Low Visual Distortion and Robust Morphing Attacks Based on Partial Face Image Manipulation

Le Qin, Fei Peng, *Member, IEEE*, Sushma Venkatesh, Raghavendra Ramachandra, *Senior Member, IEEE*, Min Long, and Christoph Busch

***Abstract*—Face verification is a popular way for verifying identities in access control systems. In this work, a partial face manipulation-based morphing attack (MA) is proposed to compromise the uniqueness of face templates. Different from existing research, this work changes MA from a holistic face level to component level, and only the most effective facial components (eyes and nose) are used. Therefore, a manipulated face is more similar to a bona fide one in terms of visual quality, texture, and noise characteristics. To validate the effectiveness of the proposed attack, a novel metric called actual mated morph presentation match rate (AMPMR) is proposed to evaluate MA performance under real-world conditions. With a collected dataset containing different attack types, image qualities, and manipulation parameters, the results indicate the proposed attack has better anti-detectability compared with the existing complete, splicing, and combined MAs. Moreover, it has low visual distortion and can reach a better tradeoff among facial biometrics verification, anti-detectability, and visual differences.**

***Index Terms*—Access control, face authentication, facial manipulation, face morphing attack, morphing attack detection.**

I. Introduction

IN MODERN unsupervised access control systems of campuses, corporations, bus and train stations, face authentication has been widely deployed for identity verification. During authentication, an access control system reads a facial biometric reference (e.g., a template) stored in a user's enrolment record in a database or in his/her personal access card, and then compares it with a trusted live facial image from the same subject. When a similarity score resulting from this comparison process is higher than the system's threshold, a gate automatically opens to permit passing without human check, and it can significantly reduce users' transaction time in the access control process. Meanwhile, to simplify the enrolment process, many systems allow users to submit their own photos when applying for access cards. If those photos comply to the predefined image standards [1], on-site data collections are not required. The unsupervised capture process of the enrolment image is convenient for ordinary users, but they also provide potential attack conditions for adversaries, such as morphing attacks (MAs). The goal of face MAs is to compromise the uniqueness of facial biometric templates with non-intrusive ways, to create a template (e.g., a manipulated face) that can match with multiple subjects. Once a manipulated facial image is maliciously injected into an enrolment record as biometric reference, multiple subjects can share and use the access card, which has negative impacts on public security.

In [2], it was pointed out that a wanted criminal (or a malicious actor) may find an accomplice to generate a morphed face, which can be verified to each of them. After that, some automatic face morphing algorithms were proposed, such as complete morphing, splicing morphing, and combined morphing [3], [4]. Complete morphing can effectively integrate facial biometrics from all morphing contributors, but its blending artefacts are obvious, which can potentially increase the risk to be spotted by a trained human expert. To reduce such artefacts, splicing morphing reversely warps a morphed face into an original facial image context preserving hair and background from one of the contributing images. However, its geometric shape is constructed only based on one morphing contributor, which leads to its limited similarities to other contributors. In combined morphing, face alignment preprocessing is introduced, and a morphed face area is directly combined with an aligned original face image. Although it can combine facial biometrics from all morphing contributors and reduce blending artefacts, the whole face region is tampered. Recently, face morphing methods based on generative adversarial network (GAN) [5], [6] and auto-encoder [7] were put forward, but the images generated by them are different from standard enrolment quality photos in terms of size and face area proportion. To defend face MAs, morphing attack detection (MAD) has attracted a lot of attention, and many methods were proposed in the past few years, such as single image morphing attack detection (S-MAD) and differential morphing attack detection (D-MAD) [8], [9]. S-MAD methods are mainly based on texture, noise, and quality cues, but these characteristics can be simulated by an attacker to evade

Manuscript received June 4, 2020; revised August 31, 2020; accepted September 2, 2020. Date of publication September 10, 2020; date of current version February 12, 2021. This work was supported in part by the National Natural Science Foundation of China under Grant U1936115 and Grant 62072055, and in part by the China Scholarship Council. This article was recommended for publication by Associate Editor D. Mery upon evaluation of the reviewers' comments. *(Corresponding author: Fei Peng.)*

Le Qin and Fei Peng are with the College of Computer Science and Electronic Engineering, Hunan University, Changsha 410082, China (e-mail: qinle@hnu.edu.cn; eepengf@gmail.com).

Sushma Venkatesh, Raghavendra Ramachandra, and Christoph Busch are with Norwegian Biometrics Laboratory, Norwegian University of Science and Technology, 2815 Gjøvik, Norway (e-mail: sushma.venkatesh@ntnu.no; raghavendra.ramachandra@ntnu.no; christoph.busch@ntnu.no).

Min Long is with the College of Computer and Communication Engineering, Changsha University of Science and Technology, Changsha 410114, China (e-mail: caslongm@aliyun.com).

This article has supplementary downloadable material available at https://ieeexplore.ieee.org, provided by the authors.

Digital Object Identifier 10.1109/TBIOM.2020.3022007

detection. Differential MAD methods take advantage of trusted live captured facial images to improve detection accuracy. However, their performance is limited when live facial images have large intra-class variations.

For MAs against access control systems, the roles between an attacker and an accomplice are asymmetric. An accomplice is confronted with strict manual inspections during enrolment, while an attacker aims to bypass an unsupervised or partially supervised system during authentication. To launch a successful MA, a manipulated face image used for reference enrolment should match with both the attacker and the accomplice. Meanwhile, to evade MAD, it also should present characteristics of a bona fide face image as much as possible. Moreover, to deceive manual inspections during the enrolment processes, the appearance differences between a manipulated face and the accomplice's face should be minimized from the aspect of human perception [10], [11]. In evading MAD, the existing research suggested that there are relationships between image tampering extent and detection accuracy. For instance, in image steganography, the smaller the differences between a stego image and a cover image are, the more inconspicuous the stego image is [12]. Motivated by these findings, a partial face manipulation-based MA is proposed in this article. Different from the existing research, it changes MA from a holistic face level to face component level, and only the most effective facial components (eyes and nose) are used for face manipulation. It can narrow the gaps between a manipulated face and a bona fide one by reducing manipulation areas. The main contributions of this work are as follows.

(1) A novel MA based on partial face manipulation is proposed. By changing MAs from face level to component level (eyes and nose), the proposed MA has low visual distortions and can better evade MAD.
(2) A partial face manipulation-based MA dataset is built. It contains 17072 images from 94 subjects with variations in image qualities as well as manipulation types and parameters. Compared with the existing datasets, the proposed components-based MA is introduced in the collected dataset.
(3) A novel metric called actual mated morph presentation match rate (AMPMR) is proposed. That can be effectively used to evaluate and report the practical success probability of MAs.
(4) A novel finding is demonstrated, namely, eyes and nose are more effective for MAs than other facial components.

With respect to ethical concerns, our purpose is not to guide an attacker, but to analyze the potential threats of face authentication systems and to advocate researchers to further improve the security of such systems.

The rest of this article is organized as follows. Related work and background are introduced in Section II. The proposed attack is illustrated in Section III. The collected dataset is described in Section IV. Experimental results and analysis are presented in Section V. Finally, our conclusions are drawn in Section VI.

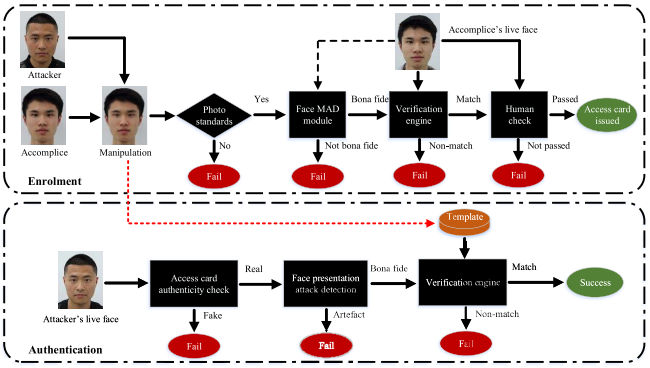

Fig. 1. Threat model.

II. Related Work and Background

A. Face Morphing Attack

It is assumed that an attacker aims to bypass an automated access control system using a valid access card, which is obtained by the help of an accomplice. The target system is based on face authentication, and the attacker should not have valid access to it. During the access card application and the biometric enrolment process, an applicant is allowed to submit a self-captured photo compliant to the predefined photo standards, but it needs to be examined by a MAD module. Meanwhile, both face authentication and human check are carried out to judge whether the photo belongs to the applicant. If a photo is predicted as a bona fide image by a MAD module and it is also judged as an image from the same applicant via a verification engine and human check, the submitted photo is stored in an access card's chip or in a machine readable travel document (MRTD) as a facial reference of the applicant. During authentication process, a trusted live face image of a subject is captured, and is compared with the facial reference image stored in the user's access or MRTD card. If subsequent to an access card authenticity check and a face presentation attack detection check, the result of the verification engine, i.e., the similarity score is higher than the system's threshold, the data subject is permitted to pass without manual inspections. The threat model is shown in Fig. 1.

To achieve the attack goal, an attacker is assisted by an accomplice who has valid access to a target access control system. The facial attributes of the attacker and the accomplice are similar to each other in terms of ethnicity, gender, and age, but their facial biometric characteristics are different such that a face verification engine would decide for a non-match. First, the face images of the attacker and the accomplice are collected, and they are used for generating a manipulated face image which needs to be compliant to the photo standards. Then, the accomplice applies for an access card (or MRTD) by submitting the manipulated face and his/her personal information. If the access card is issued, the attacker tries to pass the target system with it. The identity verification of the target system is assumed as a face verification operation, which only compares a trusted live image of a data subject with the facial reference image stored in the subject's access

TABLE I
OVERVIEW OF FACE MORPHING ATTACKS

Methods	Domains	Morphing approaches	Automatic	Visual distortions
Software [2]	spatial	manual morphing	no	imperceptible
Complete [3]	spatial	warping and blending	yes	blending artefacts
Splicing [3]	spatial	inverse warping	yes	same geometry
Combined [4]	spatial	combining	yes	imperceptible
MorGAN [5] / EMorGAN [6]	feature	model training	yes	lack of fine details / imperceptible
Style transfer [16]	spatial	morphing + model training	yes	imperceptible
StyleGAN [14]	feature	latent code averaging	yes	imperceptible

card or MRTD. Furthermore, the enrolment and authentication processes are assumed to be carried out in a normal manner. In other words, access card falsification, presentation attacks, and inside attacks to access control systems are out of scope in this work.

B. Generation of Morphed Face Images

According to different morphing processes, the existing face MAs can be classified into spatial domain morphing and feature domain morphing. In spatial domain morphing, Ferrara *et al.* proposed to manually morph faces by the use of an image manipulation program [2]. It can create a morphed face image with good visual quality and pose a threat to commercial systems, but it is time-consuming and retouching knowledge is also required. To reduce manual effort, Makrushin *et al.* proposed complete and splicing morphing algorithms to automatically morph face images [3]. Nevertheless, the shadow artefacts of complete morphed face images are easily observed by trained human eyes, and splicing morphed face images only well match one morphing contributor. After that, combined face morphing was proposed in [4]. Although it can achieve a better tradeoff between biometrics recognition and visual quality, the whole face area is tampered. Raghavendra *et al.* proposed to launch MAs with averaged faces, but they can be easily detected by MAD methods through analyzing a suspected face image [13]. An averaged face image is generated by simple pixel level averaging, while a complete morphed face image is generated by warping and triangle blending.

From the perspective of feature domain morphing, MorGAN (64×64 pixels) [5] and EMorGAN (128×128 pixels) [6] were proposed to generate morphed face images using GAN. However, the size and face area proportion of the images generated by GAN are different from those of standard enrolment quality photos and would therefore not pass a simple quality control subsystem, and its effectiveness to commercial systems also needs further investigation. Recently, an investigation of GAN-based morphs was made in [14], and it indicates that detecting GAN-based morphs in the digital domain is relatively easy compared to detecting landmark-based morphs. In [15], both spatial domain morphing and feature domain morphing are taken into account, and the principles of MAs are explained by angular distance distributions of face representations. An overview of existing face MAs is listed in Table I.

