Performance Variation of Morphed Face Image Detection Algorithms across different Datasets

Ulrich Scherhag, Christian Rathgeb and Christoph Busch
da/sec – Biometrics and Internet Security Research Group, Hochschule Darmstadt, Germany
{ulrich.scherhag,christian.rathgeb,christoph.busch}@h-da.de

***Abstract*—In past years, different researchers have shown the vulnerability of face recognition systems to attacks based on morphed face images. More recently, first morph detection subsystems have been proposed to automatically detect this kind of fraud. While some algorithms have been reported to reveal practical detection performance on individual datasets a systematic analysis of proposed detectors with respect to their robustness across different databases has remained elusive.**

In this work, we evaluate the performance of different morph detection algorithms across disjoint datasets of 2,745 bona fide and 14,337 automatically generated morphed face images. Within a generic evaluation framework a systematic robustness estimation scheme is proposed to identify reliable detection algorithms. Finally, the robustness of algorithms which have been determined as most promising is verified on another disjoint dataset. Hence, this paper represents the first attempt towards a comprehensive cross-database performance evaluation and a systematic evaluation of the robustness of morphed face image detection algorithms.

I. Introduction

Morphing techniques can be used to create artificial biometric samples, which resemble the biometric information of two (or more) individuals in image and feature domain. If morphed biometric images are infiltrated to a biometric recognition system the subjects contributing to the morphed image will both (or all) be successfully verified against that single enrolled reference data. Hence, the unique link between individuals and their biometric reference data is not warranted. In particular face recognition systems have been exposed to be highly vulnerable to attacks based on morphed face images [1], [2]. Fig. 1 shows an example of morphing two facial images.

In the recent past, researchers have presented different approaches to detect morphed face images including general purpose texture descriptors [2], digital image forensic analysis [3], keypoint descriptors and edge operators [4] or deep learning methods [5], [6]. While some of the mentioned approaches report practical detection error rates these are commonly evaluated on a dataset of bona fide and morphed face images which is extracted from a single (in-house) face database. In such an experimental setup the use of machine learning-based feature extractors or/and classifier increases the risk of overfitting, i.e. the robustness of presented morph detection algorithms with regard to images stemming from other sources is therefore questionable.

In this paper, we propose a general framework to estimate the robustness of morph detection algorithms. Four datasets referred to as *training*, *testing*, *evaluation* and *validation* which comprise bona fide and morphed face images are extracted from three different face databases. Different morph detection algorithms based on texture descriptors, keypoint extractors and gradient estimators are trained and detection errors are calculated on the test and the evaluation set. Obtained detection errors serve as input to a robustness estimation function, which in conjunction with appropriate exclusion criteria, is shown to reliably predict the detection performance of an algorithm on the disjoint validation database. Moreover, the presented evaluation strategy is generic and, hence, transferable to other pattern classification problems.

(a) Subject 1 (b) Morph (c) Subject 2

Fig. 1: Examples for bona fide and morphed face images

This paper is organized as follows: related works are summarized in Sect. II. Sect. III describes the proposed framework to evaluate the robustness of morph detection algorithms in detail. The used morph detection subsystems are listed in Sect. IV. Experimental results are reported and discussed in Sect. V. Finally, conclusions are drawn in Sect. VI.

II. Related Work

Attacks based on morphed biometric samples were first introduced by Ferrara *et al.* [1]. Motivated by security gaps in the issuance process of electronic travel documents, the authors showed that commercial face recognition software tools are highly vulnerable to such attacks, i.e. images of either subject are successfully matched against the morphed image. In their experiments, decision thresholds yielding a FMR of 0.1% have been used, according to the guidelines provided by the European Agency for the Management of Operational Cooperation at the External Borders (FRONTEX) [7]. In a further study, the authors show that morphed face images are realistic enough to fool human examiners [8].

