# Forensic Face Photo-Sketch Recognition Using a Deep Learning-Based Architecture

Christian Galea, *Student Member, IEEE*, and Reuben A. Farrugia, *Member, IEEE*

***Abstract*—Numerous methods that automatically identify subjects depicted in sketches as described by eyewitnesses have been implemented, but their performance often degrades when using real-world forensic sketches and extended galleries that mimic law enforcement mug-shot galleries. Moreover, little work has been done to apply deep learning for face photo-sketch recognition despite its success in numerous application domains including traditional face recognition. This is primarily due to the limited number of sketch images available, which are insufficient to robustly train large networks. This letter aims to tackle these issues with the following contributions: 1) a state-of-the-art model pre-trained for face photo recognition is tuned for face photo-sketch recognition by applying transfer learning, 2) a three-dimensional morphable model is used to synthesise new images and artificially expand the training data, allowing the network to prevent over-fitting and learn better features, 3) multiple synthetic sketches are also used in the testing stage to improve performance, and 4) fusion of the proposed method with a state-of-the-art algorithm is shown to further boost performance. An extensive evaluation of several popular and state-of-the-art algorithms is also performed using publicly available datasets, thereby serving as a benchmark for future algorithms. Compared to a leading method, the proposed framework is shown to reduce the error rate by 80.7% for viewed sketches and lowers the mean retrieval rank by 32.5% for real-world forensic sketches.**

***Index Terms*—Augmentation, convolutional neural network, deep learning, fusion, hand-drawn sketch, morphological model.**

## I. Introduction

ONE of the toughest heterogeneous face recognition (HFR) scenarios, involving the comparison of face images residing in different modalities, is face photo-sketch recognition. Apart from the significant modality gap, algorithms must also contend with inaccuracies in sketch images arising from memory and communication deficiencies when an eyewitness provides the description of a suspect to a sketch artist [1]–[3]. These issues cause traditional face reocgnition systems (FRSs) to perform poorly when tasked with identifying the subject in a sketch given a gallery of photo images, leading to the development of algorithms specifically designed for face photo-sketch recognition. Although several methods described in literature have reportedly achieved high retrieval rates, these were typically achieved by using sketches that bear a very high resemblance to the original photos (therefore, ignoring distortions caused by the memory and communication gaps) and failing to use an extended gallery to simulate the mug-shot galleries maintained by law enforcement agencies.

Several state-of-the-art methods utilise hand-crafted features, such as the scale-invariant feature transform (SIFT) and multiscale local binary pattern (MLBP) [2]–[5]. However, these features are likely not optimal since they were not developed for intermodality face recognition [6], and it would therefore be desirable to employ descriptors that are better suited for the task of face photo-sketch recognition. An approach that can be used to learn appropriate descriptors is via *deep learning*, which has proven to be successful in numerous domains including traditional face recognition [7]–[9]. However, there has been limited work in using deep learning for face photo-sketch recognition. One of the main reasons is the need of a large number of examples to robustly train deep networks and avoid issues, such as over-fitting and local minima [9], [10], but there are relatively few photo-sketch pairs that are publicly available. Moreover, there is typically only one sketch per subject, making it hard for a deep network to learn robust features [11]. The contributions of this letter are thus:

1) To circumvent the single-sketch-per-subject problem, a three-dimensional (3-D) Morphable model is employed to vary facial attributes and automatically synthesise a new large set of images.
2) The synthetic images are used to tune a state-of-the-art deep network (pretrained on face photos) for the task of face photo-sketch recognition via transfer learning.
3) Since forensic sketches often contain several inaccuracies, the synthetic sketches can bear a better liking to the matching photo than the original sketch. In fact, performance is improved when multiple sketches for each subject are used for comparison with the gallery photos.
4) The fusion of the proposed architecture with a leading algorithm is shown to yield further improved performance on both viewed and forensic sketches.

The rest of this letter includes a summary of related work in Section II followed by descriptions of the proposed methods in Section III, which are then evaluated in Section IV. Directions for future work and concluding remarks are given in Section V.

## II. Related Work

Several face hallucination (FH) techniques [13] have been proposed in literature, which encompass *intra-modality* methods that transform photos and sketches such that comparison can be done within the same domain. Prominent methods include Eigen-transformation (ET) [14] that performs synthesis

Manuscript received May 25, 2017; revised July 22, 2017; accepted August 21, 2017. Date of publication September 5, 2017; date of current version September 22, 2017. The research work disclosed in this publication is partly funded by the Malta Government Scholarship Scheme. *(Corresponding author: Christian Galea.)*

The authors are with the Department of Communications and Computer Engineering, University of Malta, Msida, MSD 2080, Malta (e-mail: christian.galea.09@um.edu.mt; reuben.farrugia@um.edu.mt).

