# A Large-Scale Software-Generated Face Composite Sketch Database

Christian Galea
Department of Communications & Computer Engineering
University of Malta
Msida, MSD2080
Malta, Europe
Email: christian.galea.09@um.edu.mt

Reuben A. Farrugia
Department of Communications & Computer Engineering
University of Malta
Msida, MSD2080
Malta, Europe
Email: reuben.farrugia@um.edu.mt

***Abstract*—Numerous algorithms that can identify suspects depicted in sketches following eyewitness descriptions of criminals are currently being developed because of their potential importance in forensics investigations. Yet, despite the prevalent use of software-generated composite sketches by law enforcement agencies, there still exist few such sketches which can be used by researchers to adequately evaluate face photo-sketch recognition algorithms when using these composites. The main contribution of this paper is the creation of the University of Malta Software-Generated Face Sketch (UoM-SGFS) database that is publicly available and which contains the largest number of viewed software-generated sketches, that also exhibit several deformations and exaggerations to mimic sketches obtained in real-world investigations. Further, in contrast to other databases, all sketches in this new database are represented in colour. Lastly, state-of-the-art recognition algorithms are found to perform worse on the software-generated composites than on hand-drawn sketches, while recognition accuracies still lag far behind those achieved for traditional photo-to-photo comparisons.**

***Index Terms*—Software-generated composite sketches, dataset, face photos.**

## I. Introduction

Sketches created from eyewitness descriptions of criminals are typically disseminated to the media and law enforcement officers so that any persons recognising the suspect come forward with information leading to an arrest. Automated methods that can compare these sketches with the mug-shot photographs maintained by law enforcement agencies can improve apprehension times, chances of success and utilisation of readily available resources [1]. Sketches can either be hand-drawn by forensic sketch artists, thereby termed *hand-drawn* composite sketches, or created with computer software, thus known as *software-generated* composite sketches. In addition, there are three types of composite sketches [2], [3]: the first is *forensic* sketches as obtained from eyewitness descriptions in real-world criminal investigations, whose availability for researchers is typically limited. This problem is circumvented by creating sketches whilst viewing a subject or face photo, termed *viewed* sketches, of which a virtually unlimited number can be generated and have in fact been the predominant means of evaluating face photo-sketch recognition algorithms. Lastly, *semi-forensic* sketches are created when forensic sketch artists view a subject or face photograph for a few minutes and then create the sketch from memory.

Unfortunately, most publicly available datasets contain the less popular hand-drawn sketches rather than software-generated sketches [4], [2]. Thus, the contributions of this paper are twofold: first, the University of Malta Software-Generated Face-Sketch (UoM-SGFS) database containing 600 software-generated face sketches of 300 subjects is described, with sketches having several deformations and exaggerations to mimic real-world forensic sketches. To the best of the authors' knowledge, this new database (i) is the largest publicly available software-generated sketch database, and (ii) is the only dataset containing sketches represented in colour, enabling the use of colour information for recognition. The final contribution is the evaluation of leading algorithms on this new dataset, where it is shown that sketch recognition still lags behind photo-to-photo face recognition.

The rest of this paper is organised as follows: in Section II, an overview of related databases and algorithms designed for face photo-sketch recognition are outlined, followed by a description of the new database in Section III. The methodology used to evaluate algorithms on the UoM-SGFS database and their results are given in Section IV. Concluding remarks and directions for future work are finally given in Section V.

## II. Related Work

A brief overview of face photo-sketch recognition algorithms and the datasets on which they are evaluated will now be given hereunder.

### A. Existing face photo-sketch databases

There exist five main publicly available databases for face photo-sketch recognition. The CUFS [5], [6] and CUFSF databases [7] contain 606 and 1194 viewed hand-drawn sketches, respectively, with the subjects in the CUFSF database taken from the FERET database (which has been superseded by the Color FERET database[3] [8]). Also, sketches in the CUFSF database contain several deformations and exaggerations to make them closer to real-life sketches [7]. The IIIT-D sketch database [3] contains 238 viewed, 190 forensic and

