Face Photo-Sketch Recognition using Local and Global Texture Descriptors

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

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

***Abstract*—The automated matching of mug-shot photographs with sketches drawn using eyewitness descriptions of criminals is a problem that has received much attention in recent years. However, most algorithms have been evaluated either on small datasets or using sketches that closely resemble the corresponding photos. In this paper, a method which extracts Multi-scale Local Binary Pattern (MLBP) descriptors from overlapping patches of log-Gabor-filtered images is used to obtain cross-modality templates for each photo and sketch. The Spearman Rank-Order Correlation Coefficient (SROCC) is then used for template matching. Log-Gabor filtering and MLBP provide global and local texture information, respectively, whose combination is shown to be beneficial for face photo-sketch recognition. Experimental results with a large database show that the proposed approach outperforms state-of-the-art methods, with a Rank-1 retrieval rate of 81.4%. Fusion with the intra-modality approach Eigenpatches improves the Rank-1 rate to 85.5%.**

***Index terms*— face recognition, inter-modality, log-Gabor filter, Spearman correlation, hand-drawn sketches**

I. Introduction

There exist several types of Face Recognition (FR) Systems (FRSs), with traditional FRSs typically operating on photos taken in the visible light spectrum. Much research has recently been devoted to Heterogeneous FR (HFR), where processing is performed over different modalities. One important use of such algorithms is in the comparison of mug-shot photographs with sketches obtained from eyewitness accounts of criminals, which has been described as perhaps the most challenging type of scenario in HFR since face sketches often do not resemble closely the corresponding face photo. This leads to a large modality gap between the images to be compared and in fact normal Commercial Off-the-Shelf (COTS) FRSs have been shown to perform poorly in this scenario [1], [2].

Algorithms proposed in literature to tackle this problem may be classified as either intra- or inter-modality approaches [3], [4]. *Intra*-modality approaches aim to transform a sketch (photo) to a photo (sketch) so that a normal FRS can then be used to match a photo (sketch) with the synthesised photo (sketch), thus reducing the modality gap and facilitating recognition. *Inter*-modality algorithms learn or extract modality-invariant features such that inter-class separability is maximised whilst maintaining intra-class differences [5], [6].

This paper presents an inter-modality approach where photos and sketches are processed with a series of log-Gabor filters to extract *global* texture information. *Local* texture information is extracted at a second stage by applying the Multiscale Local Binary Pattern (MLBP) operator on the resulting log-Gabor filtered images to enable matching of sketches with photos using the Spearman Rank Order Correlation Coefficient (SROCC). It can be shown that the proposed approach provides statistically superior accuracy to state-of-the-art approaches using a large set of images which represents that used by law enforcement agencies better than those typically employed in literature. Hence, the contributions of this paper include (i) the combination of both local and global texture descriptors, (ii) the use of SROCC as a similarity measure, and (iii) an extensive evaluation using a number of state-of-the-art-methods.

The rest of this paper is organised as follows: an overview of related methods found in literature is given in Section II, followed by a detailed description of the proposed method in Section III and its evaluation in Section IV. Conclusions and directions for future work are finally given in Section V.

II. Related Work

Numerous approaches proposed in literature focus on intra-modality algorithms, also known as Face Hallucination (FH) techniques which encompass both face super-resolution and face-sketch synthesis [4]. Some of the best-performing and most popular methods include Eigentransformation (ET) that synthesises whole faces using a linear combination of photos (or sketches) with the assumption that a face photo and the corresponding sketch are similar in appearance, the Eigenpatches (EP) extension in [6] to perform synthesis at a local level, the use of the Locally Linear Embedding (LLE) manifold learning technique in [7] to construct a patch using a linear combination of the nearest patches, and the Multiscale Markov Random Fields (MRF) approach [8] which models the relationships among patches. A more detailed review of FH algorithms may be found in [4]. Intra-modality approaches tend to be complex since they attempt to solve a more challenging problem than recognition itself [3], [5], and FR performance depends on the quality of reconstructed sketches which often contain undesirable artefacts especially when not evaluated using sketches that resemble very closely the original photos.