This letter has supplementary downloadable material available at http://ieeexplore.ieee.org, provided by the author.

Color versions of one or more of the figures in this letter are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LSP.2017.2749266


Fig. 1. Proposed architecture, where synthetic images are created and used to train the DCNN in [9] 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 [12]. Column ‘1’ contains images fitted with a 3-D 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). As shown in the yellow box, the DCNN is first trained for classification and then tuned for verification using triplet embedding.

using a linear combination of images, the Eigen-patches (EP) extension [15] performing synthesis at a local level, and the Bayesian framework in [16] that considers relationships among neighbouring patches for model construction. A more thorough review of FH algorithms may be found in [4], [13], [15]–[17].

State-of-the-art *inter-modality* methods that learn or extract modality-invariant features include the D-RS approach [2], [18] that compares SIFT and MLBP descriptors extracted from images that are convolved with three filters, the CBR method [19], which compares MLBP features extracted from individual facial components, the FaceSketchID system in [12] which fuses D-RS with CBR, and the recent LGMS method [4] that compares MLBP features extracted from log-Gabor-filtered images using the spearman rank-order correlation coefficient. Other methods and further information can be found in [4], [6], [12], [13], [15], [19]–[23].

Few works have considered the use of deep learning for face photo-sketch synthesis and recognition, most notable being the approaches in [24]–[27]. However, these systems generally use relatively shallow networks or are primarily trained using images residing in a single modality (typically face photos).

Finally, few works consider the use of multiple sketches per subject. Most relevant to this letter is the work done in [21], [28], 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.

## III. Proposed Method

### A. Deep Convolutional Neural Network

As shown in Fig. 1, the proposed architecture consists of a deep convolutional neural network (DCNN) on which transfer learning is applied using both photos and sketches[1] to allow the network to learn the relationships between the two modalities. The choice to use a pretrained network instead of initialising a new one follows observations in literature that the former approach will enable faster convergence, mitigate the encountering of local minima, and allow better generalisation [32], [33]. The DCNN was chosen to be VGG-Face as described in [9], since 1) it was also designed for recognition of faces, albeit only photos, and 2) it is among the leading FRSs for unconstrained face recognition. Hence, it provides a good starting point from which the parameters can be tuned for face photo-sketch recognition. A similar implementation methodology to [9] is applied for network training and testing, notably bootstrapping the DCNN for classification followed by triplet embedding to enable verification, and scaling the test images to three sizes to enable multiscale testing.

### B. Data Augmentation

Very deep architectures, such as the VGG-Face network, have been shown to be more powerful than shallow networks and are thus crucial to attain good performance [34], [35]. However, such networks contain millions of trainable parameters and tend to suffer from overfitting. Consequently, such networks must be trained using not only a vast number of classes, but also numerous images for each category [8], [9]. Databases for traditional tasks, such as object and face recognition are often constructed by exploiting the numerous images available on the web, but sketches and the corresponding photos are rarely accessible and only one sketch per subject is usually available both in real-life and in publicly available databases. Hence, apart from the original photos and sketches, the 3-D Morphable model[2] in [37] is used to enable variation of the attributes for each face image and synthesise a new set of images with which to train the network. Both individual facial features (eyes, nose, mouth, and face shape) and global features (weight, age, height, and gender) are adjusted, for a total of 645 images (including the original).[3] The resultant method is named the DEEP (face) Photo-Sketch System (DEEPS).

### C. Multiple Synthetic Sketches for Recognition

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 that they may also be useful when determining the identity of a subject during testing. This exploits the observation that attribute adjustments 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 is shown in Fig. 1. Hence, the features of a subset of 199 synthetic sketches

[1]In this letter, the focus is on using hand-drawn sketches due to the availability of both viewed sketches *and* real-world forensic sketches.

