±0.38	93.03±0.43	49.00±1.00	73.80±1.69	6.82±0.33
10	*ET + EP + HAOG*	65.92±4.32	86.15±1.28	92.95±0.63	96.19±0.14	97.60±0.14	77.28±1.69	91.13±0.80	3.48±0.25
11	*LGMS*	86.15±0.94	95.52±0.75	98.84±0.14	99.50±0.25	99.67±0.14	94.36±0.63	98.84±0.38	1.09±0.35
12	*LGMS + EP*	90.46±1.15	97.84±0.72	99.59±0.38	99.67±0.29	99.75±0.25	97.93±0.38	99.59±0.38	0.58±0.14
13	*DEEPS*	82.92±0.38	96.68±0.94	99.34±0.38	99.75±0.00	**100.00±0.00**	94.03±1.32	99.34±0.38	0.84±0.13
14	*DEEPS + EP*	89.88±0.63	98.51±0.43	99.59±0.14	**100.00±0.00**	**100.00±0.00**	98.01±0.25	**99.92±0.14**	0.44±0.25
15	*LGMS + DEEPS*	95.52±0.50	**99.50±0.25**	**99.83±0.14**	**100.00±0.00**	**100.00±0.00**	**99.59±0.29**	**100.00±0.00**	0.21±0.03
16	*LGMS + DEEPS + EP*	**97.43±0.38**	**99.67±0.38**	**99.83±0.14**	**100.00±0.00**	**100.00±0.00**	**99.67±0.38**	**100.00±0.00**	**0.11±0.12**

Table B.4: Means and standard deviations over 3 train/test-set splits for algorithms evaluated on viewed hand-drawn sketches, with race demographic filtering. Methods proposed in this research are shown in italics.

#	Method	Matching Rate (%) at Rank-X					TAR@FAR=0.1%	TAR@FAR=1.0%	EER (%)
		X=1	X=10	X=50	X=100	X=150			
1	VGG-Face [93]	38.81±2.60	63.43±1.39	80.68±1.37	87.23±1.46	89.97±0.87	38.47±2.66	66.17±1.39	7.98±0.64
2	PCA [37]	4.64±0.14	14.18±1.94	33.00±0.38	43.03±2.93	48.67±2.49	4.39±0.52	10.86±0.63	20.81±0.37
3	ET (+PCA) [17]	37.40±3.00	69.32±2.45	86.15±1.90	92.12±1.12	94.86±0.57	43.86±3.13	72.22±2.44	6.39±0.34
4	LLE (+PCA) [159]	47.35±1.50	74.46±1.04	88.64±1.28	93.12±0.63	95.27±0.25	54.06±1.23	74.88±1.79	7.10±0.33
5	HAOG [10]	61.44±2.83	78.36±1.32	89.88±0.38	93.78±0.25	95.44±0.38	69.15±1.24	86.07±0.75	4.95±0.31
6	CBR [6]	30.68±1.15	61.11±1.23	81.01±1.37	87.65±1.15	91.79±0.66	37.48±3.07	67.58±1.80	7.24±0.25
7	D-RS [12,69]	81.76±1.83	93.45±1.28	97.60±0.14	98.84±0.57	99.09±0.63	89.47±1.87	97.18±0.38	1.83±0.15
8	D-RS+CBR [19]	86.65±0.94	94.61±0.57	98.42±0.52	98.92±0.76	99.09±0.63	93.28±0.90	98.67±0.76	1.25±0.49
9	*EP* (+PCA)	45.77±1.88	73.88±0.43	88.97±1.87	93.12±0.94	95.27±0.86	53.48±1.32	78.77±0.14	5.39±0.37
10	*ET + EP + HAOG*	69.07±2.99	88.89±0.87	94.69±1.01	97.01±0.66	97.84±0.52	79.10±2.40	93.03±0.66	2.96±0.15
11	*LGMS*	88.31±1.39	96.60±1.04	99.00±0.00	99.34±0.29	**99.83±0.14**	95.11±1.12	99.00±0.50	0.99±0.24
12	*LGMS + EP*	91.87±2.35	98.26±0.43	99.50±0.25	99.75±0.25	**99.92±0.14**	98.34±0.29	99.75±0.25	0.58±0.14
13	*DEEPS*	84.58±1.97	98.01±0.66	99.67±0.38	**100.00±0.00**	**100.00±0.00**	95.11±1.69	99.83±0.29	0.70±0.12
14	*DEEPS + EP*	91.21±1.15	99.25±0.43	**99.92±0.14**	**100.00±0.00**	**100.00±0.00**	98.34±0.52	**99.92±0.14**	0.33±0.15
15	*LGMS + DEEPS*	96.85±1.44	**99.50±0.00**	**99.92±0.14**	**99.92±0.14**	**99.92±0.14**	**99.75±0.25**	**100.00±0.00**	**0.12±0.11**
16	*LGMS + DEEPS + EP*	**97.43±0.87**	**99.59±0.38**	**99.92±0.14**	**99.92±0.14**	**100.00±0.00**	**99.92±0.14**	**100.00±0.00**	**0.10±0.13**

Table B.5: Means and standard deviations over 3 train/test-set splits for algorithms evaluated on viewed hand-drawn sketches, with gender and race demographic filtering. Methods proposed in this research are shown in italics.

#	Method	Matching Rate (%) at Rank-X					TAR@FAR=0.1%	TAR@FAR=1.0%	EER (%)
		X=1	X=10	X=50	X=100	X=150			
1	VGG-Face [93]	45.02±1.32	69.82±2.45	86.15±1.01	92.21±0.29	95.52±0.66	44.69±2.75	74.46±1.12	5.62±0.71
2	PCA [37]	6.55±0.14	20.15±1.99	43.70±2.80	54.89±3.38	65.59±2.61	5.56±1.01	14.51±0.52	12.65±0.25
3	ET (+PCA) [17]	44.28±1.88	76.78±1.83	91.96±1.01	96.10±0.63	97.26±0.50	51.58±2.14	80.51±2.17	4.64±0.15
4	LLE (+PCA) [159]	54.06±2.80	80.43±0.57	93.03±0.66	96.02±0.25	97.84±0.57	58.46±1.55	81.51±2.00	5.13±0.11
5	HAOG [10]	65.51±1.80	83.91±0.14	93.78±0.00	96.19±0.38	98.09±0.38	74.79±1.37	90.22±0.14	3.75±0.04
6	CBR [6]	38.31±0.25	70.40±1.55	89.05±1.24	93.62±0.76	95.27±0.66	44.44±1.88	75.46±1.04	5.13±0.30
7	D-RS [12,69]	85.49±2.26	95.27±0.50	98.59±0.14	99.09±0.52	99.50±0.50	91.79±2.28	98.18±0.14	1.34±0.18
8	D-RS+CBR [19]	89.05±0.50	96.35±0.80	98.92±0.76	99.42±0.38	99.59±0.14	95.19±0.87	99.00±0.75	1.01±0.52
9	*EP* (+PCA)	52.82±0.14	80.43±0.38	93.20±0.76	96.43±0.57	97.68±0.29	59.04±1.66	85.32±0.25	3.82±0.14
10	*ET + EP + HAOG*	74.38±2.99	92.04±0.50	96.93±0.52	98.34±0.29	99.17±0.14	83.33±2.37	95.69±0.72	2.16±0.15
11	*LGMS*	90.55±1.39	97.60±0.38	99.42±0.29	**99.83±0.29**	**99.92±0.14**	96.60±0.76	99.50±0.00	0.67±0.15
12	*LGMS + EP*	93.86±1.28	99.00±0.25	99.59±0.38	**99.83±0.29**	**99.92±0.14**	98.67±0.38	99.75±0.25	0.42±0.12
13	*DEEPS*	87.81±2.04	99.00±0.25	99.83±0.29	**100.00±0.00**	**100.00±0.00**	96.77±1.51	**99.92±0.14**	0.49±0.04
14	*DEEPS + EP*	93.78±0.90	**99.59±0.14**	**99.92±0.14**	**100.00±0.00**	**100.00±0.00**	99.17±0.38	**99.92±0.14**	0.23±0.04
15	*LGMS + DEEPS*	97.68±1.12	**99.59±0.14**	**99.92±0.14**	**100.00±0.00**	**100.00±0.00**	**99.92±0.14**	**100.00±0.00**	**0.10±0.12**
16	*LGMS + DEEPS + EP*	**98.26±0.43**	**99.67±0.38**	**100.00±0.00**	**100.00±0.00**	**100.00±0.00**	**99.92±0.14**	**100.00±0.00**	**0.10±0.13**

Table B.6: Means and standard deviations over 3 train/test-set splits for algorithms evaluated on viewed UoM-SGFS Set A software-generated sketches, with gender demographic filtering. Methods proposed in this research are shown in italics.

#	Method	Matching Rate (%) at Rank-X					TAR@FAR=0.1%	TAR@FAR=1.0%	EER (%)
		X=1	X=10	X=50	X=100	X=150			
1	VGG-Face [93]	10.53±2.47	35.87±3.31	66.93±3.11	80.80±3.25	88.00±2.71	14.27±3.11	44.40±3.11	10.92±0.76
2	PCA [37]	3.20±1.59	13.33±1.89	30.00±3.16	39.73±2.73	45.60±3.55	5.47±2.08	12.00±0.82	25.69±1.08
3	ET (+PCA) [17]	10.40±2.19	37.33±3.50	65.60±6.34	77.33±4.29	84.27±2.29	14.80±2.60	42.40±4.04	11.70±1.50
4	LLE (+PCA) [159]	8.80±2.18	29.20±2.72	55.07±1.86	70.53±2.51	76.53±2.18	12.40±1.74	30.40±4.41	17.11±1.42
5	HAOG [10]	16.40±3.00	41.73±1.67	59.87±2.38	66.80±1.97	72.27±1.01	20.53±3.00	46.80±5.22	17.45±1.50
6	CBR [6]	7.20±2.64	28.80±2.80	52.67±1.25	63.87±2.18	72.40±1.92	9.47±2.60	29.73±1.67	15.92±0.91
7	D-RS [12,69]	25.33±2.11	56.27±4.21	77.07±3.32	86.67±2.75	91.47±1.97	33.33±4.29	61.60±4.07	8.68±0.55
8	D-RS+CBR [19]	29.87±4.31	62.80±3.75	83.73±2.77	91.47±2.64	95.20±2.08	37.47±2.38	69.33±4.50	6.78±0.84
9	*EP* (+PCA)	14.53±1.91	45.47±3.25	72.93±3.88	82.67±3.50	88.80±2.88	19.60±2.14	47.33±1.56	10.72±1.37
10	*ET + EP + HAOG*	18.93±2.24	52.27±3.22	73.87±3.66	82.80±2.88	88.80±2.76	25.07±4.18	55.87±4.98	10.12±0.90
11	*LGMS*	26.13±4.51	57.20±5.30	81.07±2.14	86.67±2.26	90.40±1.67	33.07±6.78	63.87±4.91	8.75±0.96
12	*LGMS + EP*	34.53±2.92	72.80±4.98	88.27±2.56	93.33±1.49	95.73±0.89	46.40±4.15	77.20±3.87	6.42±0.59
13	*DEEPS*	33.07±1.67	69.73±2.43	89.60±1.21	95.73±2.24	98.27±1.12	43.60±3.22	75.73±1.98	5.10±0.60
14	*DEEPS + EP*	47.33±1.94	81.07±2.48	95.20±1.79	**98.80±1.10**	**99.47±0.56**	62.93±3.58	88.53±2.33	3.60±1.09
15	*LGMS + DEEPS*	48.93±2.69	82.80±3.69	94.27±1.86	97.20±0.87	98.40±1.30	62.40±4.28	87.73±3.61	3.72±0.61
16	*LGMS + DEEPS + EP*	**58.00±3.74**	**86.00±2.26**	**96.53±1.10**	**98.67±1.33**	**99.60±0.37**	**72.67±2.11**	**91.47±1.45**	**2.91±0.69**

Table B.7: Means and standard deviations over 3 train/test-set splits for algorithms evaluated on viewed UoM-SGFS Set A software-generated sketches, with race demographic filtering. Methods proposed in this research are shown in italics.

#	Method	Matching Rate (%) at Rank-X					TAR@FAR=0.1%	TAR@FAR=1.0%	EER (%)
		X=1	X=10	X=50	X=100	X=150			
1	VGG-Face [93]	12.00±3.23	39.20±4.91	69.60±3.52	82.80±3.54	88.40±3.15	13.73±3.22	42.13±5.86	11.07±0.79
2	PCA [37]	5.07±1.21	16.00±2.05	31.87±3.60	52.27±5.30	64.80±5.00	5.73±1.98	13.33±1.70	22.17±1.59
3	ET (+PCA) [17]	13.60±1.67	44.00±1.05	72.27±1.74	84.67±1.25	88.93±2.14	16.00±1.56	46.67±3.09	10.30±1.00
4	LLE (+PCA) [159]	12.80±3.21	37.73±1.80	62.93±3.42	76.67±2.11	83.07±1.01	13.33±2.05	34.80±2.02	14.74±2.37
5	HAOG [10]	20.27±2.85	47.87±1.45	66.93±1.98	75.47±2.08	80.67±2.40	22.80±1.52	50.00±3.43	15.28±1.01
6	CBR [6]	11.47±2.08	30.67±1.94	62.40±3.93	73.07±3.35	77.47±2.72	9.87±2.38	33.33±2.16	13.86±0.88
7	D-RS [12,69]	30.00±2.26	63.87±3.54	81.73±2.77	88.93±2.69	91.87±2.56	35.07±4.61	64.93±5.22	7.56±0.81
8	D-RS+CBR [19]	35.20±3.96	69.20±3.75	86.67±1.33	92.40±1.53	94.80±0.99	39.47±2.08	71.47±4.04	5.91±0.80
9	*EP* (+PCA)	18.13±3.11	51.47±3.57	78.53±2.02	88.00±2.31	91.87±2.23	21.33±1.05	52.13±1.73	9.79±0.96
10	*ET + EP + HAOG*	23.33±1.89	55.73±1.01	80.40±2.81	87.60±1.53	90.93±0.76	27.07±4.36	61.07±2.85	8.70±0.96
11	*LGMS*	32.00±6.45	62.93±4.68	84.13±2.08	90.27±2.39	92.67±1.63	35.20±5.99	65.60±4.44	7.96±0.89
12	*LGMS + EP*	40.40±5.51	76.40±3.29	91.07±2.77	95.33±1.25	96.80±1.45	50.27±2.77	79.47±3.38	5.88±0.51
13	*DEEPS*	36.13±1.59	74.00±1.83	92.00±1.25	97.47±1.10	98.93±0.37	46.27±1.80	79.07±1.12	4.56±0.23
14	*DEEPS + EP*	51.87±2.76	86.53±2.23	97.07±1.38	**99.20±1.10**	**99.60±0.37**	64.53±3.44	90.67±0.94	3.19±0.55
15	*LGMS + DEEPS*	51.33±4.27	85.33±2.54	96.67±0.82	98.67±0.82	99.33±0.47	64.93±5.11	89.33±1.94	3.36±0.42
16	*LGMS + DEEPS + EP*	**60.27±4.75**	**89.20±3.38**	**98.40±0.76**	**99.73±0.60**	**100.00±0.00**	**75.20±3.00**	**92.67±1.25**	**2.66±0.47**

Table B.8: Means and standard deviations over 3 train/test-set splits for algorithms evaluated on viewed UoM-SGFS Set A software-generated sketches, with gender and race demographic filtering. Methods proposed in this research are shown in italics.

