IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received November 18, 2021, accepted December 9, 2021, date of publication December 20, 2021,
date of current version December 28, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3136485

Single Morphing Attack Detection Using Feature Selection and Visualization Based on Mutual Information

JUAN E. TAPIA, (Member, IEEE), AND CHRISTOPH BUSCH, (Senior Member, IEEE)
da/sec-Biometrics and Internet Security Research Group, Hochschule Darmstadt, 64295 Darmstadt, Germany

Corresponding author: Juan E. Tapia (juan.tapia-farias@h-da.de)

This work is supported by the European Union's Horizon 2020 research and innovation program under grant agreement No 883356 - "image Manipulation Attack Resolving Solutions (iMARS)" and the German Federal Ministry of Education and Research and the Hessen State Ministry for Higher Education, Research and the Arts within their joint support of the National Research Center for Applied Cybersecurity ATHENE.

ABSTRACT Face morphing attack detection is a challenging task. Automatic classification methods and manual inspection are realised in automatic border control gates to detect morphing attacks. Understanding how a machine learning system can detect morphed faces and the most relevant facial areas is crucial. Those relevant areas contain texture signals that allow us to separate the bona fide and the morph images. Also, it helps in the manual examination to detect a passport generated with morphed images. This paper explores features extracted from intensity, shape, texture, and proposes a feature selection stage based on the Mutual Information filter to select the most relevant and less redundant features. This selection allows us to reduce the workload and know the exact localisation of such areas to understand the morphing impact and create a robust classifier. The best results were obtained for the method based on Conditional Mutual Information and Shape features using only 500 features for FERET images and 800 features for FRGCv2 images from 1,048 features available. The eyes and nose are identified as the most critical areas to be analysed.

INDEX TERMS Morphing, differential morphing attack detection, feature selection.

I. INTRODUCTION

In recent years, ID verification systems have been exposed to variations of presentation attacks. For instance, they compare the user selfie with a photo ID extracted from the user ID card or passport. The critical challenge is to ensure whether the ID card image has been tampered with in the digital or physical domain. Image tampering is a significant issue for such scenarios and biometric systems at large [1], [2].

One of these approaches is related to the passports, and the Morphing attack on face recognition systems based on the enrolment of a morphed face image, which is averaged from two-parent images and allowing both contributing subjects to travel with the passport [2]–[4].

Morphing attack detection is a new topic aimed to detect unauthorised individuals who want to gain access to a "valid" identity in other countries. Morphing can be understood as a technique to combine two o more look-alike facial images from one subject and an accomplice, who could apply for a valid passport exploiting the accomplice's identity. Morphing takes place in the enrolment process stage. The threat of morphing attacks is known for border crossing or identification control scenarios. It can be broadly divided into two types: (1) Single Image Morphing Attack Detection (S-MAD) techniques (a.k.a as No-Reference MAD) and Differential Morphing Attack Detection (D-MAD) methods. The S-MAD is more challenging among these two types as the decision needs to be taken on a single image without a trusted image available for the same subject [5]. S-MAD can be organised according to different approach: Textures, Shape, Quality, Hybrid features, Residual noise and Deep learning.

A morphing attack's success depends on the decision of human observers, especially a passport identification expert. The real-life application for a border police expert who compares the passport reference image of the traveller (digital extracted from the embedded chip) with the facial appearance of the traveller [6] is too hard because of the improvements of the morphing tools and because of the difficulty for the human expert to localise facial areas, in which morphing artefacts are present.

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

This work proposes to add an extra stage of feature selection after feature extraction based on Mutual Information *MI* to estimate and keep the most relevant and remove the most redundant features from the face images to separate bona fide and morphed images on a S-MAD scenario. The high redundancy between features confuses the classifier.

The contributions of this work are described as follows: a) Identify the most relevant and less redundant features from faces that allow us to separate bona fide from morphed images. b) Localise the position of the most relevant areas on the images. c) Visualise the areas that contain morphing artefacts d) Reduce the algorithm's complexity, sending fewer features to the classifier. e) Analysis of the feature level fusion, the intensity, shape, and texture information. All these contributions may help to guide the manual inspection of morphed images.

This paper is organised as follows: a summary background in features selection and *MI* is presented in section III-B. The relate work is describe in Section II. The methods are described in Section III. The database are described in section IV and the experiments and results are presented in section V and conclusion are presented in section VII.

II. RELATED WORK

Face morphing attack has captured the interest of the research community and government agencies in Europe. For instance the European Union (EU) funded the image Manipulation Attack Resolving Solutions (iMARS) project,[1] developing new techniques of manipulation and detection of morphed images.

Ferrera *et al.* [2] were the first to investigate the face recognition system's vulnerability with regards to morphing attacks. He has evaluated the feasibility of creating deceiving morphed face images and analysed the robustness of commercial face recognition systems in the presence of morphing.

Scherlag *et al.* [3] studied the literature and developed a survey about the impact of morphing images on face recognition systems. The same author [4] proposed a face representation from embedding vectors for differential morphing attack detection, creating a more realistic database, different scenarios, and constraints with four automatic morphed tools. He also reported detection performances for several texture descriptors in conjunction with machine learning techniques.

Indeed, the NIST Face Recognition Vendor Test (FRVT) applied to MORPH images [7] evaluates and reports the performances of different morph detection algorithms organised in three tiers according to the morph images quality. Tier 1 evaluates low-quality morph images; Tier 2 considers automatic morph images; and Tier 3 for high-quality images. Further, the NIST report is organised w.r.t local (crop faces) and global (passport-photos) morphing algorithms. This fact confirms and shows that morphing images is a problem considering many scenarios.

[1]https://cordis.europa.eu/project/id/883356

Most of the state-of-the-art approaches use machine learning and deep learning to detect and classify morph images. Also, they are utilising embedding vectors from deep learning approaches to detect and classify the images. This paper focused on the machine learning method. In this scenario, popular texture-based methods include local binary patterns (LBPs) [8], [9], Local Phase Quantisation (LPQ) features [10] and Binarised Statistical Image Features (BSIF) [11]–[13] has been used. Furthermore, these texture features have also been extracted for different colour channels to obtain a robust detection performance. Several variants of LBPs and a histogram of oriented gradients (HOG) have also been explored in the literature [14]. A summary of S-MAD literature was added. See Table 1. The use of texture-based methods has shown good performance on S-MAD. However, those approaches did not analyse the most relevant features and their localisation on the original images. An efficient feature selection method may help to improve this limitation.

TABLE 1. Summary S-MAD state of the art.

Author	Approach	Database	Link Dataset
Raghavendra et al. [12], [13]	Image degradation/BSIF	Utrecht In house	[15]
Debiasi et al. [16], [17]	PRNU Analysis	FRGCv2	[18]
Zhang et al. [19]	SPN analysis	Utrecht FEI	[20] [15]
Ulrich et al. [21]	PRNU analysis	FRGCv2	[18]
Debiasi et al. [22]	PRNU analysis	CelebA	Synthetics Images
Venkatesh et al. [5]	Denoise CNN	FRGC PUT	[23]
Raghavendra et al. [24]	Luminance component	MAFI	[12]
Makrushin et al. [25]	double compresion	Utrench FEI	[20] [15]
Neubert et al. [26]	double compresion	Utrench FEI	[20] [15]
Seibold et al. [27]	Reflection analysis	In house	N/A
Ojavinsu et al. [10]	LPQ	Outex	[28]
Ulrich et al. [9]	LBP	In house	N/A

Regarding feature selection, raw input data often has very high dimensionality and a limited number of samples in image understanding. In this area, feature selection plays an essential role in improving the object identification process's accuracy, efficiency, and scalability. Since relevant features are often unknown a priori in the real world, irrelevant and redundant features may be introduced to represent the domain. However, using more features implies increasing computational cost in the feature extraction process, slowing down the classification process and increasing the time needed for training and validation, leading to classification over-fitting. As is the case in most image analysis problems, with limited sample data, irrelevant features may obscure the distributions of the small set of relevant features and confuse the classifiers.

Peng *et al.* [29] develop a general framework to analyse the interaction between the redundancy and the relevance of the

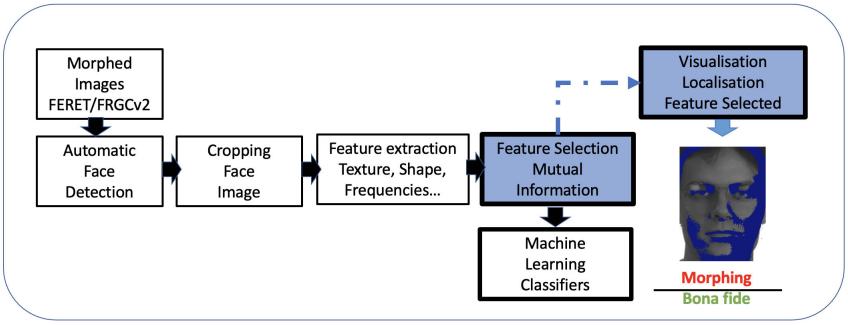

FIGURE 1. Framework proposed with feature selection stage.

features in a machine learning method to look at the most valuable features based on *MI*.

Guyon *et al.* [30] proposed the Conditional Mutual Information Maximisation (CMIM) to estimate the relationship of the relevance of the features among three pairs of features.

Vergara and Estevez [31] proposed an improvement for CMIM [30] approach based on the selection of the first relevant feature. The traditional method maximised the conditional mutual information to select relevant features. This author proposes the average of the *MI* to reduce the difference among chosen features.