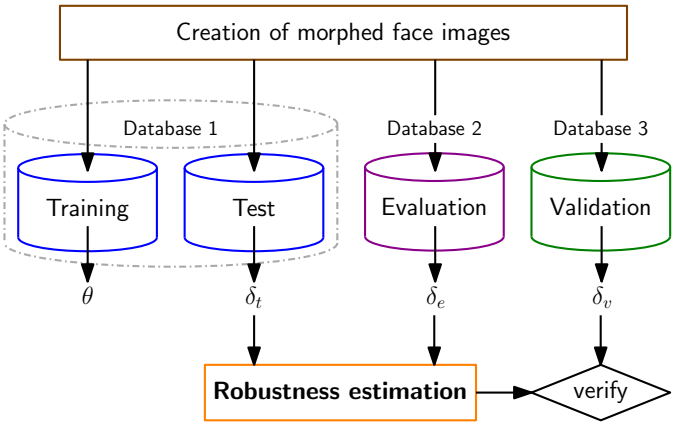


Fig. 2: Proposed robustness analysis scheme

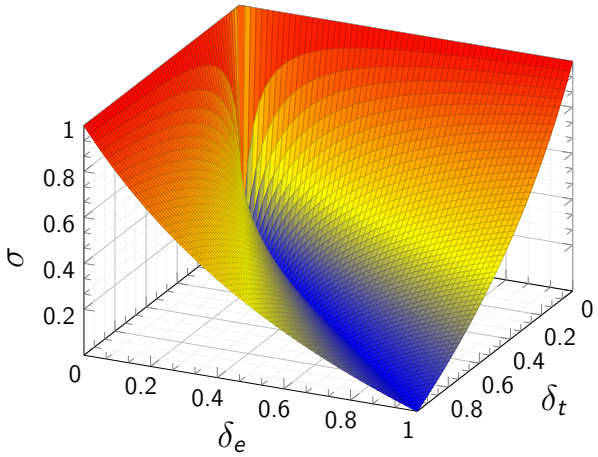


Fig. 3: Robustness estimation function

Scherhag *et al.* [2] reported moderate detection performance for benchmarking several general purpose texture descriptors used in conjunction with machine learning techniques to detect morphed face images. In [9] he demonstrated, that an improvement of the detection performance of multi-purpose image descriptors can be achieved by a fusion of multiple detection algorithms. Ferrara *et al.* [10] proposed face demorphing for morph detection employing a trusted live capture in addition to the questioned sample. Scherhag *et al.* [11] analysed multiple general purpose image feature extractors in this differential scenario. Further, Hildebrandt *et al.* [3] suggest to employ generic image forgery detection techniques, in particular multi-compression anomaly detection, to reliably detect morphed facial images. Kraetzer *et al.* [4] evaluate the feasibility of detecting facial morphs with keypoint descriptors and edge operators. The benefits of deep neural networks for detecting morphed images has been recently investigated by Ramachandra *et al.* [5] and Seibold *et al.* [6]. It is important to note that, apart from [10], published morph detection approaches are evaluated on images of a single face database. That is, the transferability of the presented scheme with respect to different datasets and morphing techniques yet needs to be analysed.

Gomez-Barrero *et al.* [12], [13] proposed the first theoretical framework for measuring the vulnerability of biometric systems to attacks. Evaluations are conducted for diverse biometric systems where expected comparison scores of attacks based on morphed images or templates are directly derived from the mated and non-mated distributions of a face, fingerprint and iris recognition system. The authors identified key factors which take a major influence on a system's vulnerability to such attacks, e.g. the shape of genuine and impostor score distributions or the FMR the system is operated at. To evaluate the vulnerability of biometric systems to attacks based on morphed images or templates Scherhag *et al.* [14] introduced metrics for vulnerability reporting, which strongly relate to the metrics defined in [15]. In addition, the authors provide recommendations on the assessment of morphing techniques. It is emphasized that unrealistic assumptions with respect to the quality of morphed biometric samples might cloud the picture regarding the performance of detection algorithms. Eventually, it is important to note that so far there is no publicly available database of morphed face images and no publicly available morph detection algorithms.

