Detecting Morphed Face Attacks Using Residual Noise from Deep Multi-scale Context Aggregation Network

Sushma Venkatesh[†], Raghavendra Ramachandra[†], Kiran Raja[$],
Luuk Spreeuwers[‡], Raymond Veldhuis [‡], Christoph Busch[†]
[†]Norwegian Biometrics Laboratory | [$]IDI, NTNU, Norway
[‡]University of Twente, Enschede, Netherlands.
{sushma.venkatesh; raghavendra.ramachandra;kiran.raja;christoph.busch} @ntnu.no
{l.j.spreeuwers;r.n.j.veldhuis@utwente.nl}

Abstract

Along with the deployment of the Face Recognition Systems (FRS), concerns were raised related to the vulnerability of those systems towards various attacks including morphed attacks. The morphed face attack involves two different face images in order to obtain via a morphing process a resulting attack image, which is sufficiently similar to both contributing data subjects. The obtained morphed image can successfully be verified against both subjects visually (by a human expert) and by a commercial FRS. The face morphing attack poses a severe security risk to the e-passport issuance process and to applications like border control, unless such attacks are detected and mitigated. In this work, we propose a new method to reliably detect a morphed face attack using a newly designed denoising framework. To this end, we design and introduce a new deep Multi-scale Context Aggregation Network (MS-CAN) to obtain denoised images, which is subsequently used to determine if an image is morphed or not. Extensive experiments are carried out on three different morphed face image datasets. The Morphing Attack Detection (MAD) performance of the proposed method is also benchmarked against 14 different state-of-the-art techniques using the ISO-IEC 30107-3 evaluation metrics. Based on the obtained quantitative results, the proposed method has indicated the best performance on all three datasets and also on cross-dataset experiments.

1. Introduction

An electronic Machine Readable Travel Document (eMRTD) is a governmental document (e.g. an electronic Passport) that stores face biometric reference images corresponding to the owner of the document. When a bona fide citizen makes the application for an eMRTD in his respective country, the applicant provides a passport photo that is taken by a photographer. Depending upon the type of the application (online or in-person), the applicant submits his/her passport photo either in digital or printed form, where printed passport photos are subsequently scanned for the digitized eMRTD production process. The submitted passport photo either in *digital* or re-digitized through scanning *i.e. print-scan*) is stored in the eMRTD.

A malicious actor in such a setting can submit a morphed face image and obtain a valid eMRTD leading to exploitation of intrinsic intra-class variation tolerance of a Face Recognition Systems (FRS), which was revealed as a serious vulnerability of FRS [13]. The morphed face image generated using the face image from an attacker and a accomplice can easily be verified against both contributing subjects with existing commercial FRS. Also a human expert such as a trained border guard can be confused [14, 15, 1, 16, 17, 18, 19, 20]. This scenario becomes critical, when attackers intentionally morph their face image with a non-blacklisted subject, in order to gain access to a protected/secured area. This poses a severe threat to the security and efficacy of border control or similar applications (using eMRTD) and thereby, it is crucial to identify such morphed face images and to prevent the attacks. A sample of morphed face image and the obtained comparison scores using a commercial FRS is illustrated in Figure 1.

Motivated by the problem, several Morphing Attack Detection (MAD) techniques to flag digital morphed face images and print-scanned morphed face images have been proposed [15, 1, 18, 19, 20, 21, 7, 10, 8]. In this work, we focus on detecting digital morphed face images as: (i) they can be easily generated in the digital domain, (ii) digital images are used in several countries like New-Zealand, Estonia, Ireland, etc. to issue/renew the documents and (iii) the constitute a low-cost attack in digital domain. Further, it has to be noted that the digital morph image is usually uploaded

Table 1: State-of-the-art digital MAD techniques

Reference	Algorithm Type	Algorithm
Raghavendra et al. [1]	Texture based method	Local binary pattern (SVM), Binary Statistical Image Features (BSIF), Image Gradient(IG)
Makrushin et al. [2]	Quantized DCT coefficients	Benford features
Hildebrandt et al. [3]	Stir trace based scenario	Multi-compressed Anomaly detection
Neubert [4]	Image degradation approach	Corner feature detector
Seibold et al. [5]	Deep learning based approach	VGG19, GoogleNet, AlexNet
Asaad and Sabah [6]	Texture based scenario	Topological data analysis approach
Scherhag et al. [7]	Texture and frequency based method	LBP, LPQ, BSIF, 2DFFT with SVM classifier
Debiasi et al. [8]	Image Quality	PRNU using Wavelet denoising
Raghavendra et al. [9]	Deep CNN based method	Feature fusion of fully connected layers of VGG19 and Alex Net
Damer et al. [10]	Deep and texture features	Feature fusion of LBP and Openface Net
Ferrara et al. [11]	Deep features	AlexNet, VGG19, VGG16, ResNet50
Venkatesh et al. [12]	Deep residual noise	Color residual noise with SRKDA

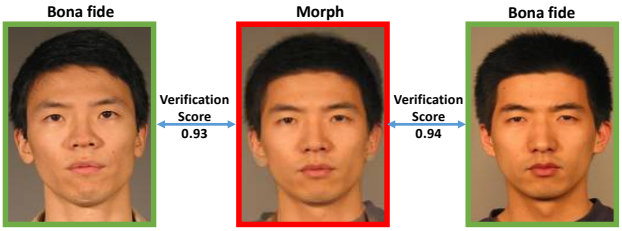

Figure 1: Illustration of successful verification with morphed image in a COTS Face Recognition System (FRS) operating at $FAR = 0.1\%$ (a) Subject 1 (b) Morphed face image (c) Subject 2

to an online passport application portal by the applicant and there is no human control to verify the authenticity of image as in a physical passport application procedure.