Tapia *et al.* [32], [33] used the measures of *MI* to guide the selection of bits from the iris code to be used as features in gender prediction. Also, in [33] used complementary information to create clusters of the most relevant features based on information theory to classify gender from faces.

According to those previous works, we believed that *MI* is suitable for detecting morphed images to localised and detect the artefact present in morphed images using an efficient number of features.

III. METHODS

Figure 1 shows the proposed framework used in this paper, where a feature selection stage is added after traditional feature extraction approaches.

A. FEATURE EXTRACTION

Three different features were extracted from the morphing face images: Intensity, Texture and Shape.

1) INTENSITY

For raw data the intensity of the values in grayscale were used and normalised between 0 and 1.

2) UNIFORM LOCAL BINARY PATTERN

For texture, the histogram of uniform local binary pattern and Binary Statistical Image Feature were used [11], [34]. LBP is a gray-scale texture operator which characterises the spatial structure of the local image texture. Given a central pixel in the image, a binary pattern number is computed by comparing its value with those of its neighbours. The original operator used a 3×3 windows size. LBP features were computed from relative pixels intensities in a neighbourhood, as is show in the following equation:

$$LBP_{P,R}(x, y) = \bigcup_{(x',y') \in N(x,y)} h(I(x, y), I(x', y')) \qquad (1)$$

where $N(x, y)$ is vicinity around (x, y), $\cup$ is the concatenation operator, P is number of neighbours and R is the radius of the neighbourhood.

The uniform Local Binary Pattern (uLBP) was used as texture information. The uLBP was introduced, extending the original LBP operator to a circular neighbourhood with a different radius size and a small subset of LBP patterns selected. In this work we use, 'U2' which refers to a uniform pattern. LBP is called uniform when it contains at most 2 transitions from 0 to 1 or 1 to 0, which is considered to be a circular code. Thus, the number of patterns is reduced from 256 to 59 bins.

The reasons for omitting the non-uniform patterns are twofold. First, most of the LBP in natural images are uniform. It was noticed experimentally that uniform patterns account for a bit less than 90% of all patterns when using the (8,1) neighbourhood. In experiments with facial images, it was found that 90.6% of the patterns in the (8,1) neighbourhood and 85.2% of the patterns in the (8,2) neighbourhood are uniform [35]. The second reason for considering uniform patterns is the statistical robustness. Using uniform patterns instead of all the possible patterns has produced better recognition results in many applications. On one hand, there are indications that uniform patterns themselves are more stable, i.e. less prone to noise and on the other hand, considering only uniform patterns makes the number of possible LBP labels significantly lower and reliable estimation of their distribution requires fewer samples. See Figure 2.

The Binary Statistical Image Feature (BSIF) was also explored as texture method. BSIF is a local descriptor designed by binarising the responses to linear filters. The filters learn from thirteen natural images using independent component analysis (ICA). The code value of pixels is considered a local descriptor of the image intensity pattern in the pixels' surroundings. The value of each element (i.e., bit) in the binary code string is computed by binarising

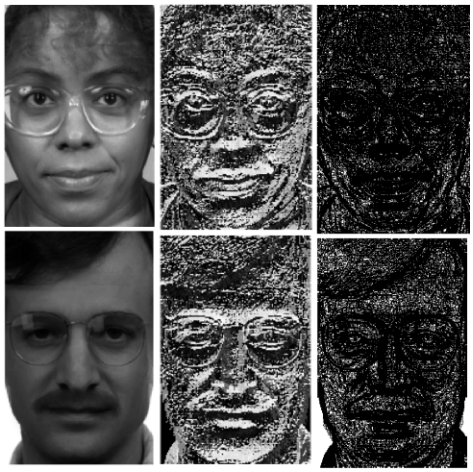

FIGURE 2. Example of LBP images. Left: Grayscale image. Middle: Traditional LBP (256 bins). Right: LBP with uniform pattern implementation (59 bins).

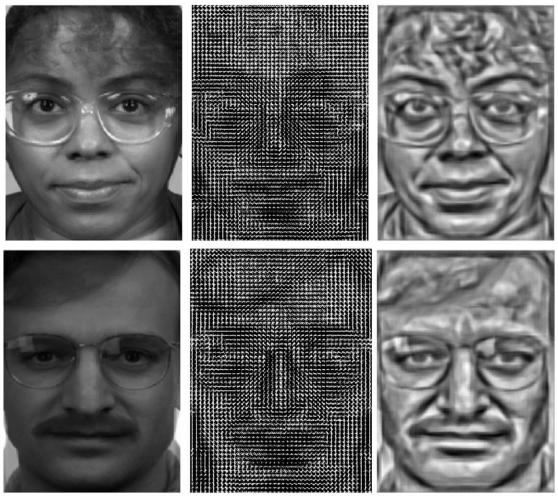

FIGURE 3. Example images of inverse HOG. Left: Morphed images. Middle: Traditional HOG. Right: Inverse HOG.

the response of a linear filter with a zero threshold. Each bit is associated with a different filter, and the length of the bit string determines the number of filters used. A grid search from the 60 filters available in BSIF implementation was explored. The filter 5×5 and 9 bits obtained the best results estimated from the baseline approach. The resulting BSIF images were used as an input of the classifiers.

3) INVERSE HISTOGRAM ORIENTED GRADIENT

From Shape, the inverse Histogram of oriented gradients [36], [37] were used. The Histogram of oriented gradient was proposed by Dalal and Triggs [37]. The distribution directions of gradients (oriented gradients) are used as features. Gradients, x, and y derivatives of an image are helpful because the magnitude of gradients is large around edges and corners (regions of abrupt intensity changes). We know that edges and corners contain more information about object shape than flat regions. However, this descriptor presents some problems. For instance, when we visualise the features for high-scoring false alarms in the object detection area, they are wrong in image space. They look very similar to true positives in feature space. To avoid this limitation that confuses the classifiers, we used the visualisation proposed by Vondrik *et al.* [36] to select the best parameters that allows us to visualise the artefacts contained in morphed images. This implementation used 10×12 blocks and 3×3 filter sizes. One example is shown in Figure 3.

B. FEATURE SELECTION

Feature selection (FS) is the process in which groups of features (Filter method) derived from image areas and textures respectively pixels (in raw images) from facial images out of a dataset are selected based on some measure or the correlation such as F-statistic, Logistic regression or *MI* between the features and the class of the labels. See Figure 4. It is closely related to feature extraction, a process in which feature vectors are created from the facial image. This takes place through domain transformation or manipulation of the data space and can be considered as selecting a subset of features.

Figure 4 shows a random morphed image with three different correlation metrics. The heat maps show the most correlated features in blue and the less correlated in red. All the features (relevant and redundant) are present in the image.

FS can be classified into three main groups: Filters, Wrappers, and Embedding methods [30]. A filter does not have a dependency with classifiers when looking for the most relevant features as it. Filters estimates the correlation values according to the *MI* values. Conversely, wrappers search for the most relevant features according to the classifier. Therefore, if the classifier changes, then the relevant features vary. The embedding method is looking to estimate an optimisation function according to the data and the classifier.

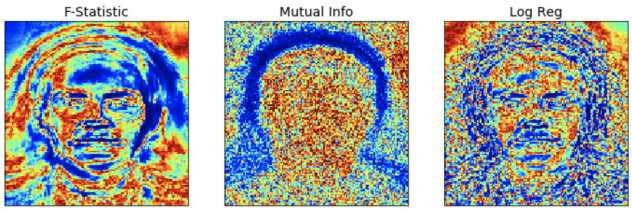

FIGURE 4. Example of morph image with different correlation metrics. Red pixels represent the less correlated features.

For this work, we propose to use a filter methods based on *MI* as correlation metrics to estimate the most relevant features to classify bona fide versus morphed face images.

C. MUTUAL INFORMATION

MI is defined as a measure of how much information is contained jointly in two variables or how much information of one variable determines the other variable [38]. *MI* is the foundation for information theoretic feature selection since it

provides a function for computing the relevance of a variable with respect to the target class [30]. The *MI* between two variables, x and y, is defined based on their joint probabilistic distribution $p(x, y)$ and the respective marginal probabilities $p(x)$ and $p(y)$ as:

$$MI(x, y) = \sum_{i,j} p(x_i, y_j) log \frac{p(x_i, y_j)}{p(x_i)p(y_j)}. \quad (2)$$

A categorical *MI* is used in this paper, which can be estimated by tallying the samples of categorical variables in the data building adaptive histograms to compute the joint probability distribution $p(x, y)$ and the marginal probabilities $p(x)$ and $p(y)$ based on the Fraser algorithm [39] for bona fide and morphing images. According to that, if more than two pairs of features reach the same value then, the information is **redundant** and must be minimised (Eq.3).

$$Red_{MI} = \frac{1}{|S|^2} \sum_{fi,fs \epsilon S} MI(fi; fs) \quad (3)$$

Conversely, if a couple of features is not contained in any, other pair of features is considered **relevant** and therefore can help to disentangle and separate the two classes. This information must be maximised (Eq.4).

$$Rel_{MI} = \frac{1}{S} \sum_{fi \epsilon S} MI(c; fi) \quad (4)$$

If a feature extracted from an image is randomly or uniformly distributed in different classes (bona fide or morph), then the *MI* between these classes is zero. If a feature is strongly differently expressed for other classes (morph), it should have a large *MI*. Thus, we use *MI* as a measure of the relevance of features presented in the images.