#	Method	Matching Rate (%) at Rank-X					TAR@FAR=0.1%	TAR@FAR=1.0%	EER (%)
		X=1	X=10	X=50	X=100	X=150			
1	VGG-Face [93]	14.27±3.64	45.20±4.51	78.93±3.45	89.20±3.25	93.33±2.21	15.60±3.00	48.67±4.71	8.13±0.60
2	PCA [37]	6.67±1.49	24.80±2.96	54.00±3.37	71.07±3.42	78.27±3.61	6.80±1.45	19.47±2.56	13.29±1.29
3	ET (+PCA) [17]	16.93±2.93	53.60±2.39	82.93±3.22	92.00±1.49	94.93±1.38	20.00±2.05	56.00±3.02	7.30±1.00
4	LLE (+PCA) [159]	16.80±2.18	47.20±2.96	76.80±2.80	86.80±1.66	93.07±1.30	16.40±2.14	43.33±2.45	9.78±1.05
5	HAOG [10]	23.60±3.39	54.40±1.80	74.93±2.14	82.53±3.07	86.80±2.88	27.07±2.61	56.27±1.92	10.13±0.35
6	CBR [6]	14.80±1.28	41.07±1.86	72.80±3.11	81.20±2.02	87.07±2.61	11.73±1.92	40.80±3.84	9.29±0.49
7	D-RS [12,69]	34.00±3.59	71.07±2.56	88.80±3.51	94.53±1.97	96.67±1.49	40.80±3.93	72.67±4.00	5.45±0.93
8	D-RS+CBR [19]	40.53±4.89	77.07±3.11	92.80±2.60	96.40±1.92	98.27±0.76	44.27±3.18	78.67±3.50	4.25±0.81
9	*EP* (+PCA)	21.60±2.56	62.53±1.73	85.87±2.18	93.07±2.52	95.47±1.73	25.47±2.18	61.33±0.82	7.21±0.56
10	*ET + EP + HAOG*	26.80±2.18	65.20±1.79	86.00±2.21	92.93±1.01	96.40±1.46	32.40±3.67	71.07±2.89	6.57±0.53
11	*LGMS*	36.80±6.52	70.13±6.24	89.73±2.85	94.80±0.87	96.67±1.05	40.67±5.23	74.67±3.71	5.71±0.34
12	*LGMS + EP*	46.13±6.24	82.53±3.96	94.93±1.30	98.53±0.30	99.20±0.73	56.67±3.74	84.80±3.41	4.39±0.33
13	*DEEPS*	38.00±1.15	78.40±0.37	94.40±0.76	99.07±0.76	**99.73±0.60**	48.67±3.40	82.40±1.80	3.82±0.25
14	*DEEPS + EP*	54.53±2.23	88.80±1.59	98.13±0.99	**99.60±0.60**	**100.00±0.00**	67.07±3.55	92.27±0.76	2.77±0.34
15	*LGMS + DEEPS*	54.27±4.07	88.13±3.11	98.40±0.76	99.07±0.60	**99.87±0.30**	68.67±3.53	91.07±2.09	**2.56±0.52**
16	*LGMS + DEEPS + EP*	**63.47±4.84**	**90.93±2.89**	**99.07±1.01**	**99.87±0.30**	**100.00±0.00**	**77.47±3.44**	**94.40±1.38**	**2.20±0.69**

Table B.9: Means and standard deviations over 3 train/test-set splits for algorithms evaluated on viewed UoM-SGFS Set B software-generated sketches, with gender demographic filtering. Methods proposed in this research are shown in italics.

#	Method	Matching Rate (%) at Rank-X					TAR@FAR=0.1%	TAR@FAR=1.0%	EER (%)
		X=1	X=10	X=50	X=100	X=150			
1	VGG-Face [93]	19.73±2.52	52.27±3.58	78.93±1.30	89.20±2.72	94.27±2.77	26.40±2.61	59.73±3.67	7.71±0.96
2	PCA [37]	5.73±2.14	14.80±2.60	30.00±3.33	40.53±1.66	46.40±2.56	6.93±1.53	14.27±1.53	25.21±1.10
3	ET (+PCA) [17]	14.00±0.82	45.73±6.97	71.73±2.65	81.87±2.64	87.47±3.03	19.87±2.72	49.33±5.35	10.09±1.01
4	LLE (+PCA) [159]	12.67±1.05	37.07±1.74	64.27±3.90	75.60±4.58	83.07±3.08	14.40±1.38	38.67±1.89	14.26±0.81
5	HAOG [10]	23.07±3.45	47.20±3.07	63.87±2.96	70.40±2.29	74.67±1.63	27.47±3.11	52.00±2.21	16.18±1.20
6	CBR [6]	11.33±2.26	33.47±1.52	57.73±1.80	68.00±2.11	76.80±1.73	13.73±3.70	36.27±1.53	14.99±1.04
7	D-RS [12,69]	44.13±1.28	75.47±1.85	91.33±1.56	96.53±1.10	97.47±0.99	55.73±3.15	82.13±2.84	4.67±0.43
8	D-RS+CBR [19]	48.67±2.62	80.13±2.28	95.07±0.89	97.73±0.76	98.80±0.87	61.60±1.74	86.93±1.92	3.49±0.73
9	*EP* (+PCA)	19.07±0.76	53.33±4.14	78.40±3.04	87.87±2.76	91.33±2.05	27.07±1.86	58.13±1.91	9.34±0.92
10	*ET + EP + HAOG*	25.87±3.69	61.47±3.81	78.00±3.27	86.40±2.65	90.67±2.21	33.87±4.04	64.27±5.22	8.56±1.27
11	*LGMS*	45.33±4.50	76.67±2.54	89.73±1.21	95.20±1.85	97.20±1.52	56.13±2.84	83.07±2.24	5.17±0.54
12	*LGMS + EP*	54.80±3.18	85.33±1.70	95.07±0.89	97.60±0.60	98.40±0.37	69.33±2.21	89.20±1.59	3.77±0.33
13	*DEEPS*	52.93±3.45	84.40±2.93	96.13±1.45	98.80±0.56	98.93±0.60	66.80±5.24	89.73±1.98	3.53±0.89
14	*DEEPS + EP*	64.93±2.03	91.33±2.26	97.87±0.87	**99.73±0.37**	**99.87±0.30**	81.33±2.16	95.60±0.76	2.40±0.38
15	*LGMS + DEEPS*	70.13±2.23	92.80±1.85	**99.47±0.87**	**100.00±0.00**	**100.00±0.00**	84.53±2.38	97.33±1.15	**1.49±0.26**
16	*LGMS + DEEPS + EP*	**74.80±3.75**	**94.67±1.41**	**99.60±0.37**	**100.00±0.00**	**100.00±0.00**	**88.93±1.67**	**98.27±0.60**	**1.42±0.24**

Table B.10: Means and standard deviations over 3 train/test-set splits for algorithms evaluated on viewed UoM-SGFS Set B software-generated sketches, with race demographic filtering. Methods proposed in this research are shown in italics.

#	Method	Matching Rate (%) at Rank-X					TAR@FAR=0.1%	TAR@FAR=1.0%	EER (%)
		X=1	X=10	X=50	X=100	X=150			
1	VGG-Face [93]	20.67±3.86	55.60±3.67	80.67±3.16	89.47±1.97	93.73±2.29	26.53±2.47	58.80±4.51	7.88±0.81
2	PCA [37]	8.00±1.94	17.87±2.08	32.67±3.23	53.33±5.01	65.07±4.91	7.07±1.53	15.20±2.47	21.78±1.54
3	ET (+PCA) [17]	18.13±2.23	53.07±2.85	80.40±2.48	87.73±2.14	92.13±2.23	22.40±2.85	55.87±2.33	8.79±1.13
4	LLE (+PCA) [159]	16.53±2.47	46.13±1.66	72.67±3.13	84.40±3.18	90.27±1.53	16.27±1.61	43.73±2.77	11.96±1.10
5	HAOG [10]	26.00±3.77	51.87±0.99	70.80±1.52	77.07±2.39	82.93±3.15	30.00±2.45	56.27±2.24	13.96±0.99
6	CBR [6]	13.87±3.25	36.27±3.15	67.47±3.28	77.20±1.66	81.47±1.73	14.80±3.69	40.13±0.87	12.47±0.83
7	D-RS [12,69]	50.13±2.76	81.07±2.85	93.33±1.56	97.07±0.60	98.40±1.12	58.80±2.56	84.80±2.47	4.12±0.54
8	D-RS+CBR [19]	54.27±2.52	85.07±2.39	94.93±1.01	98.00±1.33	98.67±0.67	62.93±2.24	89.87±1.97	3.08±0.94
9	*EP* (+PCA)	24.27±2.69	60.80±1.28	84.80±1.10	90.93±1.92	93.20±1.73	30.53±2.60	63.60±2.39	8.33±0.83
10	*ET + EP + HAOG*	32.40±2.85	66.40±3.04	83.73±2.61	90.80±1.85	93.33±0.47	39.47±4.68	69.87±2.88	7.74±0.77
11	*LGMS*	48.80±3.38	78.13±5.11	92.40±2.77	96.27±1.92	98.00±1.05	57.20±2.96	83.60±2.61	4.78±0.52
12	*LGMS + EP*	59.20±3.11	88.13±2.47	96.40±0.76	98.40±0.37	99.20±0.73	72.00±2.91	89.87±0.87	3.45±0.22
13	*DEEPS*	56.67±4.64	86.67±3.23	96.53±1.37	98.67±0.47	99.47±0.30	68.67±5.06	91.07±1.80	3.04±0.95
14	*DEEPS + EP*	68.00±2.00	94.40±1.01	98.53±0.87	99.73±0.37	99.87±0.30	83.47±2.08	96.13±0.56	2.25±0.50
15	*LGMS + DEEPS*	71.87±2.38	94.93±1.21	**99.73±0.37**	**100.00±0.00**	**100.00±0.00**	86.13±3.41	98.00±1.05	**1.27±0.27**
16	*LGMS + DEEPS + EP*	**78.27±2.29**	**97.33±0.82**	**99.87±0.30**	**100.00±0.00**	**100.00±0.00**	**90.67±1.83**	**98.67±0.47**	**1.21±0.24**

Table B.11: Means and standard deviations over 3 train/test-set splits for algorithms evaluated on viewed UoM-SGFS Set B software-generated sketches, with gender and race demographic filtering. Methods proposed in this research are shown in italics.

#	Method	Matching Rate (%) at Rank-X					TAR@FAR=0.1%	TAR@FAR=1.0%	EER (%)
		X=1	X=10	X=50	X=100	X=150			
1	VGG-Face [93]	24.67±3.86	60.53±4.36	85.73±2.03	95.07±2.81	96.40±2.65	28.53±2.42	64.27±3.52	6.15±0.60
2	PCA [37]	8.67±2.05	25.60±2.39	54.93±3.61	71.33±3.89	77.07±3.29	8.13±0.99	21.60±2.85	13.15±1.21
3	ET (+PCA) [17]	20.67±3.09	62.00±3.86	87.73±2.93	93.47±1.59	97.20±0.73	26.40±3.99	65.20±1.52	6.13±0.59
4	LLE (+PCA) [159]	20.00±2.91	54.67±3.59	82.80±3.69	91.87±0.56	95.20±0.56	19.73±1.67	51.33±3.16	8.29±0.57
5	HAOG [10]	28.80±2.64	58.00±1.05	76.80±2.33	83.47±2.80	89.07±1.80	33.87±2.08	62.27±2.34	9.36±0.64
6	CBR [6]	19.07±2.85	48.00±2.11	74.53±2.18	84.53±1.28	90.13±2.02	18.13±2.33	48.00±2.49	8.70±0.67
7	D-RS [12,69]	53.47±3.38	86.93±2.14	96.80±0.87	98.40±1.01	99.60±0.60	64.00±1.63	90.53±1.85	2.79±0.49
8	D-RS+CBR [19]	60.40±4.18	88.93±1.46	97.47±0.56	99.20±0.56	99.87±0.30	68.80±3.60	93.73±2.14	2.15±0.72
9	*EP* (+PCA)	28.53±3.18	68.00±2.05	90.00±1.25	94.80±2.02	96.53±1.19	33.47±3.60	70.40±1.12	6.12±0.71
10	*ET + EP + HAOG*	35.47±1.97	73.73±2.29	88.67±1.70	94.27±0.76	96.27±1.01	43.47±3.00	76.53±3.14	5.74±0.30
11	*LGMS*	54.00±4.45	83.47±3.38	95.73±1.61	98.67±0.82	99.33±0.47	62.13±3.54	87.47±2.72	3.44±0.56
12	*LGMS + EP*	63.60±2.77	91.33±2.16	98.13±0.56	99.20±0.56	99.33±0.67	76.40±2.89	93.07±0.37	2.51±0.23
13	*DEEPS*	58.80±4.04	87.87±2.80	97.73±0.76	99.47±0.56	99.87±0.30	71.60±4.98	92.13±1.79	2.69±0.68
14	*DEEPS + EP*	69.87±2.08	95.47±1.28	**99.33±0.67**	**99.87±0.30**	**99.87±0.30**	85.60±2.73	96.93±0.76	1.86±0.53
15	*LGMS + DEEPS*	73.20±2.51	96.53±1.52	**100.00±0.00**	**100.00±0.00**	**100.00±0.00**	86.80±3.35	**98.53±0.99**	**1.14±0.27**
16	*LGMS + DEEPS + EP*	**79.73±2.56**	**98.00±0.82**	**100.00±0.00**	**100.00±0.00**	**100.00±0.00**	**91.47±1.19**	**98.93±0.76**	**1.04±0.33**

B.3 Effect of the extended gallery

As shown in Figure B.10, algorithm performance typically decreases with an increasing number of subjects in the extended gallery [6,9]. However, an extended gallery may not make the identification task more challenging if the characteristics of the photo images in the photo-sketch databases are significantly different from the photos used to extend the gallery. In such a case, it would be easy for any algorithm to assign low similarity scores to the photos which do not have a sketch counterpart. To determine the effect on algorithm performance of the extended gallery employed in this research, which was used to evaluate all approaches in Chapter 6, the performance of the same algorithms is evaluated with the extended gallery removed. Hence, the gallery is populated using only the photo pairs of the sketches used for testing.