State-of-the-art inter-modality methods include the Prototype Random Subspace (P-RS) approach proposed in [1], where images are convolved with three filters followed by extraction of Scale-Invariant Feature Transform (SIFT) and MLBP feature descriptors from overlapping patches. The cosine kernel is then used to compare the feature descriptors of a test image with those of a set of prototypes. Feature projection is performed with Random Sampling Linear Discriminant Analysis (LDA) (RS-LDA) and final matching is done using cosine similarity. The approach in [9] was modified in [1] to use a similar methodology to P-RS and is called Direct RS (D-RS) since matching is done using the features directly. The Histogram of Averaged Orientation Gradients (HAOG) method [3] emphasises orientations of regions having high magnitudes, while it was shown in [6] that intra- and inter-modality algorithms can provide complementary information and yield improved recognition accuracy when combined.

III. Proposed Method

The system flow diagram of the proposed method is shown in Figure 1. First, all photos and sketches are aligned such that the eyes and mouth are in the same position for all images, 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 [10] and have been used in a vast number of applications [11]. Their use is motivated by the observation that these filters can model the Human Visual System (HVS) [10], [11] and have yielded good performance within their application domains. However, log-Gabor filters were proposed in [10] to better model natural images, to remove the DC component, and to reduce the number of filter banks required [11], [12]. 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. While both MLBP and log-Gabor filters extract texture information, MLBP characterises the *type* of texture present within local areas. Hence, log-Gabor filtering extracts texture information at a *global* level, while MLBP extracts *local* texture information. Following discriminant analysis, the Spearman Rank Order Correlation Coefficient (SROCC) between the resultant descriptors of the sketches and photos to be compared is found and used as a similarity measure. Whilst not often used for FR, it will be shown that SROCC outperforms popular comparison metrics. Scores are finally normalised and summed to yield the final similarity score. The proposed method is thereby named log-Gabor-MLBP-SROCC (LGMS). Further details will now be given hereunder.

A. Geometric Normalisation

All face sketches and face photos are first rotated 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. Then, all images are tightly cropped to a height of 250 pixels and a width of 200 pixels as shown in Figure 1. Eye and mouth coordinates were either provided in the databases used or marked manually.

B. Image Filtering

Each 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 [11], [12]:

$$LG_{o,s}(f,\theta) = \exp\left(-\frac{\ln^2(f/f_0)}{2\ln(\kappa_\beta)}\right)\exp\left(-\frac{(\theta-\theta_0)^2}{2\sigma_\theta^2}\right) \quad (1)$$

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 $N = 8\times4 = 32$ filters are defined. These values were chosen such that a good balance between performance and computational complexity is provided.

C. 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 therefore 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. In the proposed approach, "uniform" patterns at eight sampling locations as described in [13] are used to obtain 59D vectors 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 yielded the best performance.

Whilst SIFT or Histogram of Orientation Gradients (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, and therefore descriptors of this type do not offer much useful information since shape information is already being implicitly considered.

D. Sketch-Photo Matching

The MLBP features from each patch are concatenated into one $154 \times 236 = 36344$D vector. Discriminant analysis is then performed for each of the N 36344D feature vectors by first applying Principal Component Analysis (PCA) followed by LDA on a training set of images [14], an approach proven to be beneficial in FR [1]. The number of eigenvectors used is the upper bound $c - 1$, where c is the number of classes (subjects) as described in [14]. 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.