When 2 features are highly dependent on each other, The following protocol was used:

- Each image of size $M \times N$ was flattened to $1 \times M \times N$ for each class (bona fide and morphed).
- The matrix A is formed by K flattened images of size $1 \times M \times N$ features, and the class vector (c).
- *MI* for each pair of column of matrix A is estimated.
- The relevance (Rl) and redundancy (Rd) are estimated from matrix A.
- The trade-off between the relevance and redundancy (Rel and Red) matrices is estimated, sorted and indexes according to the *MI* values.
- A vector v with the index value of each column (feature) with the higher relevance and less redundant is formed.
- Only the N columns according to with index value are selected.
- A small matrix from A features and v images is conformed in the step of 100 features up to 1,000 features to be evaluated for the classifier.

Different implementations have been proposed in state-of-the-art [30] to estimate the trade-off between relevance and redundancy. Estimate all the combinations 2^N to remove all the redundancy is not possible because of high dimensionality problem. Then, the following methods based on *MI* and Conditional *MI* have been used and are described as follows:

D. MINIMUM REDUNDANCY MAXIMAL RELEVANCE (mRMR)

Two forms of combining relevance and redundancy operations are reported in [29]; *MI* difference (*MID*), and *MI* quotient (*MIQ*). Thus, the *mRMR* feature set is obtained by optimising *MID* and *MIQ* simultaneously. The trade-off both conditions requires to integrate them into a single criterion function [29] as follows:

$$f^{mRMR}(X_i) = MI(c; fi) - \frac{1}{S} \sum MI(fi; fs), \quad (5)$$

where, $MI(c; fi)$ measures the relevance of the feature f_i to be added for the class c, and the term $\frac{1}{S} \sum_{fi \epsilon S} MI(fi; fs)$ estimates the redundancy of the fi_{th} feature with respect to the previously selected features f_s to belong to set S.

E. NORMALISED MUTUAL INFORMATION FEATURE SELECTION (NMIFS)

Estevez *et al.* [40] proposed with the Normalised Mutual Information (NMIFS) an improved version of mRMR based on the normalised feature of *MI*. The *MI* between two random variables is bounded above by the minimum of their entropies H. As the entropy of a feature could vary greatly, this measure should be normalised before applying it to a global set of features as follows:

$$f^{NMIFS}(X_i) = MI(c; fi) - \frac{1}{S} \sum_{fi \epsilon S} MI_N(fi; fs) \quad (6)$$

where, MI_N is the normalised *MI* by the minimum entropy of both features, as defined in:

$$MI_N(fi; fs) = \frac{MI(fi; fs)}{min(H(fi), H(fs))} \quad (7)$$

F. CONDITIONAL MAXIMISATION MUTUAL INFORMATION (CMIM)

The *CMIM* criterion is a tri-variate measure of the information associated with a single feature about the class, conditioned upon an already selected feature [41]. It loops over the chosen features and assigns each candidate to feature a score based upon the lowest Conditional Mutual Information (*CMI*) between the features selected, the candidate feature, and the class [30], [41]. Then, the selected feature is the one with the maximum score.

$$CMIM = \begin{cases} arg\ max_{fi \in F} \{MI(fi; c) for\ S = \emptyset\} \\ arg\ max_{fi \in F/S} \left\{ min_{fj \in S} MI(fi; c/fj) \right\} \\ for\ S \neq \emptyset. \end{cases} \quad (8)$$

G. CONDITIONAL MAXIMISATION MUTUAL INFORMATION-2 (CMIM2)

The *CMIM* criterion selects relevant variables and avoids redundancy. However, it does not necessarily choose a variable that is complementary to the already chosen variables.

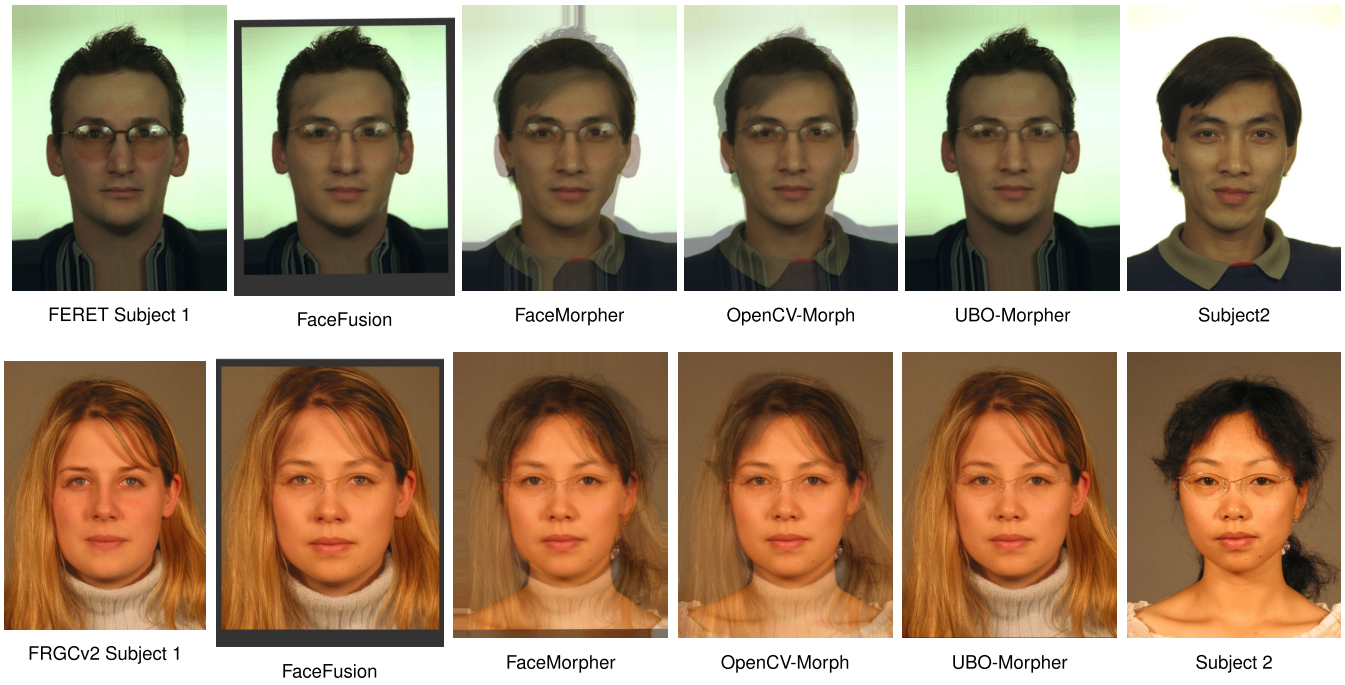

FIGURE 5. Examples of different morphing algorithms for two subjects in the FERET and FRGCv2 databases.

A variable with high complementarity information (max) to the already selected variable will be had by a high (CMI). In general, in problems where the variables are highly complementary (or dependent) to predict c, the $CMIM$ algorithm will fail to find that dependence among the variables. The $CMIM - 2$ [31] was proposed in order to improve $CMIM$ and changes the max function for the average function $(1/d)$. Then, the selected feature is the one with the average score.

$$MI(x, y) = 1/d \sum_{f,j \in S} MI(f_i; c \mid f_j). \tag{9}$$

IV. DATABASES

The FERET and FRGCv2 databases were used to create the morph images based on the protocol described by [4]. The AMSL Face Morph image dataset was also used to evaluated the best algorithm. A summary of the databases is presented in Table 2. All the images were captured in a controlled scenario and include variations in pose and illumination. FRGCv2 presents images more compliant to the passport portrait photo requirements. The images contain illumination variation, different sharpness and changes in the background. The original images have the size of 720×960 pixels. For this paper, the faces were detected, and images were resized and reduced to 180×240 pixels. These images still fulfill the resolution requirement of the intra-eye distance of 90 pixels defined by ICAO-9303-p9-2015.

The AMSL Face Morph image dataset (London DB) was created based on images from Face Research Lab London set.[2] This dataset includes genuine neutral and smiling faces as well as morphed face images. Also, all the images are ICAO compliant. The images were down-scaling to 531×413 pixels. These images present high-quality face morphed images [26].

The α value to define the contribution of each subject to morph image results was 0.5 for all the morph images. The morphing tools used are describe in Table 3.

Figure 5 shows examples of the morphing portrait images and the different output qualities with the artefact in their background. For instances FaceOpenCV implementation.

TABLE 2. Number of images used for FERET, FRGCv2 database and Face Research London dataset. Column 1, show the name of the dataset.

Database	Nº Subjects	Bona fide	Morphs
FRGCv2	533	984	964
FERET	529	529	529
Face Res. London DB	102	102	2,175

TABLE 3. Morphing tool software and number of images created by each method. The number of images is per dataset (FRGCv2/FERET).