1.1. Related Works

In this section, we summarize the existing MAD techniques in Table 1 for a quick comprehension of the reader. As observed from Table 1, the most prevalent MAD techniques can be broadly divided into four algorithm types: (a) texture-based (b) image quality based (c) deep learning-based (d) hybrid features (combined/multiple features) based detection. The first work on detecting the morphed face images based on micro-textures was presented in [1]. Following this work, several other works are reported [21, 7] using the capability of micro-texture extraction techniques that can effectively capture the variations to reflect the process of morphing, which aids the morph detection task. Lately, the use of pre-trained deep CNNs with different architectures are widely studied in [5, 9, 11]. Further, the combination of deep features with handcrafted features is proposed in [10]. Recently, the spectral analysis of Photo Response Non-Uniformity (PRNU) has been employed [8][22], to analyse modifications caused by the morphing procedure. For a quick overview of the existing state of the art based on morph attack detection are presented in [23]. In the recent past several approaches based on hybrid features and deep features are presented [24][20][25]. The combination of deep features with handcrafted features is proposed in [10]. Recently, the residual noise computed on the color channels using deep CNN based denoising is presented for the reliable face morph detection [12].

1.2. Our Contributions

Intrigued by the effectiveness of the photo-response noise and it's success in detecting morphed attacks, we investigate to detect the noise of the morphing process using a new approach. We assert that the strategy of localizing such a noise using learning approaches lead to better detection of morphing attacks. Thus, in this work, we present a novel method for the face morphing attack detection by computing the residual noise, which can be attributed to the morphing process. The intuition behind resorting to such an approach of determining the noise using a deep learning paradigm is due to three specific reasons, where the resulting noise due to the morphing process can be: (i) random (ii) non-deterministic and abrupt (iii) sparsely distributed.

Given such properties, we first focus on commonly characterized noise in the image domain and the approaches to denoise them. The widely employed denoising approaches include Wavelet Denoising (WD) [26], Block Matching and 3D filtering (BM3D) [27], Multi-resolution Bilateral Filtering (MBF) [28] and Denoising Convolutional Neural Networks (DnCNN) [29] which can intuitively cover the possible noise in morphing process. A combination of all such denoising approaches can lead to better morphing attack detection, as asserted earlier. However, the complexity in time and parameterization of each of these approaches can lead to the cumbersome effort. In the light of the recent ad-

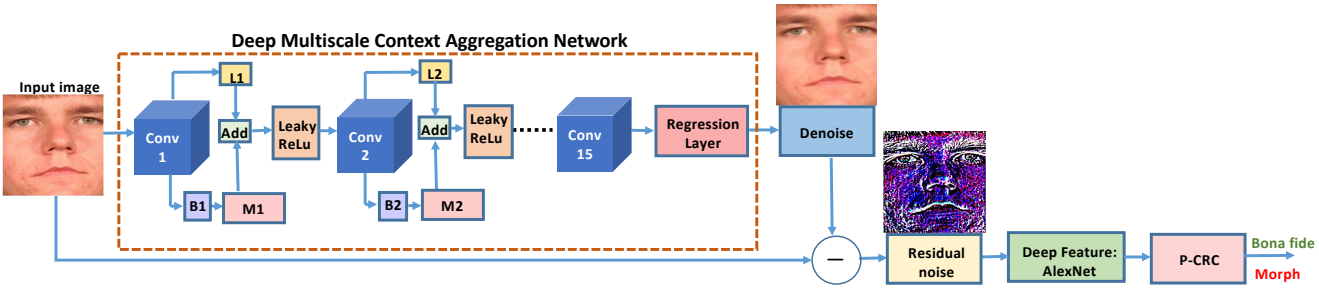

Figure 2: Block diagram of the proposed method. B denotes batch-normalization, M represents the scale layer that adjusts the strength of the batch-normalization, L corresponds to strength of the identity branch in batch-normalization.

vancements in deep learning, we propose to aggregate the denoising approaches [26, 27, 28, 29] using a deep Multi-scale Context Aggregation Network (MS-CAN) such that the noise in the morphed image can be easily determined, i.e., given the face image I, we obtain the denoised face image I_d using the MS-CAN. We then compute the residual noise I^r, which is employed to determine if the image I is morphed or not (bona fide). Given the residual noise image, we adapt the pre-trained off-the-shelf AlexNet to extract textural features. These features are then classified using a Collaborative Representative Classifier (CRC) to discriminate between the bona fide and morphed image.

The key contributions of this work can therefore be summarized as:

- We present a novel method for detecting morphed face images based on the deep textural features of residual noise from image.
- We introduce a deep Multi-scale Context Aggregation Network (MS-CAN) for aggregating four denoising methods to consider various kinds of noise characteristics.
- We present results and extensive experiments on three different face morphing datasets, and benchmark the results for our proposed approach with 14 different state-of-the-art techniques.

The rest of the paper is organized as follows: Section 2 presents the proposed method, Section 3 discusses the morphed face dataset used in this work, Section 4 discusses the quantitative performance of the state-of-the-art face Morphing Attack Detection (MAD) together with the proposed method under different evaluation protocols. Finally, Section 5 draws the conclusion.

2. Proposed Method

As noted earlier, the morphing process can involuntarily introduce noise in the resulting morphed image. The core of the proposed method is therefore to quantify the morphing noise effectively given the recent work indicating the effectiveness of noise characterization in detecting morphing attacks [8][**?**]. The motivation of this work is to explore the image denoising methods to quantify the noise and thereby detect the face morphing attacks reliably. The residual noise obtained from the image can enable reliable detection of no-reference (single image) based morph images. The proposed approach for such motivation is provided in Figure 2, which characterizes the noise pattern. The proposed method can be visualized in two main parts: (a) aggregation of multiple denoising methods realized using MS-CAN (b) feature extraction and classification, both of which are explained in the section below.