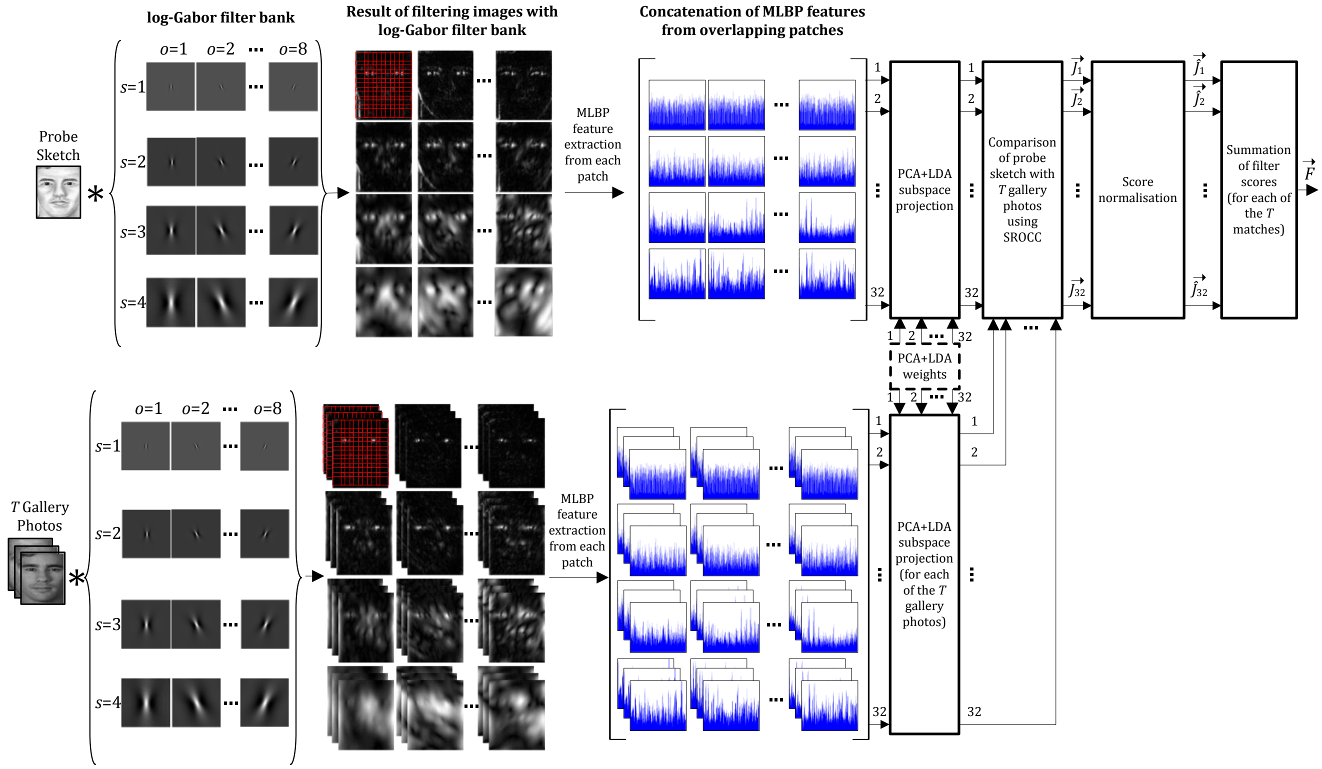

Fig. 1. System flow diagram of the proposed approach for $O = 8$ and $S = 4$. Dotted block represents data obtained from the training stage.

Given a gallery containing T subjects, there are N vectors $\vec{J}_k \in \mathbb{R}^T, k = 1, 2, \ldots, N$ containing the scores for each comparison between a probe sketch and a gallery photo to yield $N \times T$ scores for each probe sketch. Min-max normalisation is applied to obtain $\hat{\vec{J}}_k \in \mathbb{R}^T, k = 1, 2, \ldots, N$ which are finally fused using the sum-of-scores method similar to the approach in [6]. This yields a vector $\vec{F} \in \mathbb{R}^T$ containing T LGMS scores for each probe sketch, representing the similarity between the sketch and all the photos in the gallery. The choice of sum-of-scores fusion and min-max normalisation follows the observation that their combination provides some of the best results for fusion of multi-biometric systems [6], [15].

IV. Evaluation Methodology & Results

The proposed system is compared to several popular and/or state-of-the-art intra- and inter-modality algorithms using 1552 subjects, each having one photo-sketch pair, and an additional 1522 subjects each having one photo in the gallery set. The methodology used and results will now be given hereunder.

A. Evaluation Methodology

Photos form the gallery set while the probe (query) set contains the sketches of subjects whose identity is to be found. For almost all algorithms, a sub-set to train the face recogniser is required while the intra-modality algorithms need an additional set to train the synthesis algorithm. Hence, 300 subjects are selected for each training set similar to the approach in [6], [8] such 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 (when photo-to-sketch synthesis is performed). The remaining 952 subjects are used for testing. The sets are disjoint, such that each subject is used in only one of the three sets.

Since the train/test sets are constructed by random selection of subjects, each algorithm is evaluated on five train/test sets and the mean and standard deviations of results reported. In addition, given that sketches are typically used when heinous crimes are committed, much attention is dedicated by investigators to solve these cases. Hence, the top 50 to 200 subjects are often given importance equal to the best match (Rank-1) in criminal investigations [1], [5]. The Rank-retrieval rates are therefore used to evaluate the algorithms considered, along with the Area under the Receiver Operating Characteristics (ROC) Curve (AuC) that is typically used to evaluate FRSs.

