IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 9 June 2022, accepted 26 June 2022, date of publication 5 July 2022, date of current version 15 July 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3188668

RESEARCH ARTICLE

Generation and Detection of Face Morphing Attacks

MUHAMMAD HAMZA[1], SAMABIA TEHSIN[1], HANEN KARAMTI[2], AND NORAH SALEH ALGHAMDI[2]

[1]Department of Computer Science, Bahria University, Islamabad 44000, Pakistan
[2]Department of Computer Sciences, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, P.O.Box 84428, Riyadh 11671, Saudi Arabia

Corresponding author: Muhammad Hamza (01-249201-005@student.bahria.edu.pk)

This work was supported by the Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2022R192), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

ABSTRACT Failure of facial recognition and authentication system may lead to several unlawful activities. The current facial recognition systems are vulnerable to different biometric attacks. This research focuses on morphing attack detection. This research proposes a robust detection mechanism that can deal with variation in age, illumination, eye and head gears. A deep learning based feature extractor along with a classifier is adopted. Additionally, image enhancement and feature combination are proposed to augment the detection results. A versatile dataset is also developed that contains Morph-2 and Morph-3 images, created by sophisticated tools with manual intervention. Morph-3 images can give more realistic appearance and hence difficult to detect. Moreover, Morph-3 images are not considered in the literature before. Professional morphing software depicts more realistic morph attack scenario as compared to the morphs generated in the previous work from free programs and code scripts. Eight face databases are used for creation of morphs to encompass the variation. These databases are Celebrity2000, Extended Yale, FEI, FGNET, GT-DB, MULTI-PIE, FERET and FRLL. Results are investigated using multiple experimental setups and it is concluded that the proposed methodology gives promising results.

INDEX TERMS Morphing attack detection, fraudulent and forged digital identity documents, biometrics, facial recognition, access control.

I. INTRODUCTION

The world has become a global village with the introduction of modern technologies. Vast distances have now shrunk due to the availability of fast means of conveyance like airplanes, trains, ships and buses. These abundant conveyance options have given rise to a significant increase in the travelling population. With such a large number of mobile population, manual verification of travelling documents and facial authentication is not possible. Therefore, an automatic border control system is used for authentication and approval of passports [1]. Border control systems are now deployed in more than 180 airports around the world [2]. This automatic system uses face recognition system [1] to compare the live captured images of the traveller with the image of traveller that is stored in the travel agency's database system or in the form of passport or any other type of machine readable travel documents (MRTD) [3]. After face recognition system approves that both the live captured image of the traveller and the image on the passport are same, the traveller is granted travelling authorization [3]. In this way an automatic border control system [1] is implemented to deal with enormous travelling population.

Availability of image manipulation technology has also enabled the culprits to use this technology for fraudulent activities. In order to gain legal entry permission into foreign countries for unlawful activities many criminals are utilizing a technology called face morphing to trick the face recognition system. Image morphing has been around since 1980s [4] but now with the ease and abundance in availability

The associate editor coordinating the review of this manuscript and approving it for publication was Zahid Akhtar.

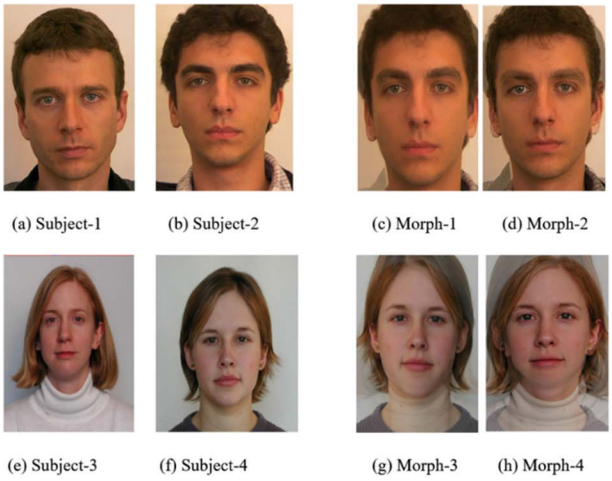

FIGURE 1. Example of morphed images (Morphs (c) and (d) are created from Subjects (a) and (b) while Morphs (g) and (h) are created from Subjects (e) and (f) using tool FotoMorph.

of software and hardware technology to the general public, creating morphed images for fraudulent activities is easier than ever. In face morphing technology the image of two or more persons can be combined or merged together in such a way that it resembles the participants of the morphed image and the facial recognition system approves the morphed image as the original image of the applicant [5]. Furthermore, the ratio of merger of different persons in the morphed image is controlled in such a way that human inspection is also extremely difficult. Example of morphed images is shown in Fig. 1 in which two separate morphed images are created from two subjects that are resembling both subjects. By using image morphing a wanted criminal who is barred from travelling can easily morph his facial image with the facial image of an accomplice and successfully acquire travel permission in an unauthorized country [5].

In order to alleviate this vulnerability of the face recognition systems several methods have been proposed in the past. These methods are categorized based on their methodology of morph detection. Single image morph attack detection and differential morph detection [5]. This study introduces a general morph attack detection model that would be able to classify a wide variety of images. Images of different types and varying features (age, expression, posture, illumination, gender, race, hair style, facial hair, head gear, eye wear) are used as different type of ID cards have different back ground colours and specifications.

Generalized images are considered to broaden the checkpoint locations by including border control checkpoints, train stations, bus stations, hotel check ins, security institutions, police stations and banks. As the quality of the input images vary in different parts of the world based on the technology and infrastructure limitations related to both hardware and software, therefore a generalized robust model is presented that would be able to be deployed in any location irrespective of the local technology and data for producing the best possible accurate results with what is available instead of overall rejection to classify morph images. Presented model is designed to be a robust, adaptable, generalized and accurate model that can be deployed for facilitating the process of morph attack detection to ensure that fraudulent activities will not go unchecked.

In this study, a unique and diverse morphed database is created manually using professional software. Morphed images created from two and three subjects are included in this work. A modern morph detection model based on deep learning based feature extractor and a machine learning based classifier is adopted to be trained and tested on the created database. Six different types of experiments are performed in this study to analyse the performance of the proposed morph attack detection model on different types of morphed and original images.

First experiment is performed by training and testing the model on the created database to analyse the impact of different feature combination techniques. Second experiment is performed to analyse the performance of different types of famous classifiers like SVM, Naive Bayes, Logistic Regression, XG Boost and majority voting. Third experiment is performed to evaluate the performance of the model after applying image enhancement techniques. Different levels of brightness and contrast are used to analyse the impact on the performance of the model in detecting morph attacks. In the fourth experiment, replication of the previous state of the art work [5] has been done to compare the performance of the model created in this study and the state of the art [5].

In the fifth experiment, the vulnerability of the morph detection model to the created diverse and unique database is elaborated by using this database as testing database to illustrate the issues and performance depreciation in training the model on one database created from single morphing tool as proposed in the previous state of the art work [5]. In the sixth experiment, the model is trained on the morphed database that is created in this research and tested on the database from the previous state of the art work [5] to illustrate that the created morphed database yields better results by training the model better as compared to previous work [5] that utilized single database and single morphing tool for creation of training database.

The following are the main contributions of this research work:

- Proving vulnerability of previous morph attack detection models to morph-3 images.
- Simulation of a realistic morph attack scenario by creation of a morph database with extreme variation in features. Creation and inclusion of morph-2 & 3 images in the database using professional morphing tool.
- Creation of a robust morph detection model using deep learning.
- Proving impact of image enhancement on morph detection.

The organization of this study is as follows. Section-2 describes the state of the art work that has been done in the field of morph attack generation and detection. Furthermore,

techniques of morph attack detection from recent research work are discussed and compared in detail along with imperative analysis. Section-3 describes the steps that are involved in the creation of the morphed databases that have been used in this research work. In Section-4, detailed information about the adopted methodology and utilized modules is discussed. Section-5 contains the presentation of results, advantages, disadvantages and limitations of different techniques. Similarly, comparative analysis of model's performance on different databases is also elaborated. Section-6 provides the conclusion drawn after the completion of this study.

II. RELATED WORK

The matter of morph attack detection has enticed a significant amount of attention from the research community in the recent years. Different studies have been conducted in this field and different approaches have been applied to effectively detect morph attacks. Variety of face databases are utilized for creation of morph image databases as sufficient morph images are not easily available for research purposes.

A. IMAGE MORPHING

In the late 1980s and 1990s, face image morphing was used for introducing visual effects in movies and animations [4]. In image morphing, features of two face images were compared and spatial relationship between the features was determined for combination. After the alignment of both images through warping, colour interpolation is applied to generate a new image. The new image is a mix of both input images [7]. The varying of warping and colour interpolation was referred to as transition control.

Different techniques of morphing like mesh warping [8], field morphing [9] and radial basis morphing have been used for creation of morphed faces. In mesh warping, meshes are used to link different landmarks or control points on the images of two subjects. The source image is morphed into the target image by freezing some parts of the image while warping others through control points. In field morphing, a pair of lines were used to map corresponding features between two images. Different points on the images were mapped based on their distance from the respective line. In radial basis functions, the features on the image were considered to be represented by a set of points. Different lines and curves that formulated a mesh on an image were considered as a set of points. Mapping was done from the two surfaces that were considered on both images.

B. METHODS OF MORPH ATTACK DETECTION

There are two basic types of morph attack detection (MAD) methods that are prevalent in the literature.

1) SINGLE IMAGE MAD METHOD

In these types of methods only the morphed image is analysed for presence of morphing attempt. Morphing an image leaves some artifacts in the image that are traced for detection of morph. Texture descriptors like binary statistical image features (BSIF) [10] are utilized for texture classification. Furthermore, ghosting or shading artifacts are also detected in such images. Similarly, deep neural network can also be trained to detect such artifacts as long as the training data contain variety of images [5].

2) DIFFERENTIAL MAD METHOD

In these types of methods both the potential morph and the live captured images are analysed, compared and processed to detect morphing attempt [5]. Feature vectors from both images are extracted for comparison [1], [5], [11]–[15]. Demorphing process is also done in some of these techniques to extract the identity of the accomplice by subtracting the live captured image from the morphed image [1], [2], [11].

C. STATE OF THE ART RESEARCH WORK

Significant amount of work has been done in the field of morph attack detection. Different tools, preprocessing methods and databases are used for morph image creation. Overview of related literature work is shown in Table 1.

Several research studies in the area of morph attack detection have reported good detection results but these studies are tested on the datasets with limited variations and lack real world scenarios. Features like variation in age, race, facial hair, head gear, eye wear, illumination, expression and posture are underutilized or not utilized at all in many studies [1], [2], [5], [11], [12], [16]. Similarly limited number of databases are utilized for morph attack detection. Furthermore, fixed contribution weights (attacker and accomplice) are used in creation of morphed images instead of random contribution weights from attacker's and accomplice's images [5]. Images in which head gear and eye wear are present, resulted in incorrect classification of original images as morph images [5]. The quality of live captured images is very high in previous studies [1], [2], [5], [12], [16], that is not applicable for all checkpoints due to variation of available resources.

