Are Adaptive Face Recognition Systems still Necessary? Experiments on the APE Dataset

Giulia Orrù
University of Cagliari (DIEE, PRA Lab)
Cagliari, Italy
giulia.orru@unica.it

Marco Micheletto
University of Cagliari, DIEE
Cagliari, Italy
marco.micheletto@unica.it

Julian Fierrez
Universidad Autonoma de Madrid (BiDA Lab)
Madrid, Spain
julian.fierrez@uam.es

Gian Luca Marcialis
University of Cagliari, DIEE
Cagliari, Italy
marcialis@unica.it

Abstract—In the last five years, deep learning methods, in particular CNN, have attracted considerable attention in the field of face-based recognition, achieving impressive results. Despite this progress, it is not yet clear precisely to what extent deep features are able to follow all the intra-class variations that the face can present over time. In this paper we investigate the performance the performance improvement of face recognition systems by adopting self updating strategies of the face templates. For that purpose, we evaluate the performance of a well-known deep-learning face representation, namely, FaceNet, on a dataset that we generated explicitly conceived to embed intra-class variations of users on a large time span of captures: the APhotoEveryday (APE) dataset[1]. Moreover, we compare these deep features with handcrafted features extracted using the BSIF algorithm. In both cases, we evaluate various template update strategies, in order to detect the most useful for such kind of features. Experimental results show the effectiveness of "optimized" self-update methods with respect to systems without update or random selection of templates.

Index Terms—self-update, face recognition, adaptive systems

I. INTRODUCTION

In recent years, CNN-based Facial Recognition (FR) approaches have had a significant impact on both research and real applications. Contemporary deep learning approaches can achieve a close-to-zero error rate on popular datasets such as LFW. On the other hand, intra-class variations such as poses, illuminations, expressions, and occlusions, still affect the performance of deep FR systems [1], [2], which may benefit from pre-processing methods in order to normalize those variability factors [3]. The excellent performance and robustness of deep FR systems might suggest that the performance of deep templates is stable over time, but some works have showed that it is not yet clear to what extent face representations are able to keep high performance over large time lapses between enrollment and testing data or large variations of other factors.

[1]https://github.com/PRALabBiometrics/APhotoEverydayDB

In the recent past, several "adaptive" biometric systems have been proposed to deal with changing context and variability factors [4]–[7]. As described in [7], any component of a biometric system is subject to time adaptation. Here we focus in updating the biometric templates [8]. The concept of template adaptation is referred to the generation of novel templates in time that can replace or can be coupled with existing ones [9], [10]. This is done automatically by exploiting samples captured during normal system operations. As highlighted by [11], template update improves the performance of facial recognition systems, including those based on deep-learning techniques.

In practice, template update (also referred to as self update from now on) has various advantages with respect to the network's re-tuning to track variations in time: (1) templates can be easily stored in the user's own device with an irrelevant memory waste, (2) no need for accessing the original data, namely the users' images, in order to retrain the network, and (3) this avoid privacy issues.

The evaluations carried out so far in related works, however, usually consider datasets collected over a short time span. Therefore, these datasets do not allow to analyze properly how the typical intra-class variations of aging are followed by template update methods.

The goal of this work is to perform an evaluation of deep features for Face Recognition over time with and without template update by using a dataset that we generated explicitly conceived to contain many intra-class variations, long acquisition time span, and temporal information: the APhotoEvery-Day (APE) dataset. In fact, exploiting the temporal information allows us to carry out an analysis that realistically simulates the normal adaptive facial recognition system operation.

In our experiments we analyze two FaceNet models in comparison with a BSIF-based handcrafted approach. Both basic and more recent template update strategies are investigated, in order to detect the most useful for such kind of features. Experimental results show that template update is helful to improve the evaluated face recognition systems over time.

The paper is organized as follows. Section II summarizes

the evolution of feature representations and the weaknesses of the approaches used in the last decades. Section III describes the adaptive methods implemented and evaluated. Section IV describe the face models and dataset used in the experimental evaluation. Section V reports the experimental results and discussions. Section VI concludes the paper.