B. Algorithms considered

LGMS is compared with (i) the intra-modality methods Eigentransformation (ET) [16], Eigenpatches (EP) [6] and the Locally-Linear Embedding (LLE) approach in [7] with PCA (Eigenfaces) [17] employed as the face recogniser, (ii) the inter-modality HAOG [3], D-RS [9] and P-RS [1] methods and the fusion of P-RS and D-RS [1], and (iii) the fusion of intra- and inter-modality algorithms (EP, ET and HAOG) proposed in [6]. The performance of PCA alone is also used as a baseline for the intra-modality methods, the results of which are reported for only photo-to-sketch but not sketch-to-photo transformation due to the former's superior performance [6].

C. Databases used

The first sketch database employed is the popular CUFS database[1] [8], [16] consisting of 606 sketches, with the corresponding frontal face photos obtained from the AR [18], XM2VTS [19] and CUHK student[1] [8], [16] databases. Another 946 sketches are derived from the CUFSF database[2] [20], with the corresponding face photos obtained from the ColorFERET database[3] [21]. Both the CUFS and CUFSF databases contain viewed hand-drawn sketches, where an artist created the sketch whilst viewing a subject or his photograph.

The gallery set is extended further with the photos of 1522 subjects to more closely mimic the mug-shot galleries maintained by law-enforcement agencies. These include 510 subjects from the MEDS-II database[4], 476 subjects from the FRGC v2.0 database[5], 337 subjects from the Multi-PIE database [22], and 199 subjects from the FEI database[6]. As a result, the test gallery set contains $T = 2474$ subjects.

D. Results

The results of the metrics considered are shown in Table I and Figure 2. While it is clear that the intra-modality approaches improve the performance of the face recogniser (PCA), they are still largely inferior to the inter-modality methods. Moreover, even though EP and LLE approach the performance of HAOG at higher ranks (ranks > 200), performance at such ranks is not important since at most only the best 200 matches will be examined by criminal investigators.

The proposed LGMS algorithm outperforms all others across virtually every performance metric considered. The only method approaching the performance of LGMS is the fusion of the state-of-the-art methods D-RS and P-RS (P-RS+D-RS), having slightly higher matching rates above approximately Rank-60. However, the differences in performance can be shown to be statistically identical at the 95% confidence level using multi-comparison Analysis of Variance (ANOVA)[7]. However, at lower ranks, the LGMS method outperforms even P-RS+D-RS with a statistically significant margin. LGMS's matching rates also exhibit the lowest standard deviation values of all inter-modality methods, showing that the proposed method is not highly affected by the train/test sets used in contrast to algorithms such as ET and P-RS+D-RS at lower ranks. This is especially impressive considering the vast number of images used that include sketches from the CUFSF database, which contain several deformations and exaggerations that make the identification task more challenging. 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

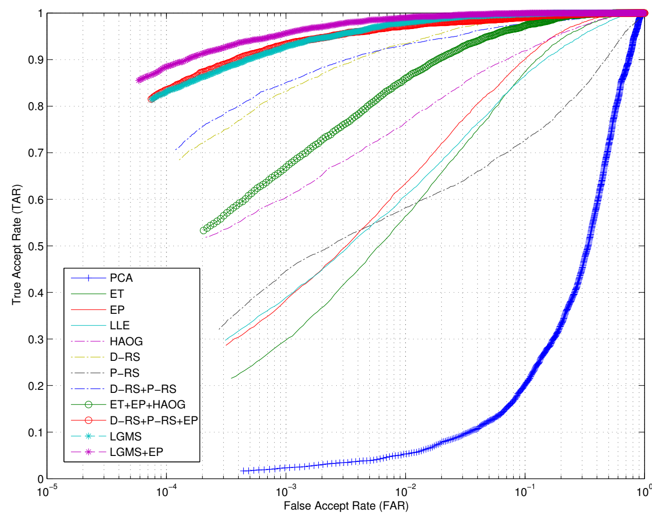

Fig. 2. ROC curve of the algorithms considered, averaged over 5 set splits

only managed a rate of 51.4% with an extended gallery and additional sketches from the CUFSF database. Moreover, LGMS outperforms P-RS and D-RS despite using LDA as the feature projection algorithm, which is theoretically inferior but less computationally intensive than RS-LDA as used in P-RS and D-RS. Of course, this also holds for the method fusing P-RS and D-RS, which is more computationally intensive than LGMS but only exhibits similar performance at ranks $\gtrapprox 50$.