Comparing Table 6.1 with Table B.12, it is evident that the use of an extended gallery negatively affects performance for all algorithms when operating on the viewed hand-drawn photo-sketch pairs. Algorithms employing a training stage tend to be more greatly affected with the use of an extended gallery, likely as a result of capturing characteristics which are not present in the photos of the extended gallery. However, algorithms which do not utilise training process can still be negatively

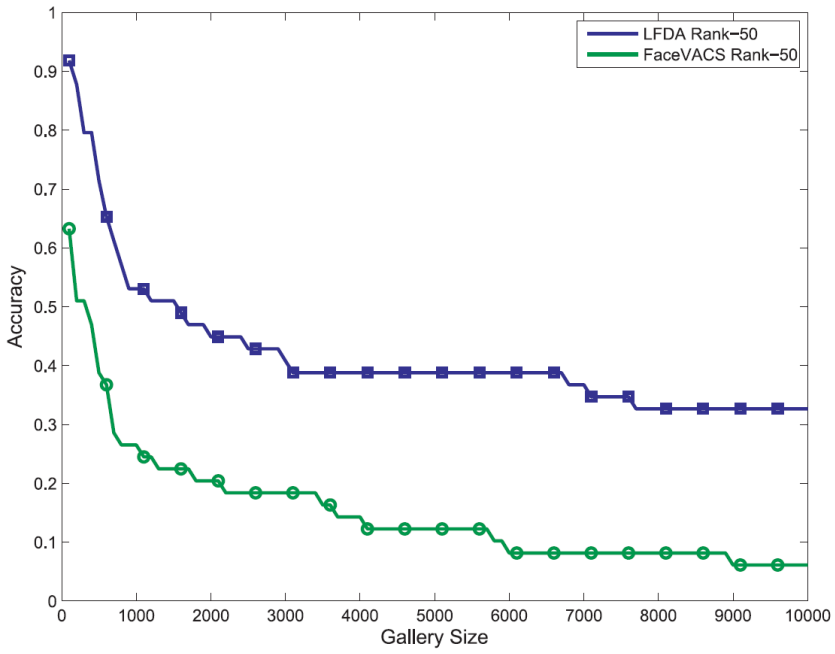


Figure B.10: Effect of the extended gallery for a commercial FRS and a face photo-sketch recognition algorithm [9]

affected. This is demonstrated by VGG-Face, which is employed by utilising the default model trained on face photos and exhibits loss of approximately 10-20% in terms of rank retrieval rates. The proposed LGMS+DEEPS methods appear relatively robust to the additional photos in the extended gallery, which exhibits a loss in performance of less than 2% at Rank-1 and even lower decreases across other ranks.

Similar observations can be obtained from comparing Table 6.2 with Table B.13 and Table 6.3 with Table B.14, for the case of the viewed software-generated sketches, although the differences in performance of all algorithms is noticeably higher than in the case of the hand-drawn sketches. This is likely a result of the relative sizes of the original and extended galleries: when algorithms operate on the hand-drawn sketches, the gallery size is increased with approximately 4.8 times as many subjects, from 402 subjects to 1923 subjects; in the case of the software-generated sketches, the gallery size is increased from 150 subjects to 1671 subjects, representing an increase of approximately 11.1 times as many subjects. Hence, the relative increase of the gallery in the case of the software-generated sketches is roughly twice more than the hand-drawn sketch case, making the identification task harder since algorithms must search through a relatively higher amount of subjects.

Finally, algorithms also perform noticeably worse with an extended gallery when operating on forensic sketches as observed when comparing Table 6.5 with Table B.15, also due to the relative gallery size difference as described above which is even more significant due to the low number of forensic sketches (gallery is increased from 47 subjects to 1568 subjects, i.e. 33.4 times as large). Moreover, certain trends also differ; most significant is the performance of the intra-modality methods, which perform better than the PCA face recogniser when no extended gallery is used whereas they perform worse with the extended gallery. This highlights the importance of using an evaluation set-up reflecting real-world conditions, i.e. the use of challenging sketches and an extended gallery.

Table B.12: Means and standard deviations over 3 train/test-set splits for algorithms evaluated on viewed hand-drawn sketches, without the extended gallery. Methods proposed in this research are shown in italics.

#	Method	Matching Rate (%) at Rank-X					TAR@FAR=0.1%	TAR@FAR=1.0%	EER (%)
		X=1	X=10	X=50	X=100	X=150			
1	VGG-Face [93]	34.66±1.83	64.59±1.44	86.90±1.60	93.95±1.28	96.60±1.15	34.66±1.83	55.47±0.86	12.06±0.69
2	PCA [37]	3.40±0.29	11.36±1.37	25.54±2.26	34.99±1.87	46.19±1.76	0.00±0.00	6.30±0.63	43.45±0.40
3	ET (+PCA) [17]	29.77±4.91	64.93±2.28	86.65±1.15	94.78±0.75	97.18±0.14	30.60±4.56	52.57±2.36	11.76±0.46
4	LLE (+PCA) [159]	41.04±1.14	70.56±0.72	88.81±0.90	94.36±0.76	96.35±0.14	41.63±1.15	59.87±1.37	11.38±0.58
5	HAOG [10]	56.72±2.62	74.30±0.80	87.40±0.63	92.45±0.63	95.27±0.75	57.05±2.61	72.14±1.51	10.25±0.36
6	CBR [6]	22.89±2.87	52.82±0.14	75.79±1.37	84.74±0.63	91.38±0.63	22.89±2.87	42.45±3.49	18.46±0.37
7	D-RS [12,69]	83.58±0.25	95.19±0.29	98.59±0.57	99.17±0.52	99.75±0.25	87.81±1.39	96.68±0.57	2.06±0.40
8	D-RS+CBR [19]	84.25±1.28	94.78±0.66	98.76±0.90	99.34±0.38	99.42±0.38	87.56±1.39	95.85±0.94	2.18±0.65
9	*EP* (+PCA)	40.38±2.74	72.14±1.00	90.30±0.25	95.36±0.57	97.68±0.29	41.13±2.88	62.60±1.80	10.41±0.44
10	*ET + EP + HAOG*	61.19±3.45	84.58±1.14	94.20±0.29	97.51±0.25	98.76±0.25	62.02±3.57	81.67±2.23	6.24±0.38
11	*LGMS*	89.14±0.87	97.68±0.38	**99.75±0.25**	**99.92±0.14**	**99.92±0.14**	93.45±0.80	98.59±0.29	1.10±0.12
12	*LGMS + EP*	92.12±1.28	99.00±0.75	**99.83±0.29**	**99.92±0.14**	**99.92±0.14**	96.68±0.57	99.50±0.25	0.70±0.18
13	*DEEPS*	85.07±1.29	98.92±0.52	**99.83±0.29**	**100.00±0.00**	**100.00±0.00**	91.29±1.24	99.17±0.38	0.97±0.23
14	*DEEPS + EP*	90.22±0.87	99.59±0.14	**99.92±0.14**	**100.00±0.00**	**100.00±0.00**	95.77±0.66	**99.92±0.14**	0.58±0.13
15	*LGMS + DEEPS*	96.85±0.76	**99.83±0.14**	**99.92±0.14**	**100.00±0.00**	**100.00±0.00**	**99.67±0.38**	**99.92±0.14**	**0.14±0.15**
16	*LGMS + DEEPS + EP*	**97.51±0.90**	**99.92±0.14**	**100.00±0.00**	**100.00±0.00**	**100.00±0.00**	**99.75±0.25**	**99.92±0.14**	**0.16±0.10**

Table B.13: Means and standard deviations over 5 train/test-set splits for algorithms evaluated on UoM-SGFS Set A sketches, without the extended gallery. Rank-150 results not shown due to all algorithms achieving a rate of 100% (since the gallery contains 150 subjects). TAR@FAR=0.1% results also omitted due to inaccurate results as a consequence of a low number of subjects.

#	Method	Matching Rate (%) at Rank-X				TAR@FAR=1.0%	EER (%)
		X=1	X=10	X=50	X=100		
1	VGG-Face [93]	26.53±3.57	70.93±1.53	93.47±1.28	98.80±0.56	35.47±1.91	14.74±1.50
2	PCA [37]	5.20±1.97	17.33±1.33	46.00±1.25	75.87±1.45	6.67±2.26	43.44±0.99
3	ET (+PCA) [17]	13.33±2.16	44.40±1.80	83.20±2.51	95.07±1.53	17.47±2.28	25.03±1.40
4	LLE (+PCA) [159]	15.07±3.04	41.60±3.22	79.07±2.85	94.00±0.67	18.93±2.61	28.79±1.40
5	HAOG [10]	16.67±3.09	42.80±1.52	69.20±2.72	89.20±2.38	20.53±2.84	30.81±1.62
6	CBR [6]	11.60±2.09	41.20±2.28	74.67±3.80	91.60±2.09	15.87±1.73	28.38±1.68
7	D-RS [12,69]	45.20±1.73	81.20±2.38	97.33±0.82	99.60±0.60	59.20±3.75	10.67±1.25
8	D-RS + CBR [19]	48.40±3.22	83.60±2.52	97.87±0.87	99.87±0.30	61.73±2.48	9.26±0.86
9	*EP* (+PCA)	17.20±2.72	55.20±1.97	88.80±1.79	95.47±1.37	22.53±2.64	20.58±1.74
10	*ET + EP + HAOG*	19.33±3.56	57.07±1.86	85.60±2.34	95.47±1.79	25.73±4.13	20.27±1.39
11	*LGMS*	41.33±3.65	80.13±4.38	95.20±0.56	99.20±0.56	56.00±3.97	11.77±1.06
12	*LGMS + EP*	46.93±4.82	85.87±1.91	97.73±0.60	**99.60±0.37**	62.93±4.56	10.28±1.02
13	*DEEPS*	58.13±3.87	93.87±1.45	**99.87±0.30**	**100.00±0.00**	75.07±2.39	4.90±0.36
14	*DEEPS + EP*	61.87±5.00	**96.40±1.98**	**100.00±0.00**	**100.00±0.00**	81.07±3.52	4.19±0.27
15	*LGMS + DEEPS*	69.33±1.25	**96.53±0.73**	**99.87±0.30**	**100.00±0.00**	**88.67±2.11**	3.61±0.43
16	*LGMS + DEEPS + EP*	**71.20±2.18**	**96.53±0.56**	**100.00±0.00**	**100.00±0.00**	**89.33±3.16**	**3.37±0.42**

Table B.14: Means and standard deviations over 5 train/test-set splits for algorithms evaluated on UoM-SGFS Set B sketches, without the extended gallery. Rank-150 results not shown due to all algorithms achieving a rate of 100% (since the gallery contains 150 subjects). TAR@FAR=0.1% results also omitted due to inaccurate results as a consequence of a low number of subjects.

#	Method	Matching Rate (%) at Rank-X				TAR@FAR=1.0%	EER (%)
		X=1	X=10	X=50	X=100		
1	VGG-Face [93]	40.93±0.89	81.33±1.49	97.73±1.21	99.87±0.30	54.80±2.68	10.21±1.15
2	PCA [37]	7.33±2.26	20.00±2.45	48.00±0.67	76.93±1.98	8.67±1.94	42.73±0.83
3	ET (+PCA) [17]	16.27±1.21	55.07±3.25	87.33±2.36	96.67±1.56	23.47±3.35	21.24±0.70
4	LLE (+PCA) [159]	18.00±3.02	53.33±2.62	87.07±2.77	96.27±1.01	23.33±2.05	23.52±1.34
5	HAOG [10]	24.00±3.97	47.60±1.86	72.00±2.26	91.20±2.23	27.07±4.07	28.51±1.94
6	CBR [6]	15.07±1.01	45.33±1.70	78.00±2.36	93.60±0.76	20.00±1.33	25.68±1.62
7	D-RS [12,69]	66.00±1.70	95.07±0.60	99.60±0.60	100.00±0.00	83.60±1.53	4.34±0.62
8	D-RS + CBR [19]	67.20±0.73	94.00±1.25	99.60±0.37	100.00±0.00	84.00±2.62	4.81±0.89
9	*EP* (+PCA)	22.93±2.34	63.60±3.55	91.87±2.02	97.07±1.12	31.60±1.98	17.86±1.06
10	*ET + EP + HAOG*	26.93±2.77	66.27±2.61	89.33±2.91	97.20±1.19	34.13±2.64	17.01±1.48
11	*LGMS*	62.80±2.28	90.93±1.12	98.40±0.37	**99.60±0.37**	77.87±3.60	6.68±0.76
12	*LGMS + EP*	67.33±2.87	93.47±0.99	98.67±0.47	**100.00±0.00**	83.20±2.08	5.25±0.27
13	*DEEPS*	73.20±2.38	97.60±1.01	**100.00±0.00**	**100.00±0.00**	90.13±1.85	3.43±0.27
14	*DEEPS + EP*	77.07±2.73	98.67±0.94	**100.00±0.00**	**100.00±0.00**	91.73±1.12	2.83±0.57
15	*LGMS + DEEPS*	**85.73±1.46**	**99.20±0.30**	**100.00±0.00**	**100.00±0.00**	**97.33±0.82**	**1.47±0.30**
16	*LGMS + DEEPS + EP*	84.93±1.12	**99.47±0.30**	**100.00±0.00**	**100.00±0.00**	**97.47±1.19**	**1.49±0.48**

Table B.15: Average values over all ranks of the 47 subjects in the PRIP-HDC [19] forensic sketch database, for the algorithms considered after averaging over 3 set splits, without the extended gallery.