III. Robustness Analysis

In the following subsections, the constituting processing steps of the proposed analysis and the employed robustness estimation are described in detail.

A. Processing steps

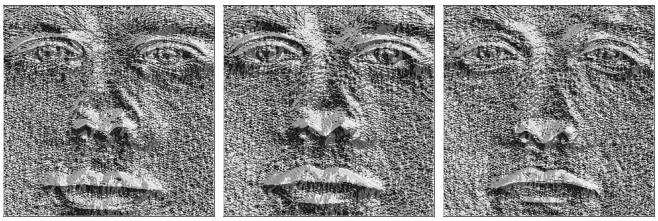
The proposed robustness analysis which is depicted in Fig. 2 comprises four major processing steps:

1) *Data preparation*: in the first step three face databases are used to create face morphs of high quality; the first database is divided in a *training* set which is employed to train classifiers for each feature extractor and a *test* set. The remaining two datasets are referred to as *evaluation* and *validation* set.
2) *Detection error estimation*: first detection error $\boldsymbol{\delta} \in [0, 1]$ is estimated on the training set and the optimal decision threshold θ is calculated. Then corresponding detection errors are evaluated on the test and evaluation set using the predefined decision threshold θ. The resulting detection errors are denoted by $\boldsymbol{\delta}_t$ and $\boldsymbol{\delta}_e$.
3) *Robustness estimation*: the robustness (or fragility) is estimated as a function of $\boldsymbol{\delta}_t$ and $\boldsymbol{\delta}_e$. In addition, appropriate exclusion criteria are defined in order to discard impractical morph detectors with regard to the employed algorithms and used parameters.
4) *Robustness validation*: finally, the robustness of selected morph detection algorithms is verified on the validation set.

The presented procedure is designed to minimize the risk of overfitting. Due to the fact that the proposed robustness estimation is performed on disjoint datasets overfitting caused by dataset specific properties, e.g. compression artefacts, is avoided. Further, potential shifts in the feature space are considered by the use of a fixed decision threshold.

(a) Subject 1 (b) Morph (c) Subject 2

Fig. 4: Examples for bona fide and morphed face images



(a) Subject 1 (b) Morph (c) Subject 2

Fig. 5: Example for BSIF responses

B. Robustness estimation

Let $\boldsymbol{\delta}_t$ and $\boldsymbol{\delta}_e$ denote the detection errors obtained during performance evaluations on the employed *test* and *evaluation* set, respectively. We suggest to estimate the fragility $\boldsymbol{\sigma} \in [0, 1]$ of a detection algorithm as,

$$\boldsymbol{\sigma} = \begin{cases} \|2\boldsymbol{\delta}_e/(\boldsymbol{\delta}_t + \boldsymbol{\delta}_e) - 1\| & \text{if } \boldsymbol{\delta}_t + \boldsymbol{\delta}_e \neq 0, \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

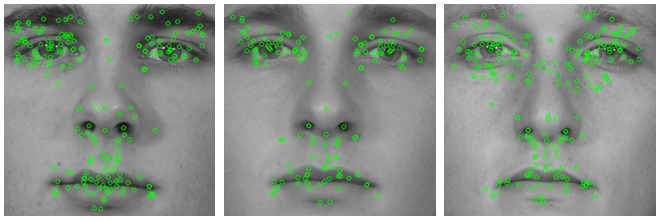
The behaviour of this function is plotted in Fig. 3 where blue regions indicate robust detection performance. If the relative difference of detection performance obtained on the test and evaluation set is small the algorithm is assume to be robust. Further, it is suggested to define two thresholds, $\mathbf{t}_\delta$ and $\mathbf{t}_\sigma$, as exclusion criteria. If $\max\{\boldsymbol{\delta}_t, \boldsymbol{\delta}_e\} < \mathbf{t}_\delta \wedge \boldsymbol{\sigma} < \mathbf{t}_\sigma$ applies the detection algorithm is identified as robust. The maximum acceptable detection error is represented by $\mathbf{t}_\delta$. Similarly, $\mathbf{t}_\sigma$ provides an upper boundary for the maximum relative difference between detection errors on the test and evaluation set.