From the results of algorithm 12 in Table I, the effectiveness of log-Gabor filtering is clear given that the intensity features alone are able to provide good performance. Nonetheless, the use of MLBP yields noticeable improvements. However, the poor results when using MLBP on the un-filtered 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 authors of [6] showed that fusion of intra- and inter-modality algorithms can improve performance using ET, EP and HAOG. The fusion of these methods on the large dataset considered in this paper also yields improvement in performance at almost all ranks. Hence, LGMS is combined with the EP+PCA approach using min-max normalisation and sum-of-scores fusion to determine if any performance increase can be obtained. The resultant LGMS+EP method achieves the best performance despite EP+PCA not being a state-of-the-art intra-modality approach. Fusion of EP with the next-best method, D-RS+P-RS, also achieves noticeable gains but the resultant method is still inferior to LGMS+EP at low ranks and in terms of the AuC, whilst also being slower to compute.

Lastly, the used images contain high variations in terms of lighting, pose, scale, and expression since they have been obtained from multiple databases. The good performance of LGMS indicates that it is also robust to these factors.

[1] Available at: http://mmlab.ie.cuhk.edu.hk/archive/facesketch.html
[2] Available at: http://mmlab.ie.cuhk.edu.hk/archive/cufsf/
[3] Available at: http://www.nist.gov/itl/iad/ig/colorferet.cfm
[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
[7] Additional results available at: http://wp.me/P6CDe8-2E

TABLE I
MEANS AND STANDARD DEVIATIONS OVER FIVE RANDOM TRAIN/TEST SET-SPLITS OF RANK RETRIEVAL RATES AND AUC. LG = LOG-GABOR.

#	Algorithm	Matching Rate (%) at Rank-N						AuC
		N=1	N=10	N=50	N=100	N=150	N=200	
1	PCA [17]	1.62±0.16	5.25±0.41	9.54±0.39	13.19±0.62	14.94±0.48	18.28±0.67	0.6443±0.0032
2	ET (+PCA) [16]	20.90±1.62	47.27±2.59	68.59±2.09	76.97±1.50	81.55±1.65	85.00±1.83	0.9569±0.0029
3	EP (+PCA) [6]	28.03±1.11	54.52±0.81	73.66±0.25	81.16±1.35	85.44±0.97	88.15±0.57	0.9670±0.0024
4	LLE (+PCA) [7]	29.01±1.19	54.45±0.80	73.15±0.84	80.90±1.06	85.34±1.02	88.19±0.91	0.9510±0.0065
5	HAOG [3]	51.39±0.99	66.97±0.85	79.26±1.28	85.15±0.47	88.11±0.35	89.68±0.54	0.9739±0.0011
6	D-RS [1], [9]	68.45±1.62	86.97±1.02	95.78±0.48	97.67±0.34	98.55±0.19	98.97±0.20	0.9946±0.0017
7	P-RS [1]	32.04±2.20	63.42±3.42	83.49±2.05	89.81±1.43	92.33±0.87	93.78±0.80	0.8809±0.0101
8	P-RS + D-RS [1]	70.63±2.79	90.25±1.32	97.23±0.62	98.76±0.38	99.35±0.20	99.50±0.16	0.9900±0.0031
9	ET + EP + HAOG [6]	52.71±2.74	76.64±1.25	88.51±1.26	91.95±0.54	93.87±0.57	95.04±0.44	0.9902±0.0011
10	P-RS + D-RS + EP	81.53±0.60	94.41±0.71	99.05±0.22	99.56±0.09	99.73±0.06	99.77±0.09	0.9971±0.0008
11	MLBP & SROCC	39.12±1.22	61.37±0.81	76.89±2.06	83.79±1.59	87.39±1.27	89.77±1.09	0.9633±0.0090
12	LG & SROCC	71.43±0.78	85.92±0.68	92.65±0.67	95.11±0.20	96.20±0.20	96.91±0.25	0.9937±0.0006
13	LG & MLBP & Euclidean	52.84±1.14	67.42±1.47	79.87±0.91	84.66±0.94	87.35±0.72	89.45±0.85	0.9776±0.0011
14	LG & MLBP & Cosine	73.61±1.62	90.53±1.00	96.68±0.48	98.17±0.35	98.74±0.32	99.03±0.34	0.9969±0.0010
15	*LG & MLBP & SROCC (LGMS)*	81.37±0.42	93.72±0.39	97.46±0.37	98.49±0.34	98.89±0.24	99.22±0.22	0.9989±0.0003
16	*LGMS + EP*	85.53±0.48	95.97±0.40	98.74±0.31	99.33±0.12	99.52±0.09	99.58±0.11	0.9995±0.0001