#	Algorithm	Average rank
1	VGG-Face [93]	14.38
2	PCA [37]	21.66
3	ET (+PCA) [17]	19.34
4	LLE (+PCA) [34]	17.46
5	HAOG [10]	18.57
6	CBR [6]	19.57
7	D-RS [12,69]	16.52
8	CBR + D-RS [12]	16.19
9	*EP* (+PCA)	17.59
10	*ET + EP + HAOG*	16.96
11	*LGMS*	14.86
12	*LGMS + EP*	13.16
13	*DEEPS*	13.52
14	*DEEPS + EP*	13.38
15	*DEEPS + LGMS*	**11.79**
16	*DEEPS-M*	13.45
17	*DEEPS-M + LGMS*	**11.91**
18	*DEEPS + LGMS + EP*	**11.48**
19	*DEEPS-M + LGMS + EP*	**11.56**

B.4 Statistical Significance

The multi-comparison Analysis of Variance (ANOVA) results will now be presented, where -1/0/1 indicate that the algorithm in the row is statistically inferior/identical/superior to the algorithm in the column, respectively, at the 95% confidence level.

Since an algorithm in a row is deemed to be inferior with a value of -1, while it is superior with a value of 1 and statistically identical with a value of 0, then the computation of a simple summation of each row yields a compact representation of the statistical performance of each algorithm; a high value is desirable since it indicates a high number of '1's' as a result of an algorithm being statistically superior to a high number of other methods. These values are also shown in the results hereunder and are normalised by dividing the addition result with the maximal value that can be achieved by each algorithm (i.e. number of algorithms$-1 = 15$) such that the values lie within [-1, 1], where -1 indicates that the algorithm is inferior to all the other algorithms and 1 indicates that the algorithm is superior to all the other algorithms. These values are denoted by Normalised Sum (NS); more formally, the NS score of a method i is given by:

$$\mathrm{NS}_i = \frac{\sum_{m=1}^{M} S_m}{M-1} \tag{B.1}$$

where $M = 16$ is number of algorithms compared, and $S_m \in \{-1, 0, 1\}$ is the statistical significance result as determined by the multi-comparison ANOVA for algorithm m.

It should be noted that multiple comparison analysis tests are known to generally be conservative, such that Type I errors are reduced at the expense of higher Type II errors. A Type I error occurs when an algorithm A is deemed to be superior to another algorithm B, but algorithm B is actually better than A [171,172]. Future work can thus include the analysis of other multiple-comparison testing methods to determine better ways to evaluate the statistical significance of evaluation metrics used for face photo-sketch recognition.

From Tables B.16 to B.22 in the case of the viewed hand-drawn sketches, it can be observed that the intra-modality methods are generally statistically superior to the

face recognisers and the proposed EP is also statistically superior to ET across all performance measures. Moreover, LGMS and DEEPS are statistically superior to all algorithms proposed in literature, except D-RS+CBR in some cases. However, this is likely a result of the conservative nature of ANOVA given the clear advantage of both methods over D-RS+CBR. ANOVA also does not take into consideration that all algorithms were applied on the same training and tests sets, and thus the standard deviation values serve primarily as a measure of the consistency of algorithms rather than as a means of performance generalisation. These observations are particularly evident at higher ranks, where values approach 100% retrieval rate and are thus quite tightly clustered, such that multi-comparison ANOVA indicates that most algorithms are statistically identical. The proposed LGMS+DEEPS system (algorithm 15) and LGMS+DEEPS+EP system (algorithm 16) overall obtain the best performance of all algorithms considered.

Similar conclusions can be derived in the case of the software-generated sketches, with results shown in Tables B.23 to B.29 for the Set A sketches and in Tables B.30 to B.36 for the Set B sketches. However, the proposed LGMS+DEEPS system demonstrates a clearer advantage with respect to other methods on this type of sketches.

Table B.16: Multi-comparison ANOVA for Rank-1 retrieval rates when using viewed hand-drawn sketches. Corresponding algorithm names shown in Table 6.1.

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**NS**
1	0	1	0	-1	-1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-0.7333
2	-1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1.0000
3	0	1	0	-1	-1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-0.7333
4	1	1	1	0	-1	1	-1	-1	0	-1	-1	-1	-1	-1	-1	-1	-0.4000
5	1	1	1	1	0	1	-1	-1	1	0	-1	-1	-1	-1	-1	-1	-0.1333
6	0	1	0	-1	-1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-0.7333
7	1	1	1	1	1	1	0	0	1	1	-1	-1	0	-1	-1	-1	0.2000
8	1	1	1	1	1	1	0	0	1	1	0	-1	0	0	-1	-1	0.3333
9	1	1	1	0	-1	1	-1	-1	0	-1	-1	-1	-1	-1	-1	-1	-0.4000
10	1	1	1	1	0	1	-1	-1	1	0	-1	-1	-1	-1	-1	-1	-0.1333
11	1	1	1	1	1	1	1	0	1	1	0	0	0	0	-1	-1	0.4667
12	1	1	1	1	1	1	1	1	1	1	0	0	1	0	-1	-1	0.6000
13	1	1	1	1	1	1	0	0	1	1	0	-1	0	-1	-1	-1	0.2667
14	1	1	1	1	1	1	1	0	1	1	0	0	1	0	-1	-1	0.5333
15	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	**0.9333**
16	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	**0.9333**

Table B.17: Multi-comparison ANOVA for Rank-10 retrieval rates when using viewed hand-drawn sketches. Corresponding algorithm names shown in Table 6.1.

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**NS**
1	0	1	-1	-1	-1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-0.8000
2	-1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1.0000
3	1	1	0	-1	-1	1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-0.6000
4	1	1	1	0	-1	1	-1	-1	0	-1	-1	-1	-1	-1	-1	-1	-0.4000
5	1	1	1	1	0	1	-1	-1	1	-1	-1	-1	-1	-1	-1	-1	-0.2000
6	0	1	-1	-1	-1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-0.8000
7	1	1	1	1	1	1	0	0	1	1	-1	-1	-1	-1	-1	-1	0.1333
8	1	1	1	1	1	1	0	0	1	1	0	-1	0	-1	-1	-1	0.2667
9	1	1	1	0	-1	1	-1	-1	0	-1	-1	-1	-1	-1	-1	-1	-0.4000
10	1	1	1	1	1	1	-1	-1	1	0	-1	-1	-1	-1	-1	-1	-0.0667
11	1	1	1	1	1	1	1	0	1	1	0	0	0	0	-1	-1	0.4667
12	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0.6667
13	1	1	1	1	1	1	1	0	1	1	0	0	0	0	-1	-1	0.4667
14	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0.6667
15	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0	0	**0.8000**
16	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0	0	**0.8000**

Table B.18: Multi-comparison ANOVA for Rank-50 retrieval rates when using viewed hand-drawn sketches. Corresponding algorithm names shown in Table 6.1.

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**NS**
1	0	1	-1	-1	-1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-0.8000
2	-1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1.0000
3	1	1	0	-1	-1	1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-0.6000
4	1	1	1	0	-1	1	-1	-1	0	-1	-1	-1	-1	-1	-1	-1	-0.4000
5	1	1	1	1	0	1	-1	-1	1	-1	-1	-1	-1	-1	-1	-1	-0.2000
6	0	1	-1	-1	-1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-0.8000
7	1	1	1	1	1	1	0	0	1	1	-1	-1	-1	-1	-1	-1	0.1333
8	1	1	1	1	1	1	0	0	1	1	0	-1	-1	-1	-1	-1	0.2000
9	1	1	1	0	-1	1	-1	-1	0	-1	-1	-1	-1	-1	-1	-1	-0.4000
10	1	1	1	1	1	1	-1	-1	1	0	-1	-1	-1	-1	-1	-1	-0.0667
11	1	1	1	1	1	1	1	0	1	1	0	0	0	0	0	0	0.6000
12	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	**0.6667**
13	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	**0.6667**
14	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	**0.6667**
15	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	**0.6667**
16	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	**0.6667**

Table B.19: Multi-comparison ANOVA for Rank-100 retrieval rates when using viewed hand-drawn sketches. Corresponding algorithm names shown in Table 6.1.

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**NS**
1	0	1	-1	-1	-1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-0.8000
2	-1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1.0000
3	1	1	0	-1	-1	1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-0.6000
4	1	1	1	0	-1	1	-1	-1	0	-1	-1	-1	-1	-1	-1	-1	-0.4000
5	1	1	1	1	0	1	-1	-1	1	-1	-1	-1	-1	-1	-1	-1	-0.2000
6	0	1	-1	-1	-1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-0.8000
7	1	1	1	1	1	1	0	0	1	1	0	0	0	-1	-1	-1	0.3333
8	1	1	1	1	1	1	0	0	1	1	0	0	0	0	0	0	0.5333
9	1	1	1	0	-1	1	-1	-1	0	-1	-1	-1	-1	-1	-1	-1	-0.4000
10	1	1	1	1	1	1	-1	-1	1	0	-1	-1	-1	-1	-1	-1	-0.0667
11	1	1	1	1	1	1	0	0	1	1	0	0	0	0	0	0	0.5333
12	1	1	1	1	1	1	0	0	1	1	0	0	0	0	0	0	0.5333
13	1	1	1	1	1	1	0	0	1	1	0	0	0	0	0	0	0.5333
14	1	1	1	1	1	1	1	0	1	1	0	0	0	0	0	0	**0.6000**
15	1	1	1	1	1	1	1	0	1	1	0	0	0	0	0	0	**0.6000**
16	1	1	1	1	1	1	1	0	1	1	0	0	0	0	0	0	**0.6000**

Table B.20: Multi-comparison ANOVA for Rank-150 retrieval rates when using viewed hand-drawn sketches. Corresponding algorithm names shown in Table 6.1.

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**NS**
1	0	1	-1	-1	-1	1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-0.7333
2	-1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1.0000
3	1	1	0	-1	-1	1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-0.6000
4	1	1	1	0	0	1	-1	-1	0	-1	-1	-1	-1	-1	-1	-1	-0.3333
5	1	1	1	0	0	1	-1	-1	1	-1	-1	-1	-1	-1	-1	-1	-0.2667
6	-1	1	-1	-1	-1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-0.8667
7	1	1	1	1	1	1	0	0	1	1	0	0	0	0	0	0	0.5333
8	1	1	1	1	1	1	0	0	1	1	0	0	0	0	0	0	0.5333
9	1	1	1	0	-1	1	-1	-1	0	-1	-1	-1	-1	-1	-1	-1	-0.4000
10	1	1	1	1	1	1	-1	-1	1	0	-1	-1	-1	-1	-1	-1	-0.0667
11	1	1	1	1	1	1	0	0	1	1	0	0	0	0	0	0	**0.5333**
12	1	1	1	1	1	1	0	0	1	1	0	0	0	0	0	0	**0.5333**
13	1	1	1	1	1	1	0	0	1	1	0	0	0	0	0	0	**0.5333**
14	1	1	1	1	1	1	0	0	1	1	0	0	0	0	0	0	**0.5333**
15	1	1	1	1	1	1	0	0	1	1	0	0	0	0	0	0	**0.5333**
16	1	1	1	1	1	1	0	0	1	1	0	0	0	0	0	0	**0.5333**

Table B.21: Multi-comparison ANOVA for TAR@FAR=0.1% values when using viewed hand-drawn sketches. Corresponding algorithm names shown in Table 6.1.

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**NS**
1	0	1	0	-1	-1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-0.7333
2	-1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1.0000
3	0	1	0	-1	-1	1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-0.6667
4	1	1	1	0	-1	1	-1	-1	0	-1	-1	-1	-1	-1	-1	-1	-0.4000
5	1	1	1	1	0	1	-1	-1	1	-1	-1	-1	-1	-1	-1	-1	-0.2000
6	0	1	-1	-1	-1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-0.8000
7	1	1	1	1	1	1	0	-1	1	1	-1	-1	-1	-1	-1	-1	0.0667
8	1	1	1	1	1	1	1	0	1	1	0	-1	0	-1	-1	-1	0.3333
9	1	1	1	0	-1	1	-1	-1	0	-1	-1	-1	-1	-1	-1	-1	-0.4000
10	1	1	1	1	1	1	-1	-1	1	0	-1	-1	-1	-1	-1	-1	-0.0667
11	1	1	1	1	1	1	1	0	1	1	0	0	0	0	-1	-1	0.4667
12	1	1	1	1	1	1	1	1	1	1	0	0	1	0	0	0	0.7333
13	1	1	1	1	1	1	1	0	1	1	0	-1	0	-1	-1	-1	0.3333
14	1	1	1	1	1	1	1	1	1	1	0	0	1	0	0	0	0.7333
15	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0	0	**0.8000**
16	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0	0	**0.8000**

Table B.22: Multi-comparison ANOVA for TAR@FAR=1.0% values when using viewed hand-drawn sketches. Corresponding algorithm names shown in Table 6.1.