Existing morph detection datasets have another very major problem. These datasets have considered the morph of two persons only (morph-2 images), leading to easy morph detection [1], [2], [5], [12], [16]. Furthermore, low quality programming script based morphing tools like FaceMorpher, OpenCV, FaceFusion [3] are used that generate morphed images automatically and majority of created morphed images are easily detectable through visual inspection by a human. Therefore, these techniques are rarely used by criminals, hence not depicting the real world scenarios. Methods tested on the datasets with the discussed limitations, can give very high detection rates but will not be very successful in real scenarios. Morphs of high quality and high variance are still very difficult to classify properly [5].

Several approaches with different benchmarks are proposed in the literature. Previous work has succeeded in achieving high accuracy but the results were achieved on databases having limited features. The accuracy of results

TABLE 1. Overview of the related literature work on morph attack detection.

Publication	Author Name	Methodology	Results	Limitations
Differential MAD Methods				
FACE DEMORPHING [2]	Matteo Ferrara, Annalisa Franco and Davide Maltoni.	Extraction of accomplice's image from morphed image through corresponding points.	Morphed image acceptance rate by facial recognition system dropped from 66.4% to 6.1%.	In advance knowledge about morphing factor must be known to set specific range of demorphing factor to ensure proper extraction of accomplice's image. Furthermore, manual removal of artifacts was required in many cases. Only morph-2 (merging of two persons only) images were used. Limited morphing tools.
FD-GAN Face De-Morphing Generative Adversarial Network for Restoring Accomplice's Facial Image [1]	Fei Ping, Le-Bing Zhang and Min Long.	Extraction of accomplice's image through dual network architecture along with two levels of restoration losses.	Accuracy of restoration increased from 49.82% to 87.5% in simple cases without expression and occlusion. Accuracy increased from 46.91% to 64.9% in case of presence of expression and occlusion.	Accuracy of restoration of images suffered in case of images in which subjects were displaying expression, posture or there was some occlusion in the image. Only morph-2 images were used. Limited morphing tools.
Border Control Morphing Attack Detection with a Convolutional Neural Network De-Morphing Approach [11]	David Ortega-Decamp, Cristina Conde, Daniel Palacios-Alonso and Enriqu Cabello.	Convolutional neural network de-morphing approach for morph attack detection.	Accuracy of 98.1% to 98.7% was achieved in morph attack detection. D-EER of 0.78% to 20.7% was achieved.	Expressions, posture, illumination and variety in databases were not considered in this study. Only morph-2 images were used. Limited morphing tools.
Deep Face Representations for Differential Morphing Attack Detection [5]	Ulrich Scherhag, Christian Rathgeb, Johannes Merkle and Christian Busch.	Deep learning neural network was used for feature extraction and machine learning classifier was used for classification of morphed images of varying quality created from four tools.	D-EER of 1% to 7% was achieved while other techniques achieved D-EER of 2.9% to 51%.	Morphs of high quality and morphs that bear high resemblance to live captured images were not properly classified. Similarly, images having variation in facial expression, headgear, eye wear, variations in illumination and focus also contributed to misclassification. Lack of realistic database depicting real world scenario. Only morph-2 images were used. Low quality morphing tools.
Single Image MAD Methods				
Accurate and Robust Neural Networks for Face Morphing Attack Detection [12]	Clemens Seibold, Wojciech Samak, Anna Hilsmann and Peter Eisert.	Four training methods to teach specific data to the network and Layer wise level propagation was used for analyzing and controlling the decision-making process of neural networks.	D-EER of 2.8% to 3.1% was acquired as compared to D-EER of 15% to 31% acquired in previous work. Robustness of morph attack detection increased from 20% to 87%.	Training and testing were done only on specific selected images with neutral expression, illumination and pose. Images with variety in pose, illumination and expression were not considered and therefore the acquired accuracy was only for such images. Testing is required for images that exhibit variety in expression, pose, illumination and other features like eye wear, head gear and age. Only morph-2 images were used. Limited morphing tools.
Detecting Morphed Face Images [15]	Raghavendra R, Raja KB and Busch C.	Features from input image were extracted using BSIF filters and SVM classifier was used to classify the image as morph or bonafide.	ACER (Average classification error rate) dropped from 37.55% to 1.73% as compared to previous work.	Uniform pose, illumination and expression were present in all images and therefore the acquired accuracy is only for such images. Testing is required for images that exhibit variety in expression, pose, illumination and other features like head gear and age. Only morph-2 images were used. Limited morphing tools.
Transferrable Deep-CNN features for detecting digital and print-scanned morphed face images [13]	Raghavendra R, Raja KB, Venkatesh S, and Busch C.	Pretrained deep learning models of AlexNet and VGG19 were used for feature extraction from input images and P-CRC classifier was used to classify the images as morph or bonafide.	Results of proposed model were significantly better than previous BSIF model. D-EER of proposed model was 15.05% as compared to previous D-EER of 26.70%.	Normal face dataset with neutral pose, expression and illumination were utilized therefore accuracy of the model is not proven on datasets having variety in features. Testing is required for images that exhibit variety in expression, pose, illumination and other features like head gear and age. Only morph-2 images were used. Limited morphing tools.
Detecting Morphing Face Attacks Using Residual Noise from Deep Multi-Scale Context Aggregation [14]	Sushma Venkatesh, Raghavendra Ramachandra, Kiran Raja, Luuk Spreeuwers and Raymond Veldhuis.	Detection of morph images by learning from features of residual noise in morph and bonafide images by using deep learning and P-CRC classifier.	D-EER of proposed method was 2.6% to 8%, which was significantly better as compared to D-EER of 3.83% to 42.2% in previous models. Similarly, at various settings of APCER the BPCER value was also significantly better than conventional model. Computational cost also improved by a factor of 4.	Normal face datasets with neutral pose, expression and illumination were utilized therefore accuracy of the model is not proven on datasets having variety in features. Testing is required for images that exhibit variety in expression, pose, illumination and other features like head gear and age. Only morph-2 images were used. Limited morphing tools.

declined whenever there were some additional features in the database related to facial hair, eye wear, cosmetics, hair style, expression and posture [1], [5].

Similarly, the age of subjects considered for analysis does not differ more than two years in the previous studies [1], [2], [5], [11], [16], but in reality, the electronic machine readable document (MRTD) is valid for up to ten years as specified in the International Civil Aviation Authority (ICAO) protocols about travel documents [17]. So, the studies should contain subjects having age difference between 5 to 9 years as the person presenting the travel document at the border control system could be incorrectly classified as an attacker if his

TABLE 2. Details and features of images in all the source databases.

Databases	Total images	Number of subjects	Images per subject	Variation in features
Celebrity2000 [18]	163,446	2000	Variable	Age (16 – 62), illumination, expression, posture, eye wear, head gear, hair style, race, gender, facial hair.
Extended Yale B [19], [6]	16,128	28	585	64 Illumination conditions with grayscale, posture, expression, race, gender.
FEI (Faculty of industrial engineering) [20], [21]	2800	200	14	Posture, expression, illumination, eye wear, race, gender.
FGNET (Face and gesture recognition working group) [22]	1002	82	Variable	Age (0-69), illumination, colour / grayscale, expression, posture, hair style, eye wear, facial hair, gender.
GT-DB (Georgia tech face database) [6]	750	50	15	Expression, illumination, posture, eye wear, gender.
CMU MULTI-PIE (Carnegie mellon university pose, illumination and expression) [23], [24]	130,000	250	520	19 illumination conditions, gender, race, eye wear, expression, posture.
FERET (Face recognition technology) [25], [26]	14,126	1199	Variable	Colour / grayscale, illumination, expression, posture, gender, race, facial hair, hair style.
FRLL (Face research lab London) [27], [28]	204	102	2	Gender, race, facial hair, hair style.

age difference is not properly taken into consideration by the facial recognition system. To cater these limitations a morph attack detection model is required that is accurate, adaptable and generalized. Moreover, the detection model should be trained and tested on the data with all the variations.

III. MATERIALS AND DATA

For implementing this proposed model, total eight face image database are used in this study and an additional FRGC morph database has been used for evaluation during the concluding experiments. Subsets of these databases are used that are most suitable for this proposed model. Images that exhibit wide variety of features are selected. Databases are divided into different sections for training and testing. Famous morphing tools are used and some of them were also used in the previously done research work [5]. Details and features of the images in all the source databases are represented in Table 2.

A. CREATION OF MORPHED DATABASE

A specialized morphed image database has been created for this study by following the proposition of a renowned researcher Andrew Ng from an interview [42]. Andrew Ng proposed that a data centric approach should be adopted to improve the performance of deep learning based models as the architecture of the models is sufficiently tweaked and improved to achieve maximum results. Therefore, in this research the main focus is allocated towards creating a quality database to facilitate the model to learn and perform better.

TABLE 3. Details and features of images in the morphed database.

Source databases	Quantity of morph-3 images	Quantity of morph-2 images	Normal images
Celebrity2000	827	148	2777
Extended Yale	21	17	97
FEI	63	23	235
FGNET	34	24	150
GTDB	18	18	90
PIE	82	39	324
FERET [29]	0	1058	1058
FRLL [29]	0	570	570
FRGC [29]	0	1928	1928
Total	1045	3825	7229
Grand total	12099		

This database has been created manually using different morphing software. In this database, morphed images are created by utilizing the images from the first eight databases that are mentioned in Table 2. Morphs created from FERET, FRGC and FRLL databases are borrowed from other research work and those images were created using programming algorithms of OpenCV and FaceMorpher.

Two types of morphed images are created in this research:

1. Morph-3, Images that are created by mixing facial images of three different persons.
2. Morph-2, Images that are created by mixing facial images of two persons only.

The detailed information about the morphed database is displayed in Table 3.

Total 1045 morph-3 images are created manually using high quality software [3] FantaMorph and 269 morph-2 images are created using FotoMorph. While 570 morph-2 images (borrowed) were created from database FRLL and 1058 morph-2 images (borrowed) were created from FERET using programming algorithms of OpenCV and FaceMorpher. FERET and FRGC databases were also used in the training and testing of model in the previous work [5]. Normal images in the morphed database contain the source images that were used to create morphed images and normal images also contain live captured images. A portion of normal images is used as original images and the second portion as live captured images during training and testing. Number of live captured images vary for different subjects. Maximum amount of live captured images is used for each subject to introduce variation in the training and testing in order to bring this study as close to a realistic scenario as possible.

Databases have different qualities and different types of features therefore images from the same database were used for creating morphed images. Each database has been divided into training and testing sets. Source original images

and created morphed images from those source images are placed in their respective training and testing sections. Approximately twenty percent images from each database are used for testing while eighty percent images are used for training of model. In some databases twenty percent source images are used for creation of more than twenty percent morphed images so in that case all the morphed images created from source images are placed in respective test set in order to avoid the appearance of training images in testing set.