2.1. Aggregation of multiple denoising methods realized using MS-CAN

Figure 3 shows the block diagram for realizing the aggregation of multiple denoising methods through deep MS-CAN. Given the RGB color image I, the first step is to perform the denoising operation. Among several types of image denoising methods, we choose four complementary methods by considering their performance and also the mode of operation (spatial/frequency/sparse). To this extent, we have used the selected denoising methods that namely Wavelet Denoising (WD) [26], Block Matching and 3D filtering (BM3D) [27], Multiresolution Bilateral Filtering (MBF) [28] and DeNoising Convolutional Neural Networks (DnCNN) [29]. Let I_{D1}, I_{D2}, I_{D3} & I_{D4} represent the denoised images corresponding to WD, BM3D, MBF and DnCNN respectively. In the next step, we perform the aggregation to obtain a single denoised image that can represent the best of all four denoising techniques. The aggregation of best-denoised parts within the image is carried out through the wavelet-based image fusion technique, where each denoised image is decomposed into sub-bands. As the pixel values in the sub-bands from different denoising approaches are multiple. We employ the criteria for selecting the best sub-band with the highest energy values for reconstructing the final denoised image (using the inverse wavelet

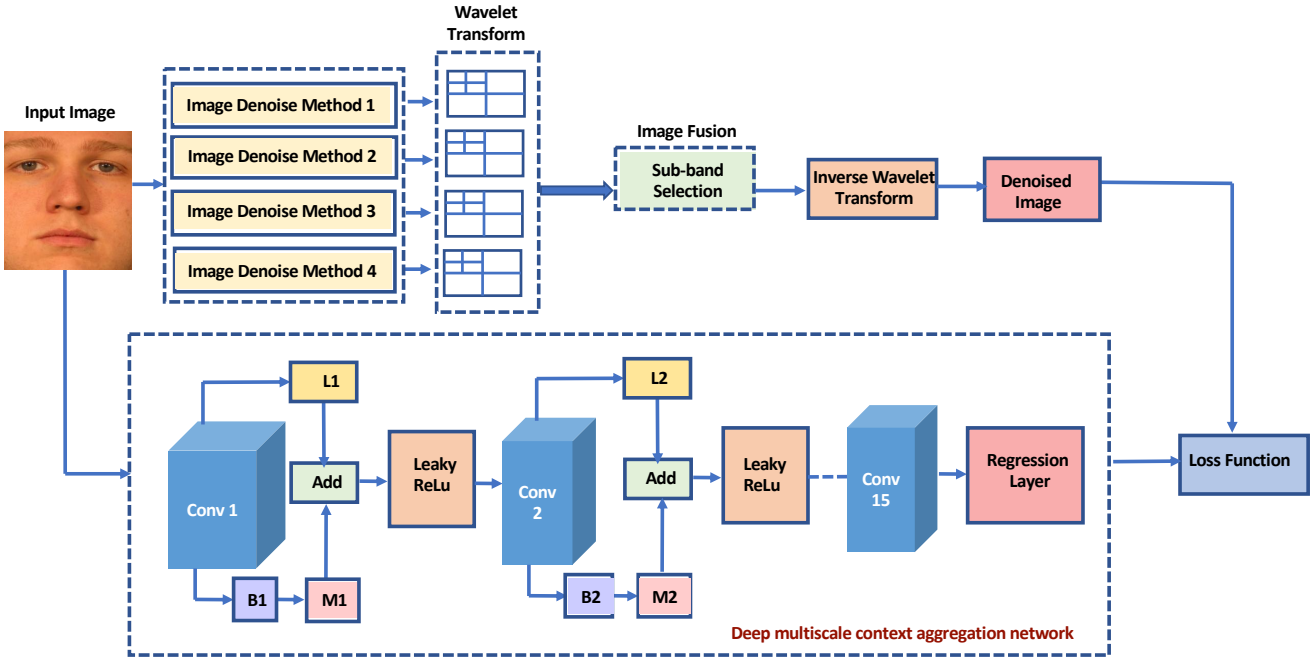

Figure 3: Realizing the multiple-denoising approach using a deep Multi-scale Context Aggregation Network (MS-CAN). B denotes batch-normalization, M represents the scale layer that adjusts the strength of the batch-normalization, L corresponds to strength of the identity branch in batch-normalization.

transform). We are motivated to use wavelet-based image fusion as it can handle multi-resolution images. Further, the image fusion strategy based on the selection of sub-bands with the highest energy allows us to retain the edge components preserved from multiple denoising methods.

Given the denoised image $I_{Di}, \forall i = \{1, \dots N\}$ where N represents the number of denoising methods. The corresponding wavelet decomposition (with level 2) of I_{Di} results in four different sub-band images such as approximate sub-band $\{a^N\}$, horizontal sub-band $\{h_1^N, h_2^N\}$, vertical sub-band $\{v_1^N, v_2^N\}$ and diagonal sub-band $\{d_1^N, d_2^N\}$. In the next step, we compute the energy corresponding to each sub-band that can be represented as $\{E_a^N\}$, $\{E_{h1}^N, E_{h2}^N\}$, $\{E_{v1}^N, E_{v2}^N\}$, $\{E_{d1}^N, E_{d2}^N\}$. The image fusion is performed by selecting the sub-band that corresponds to the highest energy as: $S_{h1} = h_1^{N(k)}$, where $k = \max_{i=1}^{N} \{E_{h1}^N\}$ is the index that corresponds to the highest energy. For example, if the highest energy for the horizontal sub-band h_1 is noted with the N^{th} image denoising method, then it is selected. We follow the same procedure for the remaining sub-bands to obtain $S_{h2}, S_{v1}, S_{v2}, S_{d1}, S_{d2}$ and S_a. Finally the fused denoised image I_F is obtained by taking the inverse wavelet transform.