In the existing research, facial components replacement and morphing were utilized in face de-identification [17], image steganography [18], and digital attacks of facial recognition [19]. The main differences between this work and [17], [19] are as follows.

1) The research field of this work is different from [17], [19]. The goal of [17], [19] is to evade face recognition, while the goal of this work is to compromise the uniqueness of facial references.
2) The attacks are launched in different ways. For [19], the focus is on digital attacks that are intrusively launched during authentication, while the proposed attack is non-intrusively launched during enrolment.
3) The roles of manipulated faces are different. A tampered face is used as a probe face in [19], while a manipulated face is used as a biometric reference in this work.
4) The implementation details are different. In [19], only eyes, nose, and mouth are considered, and they are extracted by a rectangular box. While in this work, 9 different facial components and regions are considered, and they are extracted by a polygon box. Moreover, the postprocessing of this work is implemented by Poisson image editing rather than Gaussian filtering.
5) Evaluation metrics are different. The classical metrics of face recognition are used in [19], and only the vulnerability of deep learning models is evaluated. In contrast, a novel metric is proposed in this work, and both the vulnerability of deep learning models and commercial systems is evaluated.

C. Morphing Attack Detection

Based on different detection conditions, the existing MAD methods can be classified into single image morphing attack detection (S-MAD) and differential (or reference image-based) morphing attack detection (D-MAD). S-MAD is carried out with only a suspected image itself, while differential based methods also use other detection cues, such as a trusted live face image captured by systems (example: kiosk or ABC gates). According to different detection mechanisms, D-MAD methods can be further classified into landmark-based methods and de-morphing. The detection cues of landmark-based methods are facial landmark shifts between a morphed face image and a live face image [20]. For face de-morphing, with a trusted live face image, a morphed face can be reversed by classical method [21] or GAN [22], and face verification

TABLE II
OVERVIEW OF FACE MORPHING ATTACK DETECTION METHODS

Ref.	Scenarios	Detection cues	Approaches	Generation of morphs
[24]	non-reference	texture	BSIF descriptor	GIMP program
[27]	non-reference	quality	JPEG compression analysis	complete and splicing morphing
[29]	non-reference	deep learning	transferable CNN features	GIMP program
[30]	non-reference	deep learning	end to end learning	warping and blending
[21]	differential	live faces	de-morphing	progressive morphing database and Sqirlz Morph
[32]	non-reference	specular highlights	reflection analysis	warping and blending
[20]	differential	landmarks	landmarks shifting patterns	warping and blending
[28]	non-reference	hybrid	texture and frequency statistics	complete, splicing, and combined morphing
[26]	non-reference	residual-noise	photo response non-uniformity	OpenCV, FaceMorpher, FaceFusion, and UBO tool
[23]	differential	live faces	deep face representations	FaceMorpher, FaceFusion, OpenCV, and UBO tool

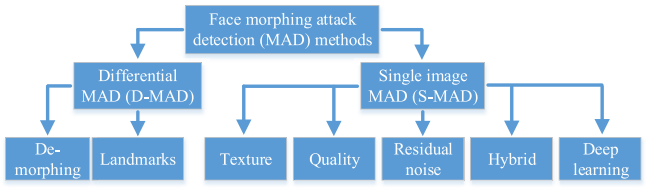

Fig. 2. A taxonomy of face morphing attack detection (MAD) methods.

is performed to detect MAs. Another D-MAD method was proposed in [23], where face representations of a suspected image and a trusted live face image are extracted as feature vectors. According to different detection cues, S-MAD methods can be further classified into texture-based methods, residual-noise based methods, quality-based methods, hybrid features based methods, and deep learning-based methods. For texture-based methods, texture descriptors are used for characterizing the textural differences between a morphed face image and a bona fide one, such as binarized statistical image features (BSIF) [24]. Residual-noise based methods focus on analyzing the noise spectrum discrepancies of different face images, such as Fourier spectrum of sensor pattern noise (FSSPN) [25] as well as spatial and spectral features of photo response non-uniformity [26]. The detection cues of quality-based methods are the distortions caused by morphing manipulations, such as JPEG compression degradation analysis [27]. In hybrid features based methods, multiple detection cues are combined to counter face MAs, such as hybrid features of texture and frequency statistics [28]. Deep learning based methods adaptively learn the detection features to discriminate morphed faces from bona fide ones, such as transferable deep convolutional neural network (CNN) features [29], end to end deep learning [30], and multiclass pre-training with complex morphs [31]. Apart from the above methods, a MAD method based on inconsistencies of reflections was proposed in [32], and it is implemented by comparing synthesized specular highlights and detected specular highlights in suspected images. A taxonomy and an overview of MAD methods are shown in Fig. 2 and Table II, respectively.

On the basis of the above analysis, from the aspect of MAs, complete morphing method will likely cause blending artefacts, splicing morphing cannot match well with all attack contributors, and combined morphing tampers a whole face area. With respect to countermeasures of MAs, the existing MAD methods are mainly focused on texture, noise, and quality cues. Therefore, the proposed attack only uses the most effective facial components for launching a MA, and it can reduce visual differences while reaching a match decision. In addition, by reducing manipulation regions, the discrepancy between a manipulated face image and a bona fide one is minimized, which is helpful for improving its robustness against MAD.

III. MORPHING ATTACKS USING PARTIAL FACE IMAGE MANIPULATION

A. Motivation

For a given attacker's face image I_{Crim} and an accomplice's face image I_{Acco}, which are morphed with a morphing weight α, a resulting manipulated face image I_{Morp} can be obtained as [33]

$$I_{Morp} = \alpha I_{Crim} + (1-\alpha) I_{Acco}, \tag{1}$$

where $\alpha \in [0, 1]$. To launch a successful face MA, I_{Morp} should first evade a face MAD module F_{mad}, and we obtain

$$F_{mad}(I_{Morp}) > th_{mad}, \tag{2}$$

where th_{mad} is a face MAD threshold. If the predicted score from F_{mad} is higher than th_{mad}, the suspected face image is judged as bona fide. Otherwise, it is judged as a manipulated one. Meanwhile, during enrolment, I_{Morp} should also deceive human check F_{hc}, and we get

$$F_{hc}(I_{Morp}, I_{Ac\_l}) = 1, \tag{3}$$

where I_{Ac_l} is the accomplice's live face image captured on-site. If the output of F_{hc} is 1, it means the two input face images are from the same person. Otherwise, it means that they are from different subjects. In addition, to spoof a face authentication system F_{fa}, I_{Morp} should match with any probe facial image of the attacker and the accomplice, and then we have

$$F_{fa}(I_{Morp}, I_{Ac\_l}) > th_{fa}, \tag{4}$$

$$F_{fa}(I_{Morp}, I_{Cr\_l}) > th_{fa}, \tag{5}$$

where I_{Cr_l} is the attacker's probe face image captured by a trusted live capture process in an access control system, th_{fa} is a face authentication threshold. If the similarity score from F_{fa} is higher than th_{fa}, it means that the two input face images are from the same person. Otherwise, it means they belong to different subjects.

To meet the requirements of Eq. (2) to Eq. (5), a MA is launched at component level in this work, and the motivations are summarized as follows.

(1) From the aspect of face recognition, the existing research demonstrated that partial face tampering can increase false match rate (FMR) in face recognition [19]. Therefore, it is reasonable to assume that partial face manipulation can take advantage of the increased FMR to create a manipulated face that can match the attacker and the accomplice.

(2) To cheat manual inspections during enrolment, a manipulated face image should look similar to the accomplice from the perspective of human perception, and the visual differences between a manipulated face image and a bona fide one should also be minimized. In this aspect, the existing work indicated that facial components replacement can hardly be observed by human eyes [17]. Thus, compared with holistic face area-based manipulation, facial components-based manipulation is more difficult to be identified by manual inspections.

(3) For evading MAD, the existing work suggested that forensics accuracy is related to the tampering degree. For instance, in image steganography, a stego image is more inconspicuous when it is more similar to a cover image [12]. Since only partial face area is tampered, the characteristics of a components-based manipulated face image in terms of image quality, noise, and texture are more similar to those of a bona fide one. As a result, partial face manipulation is more robust to MAD compared with holistic face area manipulation.

B. The Proposed Scheme

For components-based manipulation, an intuitive implementation is that a facial component from a source image is directly morphed into a target face image. However, this approach does not fully consider the facial geometric factors. Thus, a morphed face image cannot well match with all attack contributors. To overcome this challenge, in this work, complete morphing [3] is first made with the face images of an attacker and an accomplice, and then a component is extracted from the complete morphed face image. After that, the extracted component is morphed into the corresponding area of the accomplice's face image to generate a components-based manipulated face image. The proposed scheme is illustrated in Fig. 3.

Here, 9 different facial components are used for investigating partial face image manipulation based MAs, and they are eyes, nose, mouth, forehead, combination of eyes and nose, upper face, midface, lower face, and central face, and they are denoted as $Comp = \{$eyes, nose, mouth, forehead, eyes & nose, up, mid, low, central$\}$. The areas of those components are determined by dlib landmarks [34]. For a given attacker's face image I_{Crim} and an accomplice's face image I_{Acco}, the steps

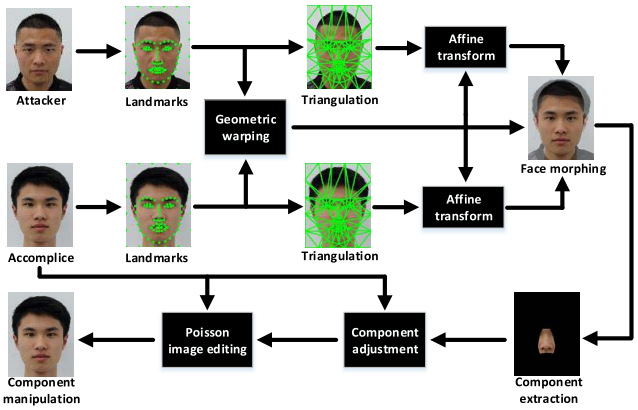

Fig. 3. The proposed partial face image manipulation scheme.

for generating a components-based manipulated face image I_{M_Comp} are described in the following.

(1) Facial landmarks of I_{Crim} and I_{Acco} are first detected, and they are combined with 20 auxiliary background points to respectively build landmark sets $P_{Crim} = \{(p_{cx1}, p_{cy1}), (p_{cx2}, p_{cy2}), \ldots, (p_{cxNp}, p_{cyNp})\}$ and $P_{Acco} = \{(p_{ax1}, p_{ay1}), (p_{ax2}, p_{ay2}), \ldots, (p_{axNp}, p_{ayNp})\}$, where N_p is the number of landmarks.

(2) Face warping is carried out based on P_{Crim} and P_{Acco}, and a warped landmark set $P_{Morp} = \{(p_{mx1}, p_{my1}), (p_{mx2}, p_{my2}), \ldots, (p_{mxNp}, p_{myNp})\}$ is obtained. Specifically, for given i–th landmarks (p_{cxi}, p_{cyi}) and (p_{axi}, p_{ayi}) from P_{Crim} and P_{Acco}, a corresponding warped landmark (p_{mxi}, p_{myi}) is calculated as

$$\begin{cases} p_{mxi} = \alpha_1 p_{cxi} + (1-\alpha_1) p_{axi} \\ p_{myi} = \alpha_1 p_{cyi} + (1-\alpha_1) p_{ayi} \end{cases}, \tag{6}$$

where $\alpha_1 \in [0, 1]$ is the warping weight.

(3) Delaunay triangulation is made with P_{Morp} to build a triangle vertex set $Tri = \{(t_{a1}, t_{b1}, t_{c1}), (t_{a2}, t_{b2}, t_{c2}), \ldots, (t_{aNt}, t_{bNt} t_{cNt})\}$, where N_t is the number of Delaunay triangles.

