Fusion of Face Demorphing and Deep Face Representations for Differential Morphing Attack Detection

E. Shiqerukaj*†, C. Rathgeb*‡, J. Merkle*, P. Drozdowski*, B. Tams*
*secunet Security Networks, Essen, Germany
†Technische Universität Dortmund, Germany
‡Hochschule Darmstadt, Germany
{elidona.shiqerukaj,christian.rathgeb,johannes.merkle}@secunet.com

2022 International Conference of the Biometrics Special Interest Group (BIOSIG) | 978-1-6654-7666-9/22/$31.00 ©2022 IEEE | DOI: 10.1109/BIOSIG55365.2022.9897023

***Abstract*—Algorithm fusion is frequently employed to improve the accuracy of pattern recognition tasks. This particularly applies to biometrics including attack detection mechanisms. In this work, we apply a fusion of two differential morphing attack detection methods, i.e. Demorphing and Deep Face Representations. Experiments are performed in a cross-database scenario using high-quality face morphs along with realistic live captures. Obtained results reveal that a weighted sum-based score-level fusion of Demorphing and Deep Face Representations improves the morphing attack detection accuracy. With the proposed fusion, a detection equal error rate of 4.9% is achieved, compared to detection equal error rates of 5.6% and 5.8% of the best individual morphing attack detection methods, respectively.**

***Index Terms*—Face recognition, morphing attack detection, fusion, demorphing, deep face representations.**

I. Introduction

Face recognition systems have been found to be vulnerable to so-called morphing attacks [1], [2]. In said attacks, the facial images of two (or more) individuals are combined into one image using image morphing techniques. The resulting morphed facial image is then presented during enrolment as a biometric reference. An example of a morphed facial image (hereinafter referred to as "morph") is shown in figure 1. If the morph is accepted at enrollment, it is likely that all individuals who contributed to the morph can be successfully authenticated against it. Morphing attacks thus pose a serious threat to facial recognition systems, in particular in border control scenarios, where the reference image is often provided in printed form by the applicant.

In response to the above described vulnerability, *Morphing Attack Detection* (MAD) approaches have been developed in the recent past. The goal of these methods is to reliably differentiate between morphs and bona fide (*i.e.* genuine) facial images. MAD schemes can be categorised into single image and differential MAD. The former type of methods examine single images with the aim of detecting traces of image morphing. In contrast, the latter type of methods analyse a potentially morphed face image together with a trusted live capture, *e.g.* by estimating differences between them. For comprehensive reviews of published MAD methods the interested reader is referred to [1], [3].

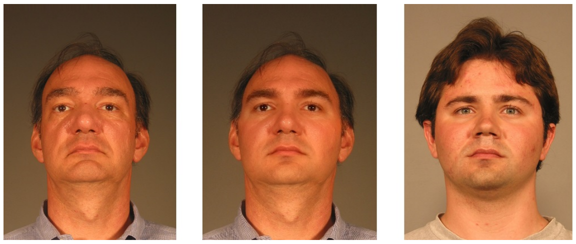

Fig. 1. Example of face morphing: two face images (left and right) are combined to obtain a face morph (middle) (the outer facial region, eyes and nostrils of the left face image are retained).

In order to improve the detection performance of MAD, some researchers proposed to combine conceptually different MAD methods. For instance, Scherhag *et al.* [4] single image MAD a score-level fusion of individual MAD methods based texture descriptors, keypoint extractors, gradient estimators, and deep learning-based features. The authors reported a significant improvement in terms of detection error rates. Similarly, Venkatesh *et al.* [5] employed a sum-rule to fuse scores obtained from several single image MAD methods that utilise conceptually different feature extractors. Another similar approach was presented by Makrushin *et al.* [6]. They showed that a more sophisticated score-level fusion algorithm based on the Dempster-Shafer theory outperforms a simple sum rule-based fusion.