II. Feature representation in Face Recognition

Over the past years, many methods have been developed to address the issues and challenges of facial recognition systems. In particular, the non-stationarity of the face appearance, which presents many intra-class variations such as aging, lighting, pose, and occlusions, is often addressed by implementing adaptive systems. These biometric systems "adapt" continuously the gallery templates to the variations of the input data, without the need of human intervention.

The performance of adaptive systems depends on the expressive power of the face representation methods adopted. These can be grouped into four broad categories (Table I) that have been developed in different periods of the past thirty years [12]. In holistic approaches, widespread since the 90s, the complete face region is taken into account as input data and the face information is represented by a small number of features [13], [14]. The main issue of this approach is that these methods fail to address uncontrolled facial changes.

To solve this problem, local features of faces were proposed to select a number of features to uniquely identify individuals, the so-called local handcrafted methods, in the early 2000s. Among them, Local Binary Patterns (LBPs) [15], [16], Gabor [17], BSIF [18], and their variants [19] exhibit as limitation the lack of compactness and in some cases a low distinctiveness among individuals.

The problem of the low compactness level was faced at the beginning of the 2010s; in particular, shallow learning methods were introduced. These methods, also known as learning-based local descriptors, use unsupervised learning methods to encode the local microstructures of the face into discriminative codes [20]. Unfortunately, shallow learning methods cannot handle complex nonlinear facial appearance variations such as self-occlusions and pose variations [21].

After 2012, deep learning methods began to spread and achieved state-of-the-art results in many problems. These methods use a cascade of processing layers to learn representations of data with multiple levels of abstraction. From 2014 also the Face Recognition community adopted deep learning methods, such as convolutional neural networks, for facial feature extraction and transformation. Among others, DeepFace [22], FaceNet [23], VGGFace [24] and VGGFace2 [25]. Some of them achieved state-of-the-art performance on the most challenging datasets known, such as LFW [26], IJB-A [27], IJB-B [28], etc.

Although the good results obtained, deep learning approaches work as "black-box" feature extractors from face images [29]. For this reason it is not yet clear how these systems can handle data that present significant time changes.

Moreover, the datasets mentioned above do not contain temporal information in order to build a realistic time sequence of facial images of the same people across months or years.

As stated in the Introduction, our main aim is to understand if the compact and powerful representation obtained through deep learning methods is able to work in situations of long-term use in which the temporal variability of the biometric data is more accentuated.

III. Adaptive Methods for Template Update

First of all, we implemented standard facial recognition as a sort of "ground truth", since our main claim is to clarify to what extent self update approaches are helpful over top-performing deep-learning based face recognition. In particular, we tested the "traditional" self-update system [30], the classification-selection method based on risk minimization [31] and two methods of classification-selection with limited number of templates per user, based on K-means [32] and on the semi-supervised application of RANDOM editing methods just to verify if the selection performed by the two methods above is significant or not. In fact, it is possible to categorize biometric adaptive systems into two categories:

- The traditional self-update system that only classifies input samples and adds them to the gallery if they meet the genuinity requirements.
- Classification/selection approach in which the selection phase allows to filter the redundant information [11], [33].

Traditional self-update estimates an updated threshold through the gallery and, when a *batch* [34] is available for a certain claimed user, the distances between each sample and the user's template(s) are computed. Input samples whose distance is less than the update threshold are added into the user's gallery. Classification/selection algorithms add another phase where the best samples for the gallery are chosen.

In particular, Rattani *et al.* introduced a classification/selection system based on harmonic functions and a risk minimization technique [31]. In [32] the authors present a method that keeps the number of templates constant at each iteration by setting the maximum number of images per user, namely, p, in the selection phase. This is obtained by deriving the centroid of the samples of a given subject through the K-means algorithm, and then selecting the p closest samples to that centroid.