(4) For a given triangle in Tri, the corresponding triangle areas t_{Crim} and t_{Acco} are taken from I_{Crim} and I_{Acco}, respectively. Then, according to the spatial domain of P_{Morp}, affine transformation is done to t_{Crim} and t_{Acco}, and warped triangles tw_{Crim} and tw_{Acco} are obtained. After that, a morphed triangle t_{Morp} is constructed by blending tw_{Crim} and tw_{Acco} in the pixel level as

$$t_{Morp} = \alpha_2 tw_{Crim} + (1-\alpha_2) tw_{Acco}, \tag{7}$$

where $\alpha_2 \in [0, 1]$ is the blending weight. With the above operation, all the triangles from Tri are warped and blended in turn, and they are filled into the corresponding regions on the basis of P_{Morp} to generate a complete morphed face image I_{Morp}.

(5) Facial component landmarks of I_{Morp} are detected to build a component landmark set $K_{Comp} = \{(k_{cx1}, k_{cy1}), (k_{cx2}, k_{cy2}), \ldots, (k_{cxNk}, k_{cyNk})\}$, where N_k is the number of the component landmarks. Afterwards, boundary points from K_{Comp} are adjusted based on I_{Acco}. For given top and bottom boundary values k_{cyt} and k_{cyb} from

Fig. 4. A visualization of different facial components.

TABLE III
DLIB LANDMARK INDEXES (x, y) OF DIFFERENT FACIAL COMPONENTS

Components	Dlib landmark indexes (x, y)
Eyes	(17-26,17-26), ((16+45)/2,(16+45)/2), ((15+46)/2,(15+46)/2), ((29+47)/2,(29+47)/2), (28,28), ((29+40)/2,(29+40)/2), ((1+41)/2,(1+41)/2), ((0+36)/2,(0+36)/2)
Nose	(21-22,21-22), (42,42), (54,30), ((35+54)/2,(35+54)/2), ((33+51)/2,(33+51)/2), ((31+48)/2,(31+48)/2), (48,30), (39,39)
Mouth	((33+51)/2,(33+51)/2), ((35+54)/2,(35+54)/2), ((13+54)/2,(13+54)/2), ((11+55)/2,(11+55)/2), ((8+57)/2,(8+57)/2), ((5+59)/2,(5+59)/2), ((3+48)/2,(3+48)/2), ((31+48)/2,(31+48)/2)
Eyes & nose	(17-26,17-26), ((16+45)/2,(16+45)/2), ((15+46)/2,(15+46)/2), ((29+47)/2,(29+47)/2), (54,30), ((35+54)/2,(35+54)/2), ((33+51)/2,(33+51)/2), ((31+48)/2,(31+48)/2), (48,30), ((29+40)/2,(29+40)/2), ((1+41)/2,(1+41)/2), ((0+36)/2,(0+36)/2)
Forehead	(17-26,17-26), (25,25-(33-27)), (18,18-(33-27))
Upper face	((15+16)/2,(15+16)/2), (28,46), (28,28), (28,41), ((0+1)/2,(0+1)/2), (17,17-(33-27)), (26,26-(33-27))
Midface	(1-3,1-3), ((31+48)/2,(31+48)/2), ((33+51)/2,(33+51)/2), ((35+54)/2,(35+54)/2), (13-15,13-15), ((15+16)/2,(15+16)/2), (28,46), (28,28), (28,41), ((0+1)/2,(0+1)/2)
Lower face	(3-13,3-13), ((35+54)/2,(35+54)/2), ((33+51)/2,(33+51)/2), ((31+48)/2,(31+48)/2)
Central face	(17,18), (19,19), ((19+24)/2,(19+24)/2), (24,24), (26,25), ((16+45)/2,(16+45)/2), (26,15), (26,14), ((35+54)/2,(35+54)/2), ((33+51)/2,(33+51)/2), ((31+48)/2,(31+48)/2), (17,2), (17,1) ((0+36)/2,(0+36)/2)

K_{Comp}, we have

$$k_{cyt} = \max\big(f_{mt}, k_{cyt} - ((tb - tt) - (sb - st))/2\big), \quad (8)$$

$$k_{cyb} = \min\big(f_{mb}, k_{cyb} + ((tb - tt) - (sb - st))/2\big), \quad (9)$$

where tt and tb are the top and the bottom boundary values of the component area in I_{Acco}, st and sb are the top and the bottom boundary values of the component area in I_{Morp}, f_{mt} and f_{mb} are the top and the bottom boundary values of face area in I_{Morp}. For given left and right boundary values k_{cxl} and k_{cxr} from K_{Comp}, we get

$$k_{cxl} = \max(f_{ml}, k_{cxl} - ((tr - tl) - (sr - sl))/2), \quad (10)$$

$$k_{cxr} = \min(f_{mr}, k_{cxr} + ((tr - tl) - (sr - sl))/2), \quad (11)$$

where tl and tr are the left and the right boundary values of the component area in I_{Acco}, sl and sr are the left and the right boundary values of the component area in I_{Morp}, f_{ml} and f_{mr} are the left and the right boundary values of face area in I_{Morp}.

(6) With the adjusted facial component landmark set K_{Comp}, a facial component mask operator is constructed as

$$M_{Comp}(x, y) = \begin{cases} 1, & (x, y) \in A_{Comp} \\ 0, & (x, y) \notin A_{Comp} \end{cases}, \quad (12)$$

where A_{Comp} is a component polygon area constructed by K_{Comp}, x and y are the horizontal and the vertical coordinates of M_{Comp}, respectively. Then, the corresponding facial component I_{Comp} is extracted from I_{Morp} as

$$I_{Comp} = I_{Morp} * M_{Comp}. \quad (13)$$

(7) The component's central point in I_{Acco} is calculated as

$$\begin{cases} cx = (tl + tr)/2 \\ cy = (tt + tb)/2 \end{cases}. \quad (14)$$

With (cx, cy), I_{Comp} is morphed into I_{Acco} by the use of Poisson image editing [35], and a facial components-based manipulated image I_{M_Comp} is finally obtained. A visualization and dlib landmarks of different facial components are shown in Fig. 4 and Table III, respectively.

IV. DATASET

In this work, clear frontal face images are collected from 94 Chinese subjects, most of them are university students with an age range from 18 to 26 years. All the images are captured by a Canon EOS 600D camera in an indoor environment. To simulate an enrolment scenario, a white wall is used as a background in the data collection, and only a fixed indoor lighting source is used. For the first 60 subjects, two images are collected for each person with an interval of one week. The first one is used for launching a MA, while the second one is used as a live probe face image. To be consistent with the existing research [4], all the images from the first 60 subjects are normalized into a size of 413×531 pixels in this work. To follow the photo standards [1], the distances between the left and the right ears of all the images from the first 60 subjects are normalized to 70% of the image width. For the remaining 34 subjects, their images are normalized into a size of 360×480 pixels, and the distances between the left and the right ears of the images are normalized to 306 pixels.

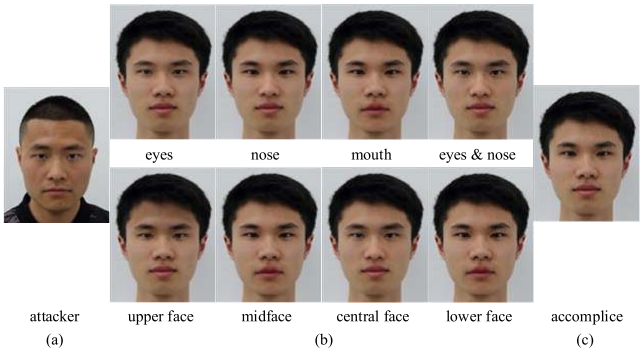

Fig. 5. Examples of partial manipulated face images. The warping weight α_1 and blending weight α_2 are set as 0.5. (a) An original face image of an attacker. (b) Manipulated face images. (c) An original face image of an accomplice.

Moreover, the left and right eyes of all the images are aligned through image rotation.

After data collection, for the first 60 subjects, one of them alternately acts as an attacker, and three subjects whose face is similar to that of the attacker are selected as accomplices. To save time, accomplices are selected on the basis of the similarity scores from Megvii Face++ [36], while an attacker can manually choose accomplices in practical. For each pair of the attacker and accomplice, 9 facial components-based manipulated face images are generated, where the warping weight α_1 and blending weight α_2 range from 0.1 to 1.0 with a step of 0.1 ($\alpha_1 = \alpha_2$). To analyze the effects of different image qualities and strategies for choosing accomplices, central face-based manipulated images are also generated using a JPEG quality factor of 50 and randomly selected accomplices. Furthermore, to investigate the transferability of the proposed attack among different groups of subjects, eyes & nose and central face-based manipulated images are generated with the remaining 34 subjects. As a result, $60 \times 3 \times 9 \times 10 + 60 \times 3 \times 1$ (JPEG quality 50) $+60 \times 4$ (randomly selected accomplices) $+34 \times 3 \times 2 = 16824$ components-based manipulated face images are generated. Some examples are shown in Fig. 5.

V. Experimental Results and Analysis

A. Experimental Setup

For vulnerability analysis of face authentication systems, three commercial off-the-shelf systems, Microsoft Azure Face [37], Neurotechnology VeriLook [38], and Cognitec FaceVACS [39] as well as four deep learning models VGG-Face [40], ResNet50 (trained on VGGFace2) [41], Light CNN-9 [42], and ArcFace [43] are used. The thresholds of Face++, VeriLook, Cognitec (the pre-set recommendations are used), VGG-Face, ResNet50, LightCNN-9, and ArcFace are set at FMR = 0.1%, while the threshold of Azure is set as the default value. For the deep learning models, the embeddings are extracted from the pre-trained models as face authentication features, and the cosine similarity between the features from two face images are used as the face authentication similarity score S_{Sim_f}

$$S_{sim\_f} = 0.5 + 0.5\frac{f_1 f_2'}{\sqrt{(f_1 f_1')(f_2 f_2')}}, \tag{15}$$

where f_1 and f_2 are two feature vectors of two face images. The threshold of the deep learning models is tuned by LFW database [44]. For bona fide face images of the collected dataset, with a threshold of FMR = 0.1%, the FMR (or false acceptance rate) and false non-match rate (FNMR, or false rejection rate) from Face++ are 0.48% and 0.00%, the FMR and FNMR from Cognitec are 1.33% and 0.00%, respectively.

To analyze the robustness of the proposed attack, four face MAD methods are used, and they are BSIF [24], transferable deep CNN features [29], deep learning (AlexNet) [30], and Fourier spectrum of sensor pattern noise [25]. Since the collected dataset only has limited number of subjects and the number of attack samples is also much more than that of bona fide ones, a MAD classifier or deep learning model cannot be well trained with it. Thus, ND-IIITD retouched faces database [45] is introduced for building training and development sets in this work. This database contains 325 subjects, and each subject has 8 bona fide face images. Among them, the faces from the first 211 subjects are used as training set, while the faces from the remaining 114 subjects are used as development set for tuning MAD thresholds. To generate the corresponding attack samples, 8 face images from each subject are respectively used for generating complete morphs, splicing morphs, and combined morphs with morphing weights of 0.3, 0.4, and 0.5.

B. Evaluation Metrics

1) Existing Metrics: To measure the vulnerability of face recognition systems to MAs, mated morph presentation match rate (MMPMR) [46] was proposed, which is calculated by

$$\mathrm{MMPMR}(th_{fa}) = \frac{1}{N}\sum_{i=1}^{N}\left(\left(\min_{j=1,\ldots,M_i} S_{ij}\right) > th_{fa}\right), \tag{16}$$

where th_{fa} is the face recognition threshold, N is the total number of MAs, M_i is the number of MA contributors for the i–th MA sample, S_{ij} is the face recognition score of the j–th contributor for the i–th MA sample. A higher value of MMPMR represents a more vulnerable face recognition system.

To measure the robustness of MAs against MAD, attack presentation classification error rate (APCER) [47] is used, and it is calculated as

$$\mathrm{APCER}(th_{mad}) = \frac{1}{N}\sum_{i=1}^{N}(S_{mad\_i} > th_{mad}), \tag{17}$$

where th_{mad} is the MAD threshold, S_{mad_i} is the MAD score of the i–th MA sample. Higher values of APCER indicate better anti-detectability and robustness of MAs. To follow the standards of biometrics presentation attacks, bona fide presentation classification error rates BPCER10 and BPCER20 at fixed security levels are also used, and they represent BPCER values when APCER equals to 10% and 5%, respectively [47]. In addition, detection error tradeoff (DET) curves are also used, and its horizontal axis and vertical axis are APCER and BPCER, respectively.