[3]Available at: http://www.nist.gov/itl/iad/ig/colorferet.cfm

©2016 Gesellschaft für Informatik e.V., Bonn, Germany

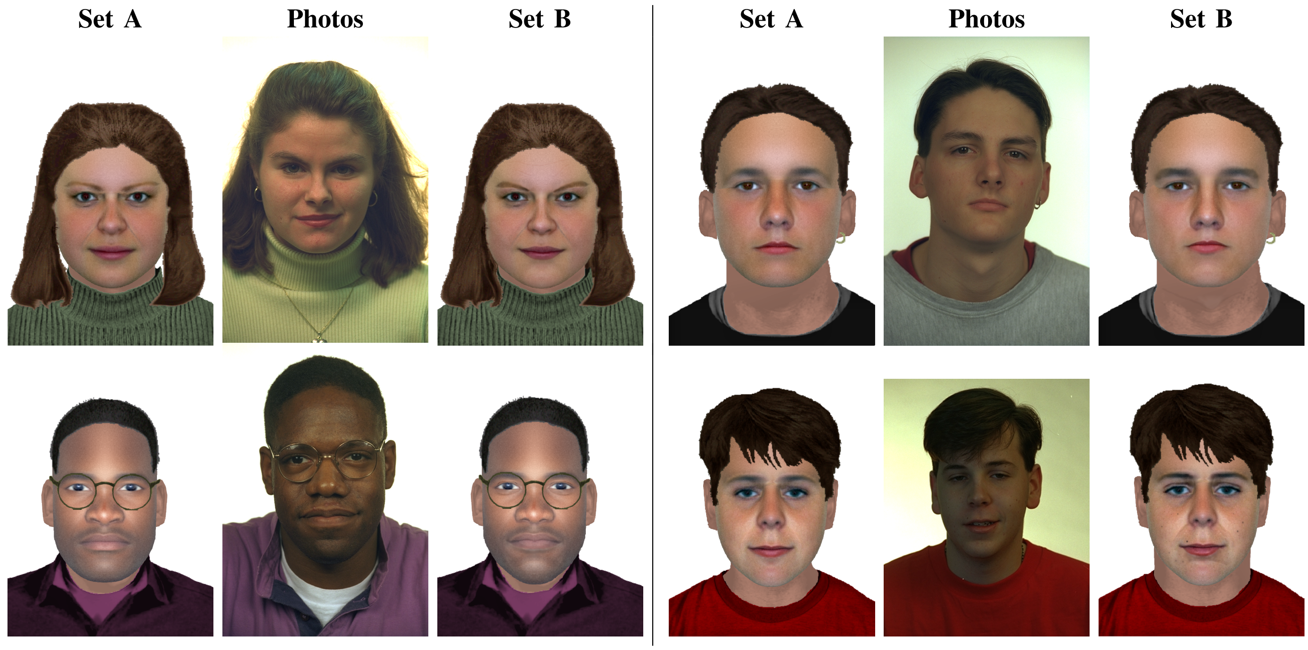



Fig. 1: Photos of four subjects from the Color FERET database and the corresponding sketches from the two sets of the UoM-SGFS database

140 semi-forensic hand-drawn sketches of subjects, although only 55 of the forensic sketches are publicly available. The Hand-Drawn Composite (PRIP-HDC) dataset [1] contains hand-drawn forensic sketches, although only 47 sketches are available, while the PRIP Viewed Software-Generated Composite (PRIP-VSGC) dataset [2], [1] contains 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 [9], [10] 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 instead use the sketches generated by a different operator and a different program for algorithm evaluation, due to better representational accuracy [2].

### B. *Face photo-sketch recognition algorithms*

Traditional Face Recognition Systems (FRSs) such as Eigenfaces [11] and the VGG-Face algorithm [12] that has achieved state-of-the-art performance for unconstrained face recognition can be used for identification of faces depicted in sketches. However, as shown in Section IV, they do not achieve good performance since the images to be compared reside in different modalities. Consequently, methods have been designed specifically for face photo-sketch recognition. An overview of leading state-of-the-art face photo-sketch recognition algorithms can be found in [7], [13], [14], [15].

## III. UoM-SGFS Database

The new UoM-SGFS database[4] contains software-generated sketches of 300 subjects in the Color FERET database, created using the EFIT-V software [16]. An overview of this software will first be given, followed by a more detailed description of the new database.