Finally, the RANDOM method simply selects p pseudo-labeled samples of each user. This method allows us to simulate what may happen by keeping a human in the loop for selecting the best templates to update. Indeed, the selection of a human supervisor is unpredictable and changes depending on the individual involved.

IV. Face Models and Dataset

A. Face Models: FaceNet and BSIF

We use two FaceNet models [23] and handcrafted Binarized Statistical Image Features (BSIFs) [18].

FaceNet uses a deep convolutional network to optimize the representation of an individual's face through a 128

78

TABLE I: Feature representation evolution in Face Recognition systems.

Feature represent.	Holistic learning	Local handcraft	Shallow learning	Deep learning
Introduction (years)	1990s	2000s	2010s	2012/2014
Issues	Fail to address uncontrolled facial changes.	Lack of distinctiveness and compactness.	Fail to address facial appearance variations.	Black-box. Requires time, powerful hardware and a large number of images for training.

bytes feature vector. In particular, faces are mapped into a 128-dimensional Euclidean space in which distances directly correspond to a measure of facial similarity. The FaceNet architecture is based on the triplet loss function [35]: the feature vectors related to the same individual have small distances, while those of distinct people have large distances. To evaluate the effectiveness of the FaceNet representation, we used an open-source implementation based on TensorFlow [36] trained on the model 20170512-110547 and on the model 20180408-102900, whose characteristics are shown in Table II. The first model has been trained on the MS-Celeb-1M dataset [37] and represents the face with a 128 byte feature vector. The second model has been trained on the CASIA-WebFace dataset [38] and uses a 512 byte feature vector.

TABLE II: FaceNet models details.

Model	20170512-110547	20180408-102900
Dataset	MS-Celeb-1M	CASIA-WebFace
Embedding size	128D	512D
Image stand.	Per image	Fixed
Data augment.	Random crop/flip	Random flip
Optimizer	RMSProp	Adam

B. APE Dataset

Using FaceNet we extracted the feature vectors related to the "APhotoEveryday" (APE) dataset [33] that contains the faces of 46 individuals with large appearance variations across time. The APE dataset was acquired by the University of Cagliari and consists of facial images extracted from YouTube Daily Photo Projects. Faces were captured in the frontal pose and most of these have a controlled expression. The images of each user are labelled with a number that indicates the temporal progression of the sequence. The APE dataset images are characterized by many temporal variations of the facial appearance thanks to the Daily Photo Project[2]. The number of images per user varies between 92 and 3892 and the acquisition time varies between less than one year and twelve years. Fig. 1 shows some facial images from the APE dataset.

V. EXPERIMENTS

A. Experimental Protocol

The experimental protocol was designed in order to exploit the full potential of the dataset, in particular the high number of samples per user and the temporal information associated with them. This is summarized as follows:

- The dataset was subdivided into ten parts maintaining the sequence time progression: the first partition is the

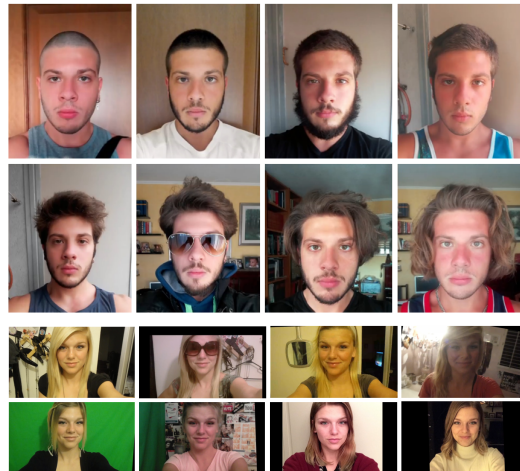[2]See for example https://www.youtube.com/watch?v=RBPYDIzEbYk

Fig. 1: Examples of face images from the APE dataset exhibiting many variations over time.

initial gallery and is composed of the first $p = 5$ images of each user; the remaining nine parts are composed of $\frac{\#samples-p}{\#adaptationsets}$ samples for each user.
- The operational points FAR and FRR are calculated and the updating threshold was estimated at FAR=1%.
- The adaptation sets were used to simulate the periodic sets of batches collected during the system's operations individually processed for updating the users' galleries.
- The test set was used to evaluate the system's performance from the state made up of the initial galleries, to the updated gallery after each updating cycle. In order to better simulate a real application and exploit time information we used the $(i + 1)^{th}$ batch as test set of the ith batch as suggested in [39].