Mixing of images from different databases for creating morphs has not been done as different features like variation in illumination causes artifacts in morphs and reduces the quality of morphs. Images of subjects having similar facial structure and features yield better resulting morphs, so similar facial images are morphed. The detailed information about the images used in training and testing is displayed in Table 4.

TABLE 4. Details and features of images used in training and testing of model.

Databases	Morph-3 train images	Morph-3 test images	Morph-2 train images	Morph-2 test images	Qty of normal train images	Quantity of normal test images
Celebrity2000	660	167	110	38	3265	836
Extended Yale	17	4	15	2	40	6
FEI	50	13	18	5	114	36
FGNET	27	7	20	4	58	20
GT-DB	14	4	15	3	42	12
MULTI-PIE	66	16	33	6	166	46
FERET	0	0	846	212	1140	286
FRLL	0	0	368	202	166	40
Total	834	211	1425	472	4991	1282
Grand total	9215					

B. QUALITY ASSESSMENT OF MORPH-3 IMAGES

Morph-3 image has several advantages over the morph-2 image as morph-3 exhibit a more realistic presentation of a human face because the third image acts as a transitioning supporter for ensuring smooth transition between the images. Colour tone is also properly normalized by combining three images and sharp or abrupt differences and artifacts between different images are reduced significantly. Morph-3 images are better at tricking the human inspector because of their additional realistic and natural look. As morph-3 images are created and introduced in this domain of morph attack detection for the first time therefore research work estimating the quality of morph-3 images is not available.

In order to prove scientifically that morph-3 images have superior quality in terms of facial features specially for facial recognition system, a face quality estimation and assessment technique from the state of the art works will be utilized. Research works on estimation of face image quality based on stochastic embedding [39], [40] proposed a technique of SER-FIQ (Stochastic embedding robustness - Face image quality). SER-FIQ is designed for face quality estimation specially for ensuring good performance of facial images during processing in different facial recognition systems. This technique outperformed previously proposed techniques in majority of cases [39]. The quality of image is assessed based on the level of variation in the stochastic embedding. High variation in stochastic embedding set indicates a low quality image. To assess and estimate the quality of morph-3 images that are created in this research, SER-FIQ technique will be used to compare the quality of morph-2 and morph-3 images.

The implementation of SER FIQ has been borrowed from [41] and it is implemented without any changes. Total 30 morph-3 and 30 morph-2 images are selected. These 30 image sets each contain one morph-2 and one respective morph-3. These sets are randomly selected from the created database. Two source images in both morph-2 and morph-3 sets are same while morph-3 have an additional third source image that further adds to its quality. The source images are those images that were used to create these morphed images. It was observed during the experiment that morph-3 image had higher quality score in 27 out of 30 cases of image sets while comparing between morph-2 and morph-3. In two cases of image sets, morph-2 images had higher quality and in one case of image set, morph-2 and morph-3 had the same quality. Based on the majority of cases, it was concluded that morph-3 images have higher quality in terms of quality of facial features. Based on the higher quality of morph-3 images, they have a higher probability of getting matched with their respective live captured images and hence successfully deceive the facial recognition system which is the main objective of creating a morph. Similarly, in the domain of morphing attack detection, after extensive testing in experiment-5 (Section-5), it was observed that around 50% morphed images were incorrectly classified when the testing database contained 80% morph-3 images.

This further illustrates that morph-3 images can easily deceive facial recognition systems that are trained on morph-2 images only. It is also important to highlight that the criminals creating the morphed images are not limited by the standard rules to create only morph-2 images only as their main aim is to deceive the facial recognition system by using any available tools and methods at their disposal.

C. SOFTWARE TOOLS USED

The following software are utilized for creating this morphed database.

1. FotoMorph [30], This software has been used for manually creating morphed images from two facial images using 20 to 45 landmarks.
2. FantaMorph [31], This software has been used for manually creating morphed images from three facial images using 127 landmarks.
3. OpenCV [32], This software has been used to create morphed images from two facial images automatically through program script code using 68 landmarks.

FIGURE 2. Morphing process in tool FotoMorph.

FIGURE 3. Morphing process in tool FantaMorph.

4. FaceMorpher [33], This software has been used to create morphed images from two facial images automatically through program script code using 77 landmarks.

Following steps are involved in the creation of morphed image from two subjects in tool FotoMorph:

1) FOTOMORPH

1. Creating corresponding points (landmarks) as shown in Fig. 2.
2. Assigning similar corresponding points to the similar regions of face on both images.
3. Selecting the most suitable frame from the animation.
4. Morphed images were created to include 50% contribution from both the images. It was ensured that the resulting image looked realistic and had minimum artifacts.
5. If the resulting image was unrealistic or had significant artifacts then the morphing contribution from images was varied to acquire the best possible morphed image. Resulting final image is shown in Fig. 2.

Following steps are involved in the creation of morphed image from three subjects in tool FantaMorph:

2) FANTAMORPH

1. Assigning similar corresponding points or face region locators to the similar regions of face on all the three images as shown in Fig. 3.
2. Selecting appropriate contribution of features and shape from all three images.
3. Morphed images were created to include 33% contribution from all the images but it was ensured that the resulting image looked realistic and had minimum artifacts.
4. If the resulting image was unrealistic or had significant artifacts then the morphing contribution from images was varied to acquire the best possible morphed image. Resulting final image is shown at the bottom in Fig. 3.

Samples of morph-3 and morph-2 images created by FantaMorph and FotoMorph respectively, are presented in Fig. 4.

FIGURE 4. Morph-3 images (top row) created manually from three subjects using tool FantaMorph and morph-2 images (bottom row) created manually from two subjects using tool FotoMorph.

IV. METHODOLOGY

After extensive analysis of the literature in Section-2, a deep learning based approach was adopted in this research as deep learning based models outperformed the other morph detection models by a significant margin. Therefore, in order to tackle the issue of morph detection, a combination of deep learning based feature extractor and a machine learning based classifier is utilized. Model with the selected combination is essential for developing an accurate and robust morph attack detection system as this type of model is capable to tackle different scenarios of morph attack detection as indicated by the previous state of the art works [5], [13]–[15]. Furthermore, most suitable live captured image is automatically selected to combine its features with the potential morph image in order to facilitate the model to only focus on discrepancies between the images.

The main morph attack detection model is shown in Fig. 5. Implementations of MTCNN (Multi task cascaded convolutional neural networks) and cosine distance calculation are borrowed from previous work [34]. This model is based on DeepFace model FaceNet (a pre-built deep learning neural network) and it is used for training and testing during this study. Some variations in hyper parameters are made in different experiments and these variations are mentioned in respective experiments. FaceNet has been selected because of its efficient performance in the presence of different poses in the images [38].

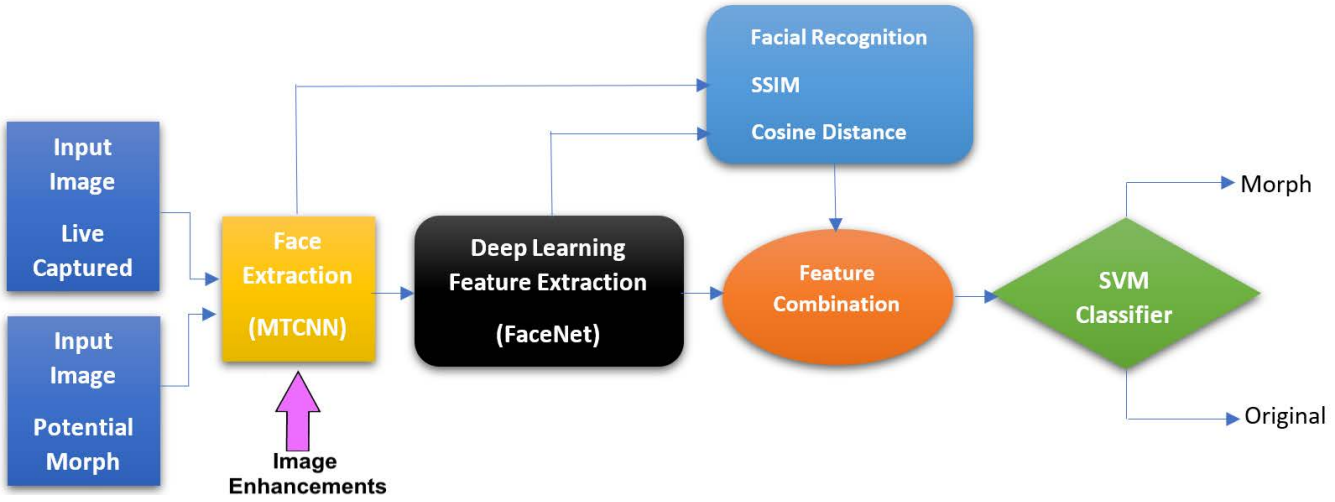

FIGURE 5. Morph attack detection model design.

The following steps are performed in the implementation phase of this study:

1) Extraction of faces from input images (morph and live captured) using MTCNN. Extracted images are converted to size of 160 × 160 × 3 for FaceNet.
2) Features are extracted from the input images using FaceNet.
3) Features from input live captured and potential morph image are combined using subtraction, addition, or concatenation. Features of potential morph and its respective live captured image are combined after verification from the facial recognition system based on structural similarity index measure (SSIM) [35] and cosine distance.
4) The combined features are forwarded to the machine learning based classifier SVM (support vector machine) for classification of input images as morphed or original images.

A. FEATURE COMBINATION

In order to combine the features of live captured and potential morphed images, the following steps are performed as represented in Fig. 6:

1) Cosine distance and SSIM score are calculated between the potential morph and the respective live captured images.
2) Cosine distance utilizes the extracted feature vectors from the input images for calculating similarity score while SSIM uses the extracted input face images for calculation of similarity score.
3) Both the cosine distance and SSIM scores are combined by averaging.
4) Features of potential morphed image are combined with the live captured image based on lowest combined cosine and SSIM similarity score.

The basic explanation of morph attack detection model has been provided in Algorithm 1.

B. ALGORITHM-1

```
procedure MORPH ATTACK DETECTION (database)
    Extracted faces = MTCNN (database)
    model = load (FaceNet, weights)
    for faces in database do
        features = model (extract features (Extracted
        faces))
    end for
    while potential morph index < total potential morphs: do
        while live image index < total live images: do
                cos = cos (features potential morph,
                features live image)
                ssim = 1 - ssim (potential morph, live
                image)
                score list = ((cos + ssim) / 2)
        end while
        smallest score = smallest value (score list)
        index = score list.index (smallest score)
        combined = (features pot.morph, features live
        (index))
        trainX = normalize (combined, L-2)
        classifier (trainX, trainY)
        predictedY = classifier.predict(testX)
        (Accuracy, DEER) = evaluation metrics (testY,
        predictedY)
    end while
end procedure
```

Live captured image with the best possible posture, illumination and other important features is selected for feature combination. If the potential morphed image is actually a morph, then the only distinguishing factor in the morphed image and the live captured image is the evidence of morphing attempt or leftover digital footprint in the morphed image's feature vector.