IV. Detection Subsystems

In the following subsection, we briefly summarize the employed morph generation, pre-processing, feature extractors and comparison.

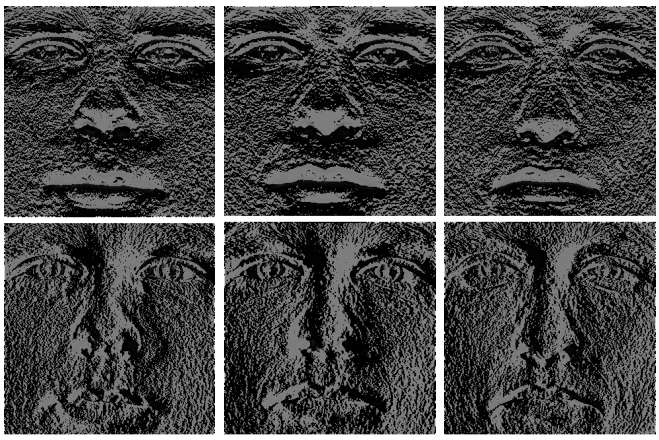
A. Morph generation and pre-processing

In order to morph two face images the *dlib* facial landmark detector [16] is applied to both images. Subsequently, a Delaunay triangulation is performed to the average of corresponding points. An affine transform is then applied to the sets of triangles in both face images resulting in two warped images which are alpha blended using a alpha value of 0.5. In the pre-processing stage an image is segmented and normalized according to eye coordinates detected by the *dlib* landmark detector [16]. Subsequently, the normalized region is



(a) Subject 1 (b) Morph (c) Subject 2

Fig. 6: Example for SURF keypoint detection



(a) Subject 1 (b) Morph (c) Subject 2

Fig. 7: Example for sharpness features (two dimensions)

cropped to 320×320 pixels using predefined offsets to ensure that the morph detection algorithm is only applied to the facial region. Finally, the cropped face part is converted to a grayscale image. Fig. 4 depicts pre-processed face images of two subjects and their corresponding morph.

B. Feature extraction

At feature extraction the pre-processed face image is optionally divided into 4×4 cells to retain local information. That is, feature extractors are applied separately on texture cells and the final feature vector is formed as a concatenation of feature vectors extracted from each cell. We employ the following three types of feature extraction methods (two algorithms are considered per type):

Texture descriptors: Local Binary Patterns (LBP) [17] and Binarized Statistical Image Features (BSIF) [18] are extracted from the cropped face images. For details on these texture descriptors the reader is referred to [17], [18]. While LBP simply processes neighbouring pixel values of each pixel, BSIF utilizes specific filters learned from a set of images. Obtained feature values are stored in a corresponding histograms. The use of these well-established general purpose texture descriptors has shown to be successful in diverse texture classification problems. Focusing on morph detection the process of image morphing is expected to cause changes in textual properties between bona fide and morphed face images. An example of BSIF applied to the images of Fig. 4 is depicted in Fig. 5. By testing different spatial sampling rates the two

TABLE I: Datasets used for experimental evaluations (a detailed list of used images will be published upon acceptance)

Gender	Database 1 (FRGCv2) Training				Database 1 (FRGCv2) Test				Database 2 (ARface) Evaluation				Database 3 (FERET) Validation			
	No. of subjects	No. of images	Bona fide images	Morphed images	No. of subjects	No. of images	Bona fide images	Morphed images	No. of subjects	No. of images	Bona fide images	Morphed images	No. of subjects	No. of images	Bona fide images	Morphed images
Male	59	2,819	1,166*	1,653	58	2,210	499	1,711	76	3,043	193	2,850	104	1,230	149	1,081
Female	40	2,073	1,332*	741	39	1,165	462	703	59	1,934	164	1,770	24	3,857	29	3.828
All	99	4,892	2,498*	2,394	97	3,375	961	2,414	135	4,977	357	4,620	128	5,087	178	4,909

* in the training set bona fide images are horizontally mirrored

best configurations for LBP and BSIF were determined, thus, 3×3 and 9×9 LBP-patches and same sized BSIF filter sets extracting 8 bit per pixel for 3×3 and 12 bit per pixel for 9×9 are employed.