2) The Proposed Metric: In a real-world scenario, to achieve the goal of MAs, a manipulated face image used for enrolment needs to evade MAD and also to match both

TABLE IV
MMPMR (%) OF DIFFERENT FACIAL COMPONENTS FROM COGNITEC*

Weights $\alpha = \alpha_1 = \alpha_2$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0	Average
Eyes	21.11	29.44	39.44	50.56	57.22	62.78	67.22	70.56	73.89	72.22	54.44
Mouth	16.11	18.89	22.22	23.89	25.56	31.11	33.33	34.44	37.78	37.22	28.06
Forehead	16.67	20.00	20.00	21.67	22.78	25.00	27.22	30.00	30.56	31.11	24.50
Upper face	24.44	40.00	58.89	74.44	80.00	85.00	86.67	87.78	88.33	88.33	71.39
Midface	20.56	26.11	37.22	45.00	49.44	51.11	54.44	54.44	53.89	50.56	44.28
Lower face	18.33	25.00	28.33	31.67	37.22	42.22	44.44	47.78	48.89	48.33	37.22
Nose (proposed)	20.00	22.78	26.67	38.33	44.44	49.44	52.78	53.89	56.11	55.00	41.94
Eyes & nose (proposed)	25.56	40.56	58.89	73.33	82.78	89.44	92.22	93.33	91.67	88.33	73.61
Central face (proposed)	27.78	50.00	68.89	89.44	95.56	96.11	97.78	96.11	92.22	89.44	80.33

*The outcome does not necessarily constitute the best the algorithm can do

probe images from an attacker and an accomplice. However, separately using MMPMR and APCER will overestimate the MA success rate. For instance, for two given morphed face images, it is assumed that the first image can match facial probe images but cannot evade MAD, while the second image can evade MAD but cannot match the facial probe images. If the MA performance of these two images is separately evaluated with MMPMR and APCER, results of MMPMR = 1/2 and APCER = 1/2 are obtained. But the actual MA success rate of these two images is 0/2, as none of them can match facial probe images and evade MAD at the same time. To address this problem, a simple but effective metric called actual mated morph presentation match rate (AMPMR) is proposed. AMPMR is defined as the proportion of MAs in all MA samples, whose face recognition scores of all MA contributors are higher than th_{fa} and the MAD score is higher than th_{mad}, for a given face recognition threshold th_{fa} and MAD threshold th_{mad}. It is calculated by

$$\mathrm{AMPMR}\left(th_{fa}, th_{mad}\right) = \frac{1}{N}\sum_{i=1}^{N}\left(\left(\left(\min_{j=1,\ldots,M_i} SC_{ij}\right) > th_{fa}\right) \mathrm{AND}\left(SC_{mad\_i} > th_{mad}\right)\right), \tag{18}$$

where N is the total number of MAs, M_i is the number of MA contributors for the i–th MA sample, SC_{ij} is the face recognition score of the j–th contributor for the i–th MA sample, SC_{mad_i} is the MAD score of the i–th MA sample. Higher values of AMPMR indicate better overall attack performance of MAs.

To quantify visual differences between a manipulated face image and an accomplice's face image, peak signal-to-noise ratio (PSNR) [48] and structural similarity (SSIM) [49] are adopted as metrics, and an accomplice's face image is used as a reference image. Since PSNR is sensitive to pixel changes, it is used for measuring local visual differences of a manipulated face image, and it is calculated by

$$\mathrm{PSNR}(I_1, I_2) = 10\log_{10}\left(\frac{\max(I_1)^2}{\mathrm{MSE}}\right), \tag{19}$$

$$\mathrm{MSE}(I_1, I_2) = \frac{1}{wh}\sum_{i=1}^{h}\sum_{j=1}^{w}(I_1(i,j) - I_2(i,j))^2, \tag{20}$$

where max () means the maximum value of image pixels, I_1 and I_2 are two images, w and h are width and height of the images, respectively. A higher value of PSNR represents small local visual differences of a manipulated face image. In contrast, SSIM is more similar to human perception and is more focused on the overall structure information of images. Thus, it is used for measuring global visual differences of a manipulated face image, and it is calculated as

$$\mathrm{SSIM}(I_a, I_b) = \frac{(2\mu_a\mu_b + C_1)(2\sigma_{ab} + C_2)}{\left(\mu_a^2 + \mu_b^2 + C_1\right)\left(\sigma_a^2 + \sigma_b^2 + C_2\right)}, \tag{21}$$

where I_a and I_b are two images, μ_a, μ_b and σ_a, σ_b are means and standard deviations of I_a and I_b, σ_{ab} is cross-covariance between I_a and I_b, C_1 and C_2 are regularization constants. A higher value of SSIM represents small global visual differences of a manipulated face image. For PSNR and SSIM, it is difficult to define a fixed value that indicates a high (or low) quality of a morphed face image, but they can still be used for objectively comparing the quality of different MAs. For instance, higher values of PSNR and SSIM represent relatively low visual distortions of a morphed face image.

C. Experimental Results

1) Vulnerability of Face Recognition Systems: To select the most effective facial components, and to demonstrate the effectiveness of the proposed attack in reaching a match for probe images of attackers and accomplices, experiments are conducted with different types of manipulated face images, and the results are listed in Table II of Appendix B (in the supplemental material), Table IV (not used for selecting components), and Table V, respectively. From the results, Face++ and VGG-Face are vulnerable to the proposed attack, and the results demonstrate face image manipulation in component level is sufficient to attack face recognition systems. By comparing the results between different facial components, relatively high MMPMR values are obtained for nose, eyes & nose, and central face-based manipulated images, which suggests those components are very relevant for face recognition systems. Meanwhile, with the increase of the weight parameters α_1 and α_2, facial characteristics from an attacker and an accomplice become more balanced in a components-based manipulated face image, and peak values of MMPMR occur when the weight parameters α_1 and α_2 are around 0.7. It is

TABLE V
MMPMR (%) OF DIFFERENT FACIAL COMPONENTS FROM VGG-FACE

Weights $\alpha = \alpha_1 = \alpha_2$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0	Average
Eyes	70.56	74.44	77.78	78.89	81.67	83.33	83.89	85.56	85.56	83.33	80.50
Mouth	70.00	73.89	72.78	75.56	76.11	74.44	74.44	75.56	73.89	75.00	74.17
Forehead	70.56	70.56	69.44	70.00	70.56	70.56	70.56	70.56	70.56	70.56	70.39
Upper face	71.67	75.56	81.11	83.33	85.56	88.33	88.89	88.33	86.11	80.56	82.95
Midface	73.89	76.11	82.22	87.78	89.44	89.44	87.22	86.67	82.78	80.00	83.56
Lower face	72.22	72.78	72.78	71.67	73.33	72.78	70.56	71.11	72.78	71.11	72.11
Nose (proposed)	73.33	75.56	80.00	86.67	87.78	88.33	90.56	88.89	87.22	87.22	84.56
Eyes & nose (proposed)	74.44	81.11	87.78	93.33	92.22	93.33	93.89	92.22	88.89	78.89	87.61
Central face (proposed)	75.00	83.89	92.78	95.00	97.22	96.11	95.00	91.11	87.22	75.56	88.89

TABLE VI
MORPHING ATTACK DETECTION RESULTS FROM BSIF BASED METHOD [24] (BPCER = 10.00%)

Metrics	Weights $\alpha = \alpha_1 = \alpha_2$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0	Average
APCER (%)	Nose (proposed)	88.89	89.44	87.78	86.67	85.56	84.44	85.56	86.67	88.33	92.22	87.56
	Eyes & nose (proposed)	85.00	82.22	73.89	74.44	71.67	76.11	80.00	85.56	92.22	96.67	81.78
	Central face (proposed)	80.56	68.89	63.33	58.89	57.22	60.00	63.89	76.11	85.56	94.44	70.89
	Combined [4]	47.78	32.22	26.67	23.89	20.00	22.78	28.89	37.22	48.89	85.00	37.33
BPCER10 (%)	Nose (proposed)	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	Eyes & nose (proposed)	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	Central face (proposed)	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	Combined [4]	100.00	100.00	33.33	18.33	16.67	21.67	26.67	100.00	100.00	100.00	61.67
BPCER20 (%)	Nose (proposed)	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	Eyes & nose (proposed)	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	Central face (proposed)	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	Combined [4]	100.00	100.00	100.00	100. 00	33.33	38.33	100. 00	100.00	100.00	100.00	83.96

observed that in a components-based manipulated face image, most of the face regions are from an accomplice's face, while facial biometric characteristics from an attacker are only represented in limited regions. Thus, larger values of α_1 and α_2 are required to compensate this imbalance. Moreover, MMPMR from VGG-Face is higher than that from Face++. The main reason is that VGG-Face has a higher FMR for the collected dataset.

According to the MMPMR results from Face++ and VGG-Face, nose, eyes & nose, and central face-based manipulated images are used in the following experiments.

2) Robustness to Morphing Attack Detection: To analyze the ability of the proposed attack for evading MAD, evaluations are made with four MAD methods, which are BSIF [24], transferable deep CNN features [29], deep learning (AlexNet) [30], and Fourier spectrum of sensor pattern noise [25]. There are 3462 (1831 morphed, 1631 bona fide) and 1799 (949 morphed, 850 bona fide) face images in the training and development sets, respectively. In this work, APCER is computed based on the equal error rate threshold tuned on the development set, while BPCER10 and BPCER20 are directly computed on the collected dataset without using the development set. The results are listed in Table VI to Table IX, respectively. From these results, it demonstrates the good ability of the proposed attack for countering MAD. Compared with other two types of manipulated face images, nose-based attack achieves relatively higher detection error rates. Because tampered regions in a nose based manipulated face image are limited, this leads to fewer MAD cues. Meanwhile, when α_1 and α_2 are around 0.5, relatively low MAD error rates are obtained. The main reason is that under those parameters, a manipulated facial component is almost an average of two subjects' components, which results in a large distortion of a manipulated face image.

3) Visual Differences: To measure visual differences between a manipulated face image and an accomplice's face image, PSNR and SSIM are calculated with different types of manipulate images, and the results are listed in Table X and Table XV, respectively. It can be found that there exist small local and global visual differences of the proposed attack. With the increases of the weight parameters α_1 and α_2, the proportion of an accomplice's face image in a manipulated image is gradually reduced, which leads to the decrease of PSNR and SSIM. Furthermore, for nose, eyes & nose, and central face-based manipulated images, the trends of PSNR and SSIM are the same as those of MAD error rates, which suggests PSNR and SSIM can effectively measure visual differences of a manipulated face image in MA.