Considering the computational effort and the parameterization of the aggregation of multiple denoising methods, we simply realize the operation of multiple denoising using a deep learning approach. It is shown in earlier works [30] [31] that, approximated image processing operations using deep MS-CAN can result in a highly accurate, robust and time-efficient technique. Inspired by such findings in [30, 31], we design our architecture in a similar fashion for our aggregated denoising approach. As indicated in Figure 3, the deep MS-CAN architecture consists of 15 layers of 3×3 convolution layers with exponentially increasing dilation factor. Thus, the dilation corresponding to the convolution layers are $1, 1, 2, 4, 8, 16, 32, 64, 128, 256, 512, 1024, 2048, 4096, 1$. Each convolution layer in the network is connected to a point-wise non-linearity using the leaky rectified linear unit (leaky-relu). Further, the adaptive normalization [31] is employed to combine both batch normalization and identity normalization. As shown in the architecture (see Figure 3), B_x (where $x = 1, 2, \dots, 15$, number of layers) represents the batch-normalization, M_x represents the scale layer that adjusts the strength of the batch-normalization, L_x represents the scale layer to adjust the strength of the identity branch. We then use the additional layer to combine both M_x and L_x. The network is trained on input-output pairs that contain images from before and after the proposed denoising operation. We further employ Mean Squared Error (MSE) within regression loss function to estimate the

learnability of the aggregation (approximation) operation.

$$L = \sum_{i} \frac{I_{Fi} - \hat{I_F}}{R} \tag{1}$$

where, R is the number of responses, I_{Fi} is the target output and $\hat{I_F}$ is the network prediction for response i.

Training details of MS-CAN

To effectively realize the generalizability of the proposed deep MS-CAN, we train the network on natural images (including photos of people, building, natural scenes, etc.) from IAPR TC-12 [1]. We further perform the proposed multiple denoising fusion approach on this dataset to obtain the denoised image. We then train the deep MS-CAN using pairs of normal-denoised image. The Adam optimizer is used with a constant learning rate of 0.0001 and the training is carried out for 250 epochs resulting in 1.2 million iterations. We subsequently use the trained deep MS-CAN to perform the denoising operation and compute the residual noise that can be used to detect a morphing attack as shown in the Figure 2.

Figure 4 illustrates qualitative results of the residual noise computed on bona fide and morphed face images using deep MS-CAN. The variation in noise intensity between bona fide and morphed image can be observed and this asserts our intuition. These qualitative results further support our approach of detecting morphing attacks based on residual noise despite learning from general image datasets.

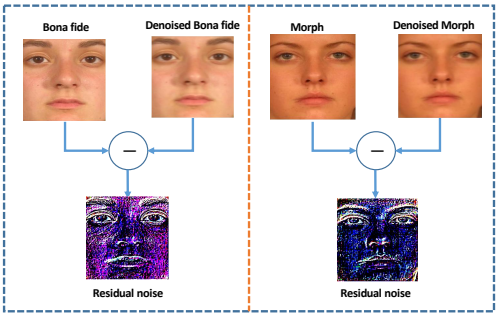

Figure 4: Illustration of residual noise computation using deep MS-CAN

2.2. Feature extraction and detection

Given the residual image, we extract the deep textural features computed using a pre-trained off-the-shelf AlexNet. We have used the features from fully connected layer $fc6$ to compute the feature from the residual noise images. These computed features are then classified using a Probabilistic Collaborative Representation Classifier (P-CRC) [32]. The P-CRC used in this work utilizes the Regularized Least Square Regression (LSR) on the learned feature vectors versus the probe feature vectors [32] formulated as:

$$\widehat{F} = argmin_{\alpha} \left\| Tr_F - \mathscr{D}\alpha \right\|_2^2 + \lambda \left\| \alpha \right\|_2^2 + \frac{\psi}{K} \left\| X\alpha - X_K\alpha_K \right\|_2^2 \tag{2}$$

Where, Tr_F is the feature vector of the test image, $\mathscr{D}$ is the learned collaborative subspace dictionary using Tr_F, α is coefficient vector, X is the collection of the training features corresponding to K classes and λ and ψ are the regularization parameter.

3. Face Morphing Datasets

The proposed approach is validated empirically using three different morphed face datasets employing different approaches for morphing and different composition representing the wide possible variation in morphing process as detailed below. The datasets are further used to benchmark the detection performance with State-Of-The-Art (SOTA) Morphing Attack Detection (MAD) methods.

3.1. Dataset-1

This database compromises 179 unique subjects that include both male and female participants from Asian and Caucasian ethnicity. This dataset is constructed using a public dataset (a subset of the FRGC face database) and a private face dataset. The whole database is divided into two partitions where the training set includes 89 disjoint and unique data subjects with multiple samples. The rest of the disjoint subjects are used in the testing set comprising 90 unique data subjects. Facial images are morphed using an open-source tool mentioned in [18]. Ultimately, the training set is composed of 709 bona fide and 1255 morphed images and the testing set is composed of 918 bona fide and 1354 morphed images. Figure 5 (a) shows example images from Dataset-1.