Keypoint extractors: Scale Invariant Feature Transform (SIFT) [19] and Speeded Up Robust Features (SURF) [20] extract sets of local keypoints. For details on keypoint detection, the extraction of keypoint descriptors and keypoint matching the reader is referred to [19], [20]. Keypoint extractors are employed, since morphed images are expected to contain fewer keypoint locations, which are defined as maxima and minima of the result of difference of Gaussians function. That is, the amount of detected keypoints is used as descriptive feature as it is also suggested in [4]. Fig. 6 shows an example of SURF keypoints detected in the images of Fig. 4.

Gradient estimators: Histogram of Gradients (HOG) and sharpness features are extracted from the normalized grayscale images. For further details to HOG the reader is referred to [21]. As a sharpness feature the mean of the gradient in two dimensions are calculated. The use of gradient-based methods is motivated by the fact that due to the morphing process high frequency changes are reduced and, hence, the steepness of gradients is decreased. An example of sharpness features extracted from the images of Fig. 4 is depicted in Fig. 7.

C. Comparison

Feature vectors are extracted from the training databases for each algorithm and support vector machines (SVMs) are trained to distinguish between bona fide and morphed face images using a disjoint training set. For a given face image the SVMs of each single algorithm generate a normalized attack detection score.

V. Experiments

In the following subsections, we describe the experimental setup, conduct a vulnerability assessment of a Custom-Of-The-Shelf (COTS) face recognition system to attacks based on the generated morphed face images, report the detection performance of the used detection systems and perform and validate the proposed robustness estimation.

A. Experimental setup

According to the described robustness analysis evaluations are carried out on three databases, namely subsets of the FRGCv2 (training and test), ARface (evaluation) and FERET face database (validation). A total number of 2,745 frontal faces have been manually chosen and ICAO compliance has been verified, i.e. the distance between the eyes of a face has to be at least 90 pixels [22]. Within these subsets 14,337 morphed faces have been automatically generated for pairs of subjects of same gender using the *OpenCV* library. Example images of bona fide and morphed face images are shown in Fig. 8 which illustrates the high quality of morphed face images being well in the quality limits set forth by ICAO and ISO/IEC standards. Details about employed databases are listed in Table I.

(a) Subject 1 (b) Morph (c) Subject 2

Fig. 8: Examples of bona fide and morphed face images of different datasets. Top to bottom: test, evaluation, validation

B. Face recognition vulnerability assessment

The vulnerability of a COTS face recognition system to attacks based on the generated morphed face images is assessed according to the metrics specified in [14], in particular, in terms of Mated Morph Presentation Match Rate (MMPMR). This metric is an adaptation of the general Impostor Attack Presentation Match Rate (IAPMR) introduced in ISO/IEC 30107-3 [15] which is defined as the proportion of attack presentations using the same presentation attack instrument species in which the target reference is matched. However, in the adaptation the MMPMR covers the fact that not only one target subject (contained in the morphed image) is matched - but both subjects who earlier contributed to the morphed