D. Performance Analysis

1) Analysis of Face Similarity Scores: To visualize the vulnerability of face recognition systems to the proposed attack, experiments are performed on Cognitec [39] to analyze face

TABLE VII
MORPHING ATTACK DETECTION RESULTS FROM TRANSFERABLE CNN FEATURES-BASED METHOD [29] (BPCER = 5.00%)

Metrics	Weights $\alpha = \alpha_1 = \alpha_2$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0	Average
APCER (%)	Nose (proposed)	91.11	88.33	86.67	84.44	85.56	85.56	85.56	88.33	93.89	96.11	88.56
	Eyes & nose (proposed)	85.56	75.56	62.22	53.33	47.22	51.11	60.56	70.56	78.89	96.67	68.17
	Central face (proposed)	81.67	65.56	47.78	38.89	36.67	41.67	49.44	60.00	75.56	94.44	59.17
	Combined [4]	48.33	18.89	7.22	1.11	1.11	1.67	6.11	17.78	40.56	81.11	22.39
BPCER10 (%)	Nose (proposed)	100.00	100.00	100.00	50.00	50.00	50.00	100.00	100.00	100.00	100.00	85.00
	Eyes & nose (proposed)	100.00	40.00	26.67	23.33	23.33	25.00	35.00	43.33	100.00	100.00	51.67
	Central face (proposed)	50.00	30.00	18.33	16.67	16.67	16.67	23.33	35.00	50.00	100.00	35.67
	Combined [4]	16.67	5.00	3.33	1.67	1.67	1.67	3.33	5.00	16.67	100.00	15.50
BPCER20 (%)	Nose (proposed)	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	Eyes & nose (proposed)	100.00	48.33	38.33	33.33	30.00	33.33	43.33	100.00	100.00	100.00	62.67
	Central face (proposed)	100.00	40.00	23.33	23.33	23.33	23.33	31.67	41.67	100.00	100.00	50.67
	Combined [4]	23.33	10.00	5.00	3.33	1.67	1.67	5.00	11.67	23.33	100.00	18.50

TABLE VIII
MORPHING ATTACK DETECTION RESULTS FROM DEEP LEARNING (ALEXNET) BASED METHOD [30] (BPCER = 11.67%)

Metrics	Weights $\alpha = \alpha_1 = \alpha_2$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0	Average
APCER (%)	Nose (proposed)	87.22	86.67	81.67	81.11	80.00	78.89	79.44	80.00	82.78	88.33	82.61
	Eyes & nose (proposed)	75.56	62.78	51.67	47.22	46.67	50.00	54.44	61.11	68.33	82.78	60.06
	Central face (proposed)	73.33	53.33	45.00	36.67	37.78	40.00	45.00	56.11	66.67	85.56	53.95
	Combined [4]	45.00	27.78	20.56	14.44	12.22	15.00	20.56	29.44	40.56	68.89	29.45
BPCER10 (%)	Nose (proposed)	80.00	80.00	80.00	81.67	80.00	81.67	81.67	81.67	85.00	90.00	82.17
	Eyes & nose (proposed)	75.00	63.33	50.00	46.67	48.33	48.33	53.33	63.33	75.00	90.00	61.33
	Central face (proposed)	71.67	51.67	46.67	45.00	41.67	46.67	48.33	55.00	65.00	90.00	56.17
	Combined [4]	46.67	30.00	18.33	18.33	16.67	18.33	23.33	30.00	50.00	80.00	33.17
BPCER20 (%)	Nose (proposed)	91.67	90.00	90.00	90.00	90.00	90.00	90.00	90.00	90.00	91.67	90.33
	Eyes & nose (proposed)	90.00	80.00	71.67	65.00	65.00	65.00	70.00	75.00	88.33	91.67	76.17
	Central face (proposed)	86.67	71.67	58.33	51.67	51.67	58.33	65.00	71.67	81.67	91.67	68.84
	Combined [4]	61.67	45.00	31.67	26.67	21.67	23.33	28.33	43.33	55.00	90.00	42.67

TABLE IX
DETECTION RESULTS FROM FOURIER SPECTRUM OF SENSOR PATTERN NOISE (FS-SPN) METHOD [25] (BPCER = 0.00%)

Metrics	Weights $\alpha = \alpha_1 = \alpha_2$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0	Average
APCER (%)	Nose (proposed)	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	Eyes & nose (proposed)	100.00	100.00	100.00	100.00	99.44	100.00	100.00	100.00	100.00	100.00	99.94
	Central face (proposed)	97.78	96.67	96.67	97.78	98.33	99.44	100.00	100.00	100.00	100.00	98.67
	Combined [4]	72.78	61.11	53.33	45.00	40.00	41.67	46.67	52.78	64.44	97.22	57.50
BPCER10 (%)	Nose (proposed)	83.33	83.33	81.67	81.67	81.67	81.67	81.67	83.33	83.33	90.00	83.17
	Eyes & nose (proposed)	48.33	45.00	43.33	43.33	43.33	43.33	41.67	45.00	48.33	88.33	49.00
	Central face (proposed)	40.00	25.00	23.33	23.33	23.33	23.33	23.33	23.33	35.00	90.00	33.00
	Combined [4]	3.33	1.67	0.00	0.00	0.00	0.00	0.00	1.67	3.33	15.00	2.50
BPCER20 (%)	Nose (proposed)	85.00	83.33	83.33	83.33	83.33	83.33	83.33	88.33	86.67	93.33	85.33
	Eyes & nose (proposed)	81.67	61.67	48.33	48.33	48.33	48.33	53.33	68.33	81.67	93.33	63.33
	Central face (proposed)	51.67	46.67	43.33	26.67	23.33	35.00	43.33	46.67	51.67	93.33	46.17
	Combined [4]	3.33	3.33	3.33	3.33	1.67	1.67	3.33	3.33	3.33	23.33	5.00

similarity scores of eyes & nose based manipulated face images, and the results are shown in Fig. 6. From the results, since a eyes & nose based manipulated face image is mainly constructed by an accomplice's face image, it can well match with the accomplice. Meanwhile, as facial biometric characteristics from attackers are also represented, 92.22% of eyes & nose based manipulated face images can match both attackers and accomplices. The results demonstrate that partial face image manipulation is effective towards face recognition systems.

2) Performance Analysis of Different File Sizes: Since many access control applications have file size requirements regarding reference image used for enrolment (e.g., 15-20kB suggested in [1]), JPEG compression with a quality factor 50 is used to generate images with a file size of 15kB. Experiments are made with images of different file

TABLE X
PSNR (dB) OF THE PROPOSED ATTACK

Weight α*	Eyes	Nose	Mouth	Eyes & nose	Forehead	Upper	Midface	Lower	Central
0.5	35.03	39.22	40.71	32.56	33.39	28.82	34.97	36.92	31.97
0.7	33.20	37.27	38.82	30.73	31.03	26.73	33.06	34.90	30.08
0.9	31.81	35.69	37.25	29.32	29.23	25.29	31.66	33.36	28.56
Average	35.19	39.22	40.66	32.85	33.54	29.04	35.22	36.95	32.20

*$\alpha_1 = \alpha_2 = \alpha$

TABLE XI
RESULTS (%) OF CENTRAL FACE BASED MANIPULATED IMAGES ($\alpha_1 = \alpha_2 = 0.6$) WITH DIFFERENT FILE SIZES

File sizes	Face++	BSIF MAD method			Deep learning (AlexNet) MAD method		
	MMPMR (%)	APCER (%)	BPCER10 (%)	BPCER20 (%)	APCER (%)	BPCER10 (%)	BPCER20 (%)
15kB	98.89	35.00	100.00	100.00	38.33	38.33	53.33
Original	98.89	60.00	100.00	100.00	40.00	46.67	58.33

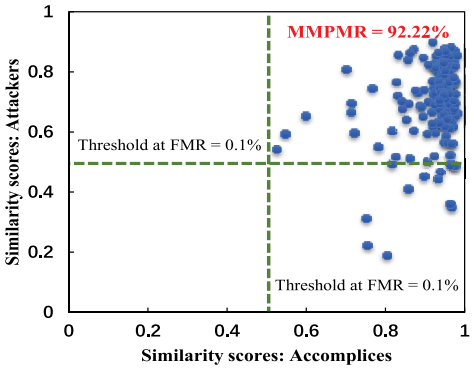

Fig. 6. Face similarity scores from Cognitec with eyes & nose based manipulated face images ($\alpha_1 = \alpha_2 = 0.7$).

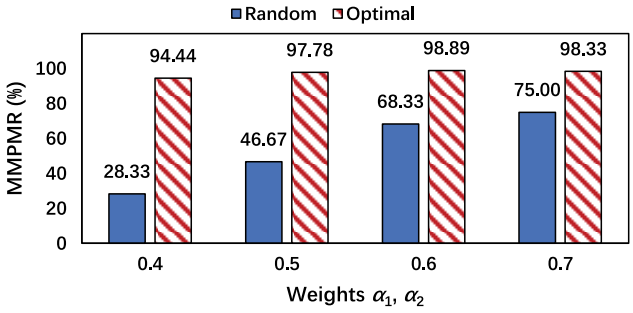

Fig. 7. MMPMR (%) from Face++ of central face based manipulated images with different combinations of attackers and accomplices.

sizes, and the results are listed in Table XI. From the results, the ability of the proposed attack to resist MAD is affected by file size changes, because JPEG compression causes image distortions. However, the change in file size has less impact on the deep learning-based MAD method. In addition, MMPMRs from Face++ with different file sizes are consistent to each other, which indicates face recognition is less affected by file size changes.

3) Performance Analysis of Random Contributors: In a real-world scenario, an attacker can hardly find many accomplices. To analyze the performance of the proposed attack under this scenario, experiments are made with random attack contributors. Specifically, the first 60 subjects in the collected dataset are randomly assigned a unique ID number range from 1 to 60, one of the subjects acts as an attacker in turn, and a subject whose ID is the next number of the attacker's ID acts as an accomplice. As a result, for each weight parameter, 60 central face-based manipulated images are generated with random attack contributors, and the results are shown in Fig. 7. From the results, MMPMR results of random attack contributors are lower than those of optimal ones. However, with appropriate weight parameters (e.g., $\alpha_1 = \alpha_2 = 0.7$), random attack contributors can still achieve a result of MMPMR = 75.00%, which demonstrates the effectiveness of the proposed attack launched by random contributors.

4) Analysis of the Transferability: To analyze the transferability of the proposed attack to unknown face recognition systems or deep learning models, evaluations are made with Azure, VeriLook, and Cognitec systems as well as ResNet50 (trained on VGGFace2) [41], Light CNN-9 [42], and ArcFace [43] models, and their thresholds are set as the default value (for Azure) and FMR = 0.1% (for the others), respectively. The results are shown in Fig. 8. From the results, with appropriate weight parameters (e.g., $\alpha_1 = \alpha_2 = 0.7$), both eyes & nose and central face-based manipulated images can achieve good performance (MMPMR > 90%) on Azure, VeriLook, and Cognitec systems. The results indicate that the facial components selected by Face++ and VGG-Face can be well generalized to other face recognition systems. Furthermore, the proposed attack also achieves competitive results on the deep learning models.

To analyze the transferability of the proposed attack among different groups of subjects, experiments are made with the remaining 34 subjects. For each pair of attacker and accomplice, eyes & nose and central face-based manipulated images are generated, where the warping weight α_1 and blending weight α_2 are both set as 0.7. As a result, $34 \times 3 \times 2 = 204$ manipulated face images are generated by the remaining 34 subjects. With Face++ (threshold at FMR = 0.1%),

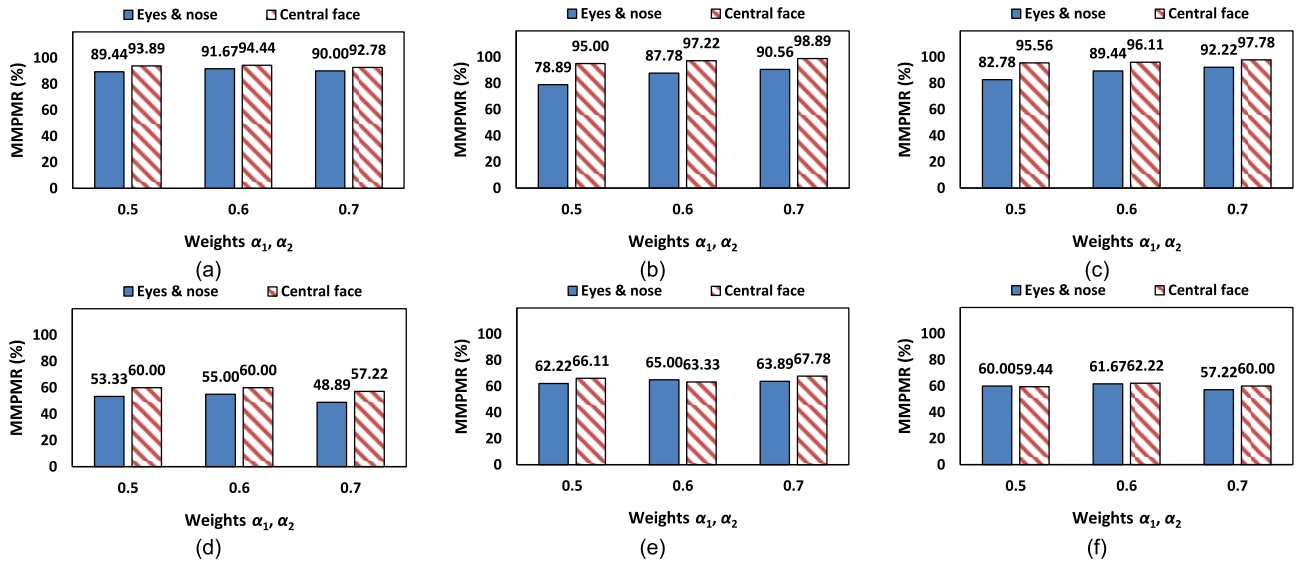

Fig. 8. MMPMR (%) of the proposed attack to unknown face recognition systems or models. (a) From Azure. (b) From VeriLook. (c) From Cognitec. (d) From ResNet50 (trained on VGGFace2). (e) From Light CNN-9. (f) From ArcFace.