3.2. Dataset-2

This morphing database is a derivative of the publicly available FRGC database that compromises of 568 subjects. The entire database is divided into two partitions that include a training set of 300 unique data subjects resulting in 300 bona fide and 3041 morphed images. The testing set consists of 268 unique data subjects resulting in 268 bona fide and 2739 morphed images. Contrary to Dataset-1, the morphing process used for this dataset is based on the automatic facial landmark and triangulation, as mentioned in [21]. It has to be noted that the face morphing is performed only on the inner part of the face excluding the silhouette of the face (i.e, hair and ear region). Examples from Dataset-2 can be seen in Figure 5 (b).

[1] https://www.imageclef.org/photodata

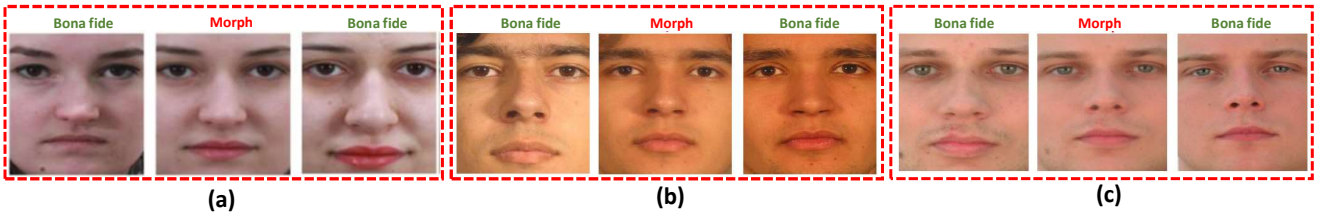

Figure 5: Example images from (a) Dataset-1 (b) Dataset-2 (c) (a) Dataset-3

Table 2: MAD performance on individual image denoising techniques and the proposed method

Algorithms	Dataset-1			Dataset-2			Dataset-3		
	D-EER(%)	BPCER@ APCER		D-EER(%)	BPCER@ APCER		D-EER(%)	BPCER@ APCER	
		=5%	=10%		=5%	=10%		=5%	=10%
BM3D [27]	15.03	40.50	22.50	25.04	55.59	42.16	14.37	32	26
WD [26]	27.96	42.83	31.50	31.35	81.34	67.53	18.01	46	34
MBF [28]	8.69	12.16	8.16	8.69	10.44	8.20	9.71	10	8
DnCNN [29]	19.82	42.83	31.50	24.96	54.85	44.77	19.83	54	38
Proposed Method	**3.24**	**3**	**1.67**	**2.63**	**1.11**	**1.11**	**7.89**	**8**	**4**

3.3. Dataset-3

This database is a derivative of the publicly available PutDB database [35] that compromises 100 subjects. The entire database is divided into two different partitions consisting of 50 training and 50 testing unique data subjects. Morphing is performed based on the automatic facial landmark and triangulation as described in [21], that results in 50 bona fide and 254 morphed samples in the training set and 50 bona fide and 244 morph samples in the testing set. Similar to Dataset-2, only the inner part of the face is morphed. Figure 5 (c) shows example images from Dataset-3.

All three datasets are developed by following the morph data preparation steps, as discussed in [18, 36]. Since all three datasets are constructed using source face images from three different face datasets, this provides an opportunity to evaluate the generalizability of the proposed method together with the SOTA methods.

Table 3: Quantitative performance of the MAD algorithms on Experiment-1 (individual dataset)

Algorithms	Dataset-1			Dataset-2			Dataset-3		
	D-EER(%)	BPCER@ APCER		D-EER(%)	BPCER@ APCER		D-EER(%)	BPCER@ APCER	
		=5%	=10%		=5%	=10%		=5%	=10%
AlexNet-SVM [11, 33, 34]	5.50	3.5	2.33	7.08	8.95	4.85	11	22	12
GoogleNet-SVM [33]	9.63	13.66	8.83	11.95	22.38	14.55	42.23	100	77.23
InceptionV3-SVM [33]	11.66	18.83	12.33	8.21	11.94	8.20	11.94	26	16
ResNet-SVM [11, 33, 34]	5.51	6.16	4	6.48	6.10	4.74	13.76	32	22
VGG16-SVM [11, 33, 9]	13.31	25	16.83	14.50	28.35	18.28	21.86	100	36
VGG19-SVM [11, 33, 34]	12.49	22.66	15	12.32	22.38	14.17	24.50	52	40
BSIF-SVM [1] [7]	26.70	53	42	12.67	25.74	14.55	20.45	44	32
Steerable pyramid-SVM [33]	26.19	65.50	50	37.97	82.08	71.64	34.00	82	70
HOG-SVM [7]	10.37	19.83	10.50	12.30	23.50	14.92	11.91	26	10
Image Gradient-SVM [1]	17.34	38	26.50	25.24	51.86	39.92	31.98	72	60
LBP-SVM [1, 21, 11, 7]	18.67	39.16	28.16	9.31	14.55	8.20	22.06	62	38
PRNU [8]	26.51	43	55.66	39.89	96.26	92.91	35.62	94	94
LPQ-SVM [1]	17.30	43.66	28.66	13.43	26.11	16.41	20.24	56	38
Deep Residual Noise [12]	3.83	3	1.5	4.85	4.85	3.35	9.71	14	8
Proposed Method	**3.24**	**3**	**1.67**	**2.63**	**1.11**	**1.11**	**7.89**	**8**	**4**