TABLE II: Robustness estimation of used morph detectors

System	Test			Evaluation			$\sigma / \max\{\delta_t, \delta_e\}$	Validation		
	APCER(θ)	BPCER(θ)	D-EER	APCER(θ)	BPCER(θ)	D-EER		APCER(θ)	BPCER(θ)	D-EER
$LBP_{1,3}$	13.3%	1.2%	5.5%	0.0%	100.0%	49.8%	0.75 / 0.50	5.1%	32.6%	23.8%
$LBP_{4,3}$	13.7%	1.2%	5.7%	0.0%	100.0%	49.7%	0.74 / 0.50	6.2%	32.6%	23.1%
$LBP_{1,9}$	16.8%	10.3%	14.0%	0.0%	99.4%	14.8%	0.57 / 0.49	63.9%	9.6%	20.9%
$LBP_{4,9}$	16.0%	9.7%	13.2%	0.0%	99.2%	24.2%	0.59 / 0.49	64.5%	9.6%	19.6%
$BSIF_{1,3}$	6.5%	1.9%	3.3%	0.0%	100.0%	49.9%	0.85 / 0.50	0.0%	79.2%	13.3%
$BSIF_{4,3}$	8.7%	2.0%	4.5%	0.0%	100.0%	49.9%	0.81 / 0.50	0.8%	45.5%	8.7%
$BSIF_{1,9}$	22.9%	17.6%	19.7%	0.0%	99.4%	20.7%	0.42 / 0.49	67.8%	0.6%	10.3%
$BSIF_{4,9}$	22.2%	21.3%	21.4%	1.7%	80.4%	24.8%	0.31 / 0.41	88.1%	0.6%	9.5%
$SIFT_1$	14.0%	22.4%	20.5%	12.3%	71.1%	43.4%	0.39 / 0.41	10.7%	32.6%	24.4%
$SIFT_4$	14.6%	13.0%	13.7%	39.8%	45.4%	41.8%	0.51 / 0.46	27.9%	21.9%	23.3%
$SURF_1$	11.1%	37.6%	15.7%	32.2%	43.4%	33.5%	0.22 / 0.37	24.0%	58.4%	29.6%
$SURF_4$	28.0%	13.6%	18.4%	76.5%	5.6%	37.2%	0.33 / 0.41	62.0%	7.9%	23.9%
$Sharp_1$	21.7%	25.7%	23.4%	26.6%	55.2%	42.0%	0.27 / 0.40	82.7%	10.1%	42.1%
$Sharp_4$	0.0%	99.8%	16.6%	50.0%	0.0%	100.0%	0.00 / 0.50	0.0%	100.0%	43.4%
HOG_1	*12.4%*	*12.0%*	*12.1%*	*8.8%*	*42.9%*	*18.8%*	*0.36 / 0.25*	*51.4%*	*1.1%*	*12.9%*
HOG_4	36.2%	28.3%	32.1%	22.9%	47.3%	32.9%	0.04 / 0.35	85.3%	7.9%	36.9%
DFF	31.1%	31.2%	30.6%	32.2%	45.7%	39.0%	0.11 / 0.38	48.0%	20.8%	29.3%

image are expected to be matched if the morphing attack is considered to be successful.

When employing the default decision threshold of the COTS face recognition system a MMPMR of 1 is obtained. This means all face images of subjects contributing to a morphed face image are successfully matched against it, hence, the attacks reveal a success chance of 100% on a COTS.

C. Robustness estimation and validation

The Attack Presentation Classification Error Rate (APCER) is defined as the proportion of attack presentations using the same presentation attack instrument species incorrectly classified as bona fide presentations in a specific scenario. The performance of the detection algorithms is reported according to metrics defined in ISO/IEC 30107-3 [15]. The Bona Fide Presentation Classification Error Rate (BPCER) is defined as the proportion of bona fide presentations incorrectly classified as presentation attacks in a specific scenario. The Detection Equal Error Rate (D-EER), i.e. the operation point where APCER = BPCER, is used as optimal operation point which is estimated during training. Based on the evaluated decision threshold θ, (APCER(θ)+BPCER(θ))/2, will be used as detection error. Alternatively, other error estimates could be employed.