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**NS**
1	0	1	-1	-1	-1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-0.8000
2	-1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1.0000
3	1	1	0	-1	-1	1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-0.6000
4	1	1	1	0	-1	1	-1	-1	0	-1	-1	-1	-1	-1	-1	-1	-0.4000
5	1	1	1	1	0	1	-1	-1	1	-1	-1	-1	-1	-1	-1	-1	-0.2000
6	0	1	-1	-1	-1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-0.8000
7	1	1	1	1	1	1	0	0	1	1	-1	-1	-1	-1	-1	-1	0.1333
8	1	1	1	1	1	1	0	0	1	1	0	0	0	0	0	0	0.5333
9	1	1	1	0	-1	1	-1	-1	0	-1	-1	-1	-1	-1	-1	-1	-0.4000
10	1	1	1	1	1	1	-1	-1	1	0	-1	-1	-1	-1	-1	-1	-0.0667
11	1	1	1	1	1	1	1	0	1	1	0	0	0	0	0	0	**0.6000**
12	1	1	1	1	1	1	1	0	1	1	0	0	0	0	0	0	**0.6000**
13	1	1	1	1	1	1	1	0	1	1	0	0	0	0	0	0	**0.6000**
14	1	1	1	1	1	1	1	0	1	1	0	0	0	0	0	0	**0.6000**
15	1	1	1	1	1	1	1	0	1	1	0	0	0	0	0	0	**0.6000**
16	1	1	1	1	1	1	1	0	1	1	0	0	0	0	0	0	**0.6000**

Table B.23: Multi-comparison ANOVA for Rank-1 retrieval rates when using UoM-SGFS Set A software-generated sketches. Corresponding algorithm names shown in Table 6.2.

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**NS**
1	0	1	0	0	0	0	-1	-1	0	-1	-1	-1	-1	-1	-1	-1	-0.5333
2	-1	0	0	0	-1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-0.8000
3	0	0	0	0	0	0	-1	-1	0	-1	-1	-1	-1	-1	-1	-1	-0.6000
4	0	0	0	0	-1	0	-1	-1	0	-1	-1	-1	-1	-1	-1	-1	-0.6667
5	0	1	0	1	0	1	-1	-1	0	0	-1	-1	-1	-1	-1	-1	-0.3333
6	0	0	0	0	-1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-0.7333
7	1	1	1	1	1	1	0	0	1	1	0	-1	-1	-1	-1	-1	0.2000
8	1	1	1	1	1	1	0	0	1	1	0	0	0	-1	-1	-1	0.3333
9	0	1	0	0	0	1	-1	-1	0	0	-1	-1	-1	-1	-1	-1	-0.4000
10	1	1	1	1	0	1	-1	-1	0	0	0	-1	-1	-1	-1	-1	-0.1333
11	1	1	1	1	1	1	0	0	1	0	0	-1	-1	-1	-1	-1	0.1333
12	1	1	1	1	1	1	1	0	1	1	1	0	0	-1	-1	-1	0.4667
13	1	1	1	1	1	1	1	0	1	1	1	0	0	-1	-1	-1	0.4667
14	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	-1	0.8000
15	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	-1	0.8000
16	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	**1.0000**

Table B.24: Multi-comparison ANOVA for Rank-10 retrieval rates when using UoM-SGFS Set A software-generated sketches. Corresponding algorithm names shown in Table 6.2.

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**NS**
1	0	1	0	0	0	1	-1	-1	0	-1	-1	-1	-1	-1	-1	-1	-0.4667
2	-1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1.0000
3	0	1	0	0	-1	1	-1	-1	0	-1	-1	-1	-1	-1	-1	-1	-0.5333
4	0	1	0	0	-1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-0.6667
5	0	1	1	1	0	1	-1	-1	0	0	-1	-1	-1	-1	-1	-1	-0.2667
6	-1	1	-1	0	-1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-0.8000
7	1	1	1	1	1	1	0	0	1	0	0	-1	-1	-1	-1	-1	0.1333
8	1	1	1	1	1	1	0	0	1	1	0	-1	-1	-1	-1	-1	0.2000
9	0	1	0	1	0	1	-1	-1	0	-1	-1	-1	-1	-1	-1	-1	-0.4000
10	1	1	1	1	0	1	0	-1	1	0	-1	-1	-1	-1	-1	-1	-0.0667
11	1	1	1	1	1	1	0	0	1	1	0	-1	-1	-1	-1	-1	0.2000
12	1	1	1	1	1	1	1	1	1	1	1	0	0	-1	-1	-1	0.5333
13	1	1	1	1	1	1	1	1	1	1	1	0	0	-1	-1	-1	0.5333
14	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	**0.8667**
15	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	**0.8667**
16	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	**0.8667**

Table B.25: Multi-comparison ANOVA for Rank-50 retrieval rates when using UoM-SGFS Set A software-generated sketches. Corresponding algorithm names shown in Table 6.2.

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**NS**
1	0	1	0	1	1	1	-1	-1	0	-1	-1	-1	-1	-1	-1	-1	-0.3333
2	-1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1.0000
3	0	1	0	1	0	1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-0.4667
4	-1	1	-1	0	-1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-0.8000
5	-1	1	0	1	0	1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-0.5333
6	-1	1	-1	0	-1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-0.8000
7	1	1	1	1	1	1	0	0	1	0	0	-1	-1	-1	-1	-1	0.1333
8	1	1	1	1	1	1	0	0	1	1	0	-1	-1	-1	-1	-1	0.2000
9	0	1	1	1	1	1	-1	-1	0	0	-1	-1	-1	-1	-1	-1	-0.2000
10	1	1	1	1	1	1	0	-1	0	0	0	-1	-1	-1	-1	-1	0.0000
11	1	1	1	1	1	1	0	0	1	0	0	-1	-1	-1	-1	-1	0.1333
12	1	1	1	1	1	1	1	1	1	1	1	0	0	-1	-1	-1	0.5333
13	1	1	1	1	1	1	1	1	1	1	1	0	0	-1	-1	-1	0.5333
14	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	**0.8667**
15	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	**0.8667**
16	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	**0.8667**

Table B.26: Multi-comparison ANOVA for Rank-100 retrieval rates when using UoM-SGFS Set A software-generated sketches. Corresponding algorithm names shown in Table 6.2.

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**NS**
1	0	1	1	1	1	1	0	-1	0	0	-1	-1	-1	-1	-1	-1	-0.1333
2	-1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1.0000
3	-1	1	0	1	1	1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-0.4667
4	-1	1	-1	0	0	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-0.7333
5	-1	1	-1	0	0	1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-0.6667
6	-1	1	-1	0	-1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-0.8000
7	0	1	1	1	1	1	0	-1	0	0	0	-1	-1	-1	-1	-1	-0.0667
8	1	1	1	1	1	1	1	0	1	1	0	0	-1	-1	-1	-1	0.3333
9	0	1	1	1	1	1	0	-1	0	0	-1	-1	-1	-1	-1	-1	-0.1333
10	0	1	1	1	1	1	0	-1	0	0	0	-1	-1	-1	-1	-1	-0.0667
11	1	1	1	1	1	1	0	0	1	0	0	-1	-1	-1	-1	-1	0.1333
12	1	1	1	1	1	1	1	0	1	1	1	0	0	-1	-1	-1	0.4667
13	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0.7333
14	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	**0.8000**
15	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	**0.8000**
16	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	**0.8000**

Table B.27: Multi-comparison ANOVA for Rank-150 retrieval rates when using UoM-SGFS Set A software-generated sketches. Corresponding algorithm names shown in Table 6.2.

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**NS**
1	0	1	1	1	1	1	0	-1	0	0	0	-1	-1	-1	-1	-1	-0.0667
2	-1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1.0000
3	-1	1	0	1	1	1	-1	-1	0	-1	-1	-1	-1	-1	-1	-1	-0.4000
4	-1	1	-1	0	0	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-0.7333
5	-1	1	-1	0	0	1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-0.6667
6	-1	1	-1	0	-1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-0.8000
7	0	1	1	1	1	1	0	-1	0	0	0	-1	-1	-1	-1	-1	-0.0667
8	1	1	1	1	1	1	1	0	1	1	1	0	-1	-1	-1	-1	0.4000
9	0	1	0	1	1	1	0	-1	0	0	0	-1	-1	-1	-1	-1	-0.1333
10	0	1	1	1	1	1	0	-1	0	0	0	-1	-1	-1	-1	-1	-0.0667
11	0	1	1	1	1	1	0	-1	0	0	0	-1	-1	-1	-1	-1	-0.0667
12	1	1	1	1	1	1	1	0	1	1	1	0	0	-1	-1	-1	0.4667
13	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0.7333
14	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	**0.8000**
15	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	**0.8000**
16	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	**0.8000**

Table B.28: Multi-comparison ANOVA for TAR@FAR=0.1% values when using UoM-SGFS Set A software-generated sketches. Corresponding algorithm names shown in Table 6.2.

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**NS**
1	0	1	0	0	0	0	-1	-1	0	-1	-1	-1	-1	-1	-1	-1	-0.5333
2	-1	0	-1	0	-1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-0.8667
3	0	1	0	0	0	0	-1	-1	0	-1	-1	-1	-1	-1	-1	-1	-0.5333
4	0	0	0	0	-1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-0.7333
5	0	1	0	1	0	1	-1	-1	0	0	-1	-1	-1	-1	-1	-1	-0.3333
6	0	0	0	0	-1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-0.7333
7	1	1	1	1	1	1	0	0	1	1	0	-1	-1	-1	-1	-1	0.2000
8	1	1	1	1	1	1	0	0	1	1	0	-1	-1	-1	-1	-1	0.2000
9	0	1	0	1	0	1	-1	-1	0	0	-1	-1	-1	-1	-1	-1	-0.3333
10	1	1	1	1	0	1	-1	-1	0	0	-1	-1	-1	-1	-1	-1	-0.2000
11	1	1	1	1	1	1	0	0	1	1	0	-1	-1	-1	-1	-1	0.2000
12	1	1	1	1	1	1	1	1	1	1	1	0	0	-1	-1	-1	0.5333
13	1	1	1	1	1	1	1	1	1	1	1	0	0	-1	-1	-1	0.5333
14	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	-1	0.8000
15	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	-1	0.8000
16	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	**1.0000**

Table B.29: Multi-comparison ANOVA for TAR@FAR=1.0% values when using UoM-SGFS Set A software-generated sketches. Corresponding algorithm names shown in Table 6.2.

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**NS**
1	0	1	0	1	0	1	-1	-1	0	-1	-1	-1	-1	-1	-1	-1	-0.4000
2	-1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1.0000
3	0	1	0	1	0	1	-1	-1	0	-1	-1	-1	-1	-1	-1	-1	-0.4000
4	-1	1	-1	0	-1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-0.8000
5	0	1	0	1	0	1	-1	-1	0	-1	-1	-1	-1	-1	-1	-1	-0.4000
6	-1	1	-1	0	-1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-0.8000
7	1	1	1	1	1	1	0	-1	1	0	0	-1	-1	-1	-1	-1	0.0667
8	1	1	1	1	1	1	1	0	1	1	1	-1	-1	-1	-1	-1	0.3333
9	0	1	0	1	0	1	-1	-1	0	-1	-1	-1	-1	-1	-1	-1	-0.4000
10	1	1	1	1	1	1	0	-1	1	0	-1	-1	-1	-1	-1	-1	0.0000
11	1	1	1	1	1	1	0	-1	1	1	0	-1	-1	-1	-1	-1	0.1333
12	1	1	1	1	1	1	1	1	1	1	1	0	0	-1	-1	-1	0.5333
13	1	1	1	1	1	1	1	1	1	1	1	0	0	-1	-1	-1	0.5333
14	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	**0.8667**
15	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	**0.8667**
16	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	**0.8667**

Table B.30: Multi-comparison ANOVA for Rank-1 retrieval rates when using UoM-SGFS Set B software-generated sketches. Corresponding algorithm names shown in Table 6.3.

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**NS**
1	0	1	0	0	0	1	-1	-1	0	-1	-1	-1	-1	-1	-1	-1	-0.4667
2	-1	0	-1	0	-1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-0.8667
3	0	1	0	0	-1	0	-1	-1	0	-1	-1	-1	-1	-1	-1	-1	-0.6000
4	0	0	0	0	-1	0	-1	-1	0	-1	-1	-1	-1	-1	-1	-1	-0.6667
5	0	1	1	1	0	1	-1	-1	1	0	-1	-1	-1	-1	-1	-1	-0.2000
6	-1	0	0	0	-1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-0.8000
7	1	1	1	1	1	1	0	0	1	1	0	-1	-1	-1	-1	-1	0.2000
8	1	1	1	1	1	1	0	0	1	1	0	-1	-1	-1	-1	-1	0.2000
9	0	1	0	0	-1	1	-1	-1	0	-1	-1	-1	-1	-1	-1	-1	-0.5333
10	1	1	1	1	0	1	-1	-1	1	0	-1	-1	-1	-1	-1	-1	-0.1333
11	1	1	1	1	1	1	0	0	1	1	0	-1	-1	-1	-1	-1	0.2000
12	1	1	1	1	1	1	1	1	1	1	1	0	0	-1	-1	-1	0.5333
13	1	1	1	1	1	1	1	1	1	1	1	0	0	-1	-1	-1	0.5333
14	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	-1	0.8000
15	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0.8667
16	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	**0.9333**

Table B.31: Multi-comparison ANOVA for Rank-10 retrieval rates when using UoM-SGFS Set B software-generated sketches. Corresponding algorithm names shown in Table 6.3.

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**NS**
1	0	1	1	1	0	1	-1	-1	0	-1	-1	-1	-1	-1	-1	-1	-0.3333
2	-1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1.0000
3	-1	1	0	1	0	1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-0.5333
4	-1	1	-1	0	-1	1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-0.7333
5	0	1	0	1	0	1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-0.4667
6	-1	1	-1	-1	-1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-0.8667
7	1	1	1	1	1	1	0	0	1	1	0	-1	-1	-1	-1	-1	0.2000
8	1	1	1	1	1	1	0	0	1	1	0	-1	0	-1	-1	-1	0.2667
9	0	1	1	1	1	1	-1	-1	0	-1	-1	-1	-1	-1	-1	-1	-0.2667
10	1	1	1	1	1	1	-1	-1	1	0	-1	-1	-1	-1	-1	-1	-0.0667
11	1	1	1	1	1	1	0	0	1	1	0	-1	-1	-1	-1	-1	0.2000
12	1	1	1	1	1	1	1	1	1	1	1	0	0	-1	-1	-1	0.5333
13	1	1	1	1	1	1	1	0	1	1	1	0	0	-1	-1	-1	0.4667
14	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	**0.8667**
15	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	**0.8667**
16	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	**0.8667**

Table B.32: Multi-comparison ANOVA for Rank-50 retrieval rates when using UoM-SGFS Set B software-generated sketches. Corresponding algorithm names shown in Table 6.3.