### A. *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 [17]. 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 if they appear similar or different to the criminal, respectively. 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.

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

EFIT-V allows features other than the facial components to be depicted, including jewellery and glasses which may not only help the eyewitness during the composite generation process but also allow the creation of sketches which can depict discriminative and therefore critical information about the perpetrators quite well. Clothes and the general appearance of the criminal's physique can also be modified. Image editing software is directly supported in EFIT-V, which can be used to fine-tune details [18].

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 [18]. As a result, the enhanced EFIT-V is nowadays used by numerous law enforcement agencies worldwide. More information may be found in [17], [16].

### B. Database details

The face photographs for the UoM-SGFS database were obtained from the Color FERET database[3] [8], 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 300 subjects considered and is partitioned into two sets, where each set contains the sketch of one subject. *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 using an image editing program to fine-tune details which cannot be easily modified with EFIT-V, yielding *Set B*. Hence, sketches in Set B are generally closer in appearance to the original face photos.

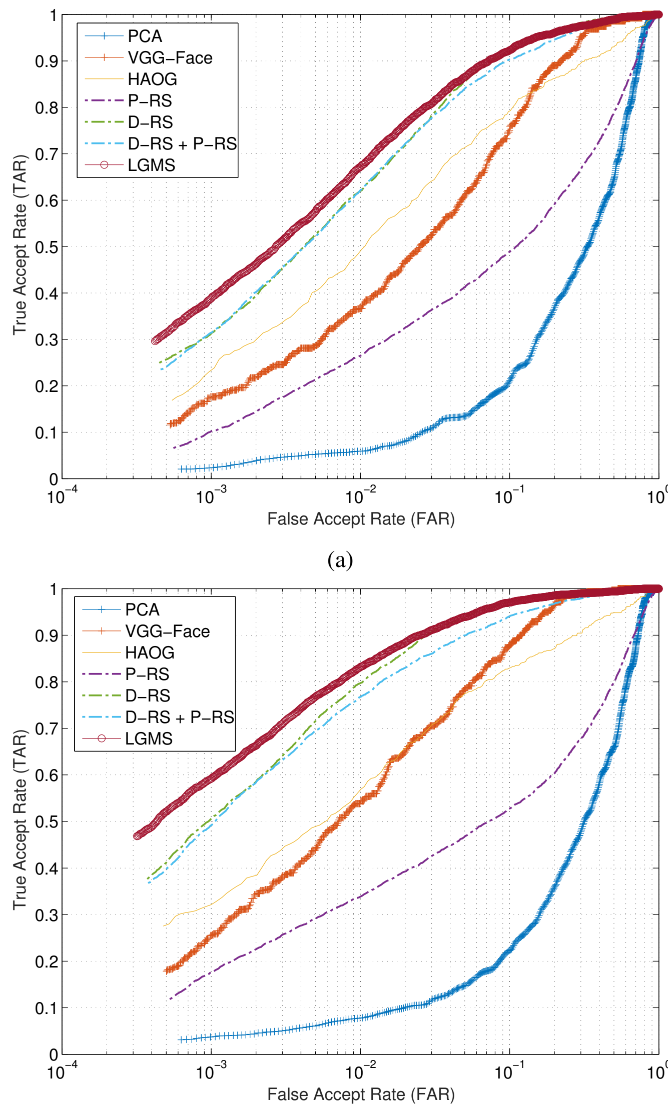
EFIT-V also allows the depiction of shoulders, which can indicate the type of clothes that the perpetrator was wearing and the physique (e.g. fat, muscular, etc.). While the type of clothing is important, more emphasis was given to the physique of the subject since it provides more salient information. 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 [19], 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 1.

## IV. Implementation Methodology & Results

Some of the most popular and state-of-the-art algorithms are evaluated on both sets of the UoM-SGFS database individually. The protocol used and results will now be given.