The performance of the analysed algorithms for all three datasets is listed in Table II. For the texture descriptors the patch size is indicated via a second sub-index, e.g. $LBP_{4,3}$ corresponds to a division into 4×4 cells and a LBP-patch size of 3×3 pixels. Significant drops in terms of APCER and BPCER can be observed on the evaluation dataset, i.e. the vast majority of detection algorithms do not seem to generalize. The detection performance on evaluation set is generally much lower than on the other databases caused by the high level of blurriness of the facial images. The significant gap between D-EER and the resulting APCER and BPCER values at a fixed threshold θ can partly be reduced by a calibration of the systems towards the images, i.e. by an adaptation of the decision threshold θ. We emphasize on the fact that an

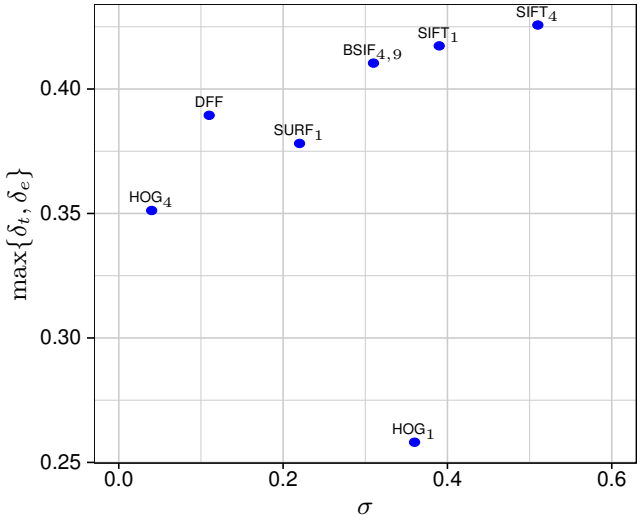


Fig. 9: Scatter plot of selected morph detectors

evaluation of morph detectors on a single database might cloud the picture over the actual detection performance. If the exclusion criteria are set to reasonable thresholds, e.g. $t_\delta = 0.3$ and $t_\sigma = 0.4$, the only algorithm identified as robust is HOG_1 (emphasized), as can be seen in Fig. 9. Even if the algorithm shows a moderate performance for both, evaluation and validation set, the fragility over the datasets is comparatively low. Moreover, the detection performance is maintained on the validation database which confirms the soundness of the proposed robustness estimation. A fusion of multiple morph detection algorithms, e.g. in [9], might be expected to be more robust with respect to the analysed scenario.

VI. Conclusion

In this work, we evaluate the performance of different morph detection algorithms across disjoint datasets of 2,745 bona fide and 14,337 automatically generated morphed face images. Within a generic evaluation framework a systematic robustness estimation scheme is proposed to identify reliable

detection algorithms. An evaluation of general purpose image descriptors employing the proposed framework shows the high database dependency. This results in alarmingly high error rates if morph detectors are evaluated across different databases. That is, for most algorithms an adaptation to each database is mandatory. In summary, this paper represents the first attempt towards a comprehensive cross-database performance evaluation and a systematic evaluation of the robustness of morphed face image detection algorithms.

ACKNOWLEDGEMENTS

This work was supported by the German Federal Ministry of Education and Research (BMBF) as well as by the Hessen State Ministry for Higher Education, Research and the Arts (HMWK) within the Center for Research in Security and Privacy (CRISP).

REFERENCES

[1] M. Ferrara, A. Franco, and D. Maltoni, "The magic passport," in *Proc. Int. Joint Conf. on Biometrics (IJCB)*, 2014, pp. 1–7.

[2] U. Scherhag, R. Raghavendra, K. B. Raja, M. Gomez-Barrero, C. Rathgeb, and C. Busch, "On the vulnerability of face recognition systems towards morphed face attacks," in *Proc. Int. Workshop on Biometrics and Forensics (IWBF)*, 2017, pp. 1–6.

[3] M. Hildebrandt, T. Neubert, A. Makrushin, and J. Dittmann, "Benchmarking face morphing forgery detection: Application of stirtrace for impact simulation of different processing steps," in *Proc. Int. Workshop on Biometrics and Forensics (IWBF)*, 2017, pp. 1–6.