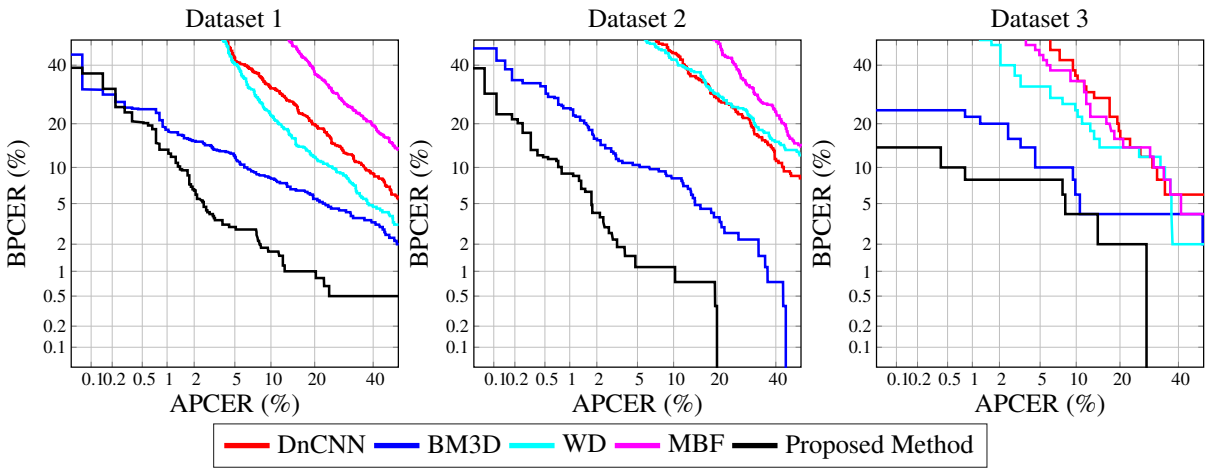

Figure 6: DET curves depicting MAD performance of the individual image denoising methods together with proposed method on different datasets

4. Experiments and Results

In this section, we present the quantitative results of the proposed method together with 14 different SOTA techniques for morphed face detection. Experimental results are presented using the ISO-IEC 30107-3 [37] metrics such as Bona fide Presentation Classification Error Rate (BPCER(%)) and Attack Presentation Classification Error Rate (APCER (%)) along with Detection-Equal Error Rate (D-EER(%)). BPCER defines the proportion of bona fide presentations incorrectly classified as morphing attack images and APCER defines attack images incorrectly classified as bona fide images [37].

In this work, we have evaluated six deep learning-based SOTA, seven non-deep learning based techniques and one hybrid method that use both deep and hand-crafted features. In case of the deep learning techniques, we have used the pre-trained network and computed the corresponding features that are further classified using a linear Support Vector Machines (SVM). To this extent, we have considered pre-trained CNN such as AlexNet [11, 33, 34], GoogleNet [33], Inception V3 [33], ResNet101 [11, 33, 34], VGG16 [11, 33, 9] and VGG19 [11, 33, 9]. The deep-learning techniques are used only as the feature extraction techniques owing to the availability of the small datasets. In case of non-deep learning techniques, texture-based techniques such as LBP [1], LPQ [1], BSIF [7], Steerable Pyramids [33] together with image distortion based features such as Image gradients [1], hybrid method [10], HoG [7] and PRNU [8] with linear SVM (except for PRNU) to compute the detection performance. To effectively evaluate the performance of the Morphing Attack Detection (MAD) schemes, we perform three different experiments such as **Experiment-1:-** designed to evaluate the performance of the MAD schemes when training and testing is carried out on the same dataset. **Experiment-2:-** designed to evaluate the MAD schemes on the merged dataset in which all three datasets are merged to one single dataset. This experiment provides an insight into the MAD performance when the dataset is increased with respect to the number of bona fide and morphed samples. **Experiment-3:-** designed to perform the cross-dataset comparison in which one of the datasets is used for training and another dataset is used for testing. This experiment will provide insights on MAD techniques that are capable to operate on unknown data.

Table 4: Quantitative performance of the MAD algorithms on Experiment-2 (merged dataset)

Algorithms	**D-EER(%)**	**BPCER@ APCER**	
		=5%	**=10%**
AlexNet-SVM [11, 33, 34]	9.70	17.32	9.36
GoogleNet-SVM [33]	10.87	21.35	11.98
InceptionV3-SVM [33]	8.69	14.59	7.51
ResNet-SVM [11, 33, 34]	7.77	9.04	4.68
VGG16-SVM [11, 33, 9]	12.83	25.49	15.03
VGG19-SVM [11, 33, 34]	12.19	24.50	15.03
BSIF-SVM [1, 7]	15.58	33.98	23.09
Steerable Pyramid-SVM [33]	36.78	77.88	68.08
HOG-SVM [7]	11.32	20.69	12.52
Image Gradient-SVM [1]	38.41	79.84	68.84
LBP-SVM [1, 21, 11, 7]	36.58	73.42	63.98
PRNU [8]	36.88	96.84	94.11
LPQ-SVM [1]	15.03	30.28	19.82
Deep Residual Noise [12]	5.35	6.31	2.50
Proposed Method	**4.96**	**5.01**	**3.05**

Table 2 indicates the performance of the proposed method and individual image denoising methods used to

build the proposed method. Figure 6 shows the DET curves for all three different morphed face datasets. To have a fair comparison, we have used the same feature extraction and comparison schemes on individual denoising schemes. Based on the results, it can be noted that the proposed method has indicated the best detection performance on all three datasets demonstrating a good robustness. The superior performance of the proposed method can be attributed to (a) the aggregated denoising and fusion scheme based on the best sub-band selection (b) the robustness of MS-CAN in obtaining the noise of images trained using natural images.