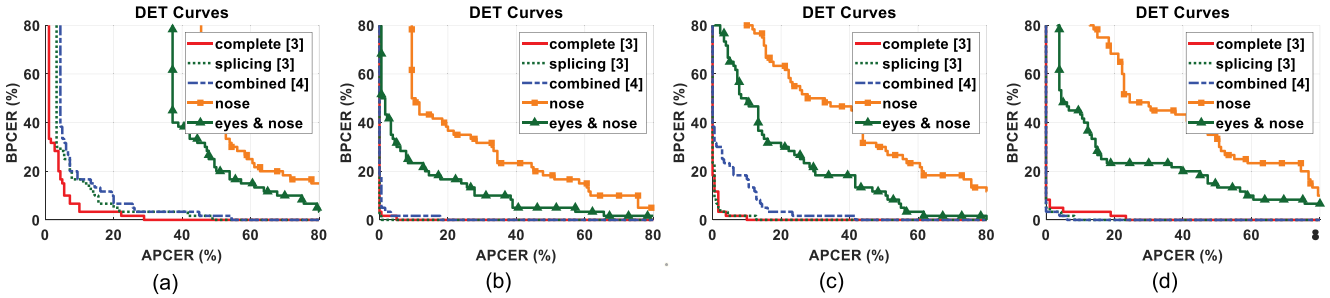

Fig. 9. DET curves of different attacks ($\alpha_1 = \alpha_2 = 0.5$). (a) From BSIF-based method. (b) From transferable deep CNN features-based method. (c) From deep learning (AlexNet) based method. (d) From Fourier spectrum of sensor pattern noise-based method.

TABLE XII
MMPMR (%) FROM VERILOOK WITH DIFFERENT FACE AUTHENTICATION THRESHOLDS

Thresholds at	FMR = 0.01%			FMR = 0.001%		
	$\alpha_1 = \alpha_2 = 0.5$	$\alpha_1 = \alpha_2 = 0.6$	$\alpha_1 = \alpha_2 = 0.7$	$\alpha_1 = \alpha_2 = 0.5$	$\alpha_1 = \alpha_2 = 0.6$	$\alpha_1 = \alpha_2 = 0.7$
Eyes & nose	36.11	54.44	62.22	7.22	21.11	33.33
Central face	70.56	86.67	90.00	33.89	60.56	70.56

MMPMR = 95.10% and 97.06% can be obtained by eyes & nose and central face-based manipulated images, respectively. The results indicate that the facial components selected by the first 60 subjects can be well generalized to other subjects.

5) Performance Analysis of Different Security Levels: To evaluate the performance of the proposed attack to face recognition systems at different security levels, experiments are conducted on VeriLook with different authentication thresholds, and the results are listed in Table XII. From the results, with appropriate weight parameters (e.g., $\alpha_1 = \alpha_2 = 0.7$), central face-based manipulated images can achieve MMPMR = 90.00% and 70.56% with the FMR of 0.01% and 0.001%, respectively. The results demonstrate the effectiveness of the proposed attack to face recognition thresholds with high security levels.

6) Performance Benchmark: To validate the effectiveness of partial face image manipulation, experiments are made to compare the performance of the proposed attack with that of complete morphing [3], splicing morphing [3], and combined morphing [4], and their performance is compared from the aspects of face recognition, countering MAD, and visual differences, respectively. For fair comparison, the existing attacks [3], [4] are applied for the collected dataset.

For detection performance benchmark of countering MAD, DET curves of different attacks are shown in Fig. 9. From the results, MAD error rates of the proposed attack are higher than those of complete morphing [3], splicing morphing [3], and combined morphing [4]. Since only partial face regions are manipulated in the proposed attack, fewer detection cues can be found in a components-based manipulated face image, and it is more similar to a bona fide face image in terms of image

TABLE XIII
AMPMR (%) OF DIFFERENT ATTACKS

Modules	Weights $\alpha = \alpha_1 = \alpha_2$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0	Average
Face++ and BSIF	Complete [3]	11.67	15.00	9.44	8.33	6.67	8.89	10.00	13.33	12.78	N/A*	10.68
	Splicing [3]	9.44	17.78	14.44	14.44	15.00	13.33	18.33	28.89	36.11	53.33	22.11
	Combined [4]	12.22	23.33	25.56	23.89	20.00	22.78	28.33	29.44	27.22	27.78	24.06
	Nose (proposed)	14.44	21.11	25.56	30.56	38.33	45.00	51.67	55.00	57.78	56.67	39.61
	Eyes & nose (proposed)	18.89	29.44	48.33	61.11	65.00	70.00	76.67	77.78	77.78	78.33	**60.33**
	Central face (proposed)	17.22	30.56	50.00	57.78	56.67	59.44	62.78	68.89	73.89	67.78	54.50
Face++ and transferable deep CNN	Complete [3]	20.56	15.00	3.33	0.00	0.00	0.56	3.89	16.11	22.22	N/A*	9.07
	Splicing [3]	2.22	0.56	0.56	0.00	0.00	0.00	0.00	0.00	1.67	31.11	3.61
	Combined [4]	11.11	8.33	7.22	1.11	1.11	1.67	6.11	15.00	27.22	28.89	10.78
	Nose (proposed)	14.44	18.33	24.44	30.00	40.00	45.56	51.11	53.89	59.44	57.78	39.50
	Eyes & nose (proposed)	17.78	26.67	40.00	42.78	43.89	46.67	57.78	65.56	66.67	78.33	**48.61**
	Central face (proposed)	17.78	28.89	35.56	36.67	36.11	40.56	48.89	56.11	66.11	66.67	43.34
Face++ and deep learning (AlexNet)	Complete [3]	17.22	16.11	6.67	1.67	0.56	1.67	5.00	16.67	20.00	N/A*	9.51
	Splicing [3]	3.33	2.78	1.67	0.00	0.56	0.56	0.56	3.33	7.78	35.00	5.56
	Combined [4]	12.22	15.00	20.00	14.44	12.22	15.00	20.56	26.11	27.22	25.00	18.78
	Nose (proposed)	13.89	18.89	23.33	29.44	37.78	43.33	48.89	49.44	53.33	53.33	37.17
	Eyes & nose (proposed)	16.11	23.33	32.78	37.78	42.78	46.67	52.78	56.67	58.33	67.78	**43.50**
	Central face (proposed)	16.67	25.00	31.67	34.44	37.22	39.44	45.00	53.89	59.44	62.22	40.50
Face++ and FS-SPN	Complete [3]	25.00	58.33	64.44	57.22	49.44	56.67	61.67	53.89	27.22	N/A*	50.43
	Splicing [3]	15.00	28.89	41.11	45.00	43.33	43.89	45.56	47.22	51.11	65.56	42.67
	Combined [4]	18.33	34.44	50.00	45.00	40.00	41.67	45.56	46.11	41.67	33.33	39.61
	Nose (proposed)	15.56	22.22	28.89	36.11	45.56	52.78	58.89	61.67	63.89	60.56	44.61
	Eyes & nose (proposed)	21.67	37.22	64.44	81.11	90.00	93.33	96.11	91.67	85.00	81.67	74.22
	Central face (proposed)	21.11	42.78	71.11	92.22	96.11	98.33	98.33	92.78	87.78	71.67	**77.22**

*When $\alpha_1 = \alpha_2 = 1.0$, a complete morphed [3] face image is the same as an original face image of an attacker

quality, texture, and noise characteristics. Thus, compared with the existing morphing attacks, the proposed attack is more robust to MAD.

For overall performance benchmark of face recognition accuracy and countering MAD, AMPMR of different attacks are listed in Table XIII. From the results, relatively high AMPMR scores are obtained with the proposed attack. It indicates that partial face image manipulation can achieve a better tradeoff between face recognition and countering MAD. Meanwhile, the proposed AMPMR metric considers attack performance against face recognition and MAD at the same time, a MA is viewed as successful only when it can simultaneously match probe images of all attack contributors and bypass MAD. Therefore, compared with the existing MMPMR [46] metric that only considers face recognition accuracy, AMPMR is more rigorous in evaluating the success rate of MAs, and it can better reflect the actual attack performance.

Performance comparison is also made with ResNet50 (trained on VGGFace2) [41], and the proposed nose-based morphs (average AMPMR = 39.48%) outperform the existing combined morphs [4] (average AMPMR = 15.11%). By comparing the results between Face++ and ResNet50 [41], relatively higher success rate is obtained on Face++. The main reason is that the FNMR of bona fide images from ResNet50 [41] (FNMR = 3.33%) is higher than that from Face++ (FNMR = 0.00%).

In addition, experiments are conducted to compare the performance of the proposed attack with that of the morphing attack used by University of Bologna (UBO) [21]. To fairly compare their performance, UBO morphing is applied for the collected dataset. The face verification scores are computed by Cognitec, and the training set and development set for MAD are the same as those used in Section V-C.2. The AMPMR results are listed in Table XIV. This experiment suggests that the proposed attack achieves a higher AMPMR result than those of UBO morphing. The results validate that partial manipulation of face images can achieve a better tradeoff between face recognition and countering MAD. Experiments are also conducted to compare the MMPMR results of the proposed attack with that of the morphing attack used by UBO [21]. The results from Cognitec show that competitive results are achieved with the eyes & nose (average MMPMR = 73.61%) and central face-based (average MMPMR = 80.33%) morphing attacks compared with the results of UBO morphing attack (average MMPMR = 79.26%).

For performance comparison of visual differences between a manipulated face image and an accomplice's face image, PSNR and SSIM results of different attacks are shown in Fig. 10 and Table XV, respectively. From the results, the proposed attack can achieve overall higher PSNR and SSIM compared to the existing attacks. The results demonstrate the proposed attack can achieve small visual distortions, and it is more difficult to be identified in human inspections.