Damer *et al.* [7] investigated score-level as well as feature-level fusion methods utilising texture descriptors and deep learning-based features. For the latter type of fusion, they performed a simple feature concatenation. Moreover, they employed training databases with variations in morphing techniques and image pairing protocols. Again, significant improvements w.r.t. MAD performance was reported.

Scherhag *et al.* [8] suggested to perform a sum rule-based score-level fusion of MAD scores obtained from single image and differential MAD methods where both employ the same feature extractor. It was shown that the fusion-based MAD system outperformed the individual single image and differential MAD schemes.

It is important to note that in all of the aforementioned works, training and test sets were obtained from a single

Gesellschaft für Informatik e.V. (GI)

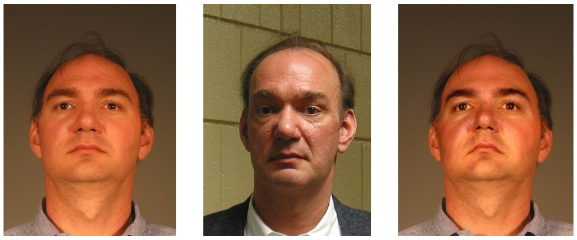

Fig. 2. Example of DM: the trusted live capture (middle) is subtracted from the morph (left) resulting in a demorphed face image (right) (morphing factor 0.25).

face image database which is known to result in overfitting and, consequently, in unrealistic performance rates. In contrast, Lorenz *et al*. [9] investigated weighted sum rule-based fusions of MAD scores obtained from single image and differential MAD methods in a cross-database evaluation. The weights of the sum rule-based fusion were adjusted using two different approaches, *i.e*. grid-search and random forest. The authors obtained an improved MAD performance for a fusion of different single image and differential MAD methods.

In summary, many works have demonstrated that MAD can benefit from algorithm fusion. However, the majority of published works exhibits two main shortcomings: on the one hand, experimental setups usually do not reflect real-world scenarios and, thus, corresponding results tend to be over-optimistic [10]; on the other hand, many works utilise MAD methods that do not represent the current state-of-the-art in MAD. Independent performance tests, such as the Face Recognition Vendor Test MORPH of the National Institute of Standards and Technology [11], revealed that in real-world scenarios, most competitive MAD performance rates are obtained in differential detection scenarios. In this type of scenario, MAD methods based on *Demorphing* (DM) [12] and *Deep Face Representations* (DFR) [13] were found to achieve the best MAD results [14].

In this work, we investigate the potential of fusing DM and DFR for differential MAD. To this end, a score-level fusion is performed based on a weighted sum of MAD scores obtained by both MAD methods. In a cross-database scenario, this simple fusion of these conceptually different MAD methods significantly outperforms both individual systems in terms of detection performance.

The rest of this paper is organised as follows: the proposed fusion-based MAD system is described in detail in section II. Subsequently, the experimental setup is summarised in section III and results are presented in section IV. Finally, conclusions are drawn in section V.

II. Proposed Fusion

The following subsections describe the used MAD methods, namely DM (subsection II-A) and DFR (subsection II-B). Afterwards, the proposed fusion strategy is presented (subsection II-C).

A. Demorphing

DM was introduced by Ferrara *et al*. [12]. In this differential MAD method, the live capture is used to revert (demorph) a potentially morphed reference image. In other words, the live capture is subtracted from the reference image with a predefined weight (demorphing factor). An example of demorphing are shown in figure 2. The resulting demorphed face image is then compared against the live capture using a generic face recognition system. This means the face recognition score between the live capture and the demorphed reference image represents the final MAD score. If this comparison results in a non-match, a morphing attack has been detected, otherwise the authentication attempt is considered to be bona fide.