B. Results

The traditional self-update [30], the method based on risk-minimization [31], the K-means [32], and the RANDOM classification/selection method were implemented and tested to evaluate the performance using feature vectors extracted with FaceNet 128 dimension embeddings (Fig. 2), FaceNet 512 dimension embeddings (Fig. 3) and with the handcrafted BSIFs (Fig. 4).

The most remarkable result emerging from Fig. 2(a) and Fig. 3(a) is that for deep features, despite their compactness and representativeness, classification/selection template update methods improve their performance compared to a solution without self-update (black line) or to the random selection of templates, which can be assimilated to the update supervised
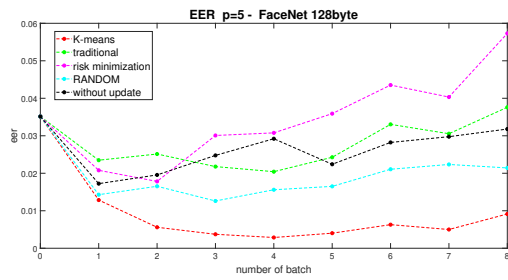
(a) EER - FaceNet 128 byte - APE dataset. For x=i: update on batch i, performance on batch $i+1$.
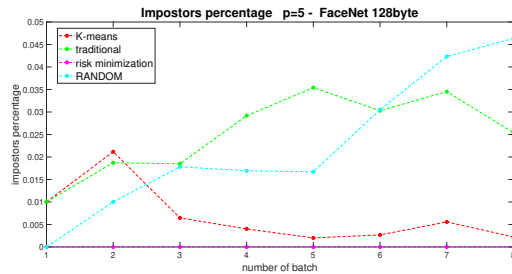
(b) Impostors percentage - FaceNet 128 byte - APE dataset.

Fig. 2: EER and percentage of impostors for different template update methods with $p=5$ for FaceNet 128D (APE dataset).

(a) EER - FaceNet 512 byte - APE dataset. For x=i: update on batch i, performance on batch $i+1$.

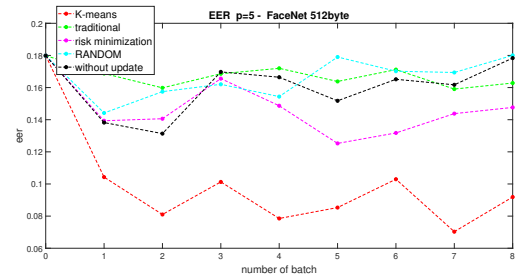(b) Impostors percentage - FaceNet 512 byte - APE dataset.

Fig. 3: EER and percentage of impostors for different template update methods with $p=5$ for FaceNet 512D (APE dataset).

by a human operator. In particular, for both FaceNet models, the K-means is the best algorithm. In fact, this classification/selection method, as demonstrated in [11], allows at the same time to keep the system error low and a limited number of templates in memory. The result is low computational complexity and stable performance over time.

The performance of the traditional self-update (i.e., adding templates to the system without selection), is worse in terms of accuracy than the system without updating, in addition to having high requirements for template storage. This is probably due to the fact that there is no filtering of introduced impostors. This drift was observed in early publications too [40] and it is shown by the increasing percentage of impostors present for all models analyzed (Figs. 2b, 3b, 4b).

Fig. 5 shows the performance of the three feature extraction approaches for facial recognition without updating, with traditional self-update and self update by classification/selection (K-means). These results highlight how the system based on 128D FaceNet performs better than the one based on 512D FaceNet and BSIFs. The better performance of the 128D model compared to the 512D model is probably due to the data used for training the neural network.