Database	Nº Subjects	Bona fide	Morphs
FaceFusion	533/529	984/529	964/529
FaceMorpher	533/529	984/529	964/529
OpenCV-Morph	533/529	984/529	964/529
UBO-Morpher	533/529	984 /529	964/529
AMSL-FaceMorph	102	204	2,175

[2]https://figshare.com/articles/dataset/Face_Research_Lab_London_Set/5047666

The following algorithms were used to create morph images:

- FaceFusion is a proprietary morphing algorithm, developing for IOS app.[3] This algorithm to create high-quality morph images without visible artifact.
- FaceMorpher is an open-source algorithm to create morph images.[4] This algorithm introduce also some artifacts in the background.
- OpenCV-Morph, this algorithm is based on the OpenCV implementation.[5] The images contain visible artefacts in the background and some areas of the face.
- Face UBO-Morpher [42]. The University of Bologna developed this algorithm. The resulting images are of high quality without artefact in the background.
- AMSL-FaceMorph [26]. The images were created on Face Research Lab London Set with an in-house method. The resulting images are of high quality without artefact in the background.

As we mentioned before, after creation of the morphed images, all the faces were cropped using a modified dlib face detector implementation.[6] Figure 6 shows examples of the FERET cropped face database. We can observe that cropped images represent a more challenging scenario because all the background artefacts of the morphing process result were removed. However, some artefacts remain and can be observed in the images, for instances for the FaceMorpher and OpenCV-Morph implementations.

FIGURE 6. Examples of FERET cropped images. From left to right: Bona fide, FaceFusion, FaceMorpher, OpenCV-Morph, UBO-Morpher implementations.

V. EXPERIMENTS AND RESULTS

This section presents the quantitative results of the proposed scheme based on feature selection for automated single-morph attack detection. In addition to the proposed system, we evaluated six different contemporary classifiers such as K-Nearest Neighbors (KNN), Logistic regression (LOGIT), Support Vector Machine (SVM), Decision Tree (DT), Random Forrest (RF), and Multilayer Perceptron (MLP). Overall, Random Forest and SVM reached the best results. See Figure 7. To compare and to estimate the baseline method, only the Random Forest classifier was used.

The experiments, tested a leave-one-out (LOO) protocol and an RF classifier with 300 trees. These datasets allow subject-disjoint results to be computed; that is, no subject has an image in both the training and the testing subset.

The FERET and FRGCv2 databases were partitioned to have 60% training and 40% testing data for feature selection. The selection of features was made using only the training set. The London DB was used only for test as an unknown scenario. The output of the four methods delivers the index of each column of the matrix A that represents the more relevant features. The number of features were evaluated in steps of 100 features up to the end of the vector.

The performance of the detection algorithms is reported according to metrics defined in ISO/IEC 30107-3. The Attack Presentation Classification Error Rate (APCER) is defined as the proportion of attack presentations using the same attack instrument species incorrectly classified as bona fide in a specific scenario. The bona fide Presentation Classification Error Rate (BPCER) is defined as the proportion of bona fide images incorrectly classified as a morphing in the system. The D-EER is the operation point where APCER = BPCER is reported for the different morphing methods.

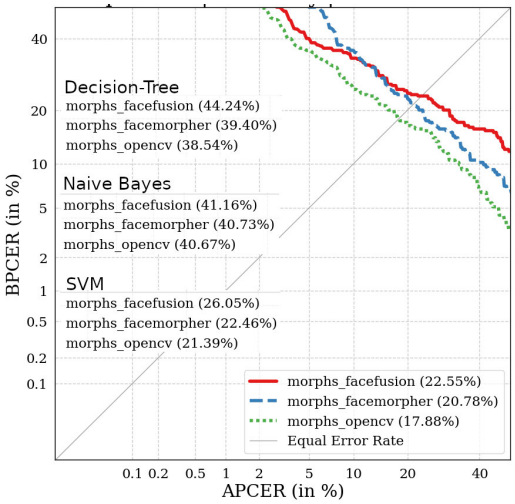

FIGURE 7. DET Curves comparing the baseline classifiers using RF. RF and SVM reached the best results. KNN, LOGIT and MLP are not showed in the curve because of poor results.

A. EXPERIMENT 1

Four different kinds of features were extracted from faces. Intensity, HOG, uLBP and BSIF. From raw images, we used the values of intensity of the pixels normalised between 0 and 1. For shape, we used the histogram of HOG. For texture, the histogram of the Uniform Local Binary Patterns (uLBP) and BSIF were used. For the uLBP all radii values were explored from uLBP81 to uLBP88. The fusion of LBPs was also investigated, concatenating the LBP81 up to LBP88 (FusionALL). The vertical (uLBP_VERT) and horizontal (uLBP_HOR) concatenation of the image divided into 8 patches also was explored. From BSIF the resulting images of filter $5 \times 5 - 9bits$ was used. After feature extraction, we fused that information at the feature level by concatenating the feature vectors from different sources

[3] www.wearemoment.com/FaceFusion/
[4] https://github.com/alyssaq/face_morpher
[5] www.learnopencv.com/face-morph-using-opencv-cpp-python
[6] https://www.pyimagesearch.com/2018/09/24/opencv-face-recognition/

TABLE 4. Baseline performance reported in % of D-EER for FERET LOO trained on FaceFusion and FaceMorpher.

Train	FACEFUSION				Train	FACEMORPHER			
Method	FACEMORPHER	OpenCV-Morph	UBO-MORPHER	Average	Method	FACEFUSION	OpenCV-Morph	UBO-MORPHER	Average
RAW	37.47	35.67	41.35	38.16	RAW	49.23	23.53	49.6	40.79
HOG	38.83	40.47	40.4	39.90	HOG4	42.03	37.14	42.07	40.41
LBP1	27.35	32.33	38.53	32.74	LBP1	45.45	32.88	42.67	40.33
LBP2	24.01	27.46	37.31	29.59	LBP2	42.8	30.65	41.79	38.41
LBP3	24.88	26.92	37.03	29.61	LBP3	40.28	26.55	40.45	35.76
LBP4	23.25	24.32	36.24	27.94	LBP4	38.76	25.29	40.91	34.99
LBP5	24.55	26.25	38.7	29.83	LBP5	36.14	30.14	38.85	35.04
LBP6	25.79	26.98	38.95	30.57	LBP6	35.71	27.49	40.27	34.49
LBP7	27.78	28.37	40.42	32.19	LBP7	37.72	26.43	42.2	35.45
LBP8	28.88	27.73	42.47	33.03	LBP8	38.01	27.21	43.59	36.27
FusionALL	23.71	26.34	38.02	29.36	FusionALL	38.59	30.05	40.33	36.32
LBP_VERT	26.76	30.1	23.98	26.95	LBP_VERT	40.99	24.81	42.68	36.16
LBP_HOR	26.59	28.66	37.69	30.98	LBP_HOR	41.95	25.06	42.71	36.57
FUSION	43.64	44.56	46.88	45.03	FUSION	44.28	32.31	46.68	41.09
BSIF(5x5-9)	32.63	30.94	31.86	31.81	FUSION	40.18	28.21	42.90	37.09

TABLE 5. Baseline performance reported in % of D-EER for FERET LOO trained on OpenCV-Morph and UBO-Morpher.

Train	OpenCV-Morph				Train	UBO-MORPHER			
Method	FACEFUSION	FACEMORPHER	UBO-MORPHER	Average	Method	FACEFUSION	FACEMORPHER	UBO-MORPHER	Average
RAW	47.45	20.21	48.82	38.83	RAW	35.55	40.45	37.46	37.82
HOG	43.17	35.7	40.032	39.63	HOG	39.77	35.74	35.82	37.11
LBP1	44.6	25.72	41.68	37.33	LBP1	42.6	27.45	32.93	34.33
LBP2	41.28	24.85	40.28	35.47	LBP2	40.28	25.26	29.86	31.80
LBP3	37.66	23.95	39.52	33.71	LBP3	35.99	24.97	25.58	28.85
LBP4	36.08	22.72	38.52	32.44	LBP4	36.03	24.33	26.99	29.12
LBP5	35.76	25.56	38.5	33.27	LBP5	34.31	26.94	28.64	29.96
LBP6	37.03	28.74	40.58	35.45	LBP6	37.74	30.36	30.21	32.77
LBP7	37.74	25.51	42.47	35.24	LBP7	38.56	32.08	31.5	34.05
LBP8	37.66	27.47	42.86	36.00	LBP8	39.79	34.58	32.16	35.51
FusionALL	42.23	22.48	41.36	35.36	FusionALL	41.58	25.78	29.81	**32.39**
LBP_VERT	40.6	24.77	42.63	36.00	LBP_VERT	38.51	29.6	31.26	33.12
LBP_HOR	41.5	23.4	42.45	35.78	LBP_HOR	38.66	29.55	31.7	33.30
FUSION	44.85	28.84	45.92	39.87	FUSION	46.17	43.57	44.05	44.60
BSIF(5x5-9)	42.23	29.53	42.43	38.06	FUSION	32.88	32.49	29.49	31.62

TABLE 6. Baseline performance reported in % of D-EER for FRGCv2 LOO trained on FaceMorpher and FaceFusion.