B. Deep Face Representations

DFR is a differential MAD method proposed by Scherhag *et al*. [13]. In this algorithm, deep neural networks initially trained for the purpose of face recognition are used to extract feature vectors (deep face representations) from the reference image and the live capture. Subsequently, both feature vectors are combined, *e.g*. by estimating a difference vector, and fed into a machine learning-based classifier, *e.g*. a Support Vector Machine (SVM), which has previously been trained to distinguish between bona fide authentications from morphing attacks. The score extracted by said classifier represents the MAD score.

C. Fusion

An overview of the proposed fusion of DM and DFR is depicted in figure 3. The previously described MAD methods are conceptually different: while the DM methods uses a live capture to revert a morphing process that has potentially been applied to the reference image, the DFR method directly extracts features from the reference image and the live capture, combines them, and classifies them using a machine learning technique. It is therefore expected that the MAD scores obtained by both algorithms exhibit a low correlation which would make them suitable candidates for a fusion-based MAD.

The MAD scores obtained from DM and DFR are first normalised with a simple min-max normalisation. The DM and DFR MAD score can then be combined in various ways. In particular, fusions based on the (weighted) sum rule, product rule, and a linear SVM are considered.

Furthermore, a nested fusion of both methods could be perform. That is, the DFR algorithm could also be applied to pairs of demorphed reference images and live captures. However, this fusion method did not result in performance improvements and is therefore not reported in this work.

III. Experimental Setup

In the next subsections, we describe the used software and databases (subsection III-A) as well as the employed evaluation metrics (subsection III-B).

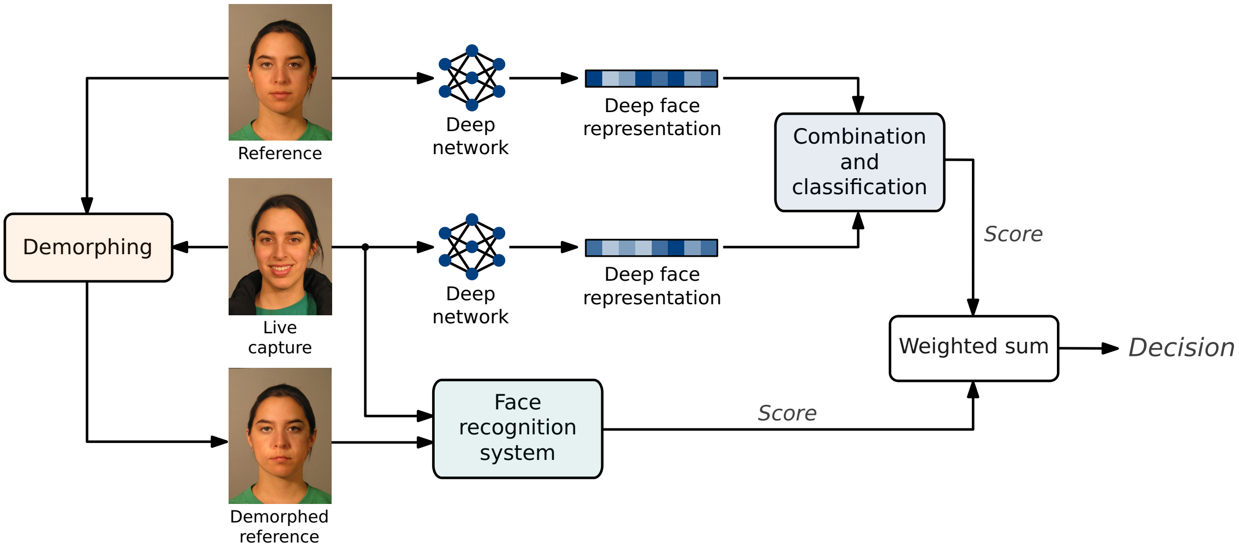

Fig. 3. Overview of the proposed fusion-based MAD system.