Table III supports this claim; the EER calculated on the initial gallery and on the first batch and the averages and standard deviations along the remaining batches (with update on the i batch and performance on the $i+1$ batch for $i>1$) are reported comparing the Face Recognition performance without update with the traditional self-update and the K-means classification-selection approach. The reported results show that the template update driven by K-means has a
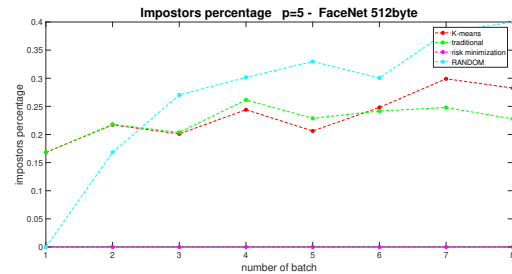
very positive effect on both global performance and system stability. In particular, for FaceNet128D the error goes from 2.54% for the method without updating to 0.62% for the K-means method. These results demonstrate that, depending on the application context, the use of template update and selection allows a substantial improvement in the performance of Face Recognition systems resulting more adequate than the intervention of a human operator.

VI. CONCLUSIONS

In this paper we analyzed the template representativeness of deep-learning features, in particular those extracted with two models of FaceNet, to understand if adaptive systems are useful in modern Face Recognition systems. In order to make a comparison between hand-crafted and deep learning features we have also analyzed the system performance with handcrafted BSIF features. To evaluate the performance across a large time lapse between enrollment and testing, and to simulate a challenging scenario including drastic appearance changes, the experiments were conducted on the APE dataset. The evidence from this study shows the benefits of update methods with classification/selection in situations where the face appearance presents many intra-class variations [41]. In fact, the use of "optimized" template update allows a substantial improvement in the performance compared to systems without updating or systems that keep the human in the loop, simulated with a random selection of templates. It is therefore possible to say that adaptive systems are useful on top of modern deep face recognition, at least in the scenarios considered here in our APE face image dataset (with time spans of months to years between enrollment and testing).
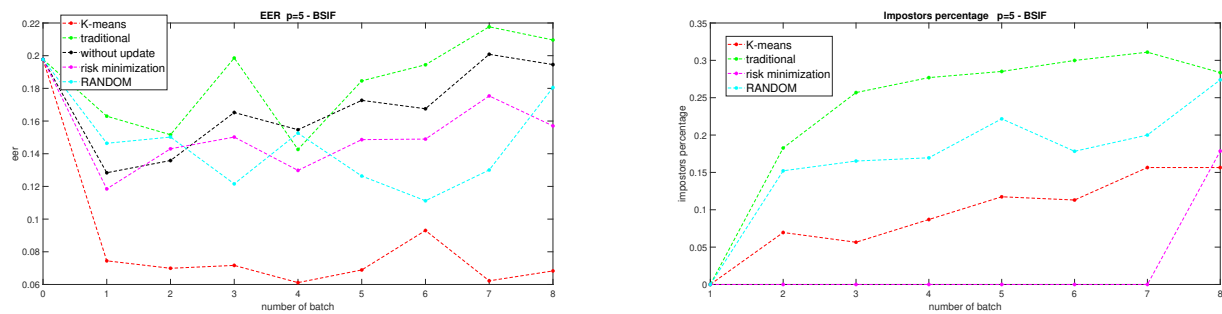
Fig. 4: EER and percentage impostors comparison among the SOTA methods with p=5 for BSIF feature vectors. On the x axis is shown the number of the batch and on the y axis the performance index.
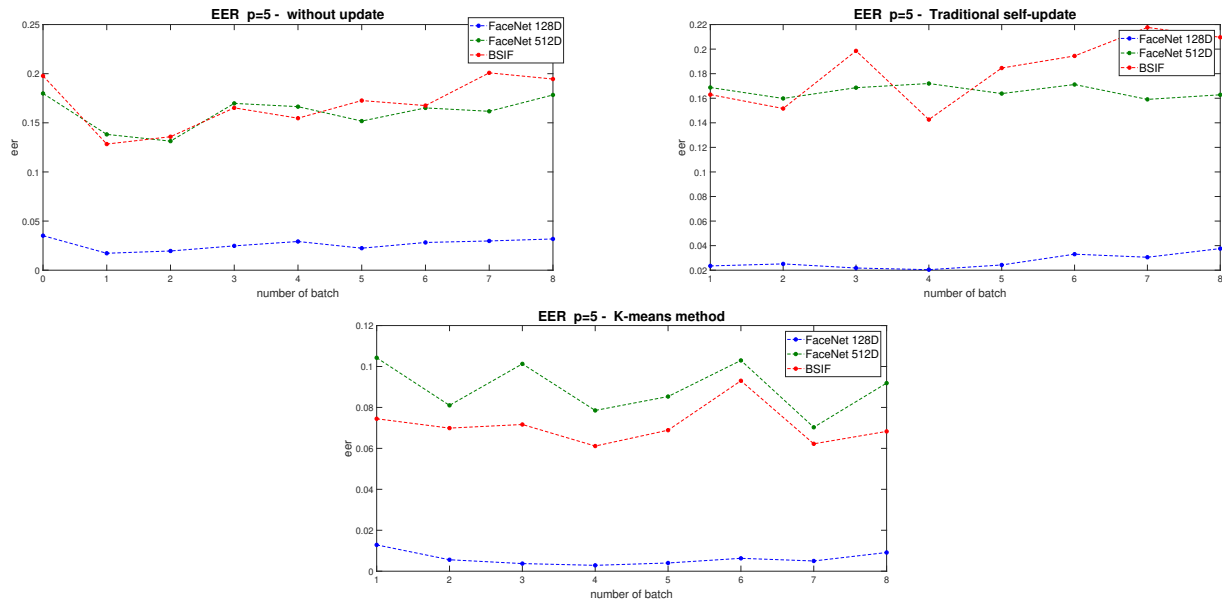