Train	FACEFUSION				Train	FACEMORPHER			
Method	FACEMORPHER	OpenCV-Morph	UBO-MORPHER	Average		FACEFUSION	OpenCV-Morph	UBO-MORPHER	Average
RAW	25.1	23.92	27.91	25.64	RAW	41.71	13.97	42.3	32.66
HOG	26.4	27.02	29.89	27.77	HOG4	30.93	24.03	32.38	29.11
LBP1	17.61	10.8	17.44	15.28	LBP1	22.6	9.57	19.77	17.31
LBP2	14.18	11.4	19.2	14.93	LBP2	20.97	13.13	19.49	17.86
LBP3	10.58	13.36	21.67	15.20	LBP3	20.46	9.53	20.9	16.96
LBP4	11.71	13.58	22.88	16.06	LBP4	20.34	10.32	23	17.89
LBP5	13.43	14.16	25.25	17.61	LBP5	20.73	10.69	25.43	18.95
LBP6	14.61	15.43	28.87	19.64	LBP6	21.38	11.37	26.74	19.83
LBP7	15.88	15.78	26.2	19.29	LBP7	21.91	11.01	26.7	19.87
LBP8	15.96	16.29	26.06	19.44	LBP8	24	12.22	27.42	21.21
FusionALL	10.05	12.38	20.36	14.26	FusionALL	22.44	7.99	21.64	17.36
LBP_VERT	13.81	14.43	20.9	16.38	LBP_VERT	20.96	11.22	22.41	18.20
LBP_HOR	13.45	13.74	19.85	15.68	LBP_HOR	20.57	11	21.1	17.56
FUSION	13.4	16.09	27.68	19.06	FUSION2	27.79	15.21	26.81	23.27
BSIF(5x5-9)	26.12	23.90	25.77	25.26	FUSION	31.54	19.19	31.43	27.38

(Intensity, HOG, and uLBP) into a single feature vector that becomes the input to the classifier (FUSION). The classifier was trained with each feature extraction method's selected features and the fused chosen features.

Tables 4, 5, and Tables 6, 7 show the baseline results for the intensity, shape and texture feature extraction methods for FERET and FRGC respectively. This baseline was estimated using a leave-one-out protocol for all the morphing methods. The intensity (Raw) and HOG reached the higher D-EER (worst results). Most of the time, the (FusionALL) obtained the lower average D-EER results (Best results).

Table 4 shows the results on the left side for the FERET database were trained with FaceFusion and tested with FaceMorpher, OpenCV-Morph, and UBO-Morpher. Right side, trained with FaceMorpher and tested with FaceFusion, OpenCV-Morph, and UBO-Morpher.

Table 5 shows the results on the left side for FERET database were trained with OpenCV-Morph and tested with FaceFusion, FaceMorpher, and UBO-Morpher. Right side, trained with UBO-Morpher and tested with FaceFusion, FaceMorpher and OpenCV-Morph. The same protocol was applied to Tables 6 and 7 with FRGCv2 database.

TABLE 7. Baseline performance reported in % of D-EER for FRGCv2 LOO trained on OpenCV-Morph and UBO-Morpher.

Train	OpenCV-Morph				Train	UBO-MORPHER			
Method	FACEFUSION	FACEMORPHER	UBO-MORPHER	Average	Method	FACEFUSION	FACEMORPHER	OpenCV-Morph	Average
RAW	40.3	13.07	41.3	31.56	RAW	22.58	23.83	22.24	22.88
HOG	31.21	22.17	32.44	28.61	HOG	27.44	25.38	26.5	26.44
LBP1	20.94	13.29	17.88	17.37	LBP1	20.54	6.32	9.92	12.26
LBP2	20.52	8.43	18.89	15.95	LBP2	20.18	7.38	9.86	12.47
LBP3	19.57	7.26	20.11	15.65	LBP3	19.59	9.54	11.76	13.63
LBP4	20.76	8.05	23.1	17.30	LBP4	18.99	10.69	12.65	14.11
LBP5	20.4	9.1	24.63	18.04	LBP5	20.38	13.28	13.64	15.77
LBP6	21.66	10.66	26.57	19.63	LBP6	21.19	15.7	15.82	17.57
LBP7	22.64	10.68	26.84	20.05	LBP7	20.26	16.85	16.47	17.86
LBP8	23.5	11.94	27.78	21.07	LBP8	22.71	19.37	19.52	20.53
FusionALL	21.55	5.79	21.49	16.28	FusionALL	18.2	7.51	9.52	11.74
LBP_VERT	21.22	9.45	22.33	17.67	LBP_VERT	18.19	13.77	14.07	15.34
LBP_HOR	20.98	9.62	21.97	17.52	LBP_HOR	18.34	13.69	13.9	15.31
FUSION	27.22	11.7	26.31	21.74	FUSION	27.59	11.24	13.22	17.35
BSIF(5x5-9)	33.08	20.54	32.98	28.86	FUSION	24.00	22.94	20.63	22.52

B. EXPERIMENT 2

This experiment explores the application of the proposed method based on feature selection. The four feature selection methods, mRMR, NMIFS, CMIM, and CMIM2, were applied in order to reduce the size of the data and estimate the position of the relevant features before entering classifiers from Intensity, HOG, uLBP and BISF. The best 5,000 from 43,200 features were extracted from the raw data (intensity). The best 1,000 from 1,048 features were extracted from HOG, and the best 400 features from 472 were selected from the fusion of uLBP (FusionALL).

Table 8 and 9 show the results for FERET and FRGCv2 databases for single morphed detection from the best feature selected from **HOG** applied to FaceFusion, FaceMorpher, OpenCV-Morph and UBO-Morpher. The results reported shown an improved in comparison to the baseline in Experiment 1 using the HOG features extracted of the images. The number of feature was reduced on average down to 10%. This reduction would enable the application in mobile devices hardware and also allow us to see the localisation of the most relevant features.

TABLE 8. D-EER in % of HOG + Fea / FERET. The figures in parenthesis represent the best number of features for each method.

	FaceFusion (bestFea)	FaceMorpher (bestFea)	OpenCV-Morph (bestFea)	UBO-Morpher (bestFea)
mRMR	17.15 (400)	7.07 (700)	6.15 (700)	15.68 (100)
NMIFS	19.98 (300)	9.74 (300)	5.84 (800)	13.88 (300)
CMIM	17.83 (300)	5.84 (600)	7.07 (500)	11.07 (400)
CMIM2	8.12 (300)	4.92 (900)	6.15 (500)	13.52 (300)

Table 10 and 11 show the results for FERET and FRGCv2 database for single morphed detection from the best feature selected from the **fusion of uLBP** (LBP8,1 up to LBP 8,8) applied to FaceFusion, FaceMorpher, OpenCV-Morph and UBO-Morpher. The results reported shown an improved in comparison to the Experiment 1 using all the features extracted of the images. The number of feature also is reduced on average down to 10% for texture features.

TABLE 9. D-EER in % of HOG + Fea / FRGCv2. The figures in parenthesis represent the best number of features for each method.

	FaceFusion (bestFea)	FaceMorpher (bestFea)	OpenCV-Morph (bestFea)	UBO-Morpher (bestFea)
mRMR	6.83 (1000)	15.06 (900)	2.17 (1000)	4.99 (500)
NMIFS	4.83 (600)	2.1 (900)	1.68 (900)	4.73 (700)
CMIM	6.50 (700)	1.83 (400)	2.17 (1000)	3.83 (500)
CMIM2	6.65 (900)	1.92 (900)	1.50 (1000)	4.02 (600)

TABLE 10. D-EER in % of fusion uLBP + Fea / FERET. The figures in parenthesis represent the best number of features for each method.

	FaceFusion (bestFea)	FaceMorpher (bestFea)	OpenCV-Morph (bestFea)	UBO-Morpher (bestFea)
mRMR	22.7 (200)	12.92 (400)	13.84 (400)	23.37 (200)
NMIFS	21.22 (100)	11.84 (300)	12.30 (400)	23,98 (200)
CMIM	21.53 (100)	12.30 (200)	11.84 (400)	22.75 (200)
CMIM2	18.45 (100)	10.45 (400)	11.68 (400)	12.11 (200)

TABLE 11. D-EER in % of fusion uLBP + Fea / FRGCv2. The figures in parenthesis represent the best number of features for each method.

	FaceFusion (bestFea)	FaceMorpher (bestFea)	OpenCV-Morph (bestFea)	UBO-Morpher (bestFea)
mRMR	9.15 (400)	1.60 (200)	8.54 (300)	10.47 (300)
NMIFS	9.65 (400)	1.49 (200	4.50 (400)	8.99 (400)
CMIM	8.52 (300)	1.09 (100)	4.17 (200)	8.70 (300)
CMIM2	7.53 (300)	1,33 (200)	3.99 (400)	8.15 (400)

Figure 8 shows the accuracy obtained for the UBO-Morpher tool when features selected were applied from **intensity** features. The UBO-Morpher constitutes a high-quality morphing implementation and then is used and analysed on FERET and FRGCv2 databases. Conversely, FaceMorpher is the more straightforward method to be detected based on the artefacts present in the images. The mRMR and NMIFS methods based on *MI* obtained the lower results. The method based

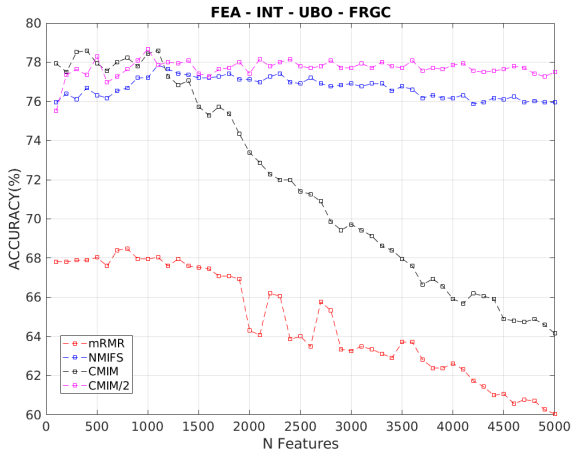

FIGURE 8. FRGCv2 feature selection for intensity features. X axis represents the number of the best features. Y axis represents the accuracy in %.