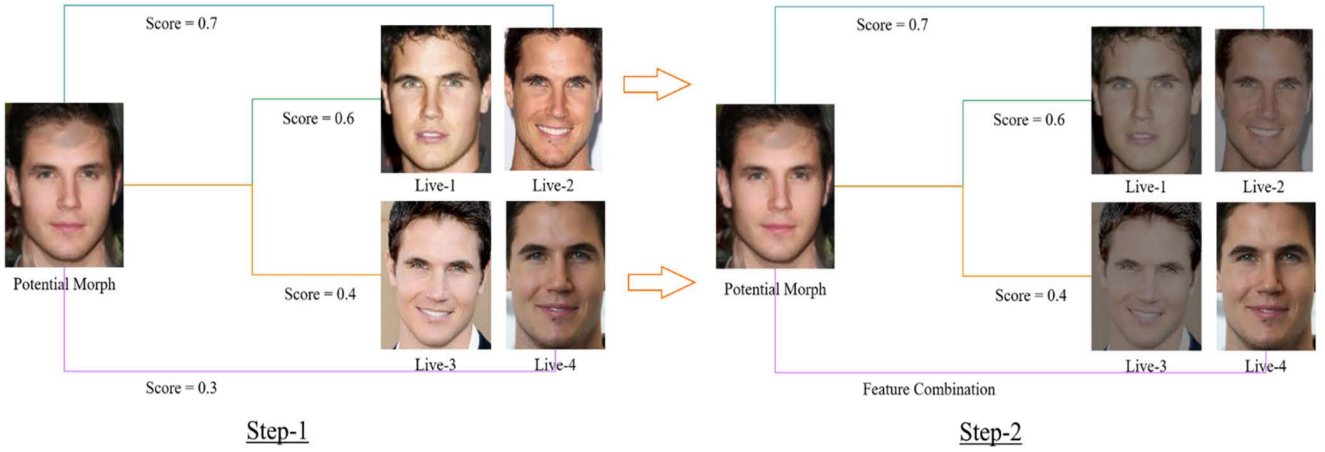

FIGURE 6. Feature combination process.

C. EVALUATION METRICS

For evaluation of the model all the relevant metrics are adopted. Fewer metrics may lead to ignore various evaluation aspects while highlighting some. Such rigorous evaluation has not been done before. General detection accuracy is reported and metrics for standard biometric systems according to the document ISO IEC 30107-3 [36] like DEER (Detection equal error rate), APCER (Proportion of attack presentations incorrectly classified as bona fide presentations), BPCER (Proportion of bonafide presentations incorrectly classified as attack presentations) and ACER (Average classification error rate) are used. The accuracy of the model is reported using the Detection equal error rate (DEER) and it is defined as the decision threshold where APCER is as high as BPCER. Similarly, APCER-10 and BPCER-10 are reported in this study. APCER-10 is defined as the point where BPCER = 10% and BPCER-10 is the point where APCER = 10%.

V. RESULTS

Six different types of experiments are performed in this study to analyse the performance of the proposed morph attack detection model on different types of morphed and original images. The following experiments are performed:

A. EXPERIMENT-1: PERFORMANCE OF THE MODEL USING DIFFERENT FEATURE COMBINATION TECHNIQUES

In this experiment the first eight databases from Table 3 are utilized excluding the ninth database FRGC. Total 2259 morphed images and 4991 source normal images are used for training and 683 morphed, 1282 normal images (original and live) are used for testing. One or more live captured images per subject are a part of 1282 normal images which are used for comparison and authentication of morphed and original images. Quantities of used images during training and testing are mentioned in Table 4. Details about the experiment's performance on SVM with linear function are represented in Table 5 and errors in each database are represented in Table 6. In Table 6, M represents the proportion of incorrectly classified morphed images and O represents proportion of incorrectly classified original images. DEER curves and confusion matrices of the experiment are represented in Fig. 7.

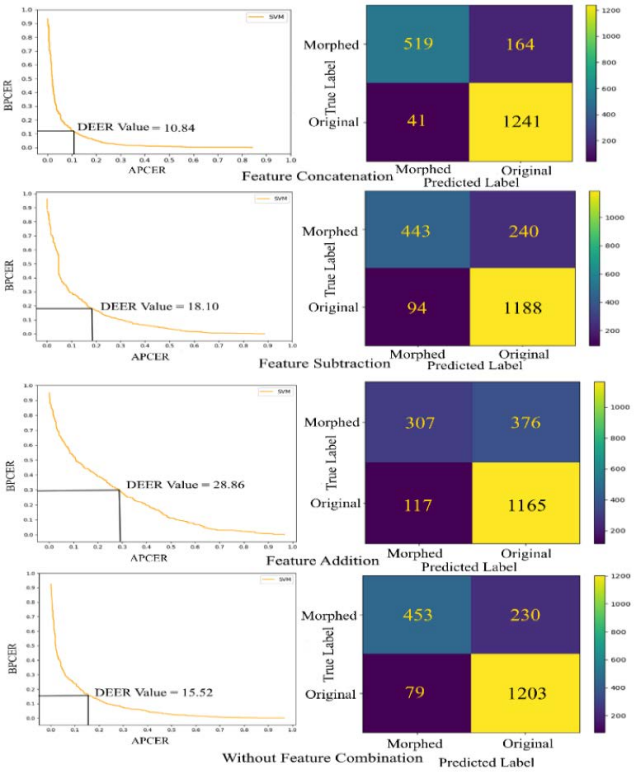

FIGURE 7. DEER curves and confusion matrices.

It can be seen from Table 5 that the best results in terms of errors like APCER, BPCER, ACER and DEER are acquired when extracted feature vectors by FaceNet are concatenated. Similarly other metrics like accuracy yielded the best

TABLE 5. Experiment-1, performance of the model using different feature combination techniques.

Feature combination strategy	Accuracy %	APCER %	BPCER %	APCER-10 %	BPCER-10 %	ACER %	DEER %
Feature concatenation	89.6	**24.0**	**3.2**	12.6	12.2	13.6	10.8
Without feature combination	84.3	33.7	6.2	23.9	23.6	20.4	15.5
Feature subtraction	83.0	35.1	7.3	30.9	29.2	21.2	18.1
Feature addition	74.9	55.1	9.1	53.1	52.3	32.1	28.9

TABLE 6. Experiment-1, performance errors proportion on each database while using feature combination techniques.

Feature combination	Celebrity		Yale		FEI		FGNET		GTDB		PIE		FERET		FRLL	
	M	O	M	O	M	O	M	O	M	O	M	O	M	O	M	O
Feature concatenation	20.5	3.8	**16.7**	**0.0**	16.7	2.8	**45.4**	**5.0**	28.6	0.0	22.7	4.3	29.7	1.4	21.3	2.5
Without combination	37.1	5.6	33.3	0.0	16.7	11.1	81.8	0.0	28.6	0.0	22.7	6.5	29.2	8.0	35.1	5.0
Feature subtraction	24.9	6.6	**0.0**	**16.7**	22.2	8.3	54.5	15.0	42.9	16.7	22.7	8.7	47.6	8.0	34.6	7.5
Feature addition	69.6	8.4	83.3	0.0	38.9	11.1	72.7	10.0	28.6	0	59.1	8.7	48.6	12.2	56.0	5.0

results in feature concatenation strategy. Concatenation of features produced the best results because majority of live captured images are very similar to morphed images and feature addition or subtraction were not able to produce a prominent pattern as compared to feature concatenation. Feature concatenation also preserved the information from morphed and live captured images in its original form and that helped the model to better learn and classify accordingly. Computationally feature concatenation is more intensive as the size of a single feature vector is 256 (dimensions 1×256) and in case of feature addition or subtraction the size of feature vector is 128 (dimensions 1×128).

Without feature combination the results are also better than the addition and subtraction methods. It also provides an additional benefit that it only requires a single morphed or original image for classification as during training and testing this method does not require a live captured image for classification. Furthermore, the model without feature combination has the lowest computational requirements as it skips the facial recognition phase and only learns from input facial images.

Utilizing both SSIM and cosine, the computational cost is almost doubled but during testing it was observed that the facial recognition phase took different time for images belonging to different databases. More time was required for high quality images but the facial recognition phase did not take more than one second for processing the best quality image. Other modules of the model take the standard default time in the processing of data as intended by the designers.

TABLE 7. Experiment-1, incorrectly classified images w.r.t morphing tools.

Morphing tool	Test images	Incorrectly classified test images	Proportion % of incorrectly classified images	Training images
FotoMorph	58	33	**60.0**	**211**
FantaMorph	211	25	**11.8**	**834**
FaceMorpher	200	50	25.0	607
OpenCV	214	56	26.2	607

From Table 6 it is evident that almost all the morphed images in each database are classified in an excellent manner except FGNET database where 45.4% images are incorrectly classified. This incorrect classification in FGNET database resulted due to the extreme levels of variations related to range of age (0 to 69 years), colour and illumination as mentioned in Table 2. Images that are present in this database are represented in Fig. 9. Best classification results

TABLE 8. Experiment-2, performance results of the model using feature concatenation technique on all classifiers.

Classifier	Accuracy %	APCER %	BPCER %	APCER-10 %	BPCER-10 %	ACER %	DEER %
SVM linear	89.6	24.0	3.2	12.6	12.2	13.6	10.8
SVM RBF default	89.6	22.8	3.8	10.2	10.0	13.3	9.5
SVM RBF gamma 0.6, C 50	90.1	**19.5**	4.8	11.1	11.8	12.1	10.5
Naive bayes	78.9	25.3	18.9	41.3	42.3	22.1	22.0
Logistic regression	88.6	26.1	3.6	13.0	14.3	14.8	11.5
XG boost	88.8	25.6	3.5	-	-	14.6	-
Majority voting	90.0	21.2	4.0	-	-	12.6	-

are acquired in Yale database because all the images of this database are grayscale and external factors like background colour and light sources are almost same in all the images. This absence of external factors allows the model to detect the variation in images that is caused from morphing process and it becomes easier for the model to detect these morphing traces with specific and undivided focus.

During experiments it was observed that 164 morphed test images were incorrectly classified and the maximum proportion of incorrectly classified images were created from tool FotoMorph as shown in Table 7. All the morphed images had significant amount of training data ranging from three to four times the test data but the reason for most incorrect classifications related to tool FotoMorph is that the training data of 211 images was spread between six databases and extreme level of variation among the databases was not properly encompassed in the training data related to FotoMorph.

Furthermore, sufficient training data from each database to ensure maximum learning (weight updation) and correct classification was not provided to the model in case of FotoMorph as evident from Table 4, otherwise the results in case of other morphing tools like FantaMorph, FaceMorpher and OpenCV are significantly better because training data contained considerable quantity from each database to ensure maximum learning (weight updation) and correct classification. This also proves that during training representation of morphed images created from each morphing tool is necessary in order to facilitate the model to efficiently detect the test morphed images created from the same tool.