Fig. 5: EER comparison between the different feature vectors tested: without update, traditional update, kmeans-based update.

TABLE III: EER comparison among the different feature vectors tested: the table shows the EER calculated by training on the initial gallery and testing on the first batch and the averages and standard deviations calculated along the remaining batches (with update on the i batch and performance on the $i+1$ batch for $i > 1$).

EER				
	Initial Gallery	**Without update**	**Traditional self-update**	**K-means**
FaceNet 128D	3.51%	2.54% ± (0.52)	2.70% ± (0.60)	**0.62% ± (0.33)**
FaceNet 512D	17.99%	15.78% ± (1.62)	16.58% ± (0.50)	8.94% ± (1.27%)
BSIF	19.78%	16.50% ± (2.54)	18.28% ± (2.75)	7.12% ± (1.00)

REFERENCES

[1] H. Proença, M. Nixon, M. Nappi, E. Ghaleb, G. Ozbulak, H. Gao, H. K. Ekenel, K. Grm, V. Struc, H. Shi, X. Zhu, S. Liao, Z. Lei, S. Z. Li, W. Gutfeter, A. Pacut, J. Brogan, W. J. Scheirer, E. Gonzalez-Sosa, R. Vera-Rodriguez, J. Fierrez, J. Ortega-Garcia, D. Riccio, and L. D. Maio, "Trends and controversies," *IEEE Intelligent Systems*, vol. 33, no. 3, pp. 41–67, 2018.

[2] E. Gonzalez-Sosa, R. Vera-Rodriguez, J. Fierrez, and J. Ortega-Garcia, "Exploring facial regions in unconstrained scenarios: Experience on ICB-RW," *IEEE Intelligent Systems*, vol. 33, no. 3, pp. 60–63, May 2018.