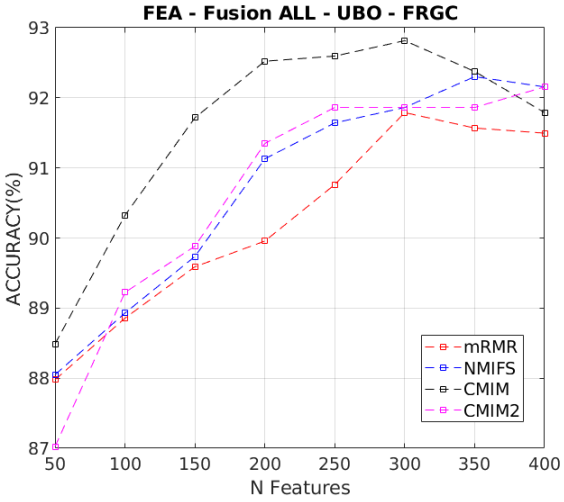

FIGURE 10. FRGCv2 feature selection for FusionALL features (uLBP fusion). X axis represents the number of the best features. Y axis represents the accuracy in %.

on conditional *MI* (CMIM and CMIM-2) reached the best results. These results show that the complementary information captures the relationship between the feature selected and the feature candidate in a better way. CMIM with only 400 features and CMIM-2 with 1,000 features reached higher accuracy and lower D-EER.

Figure 9 shows the accuracy obtained for the UBO-Morpher tool, when feature selected were applied from **HOG** features. Again, The method mRMR and NMIFS based on *MI* obtained the lower results. The method based on conditional *MI* (CMIM and CMIM-2) reached the best results with 500 and 600 features respectively.

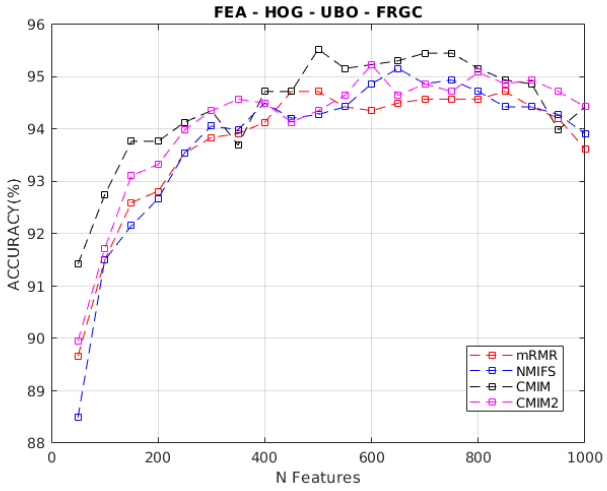

FIGURE 9. FRGCv2 feature selection for HOG features. X axis represents the number of the best features. Y axis represents the accuracy in %.

Figure 10 shows the accuracy obtained for the UBO-Morpher tool, when feature selected were applied from the **fusion of uLBP**. This time NMIFS and CMIM reached the best results with 300 and 400 features respectively. Consolidating the Conditional *MI* over traditional *MI*.

Table 12 shows the D-EER for HOG feature with the best method CMIM2. Surprisingly, the shape feature (HOG) reached the best results with the lower D-EER using CMIM and FRGCv2 database. FaceMorpher reached the lower D-EER with 1.8% with a BPCER10 of 0.3% and BPCER20 of 1.0%. Conversely, FaceFusion and London DB reached the higher D-EER of 5.8% and 9.4% respectively. The second column shows, the comparison (D-EER) between the HOG results from baseline using all the HOG features versus the proposed method with feature selected from HOG.

TABLE 12. D-EER in % for the best results reached by CMIM using HOG.

FRGCv2 - HOG	HOG / Fea+HOG (D-EER)	BPCER10	BPCER20
FaceFusion	27.7/**5.8**	3.7	7.7
FaceMorpher	29.1/**1.8**	0.3	1.0
OpenCV-Morpher	28.6/**2.0**	0.0	0.0
UBO-Morpher	26.4/**4.0**	2.0	4.4
London DB	39.8/**9.4**	2.8	9.8

Table 13 shows the D-EER for FusionALL (uLBP Fusion) feature with the best method CMIM. For FRGCv2 database the best results with the lower D-EER using CMIM-2. FaceMorpher again reached the lower D-EER with 1.3% with a BPCER10 of 0.3% and BPCER20 of 1.0%. Conversely, UBO-Morpher reached the higher D-EER of 9.4% with a BPCER10 of 2.9% and BPCER20 of 13.8%. The second column shows, the comparison (D-EER) between the FusionALL results from baseline using only the fusion of uLBP features versus the proposed method with feature selected from uLBP.

Figure 11 show the DET curves obtained for the four feature selected method for the three feature selected (Intensity, Texture and Shape). The UBO-Morpher constitutes

TABLE 13. D-EER in % for the best results reached by CMIM using FusionALL (uLBP fusion).

FRGCv2-uLBP	uLBP / Fea+uLBP (D-EER)	BPCER10	BPCER20
FaceFusion	14.2/**9.2**	7.4	20.0
FaceMorpher	17.3/**1.3**	0.3	1.0
OpenCV-Morpher	16.2/**4.0**	1.3	4.4
UBO-Morpher	11.7/**9.4**	2.9	13.8
London DB	30.9/**8.4**	3.3	10.8

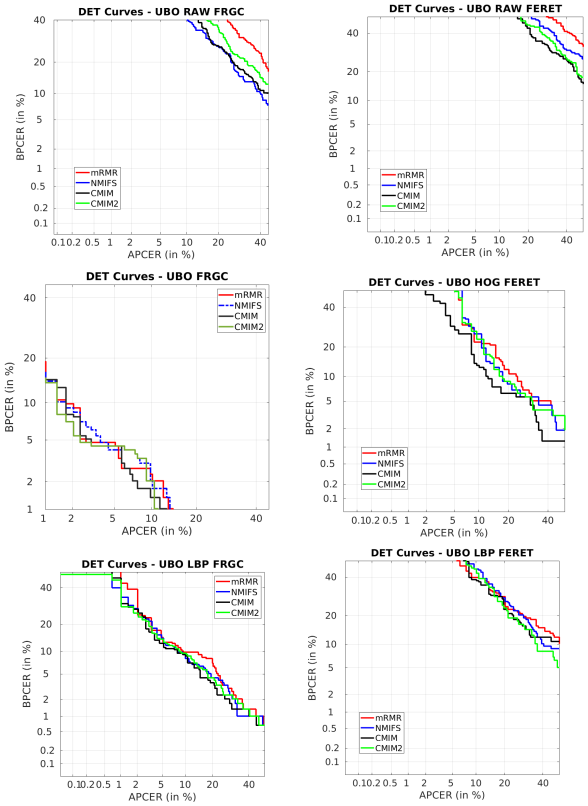

FIGURE 11. DET curves for FRGCv2 and FERET using feature selection method with 300 features. Top: RAW. Middle: HOG. Bottom: uLBP Fusion (FusionALL).

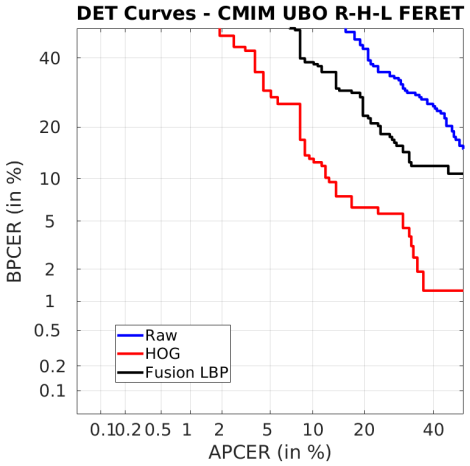

FIGURE 12. D-EER for comparison of the features selected using CMIM from intensity, shape and texture for FERET database. R: represents RAW. H: represents HOG and L, represents uLBP.

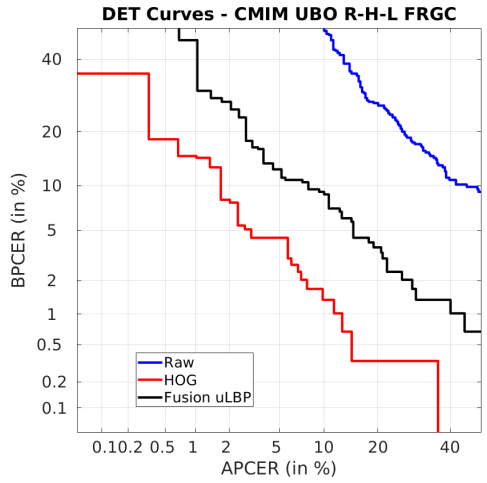

FIGURE 13. D-EER for comparison of the features selected using CMIM from intensity, shape and texture for FRGCv2 database. R: represents RAW. H: represents HOG and L, represents uLBP.

a high-quality morphing implementation and is applied on FERET and FRGCv2 databases. Conversely, FaceMorpher is the more straightforward method to be detected based on the artefacts present in the images. The features selection applied to intensities values reached the lower results. Even these results improve the baseline, the D-EER are not competitive with the literature. Conversely, uLBP and HOG improve a lot in comparison with the baseline and reached results competitive with the literature as is shown in Tables 12 and 13.