7) Subjective Evaluation of Visual Quality: To analyze the visual quality of different attacks, human experiments are conducted with three groups of participants. The first group consists of 81 inexperienced volunteers. The second group

TABLE XIV
COMPARISON OF AMPMR (%) BETWEEN UBO MORPHING AND THE PROPOSED ATTACK

Modules	Weights $\alpha = \alpha_1 = \alpha_2$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0	Average
Cognitec and BSIF	UBO morphing [21]	29.44	27.78	26.67	24.44	22.78	23.89	25.56	27.78	26.67	N/A*	26.11
	Nose (proposed)	18.89	21.11	23.89	33.33	38.33	42.22	43.89	46.11	48.33	50.00	36.61
	Eyes & nose (proposed)	22.22	35.56	42.22	53.89	57.22	67.22	73.33	78.89	83.89	85.00	**59.94**
	Central face (proposed)	21.67	33.89	41.67	51.67	55.00	58.33	62.78	72.22	77.78	83.89	55.89
Cognitec and transferable deep CNN	UBO morphing [21]	30.56	37.22	12.22	1.67	1.11	4.44	15.00	34.44	31.11	N/A*	18.64
	Nose (proposed)	17.22	20.56	23.33	33.33	38.89	43.89	45.56	46.67	52.22	52.22	37.39
	Eyes & nose (proposed)	20.56	30.00	36.67	38.89	40.56	45.00	56.11	66.11	72.22	85.00	**49.11**
	Central face (proposed)	21.67	33.33	33.89	35.00	35.56	40.56	48.33	57.78	70.56	84.44	46.11
Cognitec and deep learning (AlexNet)	UBO morphing [21]	27.22	37.22	23.33	12.78	8.89	11.11	21.67	32.78	26.11	N/A*	22.35
	Nose (proposed)	16.67	19.44	21.67	30.00	36.11	40.00	43.89	42.78	46.11	48.33	34.50
	Eyes & nose (proposed)	18.89	23.89	32.78	36.67	40.56	44.44	50.00	58.33	64.44	73.89	**44.39**
	Central face (proposed)	20.56	28.33	33.89	34.44	36.67	39.44	44.44	55.56	63.89	78.89	43.61
Cognitec and FS-SPN	UBO morphing [21]	35.00	65.00	76.11	72.78	65.00	72.78	73.89	61.11	36.11	N/A*	61.98
	Nose (proposed)	20.00	22.78	26.67	38.33	44.44	49.44	52.78	53.89	56.11	55.00	41.94
	Eyes & nose (proposed)	25.56	40.56	58.89	73.33	82.78	89.44	92.22	93.33	91.67	88.33	73.61
	Central face (proposed)	26.11	47.78	66.11	87.78	94.44	95.56	97.78	96.11	92.22	89.44	**79.33**

**When $\alpha_1 = \alpha_2 = 1.0$, a morphed face image generated by UBO morphing [21] is the same as an original face image of an attacker*

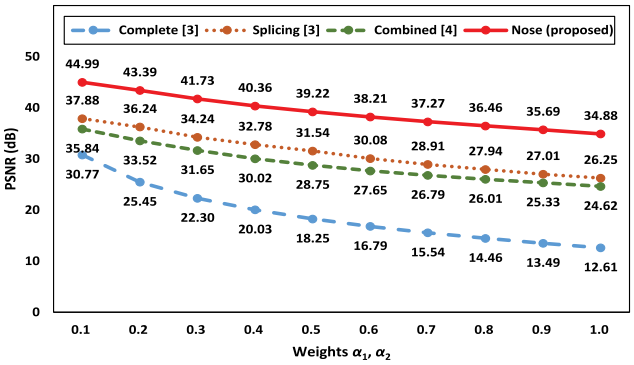

Fig. 10. PSNR (dB) of different attacks.

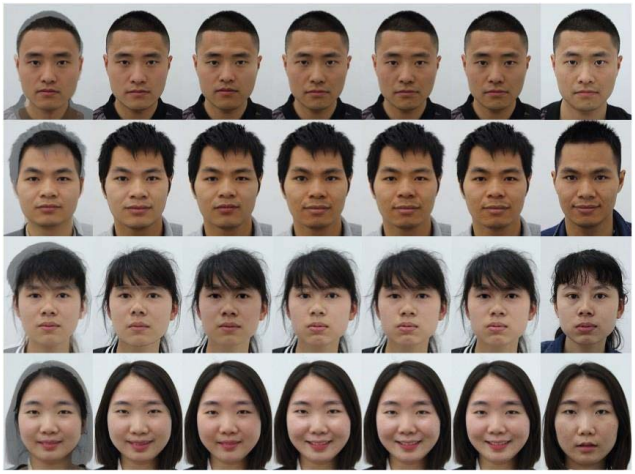

Fig. 11. Examples of complete morphs (column 1), splicing morphs (column 2), combined morphs (column 3), nose-based morphs (column 4), eyes & nose-based morphs (column 5), central face-based morphs (column 6), accomplices' live faces (column 7), $\alpha_1 = \alpha_2 = 0.5$.

consists of 139 researchers in the domain of media forensics, most of them are experienced in digital image (or video) manipulation detection. The third group consists of 46 police officers recruited from Tecent crowdsourcing platform [50], and 71.74% of them state that they have a job duty of face recognition or validating an ID photo with a live person. Each participant is randomly assigned 7 digital face images from 7 different subjects, and they are a complete morphed face image, a splicing morphed face image, a combined morphed face image, a nose-based morphed face image, an eyes & nose based morphed face image, a central face-based morphed face image, and a bona fide face image. The morphing weights of the morphed images are set as 0.5 and 0.7, and all the given images are presented in a random order. The subjective evaluation contains an authenticity check and a similarity check, and both of them are multiple choice tests. In the authenticity check, a participant is asked to judge whether a given image is bona fide. In the similarity check, the corresponding accomplice's live face image is also provided, and the participant is asked to decide whether a given face image belongs to the accomplice. To quantify the subjective evaluation results, human acceptance rate (HAR) is used, and it is calculated as

$$HAR = HP/(HP + HN), \tag{22}$$

where HP is the number of volunteers voting for "bona fide" or "belongs to the accomplice", HN is the number of volunteers voting for "not bona fide" or "does not belong to the accomplice". Examples of different attacks are shown in Fig. 11, and the subjective testing results are listed in Table XVI. From the results, in authenticity check, relatively higher HARs are achieved by the combined morphs and the proposed attack. Meanwhile, nose and eyes & nose-based morphs also achieve better performance in similarity check. By comparing the results between different groups of participants, higher HARs of bona fide face images are obtained by forensics researchers and police officers. Furthermore, the HAR of the bona fide face images is also limited. The possible reason is that the

TABLE XV
SSIM OF DIFFERENT ATTACKS

Weights $\alpha = \alpha_1 = \alpha_2$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0	Average
Complete [3]	0.960	0.912	0.870	0.834	0.800	0.769	0.739	0.709	0.681	N/A	0.808
Splicing [3]	0.992	0.991	0.988	0.986	0.984	0.980	0.976	0.973	0.968	0.965	0.980
Combined [4]	0.989	0.987	0.984	0.980	0.977	0.974	0.970	0.967	0.963	0.958	0.975
Nose (proposed)	0.998	0.997	0.997	0.997	0.997	0.996	0.996	0.996	0.995	0.994	0.996
Eyes & nose (proposed)	0.993	0.991	0.990	0.988	0.986	0.985	0.983	0.982	0.981	0.978	0.986
Central face (proposed)	0.992	0.990	0.988	0.986	0.984	0.982	0.981	0.979	0.977	0.974	0.983

When $\alpha_1 = \alpha_2 = 1.0$, a complete morphed [3] face image is the same as an original face image of an attacker

TABLE XVI
HUMAN ACCEPTANCE RATE (HAR) OF DIFFERENT ATTACKS

Weights $\alpha=\alpha_1=\alpha_2$	Participants	HAR (%)	Complete [3]	Splicing [3]	Combined [4]	Nose (proposed)	Eyes & nose (proposed)	Central face (proposed)	Bona fide
0.5	Inexperienced volunteers	Authenticity	17.28	34.57	39.51	41.98	41.98	39.51	39.51
		Similarity	17.28	27.16	33.33	30.86	23.46	34.57	28.30
	Forensics researchers	Authenticity	4.23	49.30	70.42	73.24	59.15	38.03	50.70
		Similarity	23.94	43.66	30.99	64.79	49.30	35.21	57.75
	Police officers	Authenticity	5.48	49.32	63.01	75.34	73.97	71.23	72.60
		Similarity	16.44	41.10	41.10	60.27	58.90	41.10	45.21
	All participants	Authenticity	4.86	49.31	66.67	**74.31**	66.67	54.86	61.81
		Similarity	20.14	42.36	36.11	**62.50**	54.17	38.19	51.39
0.7	Forensics researchers	Authenticity	4.41	26.47	57.35	80.88	66.18	41.18	54.41
		Similarity	8.82	30.88	26.47	66.18	48.53	32.35	60.29
	Police officers	Authenticity	9.23	30.77	58.46	67.69	61.54	61.54	60.00
		Similarity	12.31	33.85	36.92	53.85	56.92	41.54	50.77
	All participants*	Authenticity	6.77	28.57	57.89	**74.44**	63.91	51.13	57.14
		Similarity	10.53	32.33	31.58	**60.15**	52.63	36.84	55.64

*Face images with morphing weights of $\alpha_1 = \alpha_2 = 0.7$ are not evaluated by inexperienced volunteers

TABLE XVII
DETECTION RESULTS FROM PARTIAL FACE TAMPERING DETECTION NETWORK [19] (BPCER = 65.00%)

Metrics	Weights $\alpha = \alpha_1 = \alpha_2$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0	Average
APCER (%)	Complete [3]	13.33	12.22	12.22	7.78	7.78	7.22	11.67	15.56	18.33	35.00	14.11
	Splicing [3]	12.22	11.67	11.11	11.11	10.56	9.44	10.56	10.56	11.11	16.11	11.45
	Combined [4]	20.00	15.56	15.00	13.89	12.22	11.67	12.78	14.44	18.89	31.10	16.56
	Nose (proposed)	31.67	28.33	25.56	24.44	23.33	21.67	23.33	26.11	29.44	33.89	26.78
	Eyes & nose (proposed)	18.33	14.44	14.44	13.33	15.56	16.11	18.33	23.33	29.44	40.00	20.33
	Central face (proposed)	23.89	16.11	14.44	15.56	16.11	17.78	18.89	23.89	31.67	43.89	22.22

experiments are made with a mobile phone by over 80% of the participants, even though it is suggested to use a 14-inch screen.

8) Robustness to Partial Face Tampering Detection: To analyze the ability of the proposed attack for evading partial face tampering detection, evaluations are made with partial face tampering detection network [19], where the training and development sets are the same as those used in Section V-C.2. The results are listed in Table XVII. From the results, it is demonstrated the good ability of the proposed attack for countering partial face tampering detection. Meanwhile, compared with the existing morphing attacks, relatively higher detection error rate is achieved by the proposed attack.

9) Evaluation Under Known Attack Scenario: To analyze the performance of the proposed attack under known attack scenario, experiments are made with the MAD methods trained and tested on the proposed attack. To be more specific, the bona fide face images in training and development sets are the same as those used in Section V-C.2, but the corresponding morphed face images are generated by the nose, eyes & nose, and central face-based morphs with morphing weights of 0.3, 0.4, and 0.5. The results are listed in Table XVIII and Appendix E (in the supplemental material). From the results, it demonstrates the good ability of the proposed attack for evading MAD under known attack scenario. Since the proposed attack only slightly alters a face region, it is more similar to a bona fide face image compared to whole face-based MAs. Therefore, even though the MAD methods are trained with the proposed attack, they still cannot be well adapted to it.

10) Performance Analysis of Low-Quality Probes: To analyze the performance of the proposed attack to low-quality probe images, the probe images of the collected dataset are

TABLE XVIII
MORPHING ATTACK DETECTION RESULTS UNDER KNOWN ATTACK CONDITION

Methods	Metrics	Weights $\alpha = \alpha_1 = \alpha_2$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0	Average
Deep learning (AlexNet) (BPCER = 0.00%)	APCER (%)	Nose	94.44	92.78	90.00	86.11	85.00	85.56	87.22	89.44	93.33	97.22	90.11
		Eyes & nose	92.78	87.78	75.56	69.44	67.22	69.44	72.78	81.11	90.56	97.78	80.45
		Central face	92.78	87.78	72.22	66.67	63.89	66.11	71.67	80.56	88.89	96.11	78.67
	BPCER10 (%)	Nose	68.33	50.00	41.67	38.33	38.33	40.00	45.00	63.33	78.33	96.67	56.00
		Eyes & nose	58.33	41.67	25.00	18.33	21.67	25.00	38.33	45.00	73.33	91.67	43.83
		Central face	55.00	40.00	30.00	21.67	21.67	26.67	40.00	48.33	76.67	96.67	45.67

TABLE XIX
MMPMR (%) OF LOW-QUALITY PROBES FROM COGNITEC

Weights $\alpha = \alpha_1 = \alpha_2$	0.5	0.6	0.7	0.8	0.9
Nose	30.56	36.67	41.11	43.33	44.44
Eyes & nose	73.33	83.33	87.78	87.78	87.22
Central face	92.22	93.89	93.89	93.33	89.44

transformed into low-quality images, where the number of pixels between the centers of the eyes is normalized to 60 pixels. To further reduce the image quality, the probe images are also compressed with a JPEG quality factor of 50. Experiments are made with the low-quality probe images, and the results are listed in Table XIX. By comparing the results between Table IV and Table XIX, relatively low MMPMR results (from Cognitec) are obtained for the low-quality probe images. It indicates that the performance of the proposed attack is influenced by the quality of the probe images. Nevertheless, by using appropriate morphing weight parameters (e.g., $\alpha_1 = \alpha_2 = 0.7$), eyes & nose-based morphs and central face-based morphs still achieve good performance (MMPMR $> 87\%$) with the low-quality probe images. Moreover, experiments are also conducted to analyze the performance of the proposed attack to gray scale probe images and non-frontal (with a pose rotation of 22.5 degrees to the left-hand side) probe images. The results from Cognitec indicate that eyes & nose-based morphs ($\alpha_1 = \alpha_2 = 0.7$) can achieve good performance with the gray scale (MMPMR $= 92.78\%$) and non-frontal probes (MMPMR $= 83.89\%$).