With respect to effect of facial hair, eye wear and head gear on incorrect classification of original images as morphed images, the results indicated that 6 out of 31 original images were incorrectly classified due to the presence of eye wear, 3 out of 188 images due to the presence of facial hair and 0 out of 12 images due to the presence of head gear or caps. These results represent an overall improvement in correct classification of original images as compared to previous work [5]. Accuracy decreased from 87.5% to 64.9% in previous work [1] in the presence of above mentioned variations in images as compared to the accuracy of 89.6% of this research.

B. EXPERIMENT-2: PERFORMANCE OF THE MODEL USING DIFFERENT CLASSIFIERS

Training and testing of proposed model on all the first eight databases are done excluding the ninth database FRGC as shown in Table 3 to analyse the performance of different types of famous classifiers like SVM, naive bayes, logistic regression, XG boost and majority voting. Majority voting is a combination of SVM with linear function, Naive bayes and XG boost. Two out of three classifiers decided the outcome of classification. Tuning of SVM on radial basis function (RBF) has also been done by using different values of hyper parameters like gamma (influence controller) and C (decision margin controller). Details about the experiment's performance are represented in Table 8 and errors in each database are represented in Table 9. In Table 9, M represents the proportion of incorrectly classified morphed images and O represents proportion of incorrectly classified original images.

It is evident from results in Table 8 that the best results are acquired when SVM is equipped with RBF tuned function with gamma set to 0.6 and C is set to 50. While lowest DEER is observed with SVM RBF on default hyper parameters. All the classifiers performed well except naive bayes because it assumes that the features are independent in the data while some of the features in these images are dependent like morphing software leaves different artifacts in different images and the magnitude of these artifacts is dependent on the illumination, quality and light intensity in the respective images. In Table 9 SVM RBF with tuned hyper parameters is able to get the best results in classifying the morphed images in majority of databases and it is even able to achieve only 27.3 % error in classifying FGNET database in which other classifiers lagged behind significantly. Naive Bayes also produced the best results in classifying morphed images in databases like GTDB, PIE, FERET and FRLL because in these databases majority of images have small amount of variation as compared to FGNET and features

TABLE 9. Experiment-2, performance errors proportion on each database while using feature concatenation technique on all classifiers.

Classifier	Celebrity		Yale		FEI		FGNET		GTDB		PIE		FERET		FRLL	
	M	O	M	O	M	O	M	O	M	O	M	O	M	O	M	O
SVM Linear	20.5	3.8	16.7	0	16.7	2.8	45.4	5.0	28.6	0.0	22.7	4.3	29.7	1.4	21.3	2.5
SVM RBF Default	23.9	4.1	0.0	0.0	11.1	2.8	45.4	15.0	28.6	0.0	13.6	4.3	23.1	2.8	22.8	2.5
SVM RBF gamma 0.6, C 50	17.5	5.6	16.7	0.0	16.7	5.6	**27.3**	**20.0**	28.6	0.0	18.2	4.3	22.2	2.1	18.3	2.5
Naïve Bayes	51.2	11.8	33.3	33.3	16.7	47.2	45.4	20.0	**14.3**	50.0	**13.6**	21.7	**12.3**	31.1	**13.9**	37.5
Logistic Regression	24.9	3.8	16.7	0.0	16.7	5.6	54.5	5.0	28.6	0.0	22.7	2.8	30.7	8.0	22.3	2.5
XG boost	27.8	4.1	16.7	0.0	16.7	2.8	45.4	15.0	28.6	0.0	13.6	4.3	23.6	1.4	26.7	2.5
Majority Voting	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

TABLE 10. Experiment-3, performance results of the model after applying image enhancement on all the databases on SVM with linear function.

Enhancement Parameters Brightness, Contrast CLAHE Clip Limit	Accuracy %	APCER %	BPCER %	APCER-10 %	BPCER-10 %	ACER %	DEER %
0, 0, 3	89.6	24.0	3.2	12.6	12.2	13.6	10.8
1, 1, 3	88.2	26.6	3.8	17.3	21.0	15.2	13.3
5, 2, 3	82.9	35.4	7.2	29.4	27.2	21.3	17.4
5, 2, 10	83.4	36.4	6.0	28.8	31.0	21.2	17.7
5, 2, 20	83.0	37.2	6.2	30.4	28.8	21.7	18.6
10, 4, 20	81.8	39.5	6.8	32.0	35.8	23.2	20.5
15, 4, 20	81.7	40.2	6.7	62.9	36.0	23.4	20.5
20, 4, 20	82.0	40.0	6.2	33.8	37.7	23.1	20.7
30, 4, 20	81.8	40.1	6.5	34.1	35.7	23.3	20.5
30, 8, 20	77.2	50.7	7.9	44.8	45.2	29.3	22.5
30, 4, 30	81.4	40.8	6.8	35.3	36.4	23.8	20.5
30, 1, 30	85.8	32.0	4.7	21.2	24.6	18.4	15.5
85, 4, 30	80.2	41.6	8.2	37.8	37.9	24.9	20.5
100, 4, 30	79.4	44.4	7.9	39.9	37.9	26.1	21.4
100, 4, 50	79.7	43.8	7.7	38.2	37.4	25.7	21.6
100, 4, 70	79.8	43.2	7.9	39.0	37.5	25.5	21.5
100, 8, 70	75.8	59.8	5.6	46.0	45.6	32.5	25.5
150, 4, 70	78.2	51.2	6.2	42.0	45.9	28.7	22.7
300, 4, 120	65.2	100.0	0.0	-	-	50.0	-

are independent from each other due to the absence of external influences like significant variation in illumination and colour.

C. EXPERIMENT-3: PERFORMANCE OF THE MODEL USING IMAGE ENHANCEMENT

Training and testing of proposed model have been done on all the first eight databases as shown in Table 3 by applying image enhancement techniques like contrast limited adaptive histogram equalization (CLAHE) [37]. Different levels of brightness, contrast and CLAHE clip limit are used to analyse the impact on performance of the model in detecting morph attacks.

Details about the experiment's performance are represented in Table 10 for SVM with linear function and Table 12 for SVM with RBF function. Errors in each database are represented in Table 11 for SVM with linear function and Table 13 for SVM with RBF function. In Table 11, and Table 13 M represents the proportion of incorrectly classified morphed images and O represents proportion of

TABLE 11. Experiment-3, performance errors proportion on each database after applying image enhancement on SVM with linear function.

Enhancement Parameters Brightness, Contrast, CLAHE Clip Limit	Celebrity		Yale		FEI		FGNET		GTDB		PIE		FERET		FRLL	
	M	O	M	O	M	O	M	O	M	O	M	O	M	O	M	O
0, 0, 3	20.5	3.8	16.7	0.0	16.7	2.8	45.4	5.0	28.6	0.0	22.7	4.3	29.7	1.4	21.3	2.5
1, 1, 3	28.3	3.9	**0.0**	**0.0**	11.1	5.6	54.5	5.0	28.6	16.7	**13.6**	**2.2**	28.8	3.1	24.7	2.5
5, 2, 3	37.1	8.2	33.3	0.0	16.7	5.6	45.4	15.0	42.8	0.0	36.4	6.5	36.3	4.5	33.7	7.5
5, 2, 10	39.0	5.4	50.0	16.7	33.3	8.3	54.5	15.0	45.8	16.7	31.8	13.0	44.3	5.2	24.7	5.0
5, 2, 20	36.6	5.5	33.3	16.7	27.8	5.5	54.5	15.0	42.9	16.7	40.9	17.4	46.2	5.2	27.7	5
10, 4, 20	37.1	5.5	16.7	0.0	33.3	16.7	**27.3**	**10.0**	28.6	8.3	31.8	27.7	59.4	7.0	24.3	7.5
15, 4, 20	39.0	5.5	16.7	0	38.9	16.7	36.4	10.0	28.6	8.3	36.5	21.7	59.4	5.9	22.8	10.0
20, 4, 20	40.5	5.0	16.7	0.0	38.9	13.9	36.4	10.0	28.6	8.3	36.4	19.6	59.0	5.9	21.3	10.0
30, 4, 20	42.4	5.5	16.7	0.0	33.3	11.1	36.4	15.0	28.6	8.3	40.9	21.7	56.1	5.9	22.8	7.5
30, 8, 20	69.3	3.7	100.0	0.0	22.2	22.2	63.6	15.0	85.7	0.0	68.2	15.2	62.7	10.8	16.3	52.5
30, 4, 30	41.5	5.5	16.7	0.0	27.8	11.1	45.4	10.0	28.6	16.7	36.4	26.1	59.4	6.6	23.3	5.0
30, 1, 30	32.7	5.3	16.7	0.0	**11.1**	**2.8**	81.8	10.0	28.6	0.0	31.8	4.3	34.9	2.8	28.2	7.5
85, 4, 30	53.7	6.5	50.0	0.0	11.1	11.1	54.5	25.0	42.8	8.3	22.7	22.7	52.4	8.0	21.8	20.0
100, 4, 30	60.0	6.0	50.0	0.0	11.1	11.1	54.5	15.0	57.1	16.7	27.3	19.6	55.7	7.7	20.3	27.5
100, 4, 50	59.5	5.4	66.7	0.0	11.1	16.7	54.5	15.0	42.9	25.0	27.3	19.6	55.2	7.3	19.3	30.0
100, 4, 70	58.0	5.5	66.7	0.0	11.1	16.7	54.5	15.0	42.7	25.0	27.3	19.6	54.2	7.7	19.8	30.0
100, 8, 70	73.2	2.7	100.0	0.0	38.9	25.0	90.9	10.0	100	0.0	90.9	2.2	72.6	7.3	26.7	30.0
150, 4, 70	71.2	2.7	83.3	0.0	22.2	13.9	63.6	15.0	85.7	0.0	72.7	10.9	61.8	8.4	17.3	47.5
300, 4, 120	100.0	0.0	100.0	0.0	100.0	0.0	100.0	0.0	100.0	0.0	100.0	0.0	100.0	0.0	100.0	0.0

incorrectly classified original images. From results it can be formulated that medium to high level of contrast is causing a negative impact on the correct classification of morphed images while low to medium level of brightness and clip limit produce significantly better results as seen in Table 10. Indepth analysis of enhancement can be done by observing individual database results. It can be seen in Table 11 that image enhancement has different effect on different databases. The classification performance of morphed images in case of databases like Celebrity, FERET, and GTDB is continuously decreasing with the increase in image enhancement parameters because the images in these databases are already well illuminated in majority of scenarios while in databases like Yale and PIE at parameter 1, 1, 3 and in FGNET at parameter 10, 4, 20 and in FRLL at parameter 30, 8, 20 the best results of classification in case of morphed images are acquired because the images in these databases lack proper illumination and colour variance. The histograms of images after image enhancement are spread over all the colour spectrum and colour variance facilitates the model to learn more from each image and classifies them better consequently. At the end, it can be seen in the results that very high values of image enhancement parameters have resulted in very low results of classification in case of morphed images because all the details, information and patterns in the images are washed out and all the images look similar with values of each pixel reaching their maximum limit.