[4] C. Kraetzer, A. Makrushin, T. Neubert, M. Hildebrandt, and J. Dittmann, "Modeling attacks on photo-ID documents and applying media forensics for the detection of facial morphing," in *Proc. Workshop on Information Hiding and Multimedia Security (IH& MMSec)*, 2017, pp. 21–32.

[5] R. Ramachandra, K. Raja, S. Venkatesh, and C. Busch, "Transferable deep-cnn features for detecting digital and print-scanned morphed face images," in *2017 IEEE Conf. on Computer Vision and Pattern Recognition Workshop (CVPRW)*, July 2017.

[6] C. Seibold, W. Samek, A. Hilsmann, and P. Eisert, "Detection of face morphing attacks by deep learning," in *Proc. Int. Workshop on Digital Watermarking (IWDW)*, 2017, pp. 107–120.

[7] "FRONTEX – Research and Development Unit: Best practice technical guidelines for automated border control (ABC) systems," 2012, version 2.0.

[8] M. Ferrara, A. Franco, and D. Maltoni, "On the effects of image alterations on face recognition accuracy," in *Face Recognition Across the Imaging Spectrum*, T. Bourlai, Ed. Springer International Publishing, 2016, pp. 195–222.

[9] U. Scherhag, C. Rathgeb, and C. Busch, "Morph detection from single face images: a multi-algorithm fusion approach," in *International Conference on Biometric Engineering and Applications 2018 (ICBEA)*, 2018, pp. 1–7.

[10] M. Ferrara, A. Franco, and D. Maltoni, "Face demorphing," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 4, pp. 1008–1017, April 2018.

[11] 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)*, 2018, pp. 1–6.

[12] M. Gomez-Barrero, C. Rathgeb, U. Scherhag, and C. Busch, "Is your biometric system robust to morphing attacks?" in *Proc. Int. Workshop on Biometrics and Forensics (IWBF)*, 2017, pp. 1–6.

[13] ——, "Predicting the vulnerability of biometric systems to attacks based on morphed biometric information," *IET Biometrics*, 2018.

[14] U. Scherhag, A. Nautsch, C. Rathgeb, M. Gomez-Barrero, R. Veldhuis, L. Spreeuwers, M. Schils, D. Maltoni, P. Grother, S. Marcel, R. Breithaupt, R. Raghavendra, and C. Busch, "Biometric systems under morphing attacks: Assessment of morphing techniques and vulnerability reporting," in *Int. Conf. of the Biometrics Special Interest Group (BIOSIG)*, 2017, pp. 1–12.

[15] ISO/IEC JTC1 SC37 Biometrics, *ISO/IEC IS 30107-3:2017, IT – Biometric presentation attack detection – Part 3: Testing and Reporting*.

[16] D. E. King, "Dlib-ml: A machine learning toolkit," *Journal of Machine Learning Research*, vol. 10, 2009.

[17] S. Liao, X. Zhu, Z. Lei, L. Zhang, and S. Z. Li, "Learning multi-scale block local binary patterns for face recognition," in *Proc. Int'l Conference on Biometrics (ICB'07)*, 2007, pp. 828–837.

[18] J. Kannala and E. Rahtu, "BSIF: Binarized statistical image features," in *Proc. Int'l Conf. on Pattern Recognition (ICPR'12)*, 2012, pp. 1363–1366.

[19] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[20] H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool, "Speeded-Up Robust Features (SURF)," *Computer Vision and Image Understanding*, vol. 110, no. 3, pp. 346 – 359, 2008.

[21] C. Shu, X. Ding, and C. Fang, "Histogram of the oriented gradient for face recognition," *Tsinghua Science & Technology*, vol. 16, no. 2, pp. 216–224, 2011.

[22] International Organization for Standardization, "Information technology – Biometric data interchange formats – Part 5: Face image data," JTC 1/SC 37, ISO/IEC 19794-5:2005 consolidated, 2005.