V. CONCLUSION

An approach denoted LGMS for inter-modality face photo-sketch recognition has been proposed and evaluated. LGMS was shown to achieve a Rank-1 rate of 81.4% that is superior to the next-best inter-modality approach by over 10%, a margin that is statistically significant. Its fusion with an intra-modality algorithm yielded further improvements. The main contributions of this paper are (i) the use of log-Gabor filters for pre-processing, (ii) the combination of global and local texture information via log-Gabor filtering and MLBP, (iii) the use of SROCC for template matching, and (iv) an extensive evaluation using (a) several popular and state-of-the-art intra- and inter-modality algorithms, (b) sketches containing shape deformations and exaggerations, and (c) a vast number of photos to mimic mug-shot databases maintained by law enforcement agencies more closely than typically used in literature. Future work includes the use of forensic sketches and images belonging to other modalities such as Near Infra-Red, and exploitation of demographic data.

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] B. F. 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, June 2013.

[2] S. Klum, H. Han, A. K. Jain, and B. Klare, “Sketch based face recognition: Forensic vs. composite sketches,” in *Int. Conf. Biometrics (ICB)*, June 2013, pp. 1–8.

[3] H. Galoogahi and T. Sim, “Inter-modality face sketch recognition,” in *IEEE Int. Conf. Multimedia and Expo*, July 2012, pp. 224–229.

[4] N. Wang, D. Tao, X. Gao, X. Li, and J. Li, “A comprehensive survey to face hallucination,” *Int. J. Comput. Vision*, vol. 106, no. 1, pp. 9–30, Jan. 2014.

[5] 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, 2013.

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

[7] Q. Liu, X. Tang, H. Jin, H. Lu, and S. Ma, “A nonlinear approach for face sketch synthesis and recognition,” in *IEEE Conf. Comput. Vision and Pattern Recog.*, vol. 1, June 2005, pp. 1005–1010 vol. 1.

[8] 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.

[9] B. Klare and A. K. Jain, “Heterogeneous Face Recognition: Matching NIR to Visible Light Images,” in *Proc. Int. Conf. Pattern Recog.*, Aug 2010, pp. 1513–1516.

[10] 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.

[11] J. Arrospide and L. Salgado, “Log-Gabor Filters for Image-Based Vehicle Verification,” *IEEE Trans. Image Process.*, vol. 22, no. 6, pp. 2286–2295, June 2013.

[12] L. Zhang, D. Zhang, X. Mou, and D. Zhang, “FSIM: A Feature Similarity Index for Image Quality Assessment,” *IEEE Trans. Image Process.*, vol. 20, no. 8, pp. 2378–2386, August 2011.

[13] T. Ojala, M. Pietikainen, and T. Maenpaa, “Multiresolution gray-scale and rotation invariant texture classification with local binary patterns,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, July 2002.

[14] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, “Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 711–720, 1997.

[15] 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.

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

[17] M. Turk and A. Pentland, “Eigenfaces for recognition,” *J. Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.

[18] A. M. Martinez and R. Benavente, “The AR Face database,” CVC, Tech. Rep., June 1998.

[19] K. Messer, J. Matas, J. Kittler, J. Luettin, and G. Maitre, “XM2VTSdb: The Extended M2VTS Database,” in *Proc. Int. Conf. Audio and Video-base Biometric Personal Verification*, ser. Lecture Notes in Computer Science. Springer Verlag, 1999.

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

[21] 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.

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