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**NS**
1	0	1	1	1	1	1	-1	-1	0	0	-1	-1	-1	-1	-1	-1	-0.2000
2	-1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1.0000
3	-1	1	0	1	1	1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-0.4667
4	-1	1	-1	0	0	1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-0.6667
5	-1	1	-1	0	0	1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-0.6667
6	-1	1	-1	-1	-1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-0.8667
7	1	1	1	1	1	1	0	0	1	1	0	-1	-1	-1	-1	-1	0.2000
8	1	1	1	1	1	1	0	0	1	1	0	0	0	-1	-1	-1	0.3333
9	0	1	1	1	1	1	-1	-1	0	0	-1	-1	-1	-1	-1	-1	-0.2000
10	0	1	1	1	1	1	-1	-1	0	0	-1	-1	-1	-1	-1	-1	-0.2000
11	1	1	1	1	1	1	0	0	1	1	0	-1	-1	-1	-1	-1	0.2000
12	1	1	1	1	1	1	1	0	1	1	1	0	0	0	0	-1	0.6000
13	1	1	1	1	1	1	1	0	1	1	1	0	0	0	-1	-1	0.5333
14	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0.7333
15	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0	0	0.8000
16	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	**0.8667**

Table B.33: Multi-comparison ANOVA for Rank-100 retrieval rates when using UoM-SGFS Set B software-generated sketches. Corresponding algorithm names shown in Table 6.3.

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**NS**
1	0	1	1	1	1	1	-1	-1	0	0	-1	-1	-1	-1	-1	-1	-0.2000
2	-1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1.0000
3	-1	1	0	1	1	1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-0.4667
4	-1	1	-1	0	0	1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-0.6667
5	-1	1	-1	0	0	1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-0.6667
6	-1	1	-1	-1	-1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-0.8667
7	1	1	1	1	1	1	0	0	1	1	0	0	0	-1	-1	-1	0.3333
8	1	1	1	1	1	1	0	0	1	1	1	0	0	0	0	0	0.6000
9	0	1	1	1	1	1	-1	-1	0	0	-1	-1	-1	-1	-1	-1	-0.2000
10	0	1	1	1	1	1	-1	-1	0	0	-1	-1	-1	-1	-1	-1	-0.2000
11	1	1	1	1	1	1	0	-1	1	1	0	-1	-1	-1	-1	-1	0.1333
12	1	1	1	1	1	1	0	0	1	1	1	0	0	0	0	0	0.6000
13	1	1	1	1	1	1	0	0	1	1	1	0	0	0	0	0	0.6000
14	1	1	1	1	1	1	1	0	1	1	1	0	0	0	0	0	**0.6667**
15	1	1	1	1	1	1	1	0	1	1	1	0	0	0	0	0	**0.6667**
16	1	1	1	1	1	1	1	0	1	1	1	0	0	0	0	0	**0.6667**

Table B.34: Multi-comparison ANOVA for Rank-150 retrieval rates when using UoM-SGFS Set B software-generated sketches. Corresponding algorithm names shown in Table 6.3.

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**NS**
1	0	1	1	1	1	1	-1	-1	0	0	-1	-1	-1	-1	-1	-1	-0.2000
2	-1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1.0000
3	-1	1	0	1	1	1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-0.4667
4	-1	1	-1	0	1	1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-0.6000
5	-1	1	-1	-1	0	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-0.8000
6	-1	1	-1	-1	0	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-0.8000
7	1	1	1	1	1	1	0	0	1	1	0	0	0	0	0	0	0.5333
8	1	1	1	1	1	1	0	0	1	1	0	0	0	0	0	0	0.5333
9	0	1	1	1	1	1	-1	-1	0	0	-1	-1	-1	-1	-1	-1	-0.2000
10	0	1	1	1	1	1	-1	-1	0	0	-1	-1	-1	-1	-1	-1	-0.2000
11	1	1	1	1	1	1	0	0	1	1	0	0	0	-1	-1	-1	0.3333
12	1	1	1	1	1	1	0	0	1	1	0	0	0	0	0	0	0.5333
13	1	1	1	1	1	1	0	0	1	1	0	0	0	0	0	0	0.5333
14	1	1	1	1	1	1	0	0	1	1	1	0	0	0	0	0	**0.6000**
15	1	1	1	1	1	1	0	0	1	1	1	0	0	0	0	0	**0.6000**
16	1	1	1	1	1	1	0	0	1	1	1	0	0	0	0	0	**0.6000**

Table B.35: Multi-comparison ANOVA for TAR@FAR=0.1% values when using UoM-SGFS Set B software-generated sketches. Corresponding algorithm names shown in Table 6.3.

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**NS**
1	0	1	1	1	0	1	-1	-1	0	-1	-1	-1	-1	-1	-1	-1	-0.3333
2	-1	0	-1	-1	-1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-0.9333
3	-1	1	0	0	-1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-0.7333
4	-1	1	0	0	-1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-0.7333
5	0	1	1	1	0	1	-1	-1	0	-1	-1	-1	-1	-1	-1	-1	-0.3333
6	-1	0	0	0	-1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-0.8000
7	1	1	1	1	1	1	0	0	1	1	0	-1	-1	-1	-1	-1	0.2000
8	1	1	1	1	1	1	0	0	1	1	0	-1	-1	-1	-1	-1	0.2000
9	0	1	1	1	0	1	-1	-1	0	-1	-1	-1	-1	-1	-1	-1	-0.3333
10	1	1	1	1	1	1	-1	-1	1	0	-1	-1	-1	-1	-1	-1	-0.0667
11	1	1	1	1	1	1	0	0	1	1	0	-1	-1	-1	-1	-1	0.2000
12	1	1	1	1	1	1	1	1	1	1	1	0	0	-1	-1	-1	0.5333
13	1	1	1	1	1	1	1	1	1	1	1	0	0	-1	-1	-1	0.5333
14	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	-1	0.8000
15	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0.8667
16	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	**0.9333**

Table B.36: Multi-comparison ANOVA for TAR@FAR=1.0% values when using UoM-SGFS Set B software-generated sketches. Corresponding algorithm names shown in Table 6.3.

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**NS**
1	0	1	1	1	1	1	-1	-1	0	0	-1	-1	-1	-1	-1	-1	-0.2000
2	-1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1.0000
3	-1	1	0	1	0	1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-0.5333
4	-1	1	-1	0	-1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-0.8000
5	-1	1	0	1	0	1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-0.5333
6	-1	1	-1	0	-1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-0.8000
7	1	1	1	1	1	1	0	0	1	1	0	-1	-1	-1	-1	-1	0.2000
8	1	1	1	1	1	1	0	0	1	1	0	0	0	-1	-1	-1	0.3333
9	0	1	1	1	1	1	-1	-1	0	-1	-1	-1	-1	-1	-1	-1	-0.2667
10	0	1	1	1	1	1	-1	-1	1	0	-1	-1	-1	-1	-1	-1	-0.1333
11	1	1	1	1	1	1	0	0	1	1	0	-1	-1	-1	-1	-1	0.2000
12	1	1	1	1	1	1	1	0	1	1	1	0	0	-1	-1	-1	0.4667
13	1	1	1	1	1	1	1	0	1	1	1	0	0	-1	-1	-1	0.4667
14	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	**0.8667**
15	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	**0.8667**
16	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	**0.8667**

References

[1] A. K. Jain, A. Ross, and S. Prabhakar, "An introduction to biometric recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 1, pp. 4–20, Jan 2004.

[2] A. K. Jain, K. Nandakumar, and A. Ross, "Score normalization in multimodal biometric systems," *Pattern Recognition*, vol. 38, no. 12, pp. 2270 – 2285, 2005.

[3] A. K. Jain, A. Ross, and S. Pankanti, "Biometrics: a tool for information security," *IEEE Transactions on Information Forensics and Security*, vol. 1, no. 2, pp. 125–143, June 2006.

[4] D. Maltoni, D. Maio, A. K. Jain, and S. Prabhakar, *Handbook of Fingerprint Recognition*. Springer-Verlag, 2009.

[5] F. Nicolo, "Homogeneous and heterogeneous face recognition: Enhancing, encoding and matching for practical applications," Ph.D. dissertation, Morgantown, WV, USA, 2012, aAI3530478.

[6] H. Han, B. F. Klare, K. Bonnen, and A. K. Jain, "Matching composite sketches to face photos: A component-based approach," *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 1, pp. 191–204, Jan 2013.

[7] A. K. Jain, B. Klare, and U. Park, "Face matching and retrieval in forensics applications," *IEEE MultiMedia in Forensics, Security, and Intelligence*, vol. 19, no. 1, pp. 20–20, Jan 2012.

[8] J. Zuo, F. Nicolo, N. Schmid, and S. Boothapati, "Encoding, matching and score normalization for cross spectral face recognition: Matching SWIR versus visible data," in *Biometrics: Theory, Applications and Systems (BTAS), 2012 IEEE Fifth International Conference on*, Sept 2012, pp. 203–208.

[9] B. Klare, Z. Li, and A. K. Jain, "Matching forensic sketches to mug shot photos," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 3, pp. 639–646, March 2011.

[10] H. Galoogahi and T. Sim, "Inter-modality face sketch recognition," in *2012 IEEE International Conference on Multimedia and Expo (ICME)*, July 2012, pp. 224–229.

[11] S. Z. Li and A. K. Jain, *Handbook of Face Recognition*, 2nd ed. Springer, 2011.

[12] B. F. Klare and A. K. Jain, "Heterogeneous face recognition using kernel prototype similarities," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 6, pp. 1410–1422, June 2013.

[13] P. J. Phillips, J. Wechsler, H. amd Huang, and P. Rauss, "The FERET database and evaluation procedure for face recognition algorithms," *Image and Vision Computing*, vol. 16, pp. 295–3067, 1998.

[14] C. Galea and R. A. Farrugia, "A Large-Scale Software-Generated Face Composite Sketch Database," in *Int. Conf. Biometrics Special Interest Group (BIOSIG)*, Sep 2016, pp. 1–5.

[15] W. Zhang, X. Wang, and X. Tang, "Coupled information-theoretic encoding for face photo-sketch recognition," in *2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2011, pp. 513–520.

[16] A. M. Martinez and R. Benavente, "The AR Face database," CVC, Tech. Rep., June 1998.

[17] X. Tang and X. Wang, "Face sketch recognition," in *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 1, January 2004, pp. 50–57.

[18] X. Wang and X. Tang, "Face photo-sketch synthesis and recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 11, pp. 1955–1967, Nov 2009.

[19] S. J. Klum, H. Han, B. F. Klare, and A. K. Jain, "The FaceSketchID System: Matching Facial Composites to Mugshots," *IEEE Trans. Inf. Forensics Security*, vol. 9, no. 12, pp. 2248–2263, Dec 2014.

[20] Identi-Kit, Identi-Kit Solutions, last visited on Jan. 30, 2015. [Online]. Available: http://www.identikit.net/

[21] FACES 4.0, IQ Biometrix, last visited on Jan. 30, 2015. [Online]. Available: http://www.iqbiometrix.com

[22] VisionMetric, "About E-FIT," last visited on Mar. 17, 2015. [Online]. Available: http://www.visionmetric.com/products/about-e-fit/

[23] S. Ouyang, T. Hospedales, Y.-Z. Song, and X. Li, *Cross-Modal Face Matching: Beyond Viewed Sketches*. Cham: Springer International Publishing, 2015, pp. 210–225.

[24] S. Ouyang, T. M. Hospedales, Y. Z. Song, and X. Li, "ForgetMeNot: Memory-Aware Forensic Facial Sketch Matching," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016, pp. 5571–5579.

[25] N. Wang, D. Tao, X. Gao, X. Li, and J. Li, "A comprehensive survey to face hallucination," *International Journal of Computer Vision*, vol. 106, no. 1, pp. 9–30, Jan. 2014.

[26] S. Klum, H. Han, A. K. Jain, and B. Klare, "Sketch based face recognition: Forensic vs. composite sketches," in *2013 International Conference on Biometrics (ICB)*, June 2013, pp. 1–8.

[27] M. Golfarelli, D. Maio, and D. Malton, "On the error-reject trade-off in biometric verification systems," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 19, no. 7, pp. 786–796, Jul 1997.

[28] L. Ding, C. Shu, C. Fang, and X. Ding, "Computers do better than experts matching faces in a large population," in *Proceedings of the 9th IEEE International Conference on Cognitive Informatics (ICCI)*, July 2010, pp. 280–284.

[29] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proceedings of the Seventh IEEE International Conference on Computer Vision*, vol. 2, 1999, pp. 1150–1157 vol.2.

[30] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, Nov 2004.

[31] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971–987, July 2002.

[32] C. Zhou, Z. Zhang, D. Yi, Z. Lei, and S. Z. Li, "Low-resolution face recognition via simultaneous discriminant analysis," in *International Joint Conference on Biometrics (IJCB)*, Oct 2011, pp. 1–6.

[33] X. Tang and X. Wang, "Face photo recognition using sketch," in *Proceedings of the International Conference on Image Processing (ICIP 2002)*, vol. 1, 2002, pp. I–257–I–260 vol.1.

[34] Q. Liu, X. Tang, H. Jin, H. Lu, and S. Ma, "A nonlinear approach for face sketch synthesis and recognition," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005)*, vol. 1, June 2005, pp. 1005–1010 vol. 1.