### A. Evaluation Methodology

A similar evaluation methodology to that employed in [15] is adopted, with the following differences: firstly, subjects are partitioned into a testing set and a training set only. An




(a)

(b)

Fig. 2: Receiver Operating Characteristics (ROC) curves for methods operating on (a) Set A, (b) Set B

TABLE 1a: Means and standard deviations over 25 train/test set-splits of rank-retrieval rates for Set A.

| Algorithm | Matching Rate (%) at Rank-$N$ | | | | |
|---|---|---|---|---|---|
| | $N$=1 | $N$=10 | $N$=50 | $N$=100 | $N$=150 |
| Eigenfaces [11] | 2.03±0.68 | 5.33±1.23 | 9.89±1.29 | 15.65±1.84 | 20.56±1.74 |
| VGG-Face [12] | 11.57±1.51 | 33.25±2.91 | 53.65±3.91 | 65.15±3.86 | 72.21±2.85 |
| HAOG [20] | 15.92±2.70 | 44.93±2.29 | 65.20±2.14 | 73.25±1.95 | 77.92±2.05 |
| P-RS [21] | 6.56±2.17 | 25.49±3.60 | 50.24±4.36 | 64.00±4.61 | 71.47±4.37 |
| D-RS [22], [21] | 24.93±2.35 | 55.28±3.15 | 79.79±2.53 | 87.95±2.09 | 90.91±1.81 |
| D-RS+P-RS [21] | 23.55±2.66 | 59.33±4.05 | 83.31±2.41 | 90.32±1.73 | 93.28±1.62 |
| LGMS [15] | 29.63±2.64 | 60.40±3.09 | 82.29±1.89 | 87.81±1.75 | 91.20±1.72 |

TABLE 1b: Means and standard deviations over 25 train/test set-splits of rank-retrieval rates for Set B.

| Algorithm | Matching Rate (%) at Rank-$N$ | | | | |
|---|---|---|---|---|---|
| | $N$=1 | $N$=10 | $N$=50 | $N$=100 | $N$=150 |
| Eigenfaces [11] | 3.04±0.67 | 6.37±1.35 | 12.05±1.59 | 16.77±1.42 | 21.57±1.70 |
| VGG-Face [12] | 17.41±2.34 | 47.76±3.04 | 70.24±2.88 | 78.69±2.86 | 84.00±2.58 |
| HAOG [20] | 26.93±2.79 | 54.67±2.49 | 69.31±1.94 | 77.39±1.65 | 81.33±1.55 |
| P-RS [21] | 11.84±2.99 | 34.56±4.95 | 59.65±5.24 | 70.91±4.55 | 77.97±3.86 |
| D-RS [22], [21] | 37.63±3.37 | 72.29±2.52 | 91.07±2.18 | 95.71±1.28 | 97.33±0.94 |
| D-RS+P-RS [21] | 36.75±3.91 | 75.28±3.82 | 91.92±2.11 | 96.11±1.17 | 97.57±1.09 |
| LGMS [15] | 46.83±2.92 | 77.25±2.61 | 89.79±1.73 | 93.87±1.60 | 95.71±1.12 |

additional set to train intra-modality methods is not used and therefore such algorithms are not evaluated since, as shown in [14], [15], they consistently lag behind inter-modality methods. Secondly, the number of subjects used for training is 150, and the remaining 150 subjects are used for testing. Hence, the test gallery set contains $150 + 1522 = 1672$ subjects. Thirdly, twenty-five cross-validation folds are performed. Finally, the recent VGG-Face [12] FRS is added in this work to evaluate the performance of an algorithm that has achieved state-of-the-art performance for unconstrained face identification and verification. The parameters of all algorithms are the same as those described in literature, and the provided pre-trained model was used in the case of the VGG-Face algorithm.

### *B. Results*

The results of the algorithms considered are shown in Figure 2, and Tables 1a and 1b. As expected, all algorithms perform better on Set B than Set A. This is because Set B contains sketches that were modified using an image editing software program to make the sketches closer in terms of appearance to the corresponding photographs. In addition, the significant increase in performance indicates the usefulness of employing image editing software to fine-tune details according to the eye-witnesses' requests. However, the relatively low recognition rates at lower ranks indicates that Set B still poses a challenge for all algorithms, due to the deformations and shape exaggerations present.