[3] F. Alonso-Fernandez, J. Fierrez, and J. Ortega-Garcia, "Quality measures in biometric systems," *IEEE Security & Privacy*, vol. 10, no. 9, pp. 52–62, December 2012.

[4] J. Fierrez-Aguilar, D. Garcia-Romero, J. Ortega-Garcia, and J. Gonzalez-Rodriguez, "Bayesian adaptation for user-dependent multimodal biometric authentication," *Pattern Recognition*, vol. 38, no. 8, pp. 1317–1319, August 2005.

[5] ——, "Adapted user-dependent multimodal biometric authentication exploiting general information," *Pattern Recognition Letters*, vol. 26, no. 16, pp. 2628–2639, December 2005.

[6] F. Roli, L. Didaci, and G. L. Marcialis, "Adaptive biometric systems that can improve with use," in *Advances in Biometrics: Sensors, Algorithms and Systems*, N. K. Ratha and V. Govindaraju, Eds. Springer London, 2008, pp. 447–471.

[7] J. Fierrez, A. Morales, R. Vera-Rodriguez, and D. Camacho, "Multiple classifiers in biometrics. Part 2: Trends and challenges," *Information Fusion*, vol. 44, pp. 103–112, November 2018.

[8] R. Tolosana, R. Vera-Rodriguez, J. Fierrez, and J. Ortega-Garcia, "Reducing the template aging effect in on-line signature biometrics," *IET Biometrics*, vol. 8, no. 6, pp. 422–430, November 2019.

[9] U. Uludag, A. Ross, and A. Jain, "Biometric template selection and update: a case study in fingerprints," *Pattern Recognition*, vol. 37, no. 7, pp. 1533 – 1542, 2004.

[10] J. Galbally, M. Martinez-Diaz, and J. Fierrez, "Aging in biometrics: An experimental analysis on on-line signature," *PLoS ONE*, vol. 8, no. 7, p. e69897, July 2013.

[11] G. Orrù, G. L. Marcialis, and F. Roli, "A novel classification-selection approach for the self updating of template-based face recognition systems," *Pattern Recognition*, vol. 100, pp. 107–121, 2020.

[12] M. Wang and W. Deng, "Deep face recognition: A survey," *CoRR*, vol. abs/1804.06655, 2018.

[13] M. A. Turk and A. P. Pentland, "Face recognition using eigenfaces," in *Proceedings. 1991 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Jun 1991, pp. 586–591.

[14] J. Lu, J. Hu, W. Deng, and J. Guo, "Transform-invariant PCA: A unified approach to fully automatic face alignment, representation, and recognition," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 11 2013.

[15] 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.

[16] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: Application to face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 12, pp. 2037–2041, Dec 2006.

[17] C. Liu and H. Wechsler, "Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition," *IEEE Transactions on Image Processing*, vol. 11, no. 4, pp. 467–476, April 2002.

[18] J. Kannala and E. Rahtu, "BSIF: Binarized statistical image features," in *Proceedings of the 21st International Conference on Pattern Recognition*, Nov 2012, pp. 1363–1366.

[19] W. Zhang, S. Shan, W. Gao, X. Chen, and H. Zhang, "Local gabor binary pattern histogram sequence (lgbphs): a novel non-statistical model for face representation and recognition," in *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, Oct 2005, pp. 786–791.

[20] Z. Cao, Q. Yin, X. Tang, and J. Sun, "Face recognition with learning-based descriptor," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, June 2010, pp. 2707–2714.

[21] C. Ding and D. Tao, "A comprehensive survey on pose-invariant face recognition," *ACM Trans. Intell. Syst. Technol.*, vol. 7, no. 3, pp. 37:1–37:42, Feb. 2016.

[22] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "Deepface: Closing the gap to human-level performance in face verification," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, June 2014, pp. 1701–1708.

[23] 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*, June 2015, pp. 815–823.