TABLE 12. Experiment-3, performance results of the model after applying image enhancement on all databases on SVM with RBF function.

Enhancement Parameters Brightness, Contrast CLAHE Clip Limit	Accuracy %	APCER %	BPCER %	APCER-10 %	BPCER-10 %	ACER %	DEER %
0, 0, 3	89.6	22.8	3.8	10.2	10.0	13.3	9.5
1, 1, 3	88.8	27.7	2.4	13.0	12.5	15.0	11.4
5, 2, 3	85.9	31.2	4.9	20.6	20.9	18.0	13.8
5, 2, 10	85.8	33.1	4.1	17.7	17.2	18.6	12.8
5, 2, 20	85.3	34.0	4.4	19.5	18.0	19.2	13.5
10, 4, 20	84.2	36.6	4.8	24.6	27.7	20.7	16.5
15, 4, 20	84.1	37.3	4.5	24.9	28.5	20.9	17.0
20, 4, 20	84.2	37.0	4.4	26.1	28.2	20.7	17.0
30, 4, 20	83.6	38.6	4.6	26.9	27.1	21.6	16.9
30, 8, 20	80.5	45.8	5.5	35.3	40.5	25.7	21.5
30, 4, 30	84.0	37.0	4.8	26.1	28.7	20.9	17.3
30, 1, 30	86.4	31.6	3.9	18.7	18.7	17.8	13.6
85, 4, 30	82.4	40.7	5.3	27.7	30.8	23.0	18.5
100, 4, 30	81.8	43.6	4.6	28.7	38.0	24.1	19.3
100, 4, 50	82.7	40.5	4.8	28.9	36.4	22.7	19.2
100, 4, 70	82.5	41.1	4.9	28.7	36.1	23.0	18.6
100, 8, 70	78.5	50.1	6.3	40.6	43.7	28.2	22.5
150, 4, 70	79.4	47.9	6.0	39.5	41.2	26.9	21.5
300, 4, 120	65.2	100.0	0.0	-	-	50.0	-

TABLE 13. Experiment-3, performance errors proportion on each database after applying image enhancement on SVM with RBF function.

Enhancement Parameters Brightness, Contrast CLAHE Clip Limit	Celebrity		Yale		FEI		FGNET		GTDB		PIE		FERET		FRLL	
	M	O	M	O	M	O	M	O	M	O	M	O	M	O	M	O
0, 0, 3	23.9	4.1	0.0	0.0	11.1	2.8	45.4	15.0	28.6	0.0	**13.6**	**4.3**	23.1	2.8	22.8	2.5
1, 1, 3	27.3	2.4	**0.0**	**0.0**	11.1	2.8	54.5	5.0	28.6	8.3	18.2	2.2	25.9	2.4	31.7	0.0
5, 2, 3	33.2	5.1	33.3	0.0	16.7	5.6	45.4	10.0	42.9	0.0	31.8	2.2	27.8	5.2	32.7	0.0
5, 2, 10	32.7	4.2	33.3	0.0	38.9	2.8	54.5	5.0	28.3	8.3	18.2	6.5	39.1	3.8	27.2	2.5
5, 2, 20	27.3	4.7	33.3	0.0	38.9	2.8	54.5	10.0	28.6	16.7	22.7	8.8	45.7	2.8	28.2	2.5
10, 4, 20	36.1	3.6	33.3	0.0	27.8	8.3	**45.4**	**5.0**	28.6	8.3	18.2	19.6	51.9	4.9	23.8	7.5
15, 4, 20	36.1	3.3	50.0	0.0	27.8	8.3	45.4	5.0	28.6	8.3	18.2	17.4	53.8	5.2	23.8	5.0
20, 4, 20	37.6	3.1	33.3	0.0	27.8	8.3	45.4	5.0	28.6	16.7	22.7	17.4	51.9	4.9	23.3	5.0
30, 4, 20	37.6	3.7	50.0	0.0	27.8	2.8	45.4	15.0	45.9	8.3	31.8	17.4	53.3	4.9	25.2	2.5
30, 8, 20	56.6	2.5	33.3	16.7	33.3	25.0	72.7	15.0	71.4	0.0	63.6	4.3	58.5	6.6	**18.8**	40.0
30, 4, 30	36.6	4.1	33.3	0.0	27.8	2.8	45.4	10.0	42.8	16.7	31.8	17.4	50.9	4.5	23.8	5.0
30, 1, 30	25.4	4.5	16.7	0.0	16.7	2.8	90.9	10.0	28.6	0.0	22.7	2.2	37.7	2.4	31.2	2.5
85, 4, 30	44.4	3.8	50.0	0.0	22.2	5.6	63.6	10.0	42.9	0.0	27.3	15.2	56.6	7.0	21.8	12.5
100, 4, 30	49.8	3.4	33.3	0.0	16.7	5.6	72.7	0.0	57.1	8.3	31.8	15.2	60.8	5.2	21.3	15.0
100, 4, 50	47.3	3.6	50.0	0.0	11.1	5.6	63.6	0.0	57.1	8.3	22.7	15.2	56.1	5.6	19.8	15.0
100, 4, 70	47.8	3.6	33.3	16.7	11.1	5.6	63.6	0.0	57.1	8.3	27.3	15.2	56.6	5.6	20.8	15.0
100, 8, 70	61.9	2.7	50.0	50.0	27.8	19.4	63.6	15.0	71.4	0.0	63.3	17.4	69.8	7.7	**16.3**	37.5
150, 4, 70	60.0	2.7	66.7	33.3	27.8	19.4	72.7	15.0	85.7	0.0	63.6	6.5	59.4	8.0	20.3	40.0
300, 4, 120	100.0	0.0	100.0	0.0	100.0	0.0	100.0	0.0	100.0	0.0	100.0	0.0	100.0	0.0	100.0	0.0

TABLE 14. Experiment-4, replication of the state of the art work [5] (previous work) using the proposed model on the state of the art database [5].

Training database (Tool)	Test database	Accuracy %	APCER %	BPCER %	APCER-10 %	BPCER-10 %	BPCER-10 % Previous work	ACER %	DEER %	DEER % Previous work
FERET (OpenCV)	FRGC (OpenCV)	67.6	42.8	11.1	44.3	37.0	35.8	26.9	21.5	21.1
FERET (OpenCV)	FRGC (FaceMorpher)	67.4	43.1	11.1	44.7	39.1	36.2	27.1	22.1	20.1
FERET (FaceMorpher)	FRGC (OpenCV)	67.1	44.4	9.4	42.6	37.0	37.8	26.9	21.5	20.7
FERET (FaceMorpher)	FRGC (FaceMorpher)	68.6	42.1	9.4	40.1	35.5	36.4	25.7	21.5	19.8

TABLE 15. Experiment-5, training of model on the state of the art [5] database and testing on 6 databases of this research work using feature concatenation.

Training database (Tool)	Test on 6 databases	Accuracy %	APCER %	BPCER %	APCER-10 %	BPCER-10 %	ACER %	DEER %
FERET (OpenCV)	FantaMorph FotoMorph	88.2	**49.3**	1.2	20.5	21.1	25.2	14.5
FERET (FaceMorpher)	FantaMorph FotoMorph	88.3	**48.9**	1.1	20.2	22.2	25.0	14.5

The performance of SVM with RBF function is also presented in Table 12 and Table 13 and almost similar results are produced with some variation in respective databases.

Therefore, it can be concluded that these enhancements and contrast settings may vary depending upon the hardware resources.

D. EXPERIMENT-4: REPLICATION OF THE STATE OF THE ART [5] PREVIOUS WORK

In this experiment the previous work [5] has been replicated and all the work has been done using FaceNet SVM with RBF function and without feature combination. Training of the model has been done on FERET database that contains 529 morphed images created from morphing tool OpenCV and 529 morphed images created from morphing tool FaceMorpher. 1896 original images are used for training and 470 original images are used for testing. From FRGC database, 964 morphed images are used for testing that are created from tool OpenCV and 964 morphed images are used for testing that are created from tool FaceMorpher. Details about the experiment's performance are represented in Table 14. It can be seen that during replication process there is only 1 to 2% difference in DEER values of this replication and previous work [5]. Similarly, BPCER-10 value also differ not more than 3% and BPCER-10 of this replication work is 0.9% better while the training and testing is done on FaceMorpher. Replication of the previous state of the art work [5] has been done to compare the performance of the model created in this study and the state of the art [5] and it is proven that performance of the models is almost similar.

This study is a better representation of a real morph attack detection scenario as compared to previous work [5] because it is based on a highly diverse and unique morph database. This database is created manually using professional high quality tools as compared to automatically generated morphs using low quality tools in the previous work [5]. It contains morph-3 images that are created for the first time in a research study. Morph-3 images are very difficult to detect due to their realistic look and additional complexity as proven in experiment-5. This database contains images with variety in expression, posture, cosmetics, quality, age and illumination that was lacking in previous works. Furthermore, image enhancement techniques have been introduced to address the issues related to low quality images.