In order to compare and analysed which extracted feature delivers more useful information for the detection task, the Figures 12 and 13 shows a comparison of FERET and FRGCv2 for best results obtained by CMIM from intensity, shape (HOG) and texture (uLBP). Both figures have shown that HOG reached a lower D-EER in both databases. This result shows that the shape algorithms also can detect morphing images as a complement of textures. The exploration parameters to find the most representative inverse HOG features and their visualisation allows us to improve the results. This is shown in Figure 3.

C. EXPERIMENT 3

This experiment explores the application of the best result methods based on feature selection using the open-access AMSL Face dataset. Experiments 1 and 2 were performed using a leave one out protocol. Then this experiment follows the same rule. Table 12 and 13 show the results of London DB when features were extracted using the best feature selected for FRGCv2 and FERET with CMIM method. We can see that FRGCv2 generalise better than FERET with a Fusion of uLBP and HOG features. The number of images used for training makes the difference in the generalisation problem. London DB morphed scenario is very challenging with more

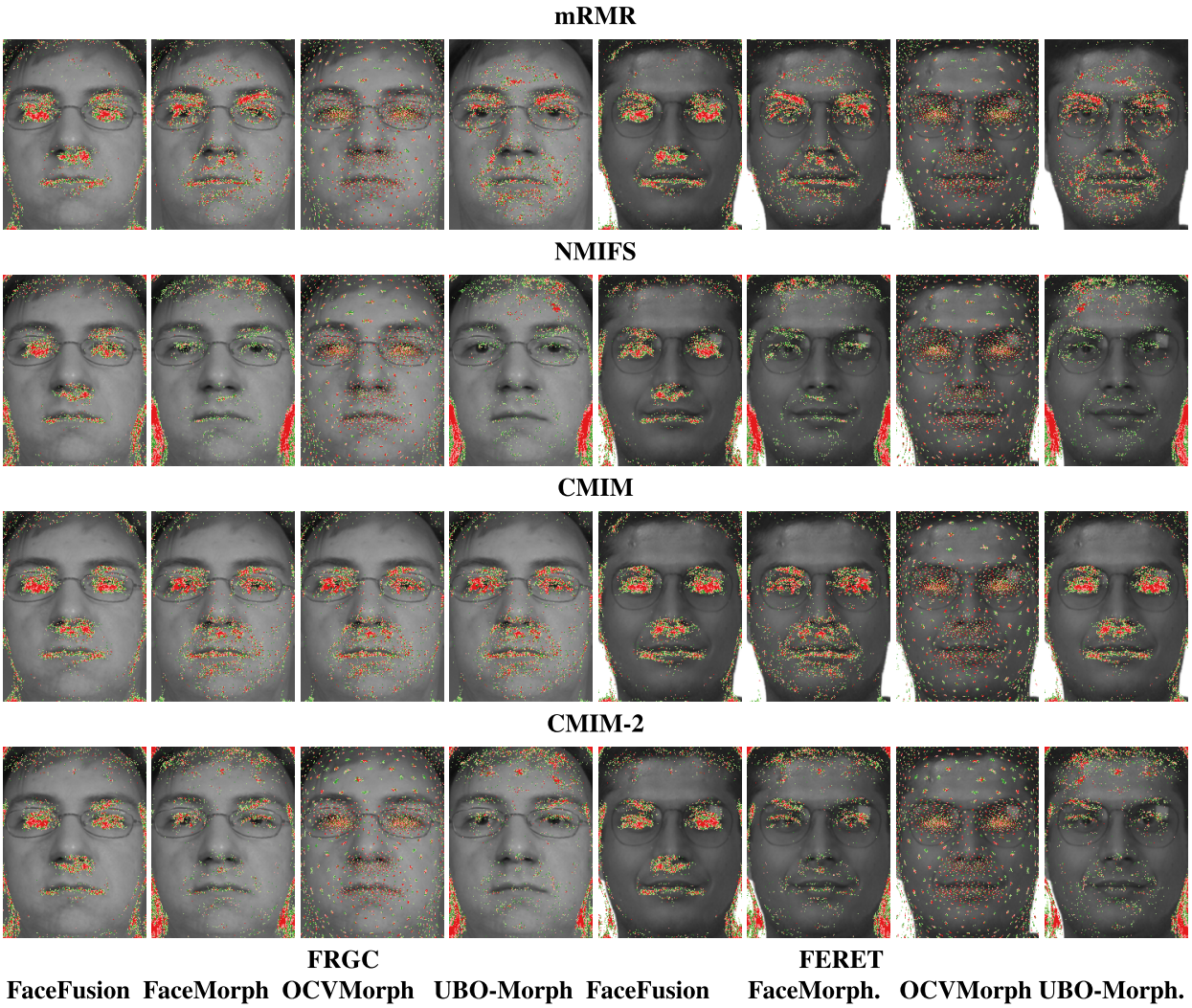

FIGURE 14. Localisation of the feature selected by mRMM, NMIFS, CMIM and CMIM2 for different morphing algorithm. Left: FRGC. Right: FERET. Each image shows the best 5.000 features.

than 2,000 morph images. Overall the quality of London DB is very similar to the UBO-morph database. For the feature selected from HOG, the D-EER decreased from 39.8% (baseline) to 9.4%, reached a BPCER10 of 2.8%. For the chosen feature from the fusing uLBP, the D-EER decrease from 30.9% (baseline) to 8.4% obtained a BPCER10 of 3.3%.

VI. VISUALISATION

Once we select the best features, it is possible to recover the coordinates of the features into the images. Then, we can visualise the attributes for each method. Figure 14 shows the localisation of the most relevant features for FRGCv2 and FERET morph random image. The 5,000 features were divided into five equal parts and assigned to five different colours. The most relevant features from 1 to 1,000 are represented as red pixels. From 1,001 to 2,000 are pink. 2,001 to 3,000 are green. 3,001 to 4,000 are light green, and 4,001 to 5,000 are represented as blue. Figure 15 show the best feature selected for CMIM method. The Red pixels represent the most relevant and less redundant features. Conversely, the green pixels are the most redundant and confuse the classifiers. It is essential to highlight that the pixels in colours represent the best features selected, which means the most relevant and less redundant from the four methods: mRMR, NMIFS, CMIM, and CMIM2, from 1,000 up to 5,000. The CMIM features are distributed in all the images and only concentrate in some areas. The CMIM focalised the features in the most relevant areas. The eyes and the nose areas are selected as relevant to detect morphed images.

Regarding the areas selected, the feature selection methods look for the best features that allow separated the two classes (Bona fide and Morphs). Many images, even crop faces are used still have some artefacts in the background or the border of faces. Then the FS found these areas as applicable to classify the images. These results can be visualised as a wrong selection because the features are outside of the faces. However, if these areas are observed in detail, the artefacts are still present, then the Bona fide and morphed

FIGURE 15. All features selected by CMIM sorted by ranking. Left: FERET morphed image. Middle: FRGC morphed image. Right: AMSL morphed image.

images can be separated. It is essential to point out that the feature selection method measures the relationship among all features. Then, the interaction of the best features selected should be analysed in context (all together) and should not be visualised individually.

Figure 15 shows all the features (43,200) according to the relevance index. This index was associated with four colours in the following order: Red, Pink, Green and Light Green. Those images clearly show the most critical areas and where the modification on the face was performed to create a morphed image for FRGCv2, FERET and London DB.

VII. CONCLUSION

After analysing all the results, we can conclude that morphing based on the FERET and London DB databases are more challenging to detect than the FRGC database. The leave-one-out protocol is essential to estimate the actual performance of the proposed method. In the literature, the test set typically contains images from the same morphing tools. The feature selection reduces the number of features used drastically to separate bona fide for morphed images and reduce the D-EER in all the cases. For the feature selected from HOG, the D-EER decreased from 26.4% (baseline) to 4.0% for UBO-Morpher, reached a BPCER10 of 2.0%. For the chosen feature from the fusing uLBP, the D-EER decrease from 11.7% (baseline) to 8.4% obtained a BPCER10 of 2.9%. These results are very competitive with the state of art. These results can be confirmed with the good performance obtained with London DB. For the feature selected from HOG, the D-EER decreased from 39.8% (baseline) to 9.4%, reached a BPCER10 of 2.8%. For the chosen feature from the fusing uLBP, the D-EER decrease from 30.9% (baseline) to 8.4% obtained a BPCER10 of 3.3%. The localisation of the features enabled us to select the most relevant and less redundant features. The nose and eyes are identified as relevant areas in the face for manual analysis of morphed images. This tool may help the border police detect morphing images and address the areas to be analysed for the artefacts. In summary, the shape feature (HOG) results outperform the texture performance as is shown in Figures 12 and 13. In future work, we will apply this method to embedding features extracted from the face-recognition system in order to choose the best features. Also, we will prepare the algorithm to be evaluated in the NIST platform.

DISCLAIMER

This text reflects only the author's views, and the commission is not liable for any use that may be made of the information contained therein.

REFERENCES

[1] S. Gonzalez, A. Valenzuela, and J. Tapia, "Hybrid two-stage architecture for tampering detection of chipless ID cards," *IEEE Trans. Biometrics, Behav., Identity Sci.*, vol. 3, no. 1, pp. 89–100, Jan. 2021.

[2] M. Ferrara, A. Franco, and D. Maltoni, "The magic passport," in *Proc. IEEE Int. Joint Conf. Biometrics*, Sep. 2014, pp. 1–7.

[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] 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.