The performance of all algorithms when using the software-generated composites can also be compared to their performance on hand-drawn sketches in the CUFSF database, as reported in [15]. Similar trends in performance between algorithms are present, i.e. the Eigenfaces FRS performs worse than the inter-modality methods and the LGMS approach is statistically superior at the 95% confidence level using multi-comparison Analysis of Variance (ANOVA) to the other algorithms at lower ranks, while performance is comparable to D-RS and the fusion of P-RS and D-RS at higher ranks. However, all algorithms achieve lower performance when using the software-generated composites. This is mainly due to: (i) the different sketch types (software-generated vs. hand-drawn), (ii) the lower number of software-generated sketches, and (iii) differences in the extent of deformations and exaggerations present in the CUFSF and UoM-SGFS databases.

Also noteworthy is the performance of the VGG-Face algorithm, which achieved over 90% accuracy for unconstrained face recognition [12]. However, the performance is significantly degraded on the composite sketches, highlighting the challenges of not only the differences in modalities but also the inaccuracies present in sketches. However, VGG-Face still outperforms P-RS on both sets of the UoM-SGFS database and achieved performance similar to HAOG at higher ranks on Set B. This is likely because the sketches generated using EFIT-V are quite photo-like, thereby facilitating the task of this state-of-the-art FRS that was designed to operate on face photographs. Nevertheless, its inferior performance compared to D-RS, D-RS + P-RS and LGMS indicate that the use of dedicated photo-sketch recognition algorithms is still the optimal approach.

## V. CONCLUSIONS & FUTURE WORK

A new database containing 600 EFIT-V-generated composite sketches of 300 subjects has been presented. This database is the largest publicly available dataset containing software-generated composite sketches, enabling more robust evaluation of face photo-sketch recognition algorithms to be performed. In addition, it is the only database containing sketches depicted in full colour rather than grayscale. Therefore, future work can include an investigation into the use of colour information for photo-sketch recognition and the use of physique information as a feature that can assist in recognition.

## Acknowledgements

The research work disclosed in this publication is partly funded by the Malta Government Scholarship Scheme. The authors would like to thank the Malta Police Force for their assistance in this research.

## References

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

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

[3] H. S. Bhatt, S. Bharadwaj, R. Singh, and M. Vatsa, "Memetically optimized mcwld for matching sketches with digital face images," *IEEE Trans. Inf. Forensics Security*, vol. 7, no. 5, pp. 1522–1535, Oct 2012.

[4] D. Mcquiston, L. Topp, and R. Malpass, "Use of facial composite systems in US law enforcement agencies," *Psychology, Crime and Law*, vol. 12, no. 5, pp. 505–517, 2006.

[5] X. Tang and X. Wang, "Face sketch recognition," in *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 1, January 2004, pp. 50–57.

[6] X. Wang and X. Tang, "Face photo-sketch synthesis and recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 11, pp. 1955–1967, 2009.

[7] W. Zhang, X. Wang, and X. Tang, "Coupled information-theoretic encoding for face photo-sketch recognition," in *IEEE Conf. Comput. Vision Pattern Recog.*, 2011, pp. 513–520.

[8] 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.

[9] 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.

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

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

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

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

[14] C. Galea and R. A. Farrugia, "Fusion of intra- and inter-modality algorithms for face-sketch recognition," in *Computer Analysis of Images and Patterns*, vol. 9257, 2015, pp. 700–711.

[15] C. Galea and R. A. Farrugia, "Face Photo-Sketch Recognition using Local and Global Texture Descriptors," in *Proc. Int. Conf. European Signal Processing Conference (EUSIPCO)*, Aug. 2016, accepted for publication.

[16] VisionMetric, About EFIT-V. [Online]. Available: http://www.visionmetric.com/products/about-e-fit/

[17] 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.

[18] 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.

[19] C. Jones and A. Abbott, "Color face recognition by hypercomplex Gabor analysis," in *Proc. Int. Conf. Automatic Face and Gesture Recognition*, Apr 2006, pp. 126–131.

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

[21] B. Klare and A. Jain, "Heterogeneous face recognition using kernel prototype similarities," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 6, pp. 1410–1422, Jun 2013.

[22] B. Klare and A. K. Jain, "Heterogeneous Face Recognition: Matching NIR to Visible Light Images," in *Proc. Int. Conf. Pattern Recognition (ICPR)*, Aug 2010, pp. 1513–1516.