[2]Model available: http://faces.cs.unibas.ch/bfm/main.php, fitted using the method in [36] available at: https://github.com/waps101/3DMM_edges

[3]The exact parameters may be found at: http://wp.me/P6CDe8-7D

TABLE I
MEANS AND STANDARD DEVIATIONS OVER 3 TRAIN/TEST-SET SPLITS FOR ALGORITHMS EVALUATED ON VIEWED HAND-DRAWN SKETCHES

| Method | Matching Rate (%) at Rank-$N$ | | | | | TAR@FAR=0.1% | TAR@FAR=1.0% | EER (%) |
|---|---|---|---|---|---|---|---|---|
| | $N=1$ | $N=10$ | $N=50$ | $N=100$ | $N=150$ | | | |
| VGG-Face [9] | 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 |
| PCA [29] | 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 |
| ET (+PCA) [14] | 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 |
| EP (+PCA) [15] | 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 |
| LLE (+PCA) [30] | 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 |
| HAOG [31] | 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 |
| CBR [19] | 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 |
| D-RS [2], [18] | 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 |
| D-RS+CBR [12] | 80.18 ± 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 |
| LGMS [4] | 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 |
| *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 |
| *LGMS [4]+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 |

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: 1) the median of the top 49 matches and the match to the original sketch, and 2) the best match among nine sketches and the original. Two distance values are thus obtained for each comparison of a sketch with a photo, that are combined using min-max normalisation and sum-of-scores fusion which are reportedly among the best fusion approaches [21], [38], [39]. This method is denoted DEEPS multisketch (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.

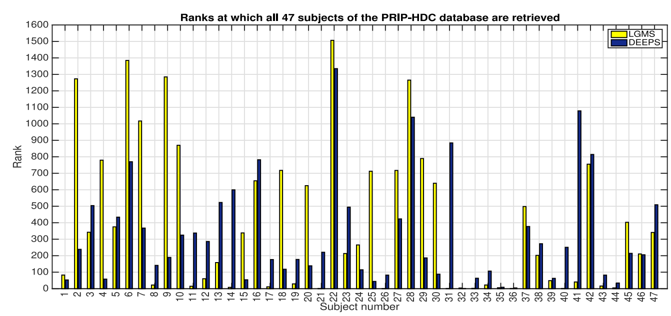


Fig. 2. Ranks of all 47 subjects in the PRIP-HDC database [12] for LGMS and DEEPS. Smaller values are desired.

### D. System Fusion

Due to their substantial quantity, viewed sketches are used to validate the performance of DEEPS with respect to several algorithms proposed in literature. Specifically, the popular CUFS database containing 606 subjects in the AR [40], XM2VTS [41], and CUHK student [14], [42] databases are used together with the sketches of 946 subjects in the CUFSF database [6] and the corresponding photos in the color FERET database [43]. While most works use only CUFS, its sketches bear a great resemblance to the corresponding photos and therefore model the memory and communication gaps inadequately. However, the CUFSF sketches were created to intentionally contain several distortions and shape exaggerations to mimic real-world sketches [6]. CUFS and CUFSF are combined, selecting: (i) 800 subjects at random for training the face recognisers and inter-modality methods, (ii) 350 subjects to train the intra-modality methods, and (iii) 402 subjects for testing. All methods are evaluated on three train/test set splits. Photos and sketches populate the gallery and probe sets, respectively, and the gallery is extended with the photos of 1521 subjects to simulate the extensive mug-shot galleries maintained by law-enforcement agencies, obtained from the MEDS-II,[4] FRGC v2.0,[5] Multi-PIE [44], and FEI[6] databases.

[4]Available at: http://www.nist.gov/itl/iad/ig/sd32.cfm
[5]Available at: http://www.nist.gov/itl/iad/ig/frgc.cfm
[6]Available at: http://fei.edu.br/∼cet/facedatabase.html

As shown in Table I, the FRSs are generally inferior to the intra-modality methods, which in turn typically lag behind the performance of the inter-modality approaches. The only exception is CBR, whose poor performance is likely a result of being designed to operate on software-generated sketches. However, its fusion with D-RS as done in [12] yields improved performance that is second only to the state-of-the-art LGMS method. The proposed approach ultimately outperforms all methods across virtually all performance metrics, except at Rank-1. However, since law enforcement agencies would still examine several tens or hundreds of top matches, the Rank-1 performance is arguably less important than other ranks. The proposed artificial expansion of the training set and application of transfer learning are clearly beneficial given the significantly improved performance compared to the VGG-Face network that was used as the basis of the proposed system.

Empirically, it was also observed that there are several cases where LGMS performs noticeably worse than DEEPS, and vice versa. As shown in Fig. 2, this phenomenon also holds true for the forensic sketches. Hence, the two approaches are combined to determine if these methods can benefit from complementary information, using also min-max normalisation and sum-of-scores fusion [38]. This is indeed the case, with the resultant system comprehensively outperforming all other methods and is able to reduce error rates by 75.7% and 80.7% compared to DEEPS and LGMS, respectively.