A. *Software and Databases*

For creating face morphs, we use an adapted version of the University of Twente (UTW) morphing algorithm described in [14]. The UTW algorithm is a landmark-based morphing method and employs the `seamlessClone` function of OpenCV to avoid unnatural skin colour transitions. In contrast to the original algorithm, we use dlib for landmark detection. Morphs are created from pairs of face images with equal weights, while the outer facial region as well as the eye and nostril regions of one of the faces are retained in order to reduce potential artefacts. While the outer facial region is defined by the convex hull of the facial landmarks, eye and nostril regions are defined by a mask that is incorporated during the blending process.

For DM, we use the Automatic Face Demorphing Tools Version 1.0 of the University of Bologna described in [12]. Again, dlib is used for landmark detection. Demorphing factors of 0.15, 0.20, 0.25, and 0.30 are considered. Note that a demorphing factor of 0.25 is suggested in [12].

DFR are extracted using the well-known open-source ArcFace [15] system using a pre-trained model (`LResNet100E-IR`). The extracted feature vectors consist of 512 floating-point values. In addition, ArcFace is used as face recognition system to extract MAD scores after the DM step. For this purpose, Euclidean distances are estimated between deep face representations.

For combination and classification in the DFR MAD, feature vectors are subtracted and classified with an SVM with an RBF kernel. Data-normalisation is applied as the feature elements of extracted feature vectors are expected to have different ranges. During training, a regularisation parameter of $C = 2$ and a kernel coefficient Gamma of $1/n$ are used, where n denotes the number of feature elements. The trained SVM generates a normalised MAD score in the range $[0, 1]$ and, thus, does not need to be normalised. In addition, a linear SVM is trained for the score-level fusion. To this end, a regularisation parameter of $C = 1$ and a kernel coefficient Gamma of $1/n$ are used.

TABLE I
OVERVIEW OF USED DATASETS.

Set	Database	Subjects	Bona fide	Morphing attacks
Training	VGGVace2	5,832	10,558	10,496
Test	FRGCv2	533	3,298	3,246

We used two datasets for training the DFR-based MAD and testing, respectively. In the training stage, we employ a manually selected subset of the VGGFace2 dataset [16] consisting of suitable reference and live images. Similarly, a subset of the FRGC face image database [17] is used during testing. This subset is equal to the one used in [13][1] and example images are shown as part of figures 1 to 3. A summary of the used databases is provided in table I.

B. *Evaluation Metrics*

MAD performance is evaluated in compliance with ISO/IEC 30107-Part 3 [18] for presentation attack detection. Specifically, we estimate:

- *Attack Presentation Classification Error Rate* (APCER), which is the proportion of attack presentations or identity attacks misclassified as bona fide presentations.
- *Bona Fide Presentation Classification Error Rate* (BPCER), which is the proportion of bona fide presentations wrongly classified as attack presentations.

Furthermore, we plot Detection Error Tradeoff (DET) curves, report BPCER values at practically relevant APCERs, and the Detection Equal Error Rate (D-EER), *i.e.* the point where APCER and BPCER are equal.

[1] The list of used face images and the used morphing software are available upon request.

TABLE II
PERFORMANCE OF INDIVIDUAL MAD METHODS (IN %).

Method	**D-EER**	**BPCER at 2% APCER**	**BPCER at 3% APCER**
DFR	5.80	11.70	9.67
DM (0.15)	5.63	15.15	11.09
DM (0.20)	5.85	15.46	11.63
DM (0.25)	6.43	17.25	12.56
DM (0.30)	6.97	19.25	15.34

TABLE III
PERFORMANCE OF DIFFERENT FUSION METHODS (IN %).