[5] S. Venkatesh, R. Ramachandra, K. Raja, L. Spreeuwers, R. Veldhuis, and C. Busch, "Morphed face detection based on deep color residual noise," in *Proc. 9th Int. Conf. Image Process. Theory, Tools Appl. (IPTA)*, Nov. 2019, pp. 1–6.

[6] S. Venkatesh, R. Ramachandra, K. Raja, and C. Busch, "Face morphing attack generation & detection: A comprehensive survey," *IEEE Trans. Technol. Soc.*, vol. 2, no. 3, pp. 128–145, Sep. 2020.

[7] M. Ngan, P. J. Grother, K. K. Hanaoka, and J. Kuo, *Face Recognition Vendor Test (FRVT) Part 4: Morph-Performance of Automated Face Morph Detection*. Gaithersburg, MD, USA: National Institute of Technology, 2020.

[8] 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.

[9] U. Scherhag, R. Raghavendra, K. B. Raja, M. Gomez-Barrero, C. Rathgeb, and C. Busch, "On the vulnerability of face recognition systems towards morphed face attacks," in *Proc. 5th Int. Workshop Biometrics Forensics (IWBF)*, Apr. 2017, pp. 1–6.

[10] V. Ojansivu and J. Heikkilä, "Blur insensitive texture classification using local phase quantization," in *Proc. Int. Conf. Image Signal Process.*, vol. 5099, Jul. 2008, pp. 236–243.

[11] J. Kannala and E. Rahtu, "BSIF: Binarized statistical image features," in *Proc. 21st Int. Conf. Pattern Recognit.*, Jan. 2012, pp. 1363–1366.

[12] 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)*, Oct. 2017, pp. 555–563.

[13] 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.
[14] C. R. U. Scherhag and C. Busch, "Performance variation of morphed face image detection algorithms across different datasets," in *Proc. Int. Workshop Biometrics Forensics*, 2018, pp. 1–6.
[15] M. El Rhazi, A. Zarghili, A. Majda, A. Bouzalmat, and A. A. Oufkir, "Facial beauty analysis by age and gender," *Int. J. Intell. Syst. Technol. Appl.*, vol. 18, nos. 1–2, pp. 179–203, 2019, doi: 10.1504/IJISTA.2019.097757.
[16] L. Debiasi, C. Rathgeb, U. Scherhag, A. Uhl, and C. Busch, "PRNU variance analysis for morphed face image detection," in *Proc. IEEE 9th Int. Conf. Biometrics Theory, Appl. Syst. (BTAS)*, Oct. 2018, pp. 1–9.
[17] L. Debiasi, U. Scherhag, C. Rathgeb, A. Uhl, and C. Busch, "PRNU-based detection of morphed face images," in *Proc. Int. Workshop Biometrics Forensics (IWBF)*, Jun. 2018, pp. 1–7.
[18] P. Phillips, P. Flynn, T. Scruggs, K. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, "Overview of the face recognition grand challenge," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 1, Jun. 2005, pp. 947–954.
[19] L.-B. Zhang, F. Peng, and M. Long, "Face morphing detection using Fourier spectrum of sensor pattern noise," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Jul. 2018, pp. 1–6.
[20] 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.
[21] 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.
[22] L. Debiasi, N. Damer, A. Moseguí-Saladié, C. Rathgeb, U. Scherhag, C. Busch, F. Kirchbuchner, and A. Uhl, "On the detection of gan-based face morphs using established morph detectors," in *Proc. 20th Int. Conf. Image Anal. Process. (ICIAP)*, in Lecture Notes in Computer Science, vol. 11752, Trento, Italy, E. Ricci, Ed. Cham, Switzerland: Springer, Sep. 2019, pp. 345–356.
[23] V. Kazemi and J. Sullivan, "One millisecond face alignment with an ensemble of regression trees," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 1867–1874.
[24] R. Raghavendra, S. Venkatesh, K. Raja, and C. Busch, "Detecting face morphing attacks with collaborative representation of steerable features," in *Proc. 3rd Int. Conf. Comput. Vis. Image Process.*, Sep. 2018, pp. 255–265.
[25] A. Makrushin, T. Neubert, and J. Dittmann, "Automatic generation and detection of visually faultless facial morphs," in *Proc. Int. Conf. Comput. Vis. Theory Appl.*, Mar. 2017, pp. 39–50.
[26] T. Neubert, A. Makrushin, M. Hildebrandt, C. Kräetzer, 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.
[27] C. Seibold, A. Hilsmann, and P. Eisert, "Reflection analysis for face morphing attack detection," in *Proc. 26th Eur. Signal Process. Conf. (EUSIPCO)*, Sep. 2018, pp. 1022–1026.
[28] T. Ojala, T. Maenpaa, M. Pietikainen, J. Viertola, J. Kyllonen, and S. Huovinen, "Outex-new framework for empirical evaluation of texture analysis algorithms," in *Proc. Object Recognit. Interact. Service Robots*, vol. 1, 2002, pp. 701–706.
[29] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 8, pp. 1226–1238, Aug. 2005.
[30] I. Guyon, S. Gunn, M. Nikravesh, and L. A. Zadeh, *Feature Extraction, Foundations and Applications, Studies in Fuzziness and Soft Computing*. Secaucus, NJ, USA: Springer-Verlag, 2006.
[31] J. Vergara and P. Estevez, "CMIM-2: An enhanced conditional mutual information maximization criterion for feature selection," *J. Appl. Comput. Sci. Methods*, vol. 2, pp. 5–20, Jan. 2010.
[32] J. Tapia, C. Perez, and K. Bowyer, "Gender classification from the same iris code used for recognition," *IEEE Trans. Inf. Forensics Security*, vol. 11, no. 8, pp. 1760–1770, Aug. 2016.
[33] J. E. Tapia and C. A. Perez, "Clusters of features using complementary information applied to gender classification from face images," *IEEE Access*, vol. 7, pp. 79374–79387, 2019.
[34] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul. 2002.
[35] J. E. Tapia, C. A. Perez, and K. W. Bowyer, "Gender classification from iris images using fusion of uniform local binary patterns," in *Computer Vision—ECCV 2014 Workshops*, L. Agapito, M. M. Bronstein, and C. Rother, Eds. Cham, Switzerland: Springer, 2015, pp. 751–763.
[36] C. Vondrick, A. Khosla, T. Malisiewicz, and A. Torralba, "HOGgles: Visualizing object detection features," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 1–8.
[37] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 1, Jun. 2005, pp. 886–893.
[38] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, C. C. O. N. Y. Donald and L. Schilling, Eds. Hoboken, NJ, USA: Wiley, 1991.
[39] A. M. Fraser and H. L. Swinney, "Independent coordinates for strange attractors from mutual information," *Phys. Rev. A, Gen. Phys.*, vol. 33, no. 2, pp. 1134–1140, Feb. 1986.
[40] P. A. Estévez, M. Tesmer, C. A. Perez, and J. M. Zurada, "Normalized mutual information feature selection," *IEEE Trans. Neural Netw.*, vol. 20, no. 2, pp. 189–201, Feb. 2009.
[41] F. Fleuret, "Fast binary feature selection with conditional mutual information," *J. Mach. Learn. Res.*, vol. 5, pp. 1531–1555, Nov. 2004.
[42] M. Ferrara, A. Franco, and D. Maltoni, "Face demorphing," *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 4, pp. 1008–1017, Apr. 2018.

JUAN E. TAPIA (Member, IEEE) received the P.E. degree in electronics engineering from Universidad Mayor, in 2004, the M.S. degree in electrical engineering from the Universidad de Chile, in 2012, and the Ph.D. degree from the Department of Electrical Engineering, Universidad de Chile, in 2016. In addition, he spent one year doing internship at the University of Notre Dame, USA. From 2016 to 2017, he was an Assistant Professor at Universidad Andrés Bello. From 2018 to 2020, he was the Research and Development Director of the area of electricity and electronics at the Universidad Tecnológica de Chile—INACAP. He is currently a Senior Researcher at Hochschule Darmstadt (HDA) and the Research and Development Director of TOC Biometrics. His main research interests include pattern recognition and deep learning applied to iris/face biometrics, vulnerability analysis, morphing, feature fusion, and feature selection. In 2016, he received the award for best Ph.D. thesis.

CHRISTOPH BUSCH (Senior Member, IEEE) is a member of the Department of Information Security and Communication Technology (IIK), Norwegian University of Science and Technology (NTNU), Norway. He holds a joint appointment with the Computer Science Faculty, Hochschule Darmstadt (HDA), Germany. He lectures biometric systems course at the Technical University of Denmark, Denmark, since 2007. On behalf of the German BSI, he has been the Coordinator for the project series BioIS, BioFace, BioFinger, BioKeyS Pilot-DB, KBEinweg, and NFIQ2.0. In the European research program, he was an Initiator of the Integrated Project 3D-Face, FIDELITY, and iMARS. Furthermore, he was/is a partner in the projects TURBINE, BEST Network, ORIGINS, INGRESS, PIDaaS, SOTAMD, RESPECT, and TReSPAsS. He is also the Principal Investigator of the German National Research Center for Applied Cybersecurity (ATHENE). Moreover, he is the Co-Founder and a member of the Board of the European Association for Biometrics (www.eab.org) that was established in 2011 and assembles in the meantime more than 200 institutional members. He has coauthored more than 500 technical papers and has been a speaker at international conferences. He is a member of the Editorial Board of the IET journal.