Table 3 indicates the performance of the proposed method together with 14 different state-of-the-art methods for Experiment-1 on all three datasets. Based on the obtained results, it can be noted that (a) the use of deep-features indicate a better performance on all three datasets when compared to non-deep feature based techniques. (2) Among the deep features, the AlexNet and ResNet101 have indicated an improved performance over other deep features. (3) Among the non-deep features, HoG-SVM has indicated the best performance. (4) The proposed method shows overall the best performance when compared to 14 different SOTA techniques on all three different datasets.

Table 4 presents the quantitative results of the proposed and existing methods for Experiment-2. Based on the obtained results, the deep features indicate better performance over non-deep techniques. Further, the proposed method has indicated the best performance with D-EER = 4.96% with BPCER = 5.01% @APCER = 5% and BPCER = 3.05% @APCER = 10%. These obtained results further justify the robustness of the proposed method to the increased number of samples with different image characteristics.

Table 5 indicates the quantitative performance of the proposed method for Experiment-3 (cross-dataset evaluation). For simplicity, we have presented the results only for the top four best performing MAD techniques based on Experiment-1 and Experiment-2. Since we have three different datasets, we get six different cases in which one dataset is enrolled and the remaining two datasets are probed. Based on the obtained results, the proposed method shows superior performance when compared with the SOTA methods in all six cases.

Thus, based on the extensive experiments carried out on three different datasets with three different performance evaluation experiments, we can conclude a superior performance over 14 different state-of-the-art methods. The evaluation results demonstrate the robustness of the proposed method, which is attributed to the proposed deep MS-CAN architecture. Further, realizing the proposed method using MS-CAN not only improved the robustness but also significantly improved computational cost by a factor of 4, as four denoising operations learnt as a single coherent operation.

Table 5: Quantitative performance of the MAD algorithms on Experiment-3 (cross Dataset) - D1-Dataset 1, D2-Dataset 2, D3- Dataset 3

Train	**Test**	**Algorithms**	**D-EER(%)**	**BPCER @ APCER**	
				=5%	**=10%**
D1	D2	AlexNet-SVM [11, 33, 34]	50	100	100
		Deep Residual Noise [12]	**7.12**	**12.31**	**5.22**
		HoG-SVM [7]	17.97	38.43	28.35
		Proposed method	10.44	16.04	10.44
D1	D3	AlexNet-SVM [11, 33, 34]	19.63	32	24
		Deep Residual Noise [12]	13.76	32	16
		HoG-SVM [7]	20.24	50	30
		Proposed method	**11.94**	**28**	**14**
D2	D1	AlexNet-SVM [11, 33, 34]	8.14	11.66	7.33
		Deep Residual Noise [12]	6.49	8.50	4.16
		HoG-SVM [7]	6.81	9	4.83
		Proposed method	**4.66**	**4.66**	**2.88**
D2	D3	AlexNet-SVM [11, 33, 34]	19.83	38	34
		Deep Residual Noise [12]	13.76	30	22
		HoG-SVM [7]	12.35	34	20
		Proposed method	**11.94**	**18**	**14**
D3	D1	AlexNet-SVM [11, 33, 34]	50	100	100
		Deep Residual Noise [12]	14.40	36.16	19.50
		HoG-SVM [7]	14.52	32	19.16
		Proposed method	**8.62**	**10.83**	**7.67**
D3	D2	AlexNet-SVM [11, 33, 34]	50	100	100
		Deep Residual Noise [12]	15.31	33.95	23.50
		HoG-SVM [7]	24.28	58.20	42.53
		Proposed method	**10.03**	**17.16**	**10.07**

5. Conclusion

We have presented a novel method to detect face morphing attacks in a reliable manner. The proposed method is based on quantifying the residual noise resulting from the effect of the morphing process. The morphing noise is quantified using a aggregation of multiple denoising methods approximated using a deep Multi-Scale Context Aggregation Network (MS-CAN). We then process the residual noise from deep MS-CAN to extract deep features computed using a pre-trained AlexNet. The final decision is computed using the Probabilistic Collaborative Representation Classifier (P-CRC) learnt using the extracted features. Extensive experiments are carried out using three different morphed face datasets with three different performance evaluation protocols. The performance of the proposed method is benchmarkd with the 14 different existing methods. The results have shown that the proposed method significantly outperforms existing methods on all three datasets for three different performance evaluation protocols.

References

[1] R. Raghavendra, K. B. Raja, and C. Busch, "Detecting Morphed Face Images," in *8th IEEE International Conference on Biometrics: Theory, Applications, and Systems (BTAS)*, pp. 1–8, 2016.

[2] A. Makrushin, T. Neubert, and J. Dittmann, "Automatic generation and detection of visually faultless facial morphs," in *Proceedings of the 12th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 6: VISAPP, (VISIGRAPP 2017)*, pp. 39–50, 2017.

[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 *International Workshop on Biometrics and Forensics (IWBF 2017)*, pp. 1–6, 2017.

[4] T. Neubert, "Face morphing detection: An approach based on image degradation analysis," in *International Workshop on Digital Watermarking*, pp. 93–106, 2017.

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

[6] A. Asaad and S. Jassim, "Topological data analysis for image tampering detection," in *International Workshop on Digital Watermarking*, pp. 136–146, 2017.

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

[8] L. Debiasi, C. Rathgeb, U. Scherhag, A. Uhl, and C. Busch, "Prnu variance analysis for morphed face image detection," in *2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, pp. 1–9, Oct 2018.

[9] R. Raghavendra, K. B. Raja, S. Venkatesh, and C. Busch, "Transferable deep-cnn features for detecting digital and print-scanned morphed face images," in *Proc. IEEE Conf. Computer Vision Pattern Recognition Workshops (CVPRW)*, pp. 1822–1830, 2017.