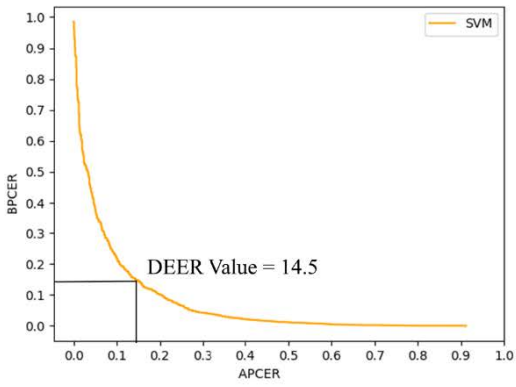

Table 15 Feret (OpenCV)
Feature Concatenation

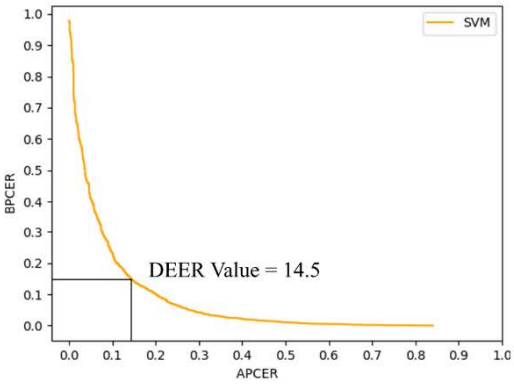

Table 15 Feret (FaceMorpher)
Feature Concatenation

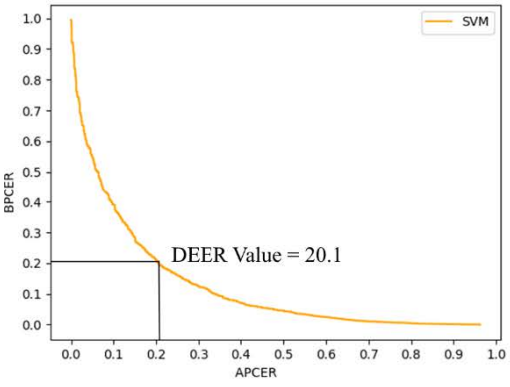

Table 16 Feret (OpenCV)
Without Feature Concatenation

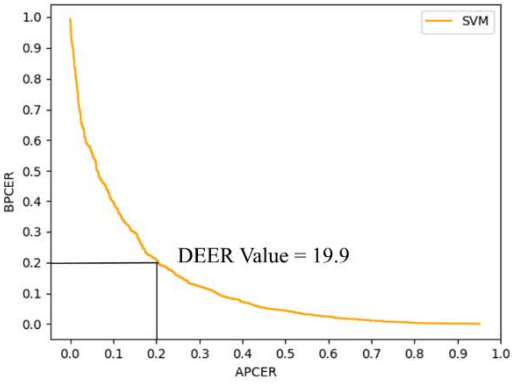

Table 16 Feret (FaceMorpher)
Without Feature Concatenation

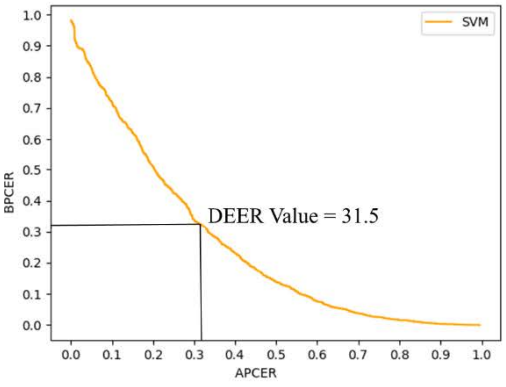

Table 17 Feret (OpenCV)
MDS

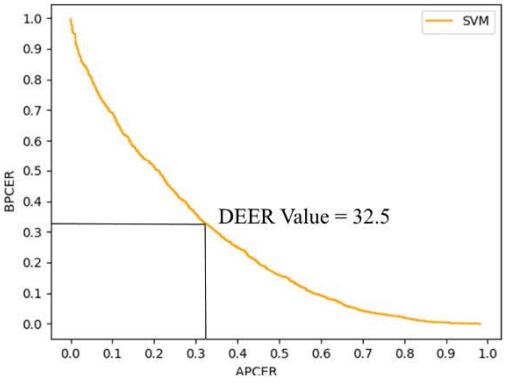

Table 17 Feret (FaceMorpher)
MDS

FIGURE 8. DEER curves of experiment-5.

E. EXPERIMENT-5: TRAINING THE MODEL ON STATE OF THE ART [5] DATABASES AND TESTING ON THE CREATED MORPHED DATABASE

To prove the unique and diverse nature of the created database as shown in Table 3, the first six databases manually created from morphing tools of FotoMorph and FantaMorph are used as testing databases to illustrate the issues and performance depreciation in training the model on one database (FERET or FRGC from previous work ([5])) created from single morphing tool. The model is trained and tested using all the techniques like feature concatenation, without feature concatenation and multidimensional scaling of difference of feature vectors. Training of the model is done on FERET database that contains 529 morphed images created in OpenCV and 529 morphed images created in FaceMorpher. FERET database contains 1426 normal images

TABLE 16. Experiment-5, training of model on the state of the art [5] database and testing on 6 databases of this research work without feature combination.

Training database (Tool)	Test on 6 databases	Accuracy %	APCER %	BPCER %	APCER-10 %	BPCER-10 %	ACER %	DEER %
FERET (OpenCV)	FantaMorph FotoMorph	85.4	**50.1**	4.5	34.3	38.9	27.3	20.1
FERET (FaceMorpher)	FantaMorph FotoMorph	85.8	**47.3**	4.8	34.2	39.0	26.0	19.9

TABLE 17. Experiment-5, training of model on state of the art [5] database and methodology (MDS of difference vector) and testing on 6 databases of this research work.

Training database (Tool)	Test on 6 databases	Accuracy %	APCER %	BPCER %	APCER-10 %	BPCER-10 %	ACER %	DEER %
FERET (OpenCV)	FantaMorph FotoMorph	80.3	**87.0**	0.7	56.1	70.8	43.8	31.5
FERET (FaceMorpher)	FantaMorph FotoMorph	80.5	**86.2**	0.6	58.2	68.5	43.4	32.5

TABLE 18. Experiment-6, training the model on 7 proposed databases of this research work and testing on state of the art [5] database using SVM with linear function.

Training on 7 databases	Test database (Tool)	Accuracy %	APCER %	BPCER %	APCER-10 %	BPCER-10 %	ACER %	DEER %
FaceMorpher FantaMorph FotoMorph	FERET (FaceMorpher)	91.8	**24.2**	2.2	14.6	15.9	13.2	12.1
FaceMorpher FantaMorph FotoMorph	FERET (OpenCV)	91.6	25.1	2.2	14.2	22.1	13.7	12.5
OpenCV FantaMorph FotoMorph	FERET (FaceMorpher)	91.7	**24.4**	2.4	13.6	16.5	13.4	12.3
OpenCV FantaMorph FotoMorph	FERET (OpenCV)	91.5	25.1	2.4	14.4	21.2	13.8	12.5

that are used for training of the model. For testing of the model 1314 morphed images and 4641 normal images from first six databases in Table 4 are used. While using SVM with RBF function the best results are shown to represent maximum performance of the model on training database. Details about the experiment's performance are represented in Table 15 for feature concatenation, Table 16 for without feature combination and Table 17 using MDS (multidimensional scaling of the difference of feature vectors to two dimensions) as per the methodology of previous work [5]. As APCER represents the total incorrectly classified morphed images therefore it is the most important evaluation metric for a morph attack detection model. It can be seen from Table 15, Table 16 and Table 17 that APCER is around

TABLE 19. Experiment-6, training the model on 7 proposed databases of this research and testing on the state of the art [5] database using SVM with RBF function.

Training on 7 databases	Test database (Tool)	Accuracy %	APCER %	BPCER %	APCER-10 %	BPCER-10 %	ACER %	DEER %
FaceMorpher FantaMorph FotoMorph	FERET (FaceMorpher)	88.9	32.5	1.8	14.2	14.9	17.2	11.5
FaceMorpher FantaMorph FotoMorph	FERET (OpenCV)	89.6	33.6	1.8	15.9	16.5	17.7	13.5
OpenCV FantaMorph FotoMorph	FERET (FaceMorpher)	90.2	**31.9**	1.6	14.6	14.2	16.8	11.8
OpenCV FantaMorph FotoMorph	FERET (OpenCV)	89.8	33.5	1.6	15.7	15.1	17.5	12.5

50% in case of feature concatenation and without feature concatenation while in case of MDS the APCER is around 86%. All of the APCER values are very high as around 50% to 87% images are incorrectly classified irrespective of the training tool. DEER curves of the experiment are represented in Fig. 8.

It is evident from the results that a single database created from a single morphing tool is not sufficient for a practical morph attack detection model. The training database must contain a vast range and variety of images and morphing tools in order for a morph attack detection model to be able to learn distinguishing features from the source database as proven in the next experiment-6.

F. EXPERIMENT-6: TRAINING THE MODEL ON THE CREATED MORPHED DATABASE AND TESTING ON THE STATE OF THE ART [5] DATABASE

In this experiment, the morph attack detection model is trained on the first seven databases from Table 3 and tested on FERET database to illustrate that the created morphed database yields better result as compared to previous work [5] that utilized single database and single morphing tool for creation of training database. Training images contain 2536 morphed images and 4845 normal images and test data contains 529 morphed images created from OpenCV and 529 morphed images created from FaceMorpher and 1426 normal images from FERET database. Results are represented in Table 18 for SVM with linear function and Table 19 for SVM with RBF function. It can be seen that APCER is around 25% in case of SVM with linear function and around 33% in case of SVM with RBF function while in previous experiment-5, the values of APCER were around 50 to 87%. So, it is evident that training the model on a database having vast variation in images and multiple high quality morphing tools yields better result in classification of morphed images.

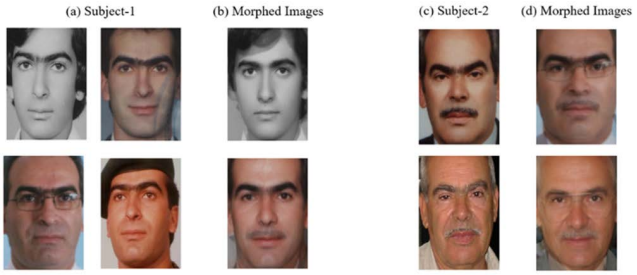

FIGURE 9. Effect of extreme variation in images.

VI. CONCLUSION

In this study, a robust and generalized morph attack detection model and a very diverse morphed database is introduced to better deal with morph attacks in a practical scenario. Different feature combination techniques are analysed and feature concatenation proved to be the best technique for morph detection. Some of the methods like feature concatenation provided better morph attack detection performance but with the increase of computational cost. Similarly, it was observed that manually created morphed images with high quality morphing tools were difficult to detect by the models that were trained on morphed databases that had low variation and were made automatically from low quality morphing tools like OpenCV and FaceMorpher using programming scripts. The training of model on manually created morphed databases with high quality tools proved to be helpful in achieving good results and the results achieved by the model on testing data improved significantly. Proposed model gives very encouraging and improved results in case of age, illumination, posture and expression variations. Testing of morphed images was also done using different machine learning based classifiers and SVM produced the best results. Different image enhancement techniques were also applied on image databases and it was observed that databases with

low variation in illumination and colour benefited from image enhancement.

Manually created morph-3 images were very difficult to detect when the model was only trained on morph-2 images created from low quality tools. After training the model on morph-3 images created from high quality tools, the performance of morph-3 detection increased significantly. It further solidifies the approach to include diverse range of morphs in the training database to improve the robustness of morph detection model. FGNET database proved to be the most difficult database of images in terms of morph detection as it can be seen in Fig. 9 that this database has a vast range of diversity in terms of age, image quality, colour variation and expression. These extreme levels of variations led to the creation of highly complex morphed images that were very difficult to classify by the morph attack detection model.

Future work that can be done to improve this model and train it for all possible morph attacks in a real world deployment scenarios will require the acquisition of real morphed images that were submitted to different organizations like airports, identity card issuing authorities, travel agencies, universities and security institutions. The model should then be trained and tested on the real images to ensure better performance. Furthermore, an adaptive morph attack detection model should be designed that automatically adapts to the input images by applying the image enhancements as per requirement. More than three images may also be used for morphing.