Fusion Method	**D-EER**	**BPCER at 2% APCER**	**BPCER at 3% APCER**
Sum	5.20	10.87	8.32
Product	5.60	11.18	8.78
Linear SVM	5.00	9.30	6.68
Weighted sum (10:90)	5.00	9.21	7.36
Weighted sum (20:80)	**4.94**	**9.42**	**6.62**
Weighted sum (30:70)	5.06	9.95	7.14
Weighted sum (40:60)	5.02	10.19	7.82
Weighted sum (60:40)	5.42	11.12	8.59
Weighted sum (70:30)	5.51	11.12	8.90
Weighted sum (80:20)	5.60	11.24	9.30
Weighted sum (90:10)	5.73	11.36	9.39

IV. RESULTS

Table II lists the results obtained for the individual MAD methods with different factors applied in the DM algorithm. It can be observed that both methods achieve similar MAD performance. For the DM approach, best results were achieved for a factor of 0.15. Therefore, this configuration of DM is used in the fusion.

Results obtained by different fusion methods are summarised in table III. Best results are obtained for a weighted sum-based fusion in which weights of 0.2 and 0.8 are assigned to the DM and DFR method, respectively. Therefore, this fusion is suggested to be preferred over the other tested techniques. DET curves of the individual MAD methods and fusions thereof are shown in figure 4.

We stress that independent test revealed that the two differential MAD methods considered in this work represent the current state-of-the-art [11], [14]. Moreover, said test have shown that differential MAD methods generally outperforms single image MAD methods. That is, a comparison other published works, in particular single image MAD scheme, is less meaningful and therefore deliberately avoided. In contrast, this work investigates the potential of improving the best known MAD methods by combining them.

V. CONCLUSION

Face morphing attacks were found to pose a severe security risk to deployments of modern face recognition systems, in particular for border control. To counteract this, MAD methods have been proposed by various biometric research groups [1]. However, in real-world scenarios, existing approaches are hardly effective as shown in different evaluation benchmarks [11], [14]. In this work, we showed that algorithm fusion can improve the MAD performance in challenging differential scenarios. To this end, two state-of-the-art MAD methods, *i.e.* DM and DFR, were combined in a score-level fusion based on a weighted sum. While the proposed fusion-based MAD method significantly improves upon the best individual MAD method, there is still room for further improvements. This means, even though fusion techniques can be employed to boost the performance of MAD, a considerable amount of future research will be needed towards robust and reliable MAD in real-world applications.

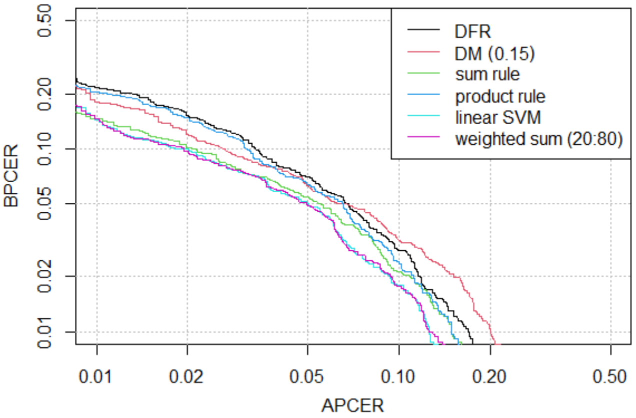

Fig. 4. DET curves of individual and fusion-based MAD methods.

ACKNOWLEDGEMENTS

This research work has partially been funded by the German Federal Ministry of Education and Research and the Hessian Ministry of Higher Education, Research, Science and the Arts within their joint support of the National Research Center for Applied Cybersecurity ATHENE.

REFERENCES

[1] U. Scherhag, C. Rathgeb, J. Merkle, R. Breithaupt, and C. Busch, "Face recognition systems under morphing attacks: A survey," *IEEE Access*, vol. 7, pp. 23 012–23 026, 2019.

[2] C. Rathgeb, R. Tolosana, R. Vera, and C. Busch, *Handbook of Digital Face Manipulation and Detection: From DeepFakes to Morphing Attacks*, 1st ed., ser. Advances in Computer Vision and Pattern Recognition. Springer-Verlag, 2022.

[3] S. Venkatesh, R. Raghavendra, K. Raja, and C. Busch, "Face morphing attack generation & detection: A comprehensive survey," *IEEE Trans. on Technology and Society*, vol. 2, no. 3, pp. 128–145, 2021.