[35] W. Liu, X. Tang, and J. Liu, "Bayesian tensor inference for sketch-based facial photo hallucination," in *Proceedings of the 20th International Joint Conference on Artifical Intelligence*, ser. IJCAI'07, 2007, pp. 2141–2146.

[36] L. Chang, M. Zhou, Y. Han, and X. Deng, "Face sketch synthesis via sparse representation," in *Proceedings of the International Conference on Pattern Recognition (ICPR)*, Aug 2010, pp. 2146–2149.

[37] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.

[38] X. Tang and X. Wang, "Face sketch synthesis and recognition," in *Proceedings of the Ninth IEEE International Conference on Computer Vision*, Oct 2003, pp. 687–694 vol.1.

[39] Chinese University of Hong Kong (CUHK), "CUHK Face Sketch Database," last visited on Feb. 2, 2015. [Online]. Available: http://mmlab.ie.cuhk.edu.hk/archive/facesketch.html

[40] L. Wiskott, J.-M. Fellous, N. Kruger, and C. von der Malsburg, "Face recognition by elastic bunch graph matching," in *Proceedings of the International Conference on Image Processing*, vol. 1, Oct 1997, pp. 129–132.

[41] X. Gao, J. Zhong, J. Li, and C. Tian, "Face Sketch Synthesis Algorithm Based on E-HMM and Selective Ensemble," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 4, pp. 487–496, April 2008.

[42] Y. hui Li, M. Savvides, and V. Bhagavatula, "Illumination tolerant face recognition using a novel face from sketch synthesis approach and advanced correlation filters," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 2, May 2006, pp. II–357 – II–360.

[43] B. V. K. V. Kumar, "Tutorial survey of composite filter designs for optical correlators," *Applied Optics*, vol. 31, pp. 4773–4801, 1992.

[44] "CMU/VASC Image Database," last visited on Apr. 21, 2015. [Online]. Available: http://vasc.ri.cmu.edu/idb/html/face/

[45] W. Zhang, X. Wang, and X. Tang, "Lighting and pose robust face sketch synthesis," in *Computer Vision ECCV 2010*, ser. Lecture Notes in Computer Science, K. Daniilidis, P. Maragos, and N. Paragios, Eds. Springer Berlin Heidelberg, 2010, vol. 6316, pp. 420–433.

[46] H. Zhou, Z. Kuang, and K.-Y. Wong, "Markov weight fields for face sketch synthesis," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2012, pp. 1091–1097.

[47] J. Zhong, X. Gao, and C. Tian, "Face Sketch Synthesis using E-HMM and Selective Ensemble," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 1, April 2007, pp. I–485–I–488.

[48] A. V. Nefian and M. H. Hayes III, "Face recognition using an embedded hmm," in *IEEE Conference on Audio and Video-based Biometric Person Authentication*, 1999, pp. 19–24.

[49] S. Moon and J.-N. Hwang, "Noisy speech recognition using robust inversion of hidden markov models," in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP 1995)*, vol. 1, May 1995, pp. 145–148 vol.1.

[50] L. Chang, X. Deng, M. Zhou, F. Duan, and Z. Wu, "Smoothness-constrained face photo-sketch synthesis using sparse representation," in *21st International Conference on Pattern Recognition (ICPR)*, Nov 2012, pp. 3025–3029.

[51] X. Gao, N. Wang, D. Tao, and X. Li, "Face sketch-photo synthesis and retrieval using sparse representation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 8, pp. 1213–1226, Aug 2012.

[52] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210–227, Feb 2009.

[53] P. N. Belhumeur, J. P. Hespanha, and D. Kriegman, "Eigenfaces vs. Fisherfaces: recognition using class specific linear projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711–720, Jul 1997.

[54] X. He, "Locality preserving projections," Ph.D. dissertation, Chicago, IL, USA, 2005, aAI3195015.

[55] X. Gao, J. Zhong, D. Tao, and X. Li, "Local face sketch synthesis learning," *Neurocomputing*, vol. 71, no. 1012, pp. 1921 – 1930, 2008, neurocomputing for Vision ResearchAdvances in Blind Signal Processing.

[56] B. Xiao, X. Gao, D. Tao, and X. Li, "A new approach for face recognition by sketches in photos," *Signal Process.*, vol. 89, no. 8, pp. 1576–1588, Aug. 2009.

[57] S. Zhang, X. Gao, N. Wang, J. Li, and M. Zhang, "Face sketch synthesis via sparse representation-based greedy search," *IEEE Transactions on Image Processing*, vol. 24, no. 8, pp. 2466–2477, Aug 2015.

[58] N. Wang, D. Tao, X. Gao, X. Li, and J. Li, "Transductive face sketch-photo synthesis," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 24, no. 9, pp. 1364–1376, Sept 2013.

[59] Y. Song, L. Bao, Q. Yang, and M.-H. Yang, "Real-time exemplar-based face sketch synthesis," in *Computer Vision ECCV 2014*, ser. Lecture Notes in Computer Science, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds. Springer International Publishing, 2014, vol. 8694, pp. 800–813.

[60] Z. Wang, A. C. Bovik, H. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, April 2004.

[61] L. Zhang, D. Zhang, X. Mou, and D. Zhang, "FSIM: A Feature Similarity Index for Image Quality Assessment," *IEEE Transactions on Image Processing*, vol. 20, no. 8, pp. 2378–2386, August 2011.

[62] C. Galea and R. A. Farrugia, "Fusion of intra- and inter-modality algorithms for face-sketch recognition," in *Computer Analysis of Images and Patterns*, ser. Lecture Notes in Computer Science, G. Azzopardi and N. Petkov, Eds. Springer International Publishing, 2015, vol. 9257, pp. 700–711.

[63] Z. Lei, S. Liao, A. K. Jain, and S. Z. Li, "Coupled discriminant analysis for heterogeneous face recognition," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 6, pp. 1707–1716, Dec 2012.

[64] B. Klare and A. K. Jain, "Sketch-to-photo matching: a feature-based approach," in *Proceedings of the SPIE Conference on Biometric Technology for Human Identification VII*, vol. 7667, 2010, pp. 766 702–766 702–10.

[65] FaceVACS Software Developer's Kit, Cognitec Systems GmbH, last visited on Jan. 30, 2015. [Online]. Available: http://www.cognitec.com/facevacs-sdk.html

[66] S. Pramanik and D. Bhattacharjee, "An Approach: Modality Reduction and Face-Sketch Recognition," *International Journal of Computational Intelligence and Informatics*, vol. 1, no. 2, 2011.

[67] B. Klare, Z. Li, and A. K. Jain, "On Matching Forensic Sketches to Mugshot Photos," Michigan State University, Tech. Rep. MSU-CSE-10-3, 2010.

[68] S. Liao, D. Yi, Z. Lei, R. Qin, and S. Z. Li, "Heterogeneous face recognition from local structures of normalized appearance," in *Proceedings of the Third International Conference on Advances in Biometrics*, ser. ICB '09. Berlin, Heidelberg: Springer-Verlag, 2009, pp. 209–218.

[69] B. Klare and A. K. Jain, "Heterogeneous Face Recognition: Matching NIR to Visible Light Images," in *20th International Conference on Pattern Recognition (ICPR)*, Aug 2010, pp. 1513–1516.

[70] R. K. Bothwell, J. C. Brigham, and R. S. Malpass, "Cross-racial identification," *Personality Social Psychology Bulletin*, vol. 15, no. 1, pp. 19–25, 1989.

[71] C. Meissner and J. Brigham, "Thirty years of investigating the ownrace bias in memory for faces: A meta-analytic review," *Psychology, Public Policy, and Law*, vol. 7, no. 1, pp. 3–35, Jan 2001.

[72] R. E. Geiselman, R. P. Fisher, D. P. MacKinnon, and H. L. Holland, "Eyewitness Memory Enhancement in the Police Interview: Cognitive Retrieval Mnemonics Versus Hypnosis," *Journal of Applied Pyschology*, vol. 70, no. 2, pp. 401–412, 1985.

[73] R. E. Geiselman, R. P. Fisher, D. P. MacKinnon, and H. L. Holland, "Enhancement of eyewitness memory with the cognitive interview," *Journal of Applied Pyschology*, vol. 99, no. 3, pp. 385–401, 1986.

[74] H. Hotelling, "Relations between two sets of variants," *Biometrika*, vol. 28, pp. 321–377, December 1936.

[75] L. Sun, S. Ji, and J. Ye, "A least squares formulation for canonical correlation analysis," in *Proceedings of the 25th International Conference on Machine Learning*, ser. ICML '08. New York, NY, USA: ACM, 2008, pp. 1024–1031.

[76] B. Li, H. Chang, S. Shan, and X. Chen, "Low-resolution face recognition via coupled locality preserving mappings," *IEEE Signal Processing Letters*, vol. 17, no. 1, pp. 20–23, Jan 2010.

[77] S. Siena, V. N. Boddeti, and B. V. K. Vijaya Kumar, "Coupled marginal fisher analysis for low-resolution face recognition," in *Proceedings of the 12th International Conference on Computer Vision - Volume 2*, ser. ECCV'12. Berlin, Heidelberg: Springer-Verlag, 2012, pp. 240–249.

[78] Z. Lei and S. Li, "Coupled spectral regression for matching heterogeneous faces," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, June 2009, pp. 1123–1128.

[79] S. Siena, V. Boddeti, and B. Kumar, "Maximum-Margin Coupled Mappings for cross-domain matching," in *Biometrics: Theory, Applications and Systems (BTAS), 2013 IEEE Sixth International Conference on*, Sept 2013, pp. 1–8.

[80] C. Peng, N. Wang, X. Gao, and J. Li, *Face Recognition from Multiple Stylistic Sketches: Scenarios, Datasets, and Evaluation.* Cham: Springer International Publishing, 2016, pp. 3–18.

[81] C. Peng, X. Gao, N. Wang, and J. Li, "Graphical representation for heterogeneous face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 2, pp. 301–312, Feb 2017.

[82] T. Chugh, M. Singh, S. Nagpal, R. Singh, and M. Vatsa, "Transfer Learning based Evolutionary Algorithm for Composite Face Sketch Recognition," in *IEEE Comput. Vision Pattern Recog. Workshop*, July 2017.

[83] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 1, June 2005, pp. 886–893 vol. 1.

[84] M.-K. Hu, "Visual pattern recognition by moment invariants," *IRE Transactions on Information Theory*, vol. 8, no. 2, pp. 179–187, February 1962.

[85] D. Gong, Z. Li, W. Huang, X. Li, and D. Tao, "Heterogeneous face recognition: A common encoding feature discriminant approach," *IEEE Transactions on Image Processing*, vol. PP, no. 99, pp. 1–1, 2017.

[86] S. Ouyang, T. Hospedales, Y.-Z. Song, X. Li, C. C. Loy, and X. Wang, "A survey on heterogeneous face recognition," *Image Vision Comput.*, vol. 56, no. C, pp. 28–48, Dec. 2016.

[87] B. F. Klare, S. S. Bucak, A. K. Jain, and T. Akgul, "Towards automated caricature recognition," in *5th IAPR International Conference on Biometrics (ICB)*, March 2012, pp. 139–146.

[88] M. Eitz, K. Hildebrand, T. Boubekeur, and M. Alexa, "An evaluation of descriptors for large-scale image retrieval from sketched feature lines," *Computers & Graphics*, vol. 34, no. 5, pp. 482–498, 2010.

[89] M. Eitz, J. Hays, and M. Alexa, "How do humans sketch objects?" *ACM Trans. Graph. (Proc. SIGGRAPH)*, vol. 31, no. 4, pp. 44:1–44:10, 2012.

[90] Q. Yu, Y. Yang, F. Liu, Y.-Z. Song, T. Xiang, and T. M. Hospedales, "Sketch-a-net: A deep neural network that beats humans," *International Journal of Computer Vision*, pp. 1–15, 2016.

[91] P. Sangkloy, N. Burnell, C. Ham, and J. Hays, "The sketchy database: Learning to retrieve badly drawn bunnies," *ACM Trans. Graph.*, vol. 35, no. 4, pp. 119:1–119:12, Jul. 2016.

[92] M. A. Nielsen, *Neural Networks and Deep Learning*. Determination Press, 2015. [Online]. Available: http://neuralnetworksanddeeplearning.com/index.html

[93] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," in *British Machine Vision Conference*, 2015.

[94] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, 2012.

[95] J. Hstad, "On the correlation of parity and small-depth circuits," *SIAM Journal on Computing*, vol. 43, no. 5, pp. 1699–1708, 2014. [Online]. Available: http://dx.doi.org/10.1137/120897432

[96] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning Representations by back-propagating errors," *Nature*, vol. 353, no. 6088, pp. 533–536, October 1986.

[97] G. E. Hinton and T. J. Sejnowski, "Learning and Relearning in Boltzmann Machines," *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, pp. 282–317, 1986.

[98] G. E. Hinton, S. Osindero, and Y. W. Teh, "A Fast Learning Algorithm for Deep Belief Nets," *A Fast Learning Algorithm for Deep Belief Nets*, vol. 18, no. 7, pp. 1527–1554, 2006.

[99] Y. Bengio, "Practical recommendations for gradient-based training of deep architectures," *CoRR*, vol. abs/1206.5533, 2012.

[100] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proceedings of the International Conference on Artificial Intelligence and Statistics (AISTATS10). Society for Artificial Intelligence and Statistics*, 2010.

[101] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Improving neural networks by preventing co-adaptation of feature detectors," *CoRR*, vol. abs/1207.0580, 2012.

[102] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *arXiv CoRR*, vol. abs/1502.03167, 2015. [Online]. Available: http://arxiv.org/abs/1502.03167

[103] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov 1998.

[104] A. Vedaldi and K. Lenc, "MatConvNet – Convolutional Neural Networks for MATLAB," in *Proceeding of the ACM Int. Conf. on Multimedia*, 2015.

[105] NVIDIA, "Nvidia cudnn." [Online]. Available: https://developer.nvidia.com/cudnn

[106] A. Ng, J. Ngiam, C. Y. Foo, Y. Mai, and C. Suen, "Ufldl tutorial," 2013. [Online]. Available: http://deeplearning.stanford.edu/wiki/index.php/UFLDL_Tutorial

[107] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "DeepFace: Closing the Gap to Human-Level Performance in Face Verification," in *IEEE Conference on Computer Vision and Pattern Recognition*, June 2014, pp. 1701–1708.