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

[25] Q. Cao, L. Shen, W. Xie, O. M. Parkhi, and A. Zisserman, "VGGFace2: A dataset for recognising faces across pose and age," *CoRR*, vol. abs/1710.08092, 2017.

[26] 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, Amherst, Tech. Rep. 07-49, Oct 2007.

[27] 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 *IEEE Conf. Computer Vision and Pattern Recognition*, June 2015, pp. 1931–1939.

[28] C. Whitelam, E. Taborsky, A. Blanton, B. Maze, J. Adams, T. Miller, N. Kalka, A. K. Jain, J. A. Duncan, K. Allen, J. Cheney, and P. Grother, "IARPA Janus Benchmark-B Face Dataset," in *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops*, July 2017, pp. 592–600.

[29] K. Grm, V. Štruc, A. Artiges, M. Caron, and H. K. Ekenel, "Strengths and weaknesses of deep learning models for face recognition against image degradations," *IET Biometrics*, vol. 7, no. 1, pp. 81–89, 2018.

[30] A. Rattani, B. Freni, G. L. Marcialis, and F. Roli, "Template update methods in adaptive biometric systems: A critical review," in *Advances in Biometrics*, M. Tistarelli and M. S. Nixon, Eds. Springer Berlin Heidelberg, 2009, pp. 847–856.

[31] A. Rattani, G. L. Marcialis, E. Granger, and F. Roli, "A dual-staged classification-selection approach for automated update of biometric templates," in *Proceedings of the 21st International Conference on Pattern Recognition*, Nov 2012, pp. 2972–2975.

[32] P. Tuveri, V. Mura, G. L. Marcialis, and F. Roli, "A classification-selection approach for self updating of face verification systems under stringent storage and computational requirements," in *18th IAPR Int. Conf. on Image Analysis and Processing*, vol. 9280, 2015, pp. 540–550.

[33] G. Orrù, G. L. Marcialis, and F. Roli, "An experimental investigation on self adaptive facial recognition algorithms using a long time span data set," in *IEEE 8th Int. Conf. Image Processing, Theory, Tools and Applications*, Nov 2018.

[34] A. Rattani, G. L. Marcialis, and F. Roli, "A multi-modal dataset, protocol and tools for adaptive biometric systems: a benchmarking study," *International Journal of Biometrics*, vol. Vol.5, pp. 266–287, 2013.

[35] K. Q. Weinberger and L. K. Saul, "Distance metric learning for large margin nearest neighbor classification," *J. Mach. Learn. Res.*, vol. 10, pp. 207–244, Jun. 2009.

[36] D. Sandberg, *Face Recognition using Tensorflow*, https://github.com/davidsandberg/facenet.

[37] Y. Guo, L. Zhang, Y. Hu, X. He, and J. Gao, "MS-Celeb-1M: A dataset and benchmark for large scale face recognition," in *European Conference on Computer Vision*. Springer, 2016.

[38] D. Yi, Z. Lei, S. Liao, and S. Z. Li, "Learning face representation from scratch," *CoRR*, vol. abs/1411.7923, 2014.

[39] A. Rattani, G. L. Marcialis, and F. Roli, "Self adaptive systems: An experimental analysis of the performance over time," in *2011 IEEE Workshop on Computational Intelligence in Biometrics and Identity Management*, 2011, pp. 36–43.

[40] F. Roli and G. L. Marcialis, "Semi-supervised PCA-based face recognition using self-training," in *Structural, Syntactic, and Statistical Pattern Recognition*, D.-Y. Yeung, J. T. Kwok, A. Fred, F. Roli, and D. de Ridder, Eds. Springer Berlin Heidelberg, 2006, pp. 560–568.

[41] E. Gonzalez-Sosa, J. Fierrez, R. Vera-Rodriguez, and F. Alonso-Fernandez, "Facial soft biometrics for recognition in the wild: Recent works, annotation and COTS evaluation," *IEEE Trans. on Information Forensics and Security*, vol. 13, no. 8, pp. 2001–2014, August 2018.