VI. CONCLUSION

In this article, a partial face manipulation-based MA is proposed to compromise the uniqueness of facial references. It only uses the most effective facial components (eyes and nose) to create a manipulated face image, and can reduce the discrepancies of a manipulated face image and a bona fide face image. In the experiments, a partial face image manipulation-based MA dataset is collected with different attack types, image qualities, and weight parameters. Meanwhile, AMPMR metric is proposed to evaluate MA performance, and image quality assessment metrics are adopted to measure the visual differences of MAs. The results indicate that the proposed attack can achieve higher AMPMR results compared with the existing attacks, and it can improve anti-detectability and visual quality of MAs. In addition, compared to the existing MMPMR metric, the proposed AMPMR metric can better reflect the practical MA performance. Our future work will be focused on implementing partial face manipulation-based MA using GAN, and investigating the countermeasures for the proposed attack.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their kind comments and suggestions for improving the paper, and also thank all the participants of the human experiment and dataset collection.

REFERENCES

[1] *Machine Readable Travel Documents*, document ICAO Doc 9303, Int. Civil Aviation Org., Montreal, QC, Canada, 2015.
[2] M. Ferrara, A. Franco, and D. Maltoni, "The magic passport," in *Proc. IEEE Int. Joint Conf. Biometrics (IJCB)*, Clearwater, FL, USA, 2014, pp. 1–7.
[3] A. Makrushin, T. Neubert, and J. Dittmann, "Automatic generation and detection of visually faultless facial morphs," in *Proc. 12th Int. Joint Conf. Comput. Vis. Imag. Comput. Graph. Theory Appl. (VISAPP)*, 2017, pp. 39–50.
[4] T. Neubert, A. Makrushin, M. Hildebrandt, C. Kraetzer, and J. Dittmann, "Extended stirtrace benchmarking of biometric and forensic qualities of morphed face images," *IET Biometrics*, vol. 7, no. 4, pp. 325–332, Jul. 2018.
[5] N. Damer, A. M. Saladie, A. Braun, and A. Kuijper, "MorGAN: Recognition vulnerability and attack detectability of face morphing attacks created by generative adversarial network," in *Proc. IEEE 9th Int. Conf. Biometrics Theory Appl. Syst. (BTAS)*, Redondo Beach, CA, USA, 2018, pp. 1–10.
[6] N. Damer, F. Boutros, A. M. Saladie, F. Kirchbuchner, and A. Kuijper, "Realistic dreams: Cascaded enhancement of GAN-generated images with an example in face morphing attacks," in *Proc. IEEE 10th Int. Conf. Biometrics Theory Appl. Syst. (BTAS)*, Tampa, FL, USA, 2019, pp. 1–10.
[7] S. Qian *et al.*, "Make a face: Towards arbitrary high fidelity face manipulation," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Seoul, South Korea, 2019, pp. 10033–10042.
[8] A. Makrushin and A. Wolf, "An overview of recent advances in assessing and mitigating the face morphing attack," in *Proc. 26th Eur. Signal Process. Conf. (EUSIPCO)*, Rome, Italy, 2018, pp. 1017–1021.
[9] 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.
[10] D. J. 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, 2017, Art. no. e0173319.
[11] R. S. Kramer, M. O. Mireku, T. R. Flack, and K. L. Ritchie, "Face morphing attacks: Investigating detection with humans and computers," *Cogn. Res. Principles Implications*, vol. 4, no. 1, p. 28, 2019.
[12] X. Zhang, F. Peng, and M. Long, "Robust coverless image steganography based on DCT and LDA topic classification," *IEEE Trans. Multimedia*, vol. 20, no. 12, pp. 3223–3238, Dec. 2018.
[13] R. Raghavendra, K. Raja, S. Venkatesh, and C. Busch, "Face morphing versus face averaging: Vulnerability and detection," in *Proc. IEEE Int. Joint Conf. Biometrics (IJCB)*, Denver, CO, USA, 2017, pp. 555–563.

[14] S. Venkatesh, H. Zhang, R. Raghavendra, K. B. Raja, N. Damer, and C. Busch, "Can GAN generated morphs threaten face recognition systems equally as landmark based morphs?—Vulnerability and detection," in *Proc. 8th Int. Workshop Biometrics Forensics (IWBF)*, Porto, Portugal, 2020, pp. 1–16.
[15] J. T. Andrews, T. Tanay, and L. D. Griffin, "Multiple-identity image attacks against face-based identity verification," 2019. [Online]. Available: arXiv:1906.08507.
[16] C. Seibold, A. Hilsmann, and P. Eisert, "Style your face morph and improve your face morphing attack detector," in *Proc. Int. Conf. Biometrics Spec. Interest Group (BIOSIG)*, Darmstadt, Germany, 2019, pp. 1–6.
[17] S. Mosaddegh, L. Simon, and F. Jurie, "Photorealistic face de-identification by aggregating donors' face components," in *Proc. Asian Conf. Comput. Vis. (ACCV)*, 2014, pp. 159–174.
[18] C. Kraetzer and J. Dittmann, "Steganography by synthesis: Can commonplace image manipulations like face morphing create plausible steganographic channels?" in *Proc. 13th Int. Conf. Availability Rel. Security (ARES)*, 2018, p. 11.
[19] P. Majumdar, A. Agarwal, R. Singh, and M. Vatsa, "Evading face recognition via partial tampering of faces," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Long Beach, CA, USA, 2019, pp. 1–10.
[20] N. Damer *et al.*, "Detecting face morphing attacks by analyzing the directed distances of facial landmarks shifts," in *Proc. German Conf. Pattern Recognit. (GCPR)*, 2018, pp. 518–534.
[21] M. Ferrara, A. Franco, and D. Maltoni, "Face demorphing," *IEEE Trans. Inf. Forensics Security*, vol. 13, pp. 1008–1017, 2018.
[22] 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.
[23] U. Scherhag, C. Rathgeb, J. Merkle, and C. Busch, "Deep face representations for differential morphing attack detection," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 3625–3639, 2020.
[24] R. Raghavendra, K. B. Raja, and C. Busch, "Detecting morphed face images," in *Proc. IEEE 8th Int. Conf. Biometrics Theory Appl. Syst. (BTAS)*, Niagara Falls, NY, USA, 2016, pp. 1–7.
[25] L. Zhang, F. Peng, and M. Long, "Face morphing detection using fourier spectrum of sensor pattern noise," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, San Diego, CA, USA, 2018, pp. 1–6.
[26] U. Scherhag, L. Debiasi, C. Rathgeb, C. Busch, and A. Uhl, "Detection of face morphing attacks based on PRNU analysis," *IEEE Trans. Biometrics Behav. Identity Sci.*, vol. 1, no. 4, pp. 302–317, Oct. 2019.
[27] T. Neubert, "Face morphing detection: An approach based on image degradation analysis," in *Proc. Int. Workshop Digit. Watermarking (IWDW)*, 2017, pp. 93–106.
[28] T. Neubert, C. Kraetzer, and J. Dittmann, "A face morphing detection concept with a frequency and a spatial domain feature space for images on eMRTD," in *Proc. ACM Workshop Inf. Hiding Multimedia Security (IH&MMSec)*, 2019, pp. 95–100.
[29] 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. Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Honolulu, HI, USA, 2017, pp. 1822–1830.
[30] C. Seibold, W. Samek, A. Hilsmann, and P. Eisert, "Detection of face morphing attacks by deep learning," in *Proc. Int. Workshop Digit. Watermarking (IWDW)*, 2017, pp. 107–120.
[31] C. Seibold, W. Samek, A. Hilsmann, and P. Eisert, "Accurate and robust neural networks for face morphing attack detection," *J. Inf. Security Appl.*, vol. 53, Aug. 2020, Art. no. 102526.
[32] C. Seibold, A. Hilsmann, and P. Eisert, "Reflection analysis for face morphing attack detection," in *Proc. Eur. Signal Process. Conf. (EUSIPCO)*, Rome, Italy, 2018, pp. 1022–1026.
[33] *Face Morph Using OpenCV*. Accessed: Aug. 1, 2019. [Online]. Available: http://www.learnopencv.com/face-morph-using-opencv-cpp-python
[34] D. E. King, "Dlib-ml: A machine learning toolkit," *J. Mach. Learn. Res.*, vol. 10, pp. 1755–1758, Jul. 2009.
[35] P. Perez, M. Gangnet, and A. Blake, "Poisson image editing," *ACM Trans. Graph.*, vol. 22, no. 3, pp. 313–318, 2003.
[36] *Megvii Face++ Compare API*. Accessed: Aug. 5, 2019. [Online]. Available: https://www.faceplusplus.com/face-comparing
[37] *Microsoft Azure Face Verification API*. Accessed: Sep. 11, 2019. [Online]. Available: https://azure.microsoft.com/en-us/services/cognitive-services/face
[38] *Neurotechnology VeriLook Face Verification SDK*. Accessed: Oct. 5, 2019. [Online]. Available: https://www.neurotechnology.com/face-verification.html
[39] *Cognitec FaceVACS SDK Version 9.4.2*. Accessed: Feb. 21, 2020. [Online]. Available: https://www.cognitec.com/facevacs-technology.html
[40] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," in *Proc. Brit. Mach. Vis. Conf. (BMVC)*, vol. 1, 2015, p. 6.
[41] Q. Cao, L. Shen, W. Xie, O. M. Parkhi, and A. Zisserman, "VGGFace2: A dataset for recognising faces across pose and age," in *Proc. Int. Conf. Autom. Face Gesture Recognit. (FG)*, 2018, pp. 67–74.
[42] X. Wu, R. He, Z. Sun, and T. Tan, "A light CNN for deep face representation with noisy labels," *IEEE Trans. Inf. Forensics Security*, vol. 13, pp. 2884–2896, 2018.
[43] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "ArcFace: Additive angular margin loss for deep face recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Long Beach, CA, USA, 2019, pp. 4690–4699.
[44] G. B. Huang, M. Mattar, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," Univ. Massachusetts, Amherst, MA, USA, Rep. 07-49, 2007.
[45] A. Bharati, R. Singh, M. Vatsa, and K. W. Bowyer, "Detecting facial retouching using supervised deep learning," *IEEE Trans. Inf. Forensics Security*, vol. 11, pp. 1903–1913, 2016.
[46] U. Scherhag *et al.*, "Biometric systems under morphing attacks: Assessment of morphing techniques and vulnerability reporting," in *Proc. Int. Conf. Biometrics Spec. Interest Group (BIOSIG)*, Darmstadt, Germany, 2017, pp. 1–7.
[47] *Information Technology—Biometric Presentation Attack Detection—Part 3: Testing and Reporting*, Standard ISO/IEC 30107-3, 2017.
[48] Q. Huynh-Thu and M. Ghanbari, "Scope of validity of PSNR in image/video quality assessment," *Electron. Lett.*, vol. 44, no. 13, pp. 800–801, Jun. 2008.
[49] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
[50] *Tecent Crowdsourcing*. Accessed: May 7, 2020. [Online]. Available: https://wj.qq.com/answer_group.html