ACKNOWLEDGMENT

This work was supported by the Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2022R192), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia. For MIT-CBCL Database "Credit is hereby given to the Massachusetts Institute of Technology and to the Center for Biological and Computational Learning for providing the database of facial images." Copyright 2003–2005 Massachusetts Institute of Technology. For Feret Database "Portions of the research in this paper use the FERET database of facial images collected under the FERET program, sponsored by the DOD Counterdrug Technology Development Program Office." The MIT License (MIT) Copyright 2022. Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the "Software"), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions: The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software. THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

REFERENCES

[1] F. Peng, L.-B. 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.

[2] M. Ferrara, A. Franco, and D. Maltoni, "Face demorphing," *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 4, pp. 1008–1017, Apr. 2018.

[3] 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.

[4] A. W. Yip and P. Sinha, "Contribution of color to face recognition," *Perception*, vol. 31, no. 8, pp. 995–1003, 2002.

[5] 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.

[6] K. Panetta, Q. Wan, S. Agaian, S. Rajeev, S. Kamath, R. Rajendran, S. P. Rao, A. Kaszowska, H. A. Taylor, A. Samani, and X. Yuan, "A comprehensive database for benchmarking imaging systems," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 3, pp. 509–520, Mar. 2020.

[7] G. Wolberg, "Image morphing: A survey," *Vis. Comput.*, vol. 14, no. 8, pp. 360–372, 1998.

[8] D. B. Smythe, "A two-pass mesh warping algorithm for object transformation and image interpolation," *Rapport Technique*, vol. 1030, p. 31, Mar. 1990.

[9] T. Beier and S. Neely, "Feature-based image metamorphosis," *ACM SIGGRAPH Comput. Graph.*, vol. 26, no. 2, pp. 35–42, Jul. 1992.

[10] J. Kannala and E. Rahtu, "Bsif: Binarized statistical image features," in *Proc. 21st Int. Conf. pattern Recognit. (ICPR2012)*, pp. 1363–1366, IEEE, 2012.

[11] D. Ortega-Delcampo, C. Conde, D. Palacios-Alonso, and E. Cabello, "Border control morphing attack detection with a convolutional neural network de-morphing approach," *IEEE Access*, vol. 8, pp. 92301–92313, 2020.

[12] C. Seibold, W. Samek, A. Hilsmann, and P. Eisert, "Accurate and robust neural networks for face morphing attack detection," *J. Inf. Secur. Appl.*, vol. 53, Aug. 2020, Art. no. 102526.

[13] 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. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 10–18.

[14] S. Venkatesh, R. Ramachandra, K. Raja, L. Spreeuwers, R. Veldhuis, and C. Busch, "Detecting morphed face attacks using residual noise from deep multi-scale context aggregation network," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2020, pp. 280–289.

[15] R. Raghavendra, K. B. Raja, and C. Busch, "Detecting morphed face images," in *Proc. IEEE 8th Int. Conf. Biometrics Theory, Appl. Syst. (BTAS)*, Sep. 2016, pp. 1–7.

[16] L. Qin, F. Peng, S. Venkatesh, R. Ramachandra, M. Long, and C. Busch, "Low visual distortion and robust morphing attacks based on partial face image manipulation," *IEEE Trans. Biometrics, Behav., Identity Sci.*, vol. 3, no. 1, pp. 72–88, Jan. 2021.

[17] D. ICAO, *9303-Machine Readable Travel Documents—Part 9: Deployment of Biometric Identification and Electronic Storage of Data in EMRTDS*, International Civil Aviation Organization (ICAO), Montreal, QC, Canada, 2015.

[18] B.-C. Chen, C.-S. Chen, and W. H. Hsu, "Face recognition and retrieval using cross-age reference coding with cross-age celebrity dataset," *IEEE Trans. Multimedia*, vol. 17, no. 6, pp. 804–815, Jun. 2015.

[19] A. S. Georghiades, P. N. Belhumeur, and D. Kriegman, "From few to many: Illumination cone models for face recognition under variable lighting and pose," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 6, pp. 643–660, Jun. 2001.

[20] E. Kussul and T. Baydyk, "Face recognition using special neural networks," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2015, pp. 1–7.
[21] C. E. Thomaz and G. A. Giraldi, "A new ranking method for principal components analysis and its application to face image analysis," *Image Vis. Comput.*, vol. 28, no. 6, pp. 902–913, 2010.
[22] R. R. Atallah, A. Kamsin, M. A. Ismail, S. A. Abdelrahman, and S. Zerdoumi, "Face recognition and age estimation implications of changes in facial features: A critical review study," *IEEE Access*, vol. 6, pp. 28290–28304, 2018.
[23] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker, "Multi-PIE," in *Proc. 8th IEEE Int. Conf. Autom. Face Gesture Recognit.*, 2008, pp. 1–8. [Online]. Available: https://ieeexplore.ieee.org/document/4813399, doi: 10.1109/AFGR.2008.4813399.
[24] T. Sim, S. Baker, and M. Bsat, "The CMU pose, illumination, and expression database," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 12, pp. 1615–1618, Dec. 2003.
[25] P. J. Phillips, H. Wechsler, J. Huang, and P. J. Rauss, "The FERET database and evaluation procedure for face-recognition algorithms," *Image Vis. Comput.*, vol. 16, no. 5, pp. 295–306, 1998.
[26] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, "The FERET evaluation methodology for face-recognition algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 10, pp. 1090–1104, Oct. 2000.
[27] L. DeBruine and B. Jones. (2017). *Face Research Lab London Set*. [Online]. Available: https://figshare.com/articles/Face_Research_Lab_London_Set/5047666 and https://www.semanticscholar.org/paper/Face-Research-Lab-London-Set-DeBruine-Jones/69ab657ab81efc25f92b2b47b299980134fbecfd#citing-papers, doi: 10.6084/m9.figshare.5047666.v3.
[28] N. Burton, M. Burton, D. Rigby, C. A. M. Sutherland, and G. Rhodes, "Best-worst scaling improves measurement of first impressions," *Cognit. Res., Princ. Implications*, vol. 4, no. 1, pp. 1–10, Dec. 2019.
[29] E. Sarkar, P. Korshunov, L. Colbois, and S. Marcel, "Vulnerability analysis of face morphing attacks from landmarks and generative adversarial networks," 2020, *arXiv:2012.05344*.
[30] *Digital Photo Morphing Software*. FotoMorph 13.9. Accessed: May 2021. [Online]. Available: https: FotoMorph.informer.com
[31] *Digital Photo Morphing Software*. Abrosoft FantaMorph 5 Professional Edition. Accessed: May 2021. [Online]. Available: https://www.FantaMorph.com
[32] S. Mallick. *Face Morph Using OpenCV–C++/Python*. Accessed: May 2021. [Online]. Available: https://learnOpenCV.com/face-morph-using-OpenCV-cpp-python/
[33] Yaopang. *Face Morpher*. Accessed: May 2021. [Online]. Available: https://github.com/yaopang/FaceMorpher/tree/master/FaceMorpher
[34] S. Hassan. *Face Classification Using Facenet and MTCNN*. Accessed: May 2021. [Online]. Available: https://github.com/saadhaxxan/Face-Classification-using-FaceNet-and-MTCNN
[35] A. Rosebrock. *How-to: Python Compare Two Images*. Accessed: May 2021. [Online]. Available: https://www.pyimagesearch.com/2014/09/15/python-compare-two-images/
[36] *ISO/IEC JTC1 SC37 Biometrics, Information Technology-Biometric Presentation Attack Detection—Part 3: Testing and Reporting*, International Organization for Standardization, Geneva, Switzerland, document ISO ISO/IEC IS 30107-3:2017, 2017.
[37] FooBar167. *Fastest Way to Increase Colour Image Contrast With OpenCV in Python*. Accessed: Aug. 2021. [Online]. Available: https://stackoverflow.com/users/7550928/foobar167
[38] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 815–823, doi: 10.1109/CVPR.2015.7298682.
[39] P. Terhörst, J. N. Kolf, N. Damer, F. Kirchbuchner, and A. Kuijper, "SER-FIQ: Unsupervised estimation of face image quality based on stochastic embedding robustness," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 5650–5659, doi: 10.1109/CVPR42600.2020.00569.
[40] P. Terhorst, J. N. Kolf, N. Damer, F. Kirchbuchner, and A. Kuijper, "Face quality estimation and its correlation to demographic and non-demographic bias in face recognition," in *Proc. IEEE Int. Joint Conf. Biometrics (IJCB)*, Sep. 2020, pp. 1–11, doi: 10.1109/IJCB48548.2020.9304865.
[41] P. Terhorst. *Faceimagequality*. Accessed: Apr. 2022. [Online]. Available: https://github.com/pterhoer/FaceImageQuality
[42] E. Strickland, "Andrew Ng: Unbiggen AI," *IEEE Spectr.*, vol. 59, no. 4, pp. 22–50, Apr. 2022. [Online]. Available: https://spectrum.ieee.org/andrew-ng-data-centric-ai and https://ieeexplore.ieee.org/document/9754503

MUHAMMAD HAMZA received the B.S. degree in electronics engineering from the International Islamic University Islamabad, Islamabad, Pakistan, and the M.S. degree in data science from Bahria University, Islamabad. His research interests include computer vision and deep learning. He was awarded the Gold Medal for having obtained first position and highest cumulative grade point average in both the B.S. and M.S. degrees.

SAMABIA TEHSIN received the M.S. and Ph.D. degrees in computer software engineering with specialization in digital image analysis from the National University of Science and Technology (NUST), Islamabad, Pakistan. She is currently working as a Senior Associate Professor with Bahria University, Islamabad. Her research interests include image analysis, machine learning, and multimedia forensics.

HANEN KARAMTI received the bachelor's degree in computer science and multimedia from the High Institute of Computer Science and Multimedia (ISIMS), University of Sfax, Tunisia, the master's degree in computer science and multimedia from the University of Sfax, and the Ph.D. degree in computer science from the National Engineering School of Sfax, University of Sfax, in cooperation with the University of La Rochelle, France, and the University of Hanoi, Vietnam. She is currently an Assistant Professor with Princess Norah Bint Abdulrahman University, Riyadh, Saudi Arabia. Her research interests include information retrieval, multimedia systems, image retrieval, health informatics, big data, and data analytics.

NORAH SALEH ALGHAMDI received the bachelor's degree in computer science from Taif University, Taif, Saudi Arabia, and the master's degree in computer science and the Ph.D. degree from the Department of Computer Science, La Trobe University, Melbourne, Australia. She is currently an Associate Professor with the Department of Computer Science, College of Computer and Information Sciences, Princess Nourah Bint Abdulrahman University (PNU), Riyadh, Saudi Arabia, where she has been the Vice-Dean of quality assurance, since 2019. Her research interests include data mining, machine learning, text analytics, image classification, bioengineering, and deep learning. She has participated in organizing the International Conference on Computing (ICC), in 2019. She is a member of the reviewer committee of several journals, such as IEEE Access, *Journal of Computer Science*, and *International Journal of Web Information Systems*.