[4] U. Scherhag, C. Rathgeb, and C. Busch, "Morph deterction from single face image: A multi-algorithm fusion approach," in *Internat'l Conf. on Biometric Engineering and Applications (ICBEA'18)*, 2018, p. 6–12.

[5] S. Venkatesh, R. Ramachandra, K. Raja, and C. Busch, "Single image face morphing attack detection using ensemble of features," in *Int'l Conf. on Information Fusion (FUSION'20)*, 2020, pp. 1–6.

[6] A. Makrushin, C. Kraetzer, J. Dittmann, C. Seibold, A. Hilsmann, and P. Eisert, "Dempster-shafer theory for fusing face morphing detectors," in *European Signal Processing Conf. (EUSIPCO'19)*, 2019, pp. 1–5.

[7] N. Damer, S. Zienert, Y. Wainakh, A. M. Saladié, F. Kirchbuchner, and A. Kuijper, "A multi-detector solution towards an accurate and generalized detection of face morphing attacks," in *Int'l Conf. on Information Fusion (FUSION'19)*, 2019, pp. 1–8.

[8] U. Scherhag, C. Rathgeb, and C. Busch, "Towards detection of morphed face images in electronic travel documents," in *13th IAPR Workshop on Document Analysis Systems (DAS'18)*, 2018, pp. 1–6.

[9] S. Lorenz, U. Scherhag, C. Rathgeb, and C. Busch, “Morphing attack detection: A fusion approach,” in *24th Intl. Conf. on Information Fusion (FUSION’21)*, 2021, pp. 703–709.

[10] C. Kraetzer, A. Makrushin, J. Dittmann, and M. Hildebrandt, “Potential advantages and limitations of using information fusion in media forensics—a discussion on the example of detecting face morphing attacks,” *EURASIP Journal on Information Security*, vol. 2021, pp. 23 012–23 026, 2021.

[11] M. L. Ngan, P. J. Grother, K. K. Hanaoka, and J. M. Kuo, “Face recognition vendor test (FRVT) part 4: MORPH - performance of automated face morph detection,” National Institute of Standards and Technology, NIST Interagency Report 8292, April 2020.

[12] M. Ferrara, A. Franco, and D. Maltoni, “Face demorphing,” *IEEE Trans. on Information Forensics and Security*, vol. 13, no. 4, pp. 1008–1017, 2018.

[13] U. Scherhag, C. Rathgeb, J. Merkle, and C. Busch, “Deep face representations for differential morphing attack detection,” *IEEE Trans. on Information Forensics and Security*, vol. 15, pp. 3625–3639, 2020.

[14] K. Raja, M. Ferrara, A. Franco, L. Spreeuwers, I. Batskos *et al.*, “Morphing attack detection - database, evaluation platform and benchmarking,” *IEEE Trans. on Information Forensics and Security*, vol. 16, pp. 4336–4351, 2020.

[15] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, “Arcface: Additive angular margin loss for deep face recognition,” in *Conf. on Computer Vision and Pattern Recognition (CVPR)*. IEEE, June 2019, pp. 4690–4699.

[16] Q. Cao, L. Shen, W. Xie, O. M. Parkhi, and A. Zisserman, “VGGFace2: A dataset for recognising faces across pose and age,” *Intl. Conf. on Automatic Face and Gesture Recognition (FG)*, 2018.

[17] J. Phillips, P. Flynn, T. Scruggs, K. Bowyer, J. Chang *et al.*, “Overview of the Face Recognition Grand Challenge,” in *Conf. on Computer Vision and Pattern Recognition (CVPR)*, vol. 1, 2005, pp. 947–954.

[18] ISO/IEC JTC1 SC37 Biometrics, *ISO/IEC 30107-3. Information Technology - Biometric presentation attack detection - Part 3: Testing and Reporting*, International Organization for Standardization, 2017.