[108] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "Web-scale training for face identification," *CoRR*, vol. abs/1406.5266, 2014. [Online]. Available: http://arxiv.org/abs/1406.5266

[109] Y. Sun, X. Wang, and X. Tang, "Deep learning face representation from predicting 10,000 classes," in *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition*, ser. CVPR '14. Washington, DC, USA: IEEE Computer Society, 2014, pp. 1891–1898.

[110] Y. Sun, Y. Chen, X. Wang, and X. Tang, "Deep learning face representation by joint identification-verification," in *Proceedings of the 27th International Conference on Neural Information Processing Systems*, ser. NIPS'14. Cambridge, MA, USA: MIT Press, 2014, pp. 1988–1996.

[111] Y. Sun, X. Wang, and X. Tang, "Deeply learned face representations are sparse, selective, and robust," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015, pp. 2892–2900.

[112] Y. Sun, D. Liang, X. Wang, and X. Tang, "Deepid3: Face recognition with very deep neural networks," *CoRR*, vol. abs/1502.00873, 2015.

[113] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015, pp. 815–823.

[114] J. C. Chen, R. Ranjan, A. Kumar, C. H. Chen, V. M. Patel, and R. Chellappa, "An End-to-End System for Unconstrained Face Verification with Deep Convolutional Neural Networks," in *IEEE International Conference on Computer Vision Workshop (ICCVW)*, Dec 2015, pp. 360–368.

[115] R. Ranjan, S. Sankaranarayanan, C. D. Castillo, and R. Chellappa, "An all-in-one convolutional neural network for face analysis," in *IEEE International Conference on Automatic Face Gesture Recognition (FG 2017)*, May 2017, pp. 17–24.

[116] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," University of Massachusetts, Tech. Rep., 2007.

[117] J. Hsu, "Finding one face in a million," July 2016. [Online]. Available: http://spectrum.ieee.org/computing/software/finding-one-face-in-a-million

[118] B. F. Klare, B. Klein, E. Taborsky, A. Blanton, J. Cheney, K. Allen, P. Grother, A. Mah, M. Burge, and A. K. Jain, "Pushing the frontiers of unconstrained face detection and recognition: IARPA Janus Benchmark A," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015, pp. 1931–1939.

[119] D. Wang, C. Otto, and A. K. Jain, "Face search at scale," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1122–1136, June 2017.

[120] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *CoRR*, vol. abs/1409.1556, 2014. [Online]. Available: http://arxiv.org/abs/1409.1556

[121] P. Mittal, M. Vatsa, and R. Singh, "Composite sketch recognition via deep network - a transfer learning approach," in *International Conference on Biometrics (ICB)*, May 2015, pp. 251–256.

[122] L. Zhang, L. Lin, X. Wu, S. Ding, and L. Zhang, "End-to-end photo-sketch generation via fully convolutional representation learning," in *Proc. ACM Int. Conf. Multimedia Retrieval*, ser. ICMR '15. New York, NY, USA: ACM, 2015, pp. 627–634.

[123] S. Saxena and J. Verbeek, "Heterogeneous Face Recognition with CNNs," *ECCV Workshops on Comput. Vision*, pp. 483–491, 2016.

[124] P. Mittal, A. Jain, G. Goswami, R. Singh, and M. Vatsa, "Recognizing composite sketches with digital face images via ssd dictionary," in *IEEE International Joint Conference on Biometrics (IJCB)*, Sept 2014, pp. 1–6.

[125] D. Zhang, L. Lin, T. Chen, X. Wu, W. Tan, and E. Izquierdo, "Content-adaptive sketch portrait generation by decompositional representation learning," *IEEE Trans. Image Proc.*, vol. 26, no. 1, pp. 328–339, Jan 2017.

[126] J. Choi, A. Sharma, D. Jacobs, and L. Davis, "Data insufficiency in sketch versus photo face recognition," in *2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, June 2012, pp. 1–8.

[127] S. J. Klum, H. Han, B. Klare, and A. K. Jain, "The FaceSketchID System: Matching Facial Composites to Mugshots," Michigan State University, Tech. Rep. MSU-CSE-14-6, 2014.

[128] National Institute of Standards and Technology (NIST), "NIST Special Database 32 - Multiple Encounter Dataset (MEDS)," 2011, last visited on Jan. 30, 2015. [Online]. Available: http://www.nist.gov/itl/iad/ig/sd32.cfm

[129] Chinese University of Hong Kong (CUHK), "CUHK Face Sketch FERET Database," last visited on Feb. 2, 2015. [Online]. Available: http://mmlab.ie.cuhk.edu.hk/archive/cufsf/

[130] D. Yi, Z. Lei, S. Liao, and S. Z. Li, "Learning face representation from scratch," *CoRR*, vol. abs/1411.7923, 2014. [Online]. Available: http://arxiv.org/abs/1411.7923

[131] C. Galea and R. A. Farrugia, "Face Photo-Sketch Recognition using Local and Global Texture Descriptors," in *European Signal Processing Conference (EUSIPCO)*, Budapest, Hungary, Aug. 2016.

[132] C. Galea and R. A. Farrugia, "Forensic face photo-sketch recognition using a deep learning-based architecture," *IEEE Signal Processing Letters*, vol. 24, no. 11, pp. 1586–1590, Nov 2017.

[133] C. Galea and R. A. Farrugia, "Matching Software-Generated Sketches to Face Photos with a Very Deep CNN, Morphed Faces, and Transfer Learning,"

IEEE Transactions on Information Forensics and Security, 2017, accepted for publication.

[134] H.-Y. Chen and S.-Y. Chien, "Eigen-patch: Position-patch based face hallucination using eigen transformation," in *IEEE International Conference on Multimedia and Expo*, July 2014, pp. 1–6.

[135] R. Snelick, M. Indovina, J. Yen, and A. Mink, "Multimodal biometrics: Issues in design and testing," in *Proceedings of the 5th International Conference on Multimodal Interfaces*, ser. ICMI '03, 2003, pp. 68–72.

[136] D. J. Field, "Relations between the statistics of natural images and the response properties of cortical cells," *Optical Society of America*, vol. 4, no. 12, pp. 2379–2394, December 1987.

[137] J. Arrospide and L. Salgado, "Log-Gabor Filters for Image-Based Vehicle Verification," *IEEE Transactions on Image Processing*, vol. 22, no. 6, pp. 2286–2295, June 2013.

[138] D. Boukerroui, J. Noble, and M. Brady, "On the choice of band-pass quadrature filters," *Journal of Mathematical Imaging and Vision*, vol. 21, no. 1-2, pp. 53–80, 2004.

[139] V. Štruc and N. Pavešić, "Gabor-based kernel partial-least-squares discrimination features for face recognition," *Informatica (Vilnius)*, vol. 20, no. 1, p. 115138, 2009.

[140] V. Štruc and N. Pavešić, "The complete gabor-fisher classifier for robust face recognition," *EURASIP Advances in Signal Processing*, vol. 2010, p. 26, 2010.

[141] L. Statistics, "Spearman's rank-order correlation," 2013. [Online]. Available: https://statistics.laerd.com/statistical-guides/spearmans-rank-order-correlation-statistical-guide.php

[142] D. J. Best and D. E. Roberts, "Algorithm as 89: The upper tail probabilities of spearman's rho," *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, vol. 24, no. 3, pp. 377–379, 1975.

[143] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295–307, Feb 2016.

[144] D. Erhan, P.-A. Manzagol, Y. Bengio, S. Bengio, and P. Vincent, "The difficulty of training deep architectures and the effect of unsupervised pre-training," in *Proc. Int. Conf. Artificial Intelligence and Statistics*, vol. 5, 2009, pp. 153–160.

[145] G. Özbulak, Y. Aytar, and H. K. Ekenel, "How Transferable Are CNN-Based Features for Age and Gender Classification?" in *Int. Conf. of the Biometrics Special Interest Group (BIOSIG)*, Sep 2016, pp. 1–6.

[146] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "Imagenet large scale visual recognition challenge," *Int. J. Comput. Vision*, vol. 115, no. 3, pp. 211–252, Dec. 2015.

[147] P. Y. Simard, D. Steinkraus, and J. C. Platt, "Best practices for convolutional neural networks applied to visual document analysis," in *Proc. Int. Conf. Document Analysis and Recognition*, Aug 2003, pp. 958–963.

[148] P. Paysan, R. Knothe, B. Amberg, S. Romdhani, and T. Vetter, "A 3D Face Model for Pose and Illumination Invariant Face Recognition," *Proc. Int. Conf. Adv. Video and Signal based Surveillance (AVSS) for Security, Safety and Monitoring in Smart Environments*, Sep 2009.

[149] A. Bas, W. A. P. Smith, T. Bolkart, and S. Wuhrer, "Fitting a 3D Morphable Model to Edges: A Comparison Between Hard and Soft Correspondences," *Proceedings of the Asian Conference on Computer Vision (ACCV) Workshops*, 2016.

[150] S. Ding, L. Lin, G. Wang, and H. Chao, "Deep Feature Learning with Relative Distance Comparison for Person Re-identification," *Pattern Recogn.*, vol. 48, no. 10, pp. 2993–3003, Oct. 2015.

[151] M. Budnik, E. L. Gutierrez-Gomez, B. Safadi, and G. Qunot, "Learned features versus engineered features for semantic video indexing," in *2015 13th International Workshop on Content-Based Multimedia Indexing (CBMI)*, June 2015, pp. 1–6.

[152] National Institute of Standards and Technology (NIST), "The Color FERET Database version 2," last visited on Mar. 17, 2015. [Online]. Available: http://www.nist.gov/itl/iad/ig/colorferet.cfm

[153] C. Jones and A. Abbott, "Color face recognition by hypercomplex Gabor analysis," in *7th International Conference on Automatic Face and Gesture Recognition (FGR 2006)*, April 2006, pp. 126–131.

[154] B. George, S. J. Gibson, M. Maylin, and C. Solomon, "EFIT-V - Interactive Evolutionary Strategy for the Construction of Photo-realistic Facial Composites," in *Proc. Ann. Conf. Genetic and Evolutionary Comput.*, 2008, pp. 1485–1490.

[155] C. D. Frowd, D. Carson, H. Ness, D. McQuiston-Surrett, J. Richardson, H. Baldwin, and P. Hancock, "Contemporary composite techniques: The impact of a forensically-relevant target delay," *Legal and Criminological Psychology*, vol. 10, no. 1, pp. 63–81, 2005.

[156] Corel, "PaintShop Pro X7 Ultimate: photo editing software from Corel," last visited on Mar. 23, 2015. [Online]. Available: http://www.paintshoppro.com/en/products/paintshop-pro/ultimate/

[157] K. Messer, J. Matas, J. Kittler, J. Luettin, and G. Maitre, "XM2VTSdb: The Extended M2VTS Database," in *Proceedings 2nd Conference on Audio and Video-base Biometric Personal Verification (AVBPA99)*, ser. Lecture Notes in Computer Science. Springer Verlag, 1999.

[158] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker, "Multi-PIE," in *8th IEEE International Conference on Automatic Face Gesture Recognition*, Sept 2008, pp. 1–8.

[159] H. Chang, D.-Y. Yeung, and Y. Xiong, "Super-resolution through neighbor embedding," in *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, June 2004, pp. I–I.

[160] N. Poh, C.-H. Chan, J. Kittler, J. Fierrez, and J. Galbally, "D3.3: Description of Metrics For the Evaluation of Biometric Performance," Biometrics Evaluation and Testing, Tech. Rep., August 2012.

[161] A. Mahendran and A. Vedaldi, "Understanding Deep Image Representations by Inverting Them," in *Proceedings of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2015.

[162] L. van der Maaten and G. Hinton, "Visualizing high-dimensional data using t-sne," *Journal of Machine Learning Research*, pp. 2579–2605, Nov 2008.

[163] P. Campisi, *Security and Privacy in Biometrics*. Springer-Verlag London Limited, 2013.

[164] L. Wolf, T. Hassner, and I. Maoz, "Face recognition in unconstrained videos with matched background similarity," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2011, pp. 529–534.

[165] A. Tewari, M. Zollhöfer, H. Kim, P. Garrido, F. Bernard, P. Pérez, and C. Theobalt, "MoFA: Model-based Deep Convolutional Face Autoencoder for Unsupervised Monocular Reconstruction," *CoRR*, vol. abs/1703.10580, 2017. [Online]. Available: http://arxiv.org/abs/1703.10580

[166] L. Best-Rowden, H. Han, C. Otto, B. F. Klare, and A. K. Jain, "Unconstrained face recognition: Identifying a person of interest from a media collection," *Trans. Info. For. Sec.*, vol. 9, no. 12, pp. 2144–2157, Dec. 2014.

[167] L. Lin, G. Wang, W. Zuo, X. Feng, and L. Zhang, "Cross-Domain Visual Matching via Generalized Similarity Measure and Feature Learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1089–1102, June 2017.

[168] W. Härdle and L. Simar, *Canonical Correlation Analysis*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2003, pp. 361–372.

[169] B. F. Klare, S. Klum, J. C. Klontz, E. Taborsky, T. Akgul, and A. K. Jain, "Suspect identification based on descriptive facial attributes," in *IEEE International Joint Conference on Biometrics*, Sept 2014, pp. 1–8.

[170] H. Han, A. K. Jain, S. Shan, and X. Chen, "Heterogeneous face attribute estimation: A deep multi-task learning approach," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PP, no. 99, pp. 1–1, 2017.

[171] M. L. McHugh, "Multiple Comparison Analysis testing in ANOVA," 2011, last visited on Feb. 4, 2016. [Online]. Available: http://www.biochemia-medica.com/2011/21/203

[172] MathWorks, "Multiple comparisons," last visited on Feb. 4, 2016. [Online]. Available: http://www.mathworks.com/help/stats/multiple-comparisons.html#bum7ugv-1