[10] D. Naser, Z. Steffen, W. Yaza, M. S. Alexandra, K. Florian, and A. Kuijper, "A multi-detector solution towards an accurate and generalized detection of face morphing attacks," in *Proceedings of the IEEE International Conference on Information Fusion*, pp. 1–8, 2019.

[11] M. Ferrara, A. Franco, and D. Maltoni, "Face morphing detection in the presence of printing/scanning and heterogeneous image sources," *CoRR*, vol. abs/1901.08811, 2019.

[12] S. Venkatesh, R. Raghavendra, K. Raja, L. Spreeuwers, R. Veldhuis, and C. Busch, "Morphed face detection based on deep color residual noise," in *ninth International Conference on Image Processing Theory, Tools and Applications (IPTA 2019)*, pp. 1–5, IEEE, 2019.

[13] M. Ferrara, A. Franco, and D. Maltoni, "The magic passport," in *IEEE International Joint Conference on Biometrics*, pp. 1–7, sep 2014.

[14] J. T. Andrews, T. Tanay, and L. D. Griffin, "Multiple-identity image attacks against face-based identity verification," *arXiv preprint arXiv:1906.08507*, 2019.

[15] M. Ferrara, A. Franco, and D. Maltoni, *Face Recognition Across the Imaging Spectrum*, ch. On the Effects of Image Alterations on Face Recognition Accuracy, pp. 195–222. Springer International Publishing, 2016.

[16] D. Robertson, R. S. Kramer, and A. M. Burton, "Fraudulent id using face morphs: Experiments on human and automatic recognition," *PLoS ONE*, vol. 12, no. 3, pp. 1–12, 2017.

[17] R. S. Kramer, M. O. Mireku, T. R. Flack, and K. L. Ritchie, "Face morphing attacks: Investigating detection with humans and computers," *Cognitive research: principles and implications*, vol. 4, no. 1, p. 28, 2019.

[18] R. Raghavendra, K. B. Raja, S. Venkatesh, and C. Busch, "Face morphing versus face averaging: Vulnerability and detection," in *IEEE International Joint Conference on Biometrics (IJCB)*, pp. 555–563, 2017.

[19] "Facial landmark based face morphing." https://www.learnopencv.com/face-morph-using-opencv-cpp-python/.

[20] N. Damer, A. M. Saladié, A. Braun, and A. Kuijper, "Morgan: Recognition vulnerability and attack detectability of face morphing attacks created by generative adversarial network," in *2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, pp. 1–10, Oct 2018.

[21] L. Spreeuwers, M. Schils, and R. Veldhuis, "Towards robust evaluation of face morphing detection," in *2018 26th European Signal Processing Conference (EUSIPCO)*, pp. 1027–1031, Sep. 2018.

[22] U. Scherhag, L. Debiasi, C. Rathgeb, C. Busch, and A. Uhl, "Detection of face morphing attacks based on prnu analysis," *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 2019.

[23] U. Scherhag, C. Rathgeb, J. Merkle, R. Breithaupt, and C. Busch, "Face recognition systems under morphing attacks: A survey," *IEEE Access*, vol. 7, pp. 23012–23026, 2019.

[24] R. Raghavendra, S. Venkatesh, K. Raja, and C. Busch, "Towards making morphing attack detection robust using hybrid scale-space colour texture features," in *IEEE International Conference on Identity, Security and Behaviour Analysis (ISBA 2019)*, pp. 1–7, 2019.

[25] F. Peng, L. Zhang, and M. Long, "Fd-gan: Face de-morphing generative adversarial network for restoring accomplice's facial image," *IEEE Access*, vol. 7, pp. 75122–75131, 2019.

[26] D. L. Donoho, "De-noising by soft-thresholding," *IEEE Transactions on Information Theory*, vol. 41, pp. 613–627, May 1995.

[27] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-d transform-domain collaborative filtering," *IEEE Transactions on Image Processing*, vol. 16, pp. 2080–2095, Aug 2007.

[28] M. Zhang and B. K. Gunturk, "Multiresolution bilateral filtering for image denoising," *IEEE Transactions on Image Processing*, vol. 17, pp. 2324–2333, Dec 2008.

[29] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a gaussian denoiser: Residual learning of deep CNN for image denoising," *CoRR*, vol. abs/1608.03981, 2016.

[30] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," *arXiv preprint arXiv:1511.07122*, 2015.

[31] Q. Chen, J. Xu, and V. Koltun, "Fast image processing with fully-convolutional networks," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2497–2506, 2017.

[32] L. Zhang, M. Yang, and X. Feng, "Sparse representation or collaborative representation: Which helps face recognition?," in *IEEE International Conference on Computer Vision (ICCV)*, pp. 471–478, 2011.

[33] R. Raghavendra, S. Venkatesh, K. Raja, and C. Busch, "Detecting face morphing attacks with collaborative representation of steerable features," in *IAPR International Conference on Computer Vision & Image Processing (CVIP-2018)*, pp. 1–7, 2018.

[34] C. Seibold, W. Samek, A. Hilsmann, and P. Eisert, "Detection of face morphing attacks by deep learning," in *Digital Forensics and Watermarking* (C. Kraetzer, Y.-Q. Shi, J. Dittmann, and H. J. Kim, eds.), pp. 107–120, Springer International Publishing, 2017.

[35] A. Kasiński, A. Florek, and A. Schmidt, "The put face database," *Image Processing & Communications*, vol. 13(3-4), pp. 59–64, 2008.

[36] U. Scherhag, A. Nautsch, C. Rathgeb, M. Gomez-Barrero, R. N. J. 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 *2017 International Conference of the Biometrics Special Interest Group (BIOSIG)*, pp. 1–7, Sep. 2017.

[37] 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.