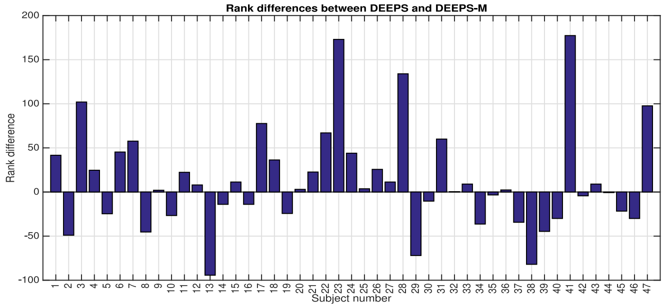


Fig. 3. Rank differences for all 47 subjects in the PRIP-HDC database [12] when comparing DEEPS with DEEPS-M.

| Method | Rank |
|---|---|
| D-RS+CBR [12] | 499.00 |
| LGMS [4] | 82.67 |
| DEEPS | 53.00 |
| DEEPS-M | 11.33 |
| LGMS [4]+DEEPS-M | 22.33 |

| Method | Rank |
|---|---|
| D-RS+CBR [12] | 880.00 |
| LGMS [4] | 342.33 |
| DEEPS | 504.00 |
| DEEPS-M | 402.00 |
| LGMS [4]+DEEPS-M | 275.33 |

Fig. 4. 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 [12].

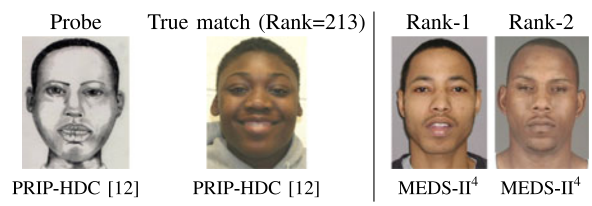


Fig. 5. Example where best matches retrieved by LGMS + DEEPS-M bear a better liking to probe than the true match.

## IV. Results

The PRIP-HDC dataset [12] is used to evaluate the methods considered on 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 available after the suspects in the sketches were identified [1], [2]. 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. Additional results are also available as part of the Supplementary Material.

### A. LGMS Versus DEEPS

Comparing DEEPS to the best-performing method in literature (LGMS, as discussed above) in terms of differences in the ranks at which the identity of the subjects are retrieved as shown in Fig. 2, 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.

### B. DEEPS Versus DEEPS-M

As shown in Fig. 3, 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.

### C. LGMS + DEEPS-M

As discussed in Section III-D, LGMS and DEEPS can provide complementary information when combined in the case of viewed sketches. Fusion of the superior DEEPS-M with LGMS can lead to smaller ranks than either approach as shown in Fig. 4, 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 Fig. 5, 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.

## V. Conclusion & Future Work

This letter presented a face photo-sketch recognition system utilising a 3-D Morphable model to vary facial features and automatically generate new images, circumventing the problem of having only a single sketch per subject thus enabling a deep convolutional neural network to learn the relationships between photos and sketches and exceed the performance of leading methods. The combination of multiple sketches at test-time was also shown to result in improved performance for real-world sketches, since the facial feature variations can yield sketches that are more similar to the matching photo than the original sketch. Hence, this is one of only few works considering deep learning for hand-drawn face sketches and multiple sketches for subject identification. The proposals were also found to be effective for real-world forensic sketches. It was further demonstrated that the fusion of the proposed system with a leading method yields further performance for both viewed and forensic sketches. Future work includes the use of a more advanced Morphable model which allows more flexibility in the variation of the facial features, and the application of the proposed approaches for other HFR tasks.

## Acknowledgment

The authors gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan X Pascal GPU used for this research, and thank Dr. K. Bugeja and Dr. A. Magro at the University of Malta for providing two NVIDIA Tesla K20c GPUs. The authors would also like to thank the Malta Police Force for their assistance in this research.

## References


[1] B. Klare, Z. Li, and A. K. Jain, "Matching forensic sketches to mug shot photos," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 3, pp. 639–646, Mar. 2011.

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

[3] S. Ouyang, T. M. Hospedales, Y. Z. Song, and X. Li, "ForgetMeNot: memory-aware forensic facial sketch matching," in *Proc. Int. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 5571–5579.

[4] C. Galea and R. A. Farrugia, "Face photo-sketch recognition using local and global texture descriptors," in *Proc. Eur. Signal Process. Conf.*, Budapest, Hungary, Aug. 2016, pp. 2240–2244.

[5] C. Galea and R. A. Farrugia, "A large-scale software-generated face composite sketch database," in *Proc. Int. Conf. Biometrics Special Interest Group*, Sep. 2016, pp. 1–5.

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

[7] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "DeepFace: closing the gap to human-level performance in face verification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 1701–1708.

[8] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 815–823.

[9] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," in *Proc. Brit. Mach. Vis. Conf.*, 2015.

[10] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[11] 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 Anal. Recognit.*, Aug. 2003, pp. 958–963.

[12] 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.

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

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

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

[16] N. Wang, X. Gao, L. Sun, and J. Li, "Bayesian face sketch synthesis," *IEEE Trans. Image Process.*, vol. 23, no. 3, pp. 1264–1274, Jan. 2017.

[17] N. Wang, S. Zhang, X. Gao, J. Li, B. Song, and Z. Li, "Unified framework for face sketch synthesis," *Signal Process.*, vol. 130, pp. 1–11, Jan. 2017.

[18] B. Klare and A. K. Jain, "Heterogeneous face recognition: Matching NIR to visible light images," in *Proc. Int. Conf. Pattern Recognit.*, Aug. 2010, pp. 1513–1516.

[19] 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.

[20] S. Ouyang, T. Hospedales, Y.-Z. Song, X. Li, C. C. Loy, and X. Wang, "A survey on heterogeneous face recognition," *Image Vis. Comput.*, vol. 56, pp. 28–48, Dec. 2016.

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

[22] 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.

[23] T. Chugh, M. Singh, S. Nagpal, R. Singh, and M. Vatsa, "Transfer learning based evolutionary algorithm for composite face sketch recognition," in *Proc. IEEE Comput. Vis. Pattern Recognit. Workshop*, Jul. 2017, pp. 619–627.

[24] P. Mittal, M. Vatsa, and R. Singh, "Composite sketch recognition via deep network – A transfer learning approach," in *Proc. Int. Conf. Biometrics*, May 2015, pp. 251–256.

[25] 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*, New York, NY, USA, 2015, pp. 627–634.

[26] S. Saxena and J. Verbeek, "Heterogeneous face recognition with CNNs," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 483–491.

[27] 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 Process.*, vol. 26, no. 1, pp. 328–339, Jan. 2017.

[28] X. Gao, N. Wang, D. Tao, and X. Li, "Face sketch-photo synthesis and retrieval using sparse representation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 8, pp. 1213–1226, Aug. 2012.

[29] M. Turk and A. Pentland, "Eigenfaces for recognition," *J. Cognitive Neurosci.*, vol. 3, no. 1, pp. 71–86, 1991.

[30] H. Chang, D.-Y. Yeung, and Y. Xiong, "Super-resolution through neighbor embedding," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, vol. 1, Jun. 2004, pp. 275–282.

[31] H. Galoogahi and T. Sim, "Inter-modality face sketch recognition," in *Proc. IEEE Int. Conf. Multimedia Expo.*, Jul. 2012, pp. 224–229.

[32] 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. Artif. Intell. Stat.*, 2009, vol. 5, pp. 153–160.

[33] G. Özbulak, Y. Aytar, and H. K. Ekenel, "How transferable are CNN-based features for age and gender classification?," in *Proc. Int. Conf. Biometrics Special Interest Group*, Sep. 2016, pp. 1–6.

[34] Y. Sun, L. Ding, X. Wang, and X. Tang, "Deepid3: Face recognition with very deep neural networks," *CoRR*, abs/1502.00873, 2015.

[35] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. Int. Conf. Learning Representations*, 2015.

[36] 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," in *Proc. Asian Conf. Comput. Vis. Workshops*, 2016, pp. 377–391.

[37] P. Paysan, R. Knothe, B. Amberg, S. Romdhani, and T. Vetter, "A 3D face model for pose and illumination invariant face recognition," in *Proc. Int. Conf. Adv. Video Signal Based Surveillance Security, Safety Monitoring Smart Environ.*, Sep. 2009, pp. 296–301.

[38] R. Snelick, M. Indovina, J. Yen, and A. Mink, "Multimodal biometrics: Issues in design and testing," in *Proc. Int. Conf. Multimodal Interfaces*, 2003, pp. 68–72.

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

[40] A. M. Martinez and R. Benavente, "The AR face database," CVC, Ohio State Univ., Columbus, OH, USA, Tech. Rep., Jun. 1998.

[41] K. Messer, J. Matas, J. Kittler, J. Luettin, and G. Maitre, "XM2VTSdb: The Extended M2VTS Database," in *Proc. Conf. Audio and Video-Based Biometric Personal Verification* 1999, pp. 72–77.

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

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

[44] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker, "Multi-PIE," in *Proc. IEEE Int. Conf. Autom. Face Gesture Recognit.*, Sep. 2008, pp. 1–8.