Face Presentation Attack Detection using Nuisance Attribute Projection

Javier Cubelos Ordás
Faculty of Electrical Engineering
University of Ljubljana
jc3872@student.uni-lj.si

January 14, 2018

Abstract

Face recognition is today the second most deployed biometric technology and its performance has improved remarkably in the last decade. However, it's also the biometric technology where most spoofing attacks are detected. Due to it, a research revolution took place and lots of countermeasures for spoofing attacks have been developed in the last few years. Nevertheless, most of these techniques have been only tested in one dataset or see its performance considerably affected when it's tested in another dataset than the original one it was build for. The Nuisance Attribute Projection (NAP) has been used since decades to compensate the 'channel' effects in speaker recognition, and since a couple of years it has started been used to other purposes. In this paper, we study the possibility of using NAP as a, or as part of a countermeasure technique, for spoofing attacks to face recognition systems. In fact, we evaluate if there is relevant information contained in the NAP subspace for detecting attacks. To address this problem, we compute the NAP subspace, normalize the testing image, extract some features from it and classify them, all of it using two different datasets: CASIA's face anti-spoofing database and OULU-NPU's mobile face presentation database with real-world variations, in order to check the performance of our technique when using it in different datasets.

1. Introduction

Nowadays, biometric [37] provides greater security and convenience than traditional methods of personal recognition. In fact, since its first works around 40 years ago, mainly oriented to automatic voice and face recognition [12], [24], [39], a new vision appeared where the user started becoming its own password or key. Since then, researchers from different fields such as computer vision, image processing or pattern recognition focused their work in this promising technology.

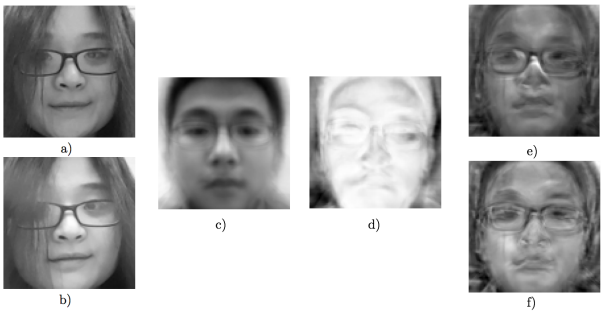

Figure 1. Examples of the whole working system for one subject. Image a) shows one image of a genuine attempt to a face recognition system from OULU-NPU dataset, whereas image b) shows a frame from a spoofing attack to a face recognition system from the same dataset. Then, in image c) we can observe the global mean of all the genuine attempts of the training set of OULU's dataset, and image d) depicts an example of a representation of the NAP directions of the NAP subspace. Finally, images e) and f) show the 'error' corresponding to the NAP artefacts found in the genuine and spoofing frames, respectively.

However, due to this unstoppable technological evolution, new concerns about the security and privacy of biometric technology emerged. In fact, public confidence and acceptance of the biometric systems will depend on the ability to demonstrate that these technologies are robust, have low error rates and are tamper-proof (*a.k.a.* biometric spoofing) [32].

A hacker might present a copy of a known person's biometric sample to the system. For example, in 2008, a hacker club published German interior minister's fingerprint [70], supporter of the collection of citizens' unique physical characteristics as a means of preventing terrorism. But this is just an example of a big list, including real criminal scenes [21], [58], [59], [60] and even attacks to big technological companies, as happened to Apple with the iPhone 5S fingerprint reader [34], or, more recently, with the iPhone X face ID [61], both hacked the first day after they hit the shelves.

These attacks have even been reported from live demonstrations in biometric conferences [28], [69].

To try to solve these security problems surrounding biometric technology, lots of researchers all around the world have started searching for the best technique (or combination of different techniques) to countermeasure biometric spoofing. Different approaches have been considered, as we will explain in the related work section, providing really good results in the used databases during the research. However, the performance of almost all these approaches decreases dramatically when they are tested in different databases than the original one.

In this paper we try to address this gap by studying if the Nuisance Attribute Projection (NAP) could be used as a countermeasure technique for face spoofing attacks, that may provide a more stable performance when working alone (or in combination with other techniques) with different databases. To do it, we have first computed the nuisance subspace, obtained the 'error' corresponding to the NAP 'channel' artefacts and then evaluated if there is relevant information contained in this subspace. To evaluate the usefulness of this information, different methods have been performed, using the full 'error' image obtained with NAP and extracting the features contained in this subspace to use them with further classification methods.

For this evaluation, we used the CASIA face anti-spoofing database [84] and the OULU-NPU mobile face presentation database with real-world variations [13], both containing genuine real accesses and masks and video replay spoofing attacks to face recognition systems, as we explain more in detail in the experimental section. An example of the 'error' resulting of the NAP projection for a spoofing and a genuine attack can be observed in the Figure 1, where also the global mean of OULU-NPU dataset and an example representation of the NAP directions are shown. In summary, the main contributions of this paper are:

- The computation of the nuisance subspace of the training set of images.
- The analysis of the information contained in this subspace ('channel' effects) with different normalization methods, such as using the 'error' images obtained through NAP or the extraction of the Histogram of Oriented Gradients (HoG) descriptors of these images, for posterior classification using and comparing different classifier as the Multilayer Perceptron, Random Forest or Support Vector Machine (SVM).
- The evaluation of this technique with two different datasets (CASIA and OULU-NPU), checking how well our system distinguished spoofing attacks from genuine accesses to face recognition systems.

The rest of paper is structured as follows: First, the related work is covered, through a brief description of the typical face spoofing techniques, the principal approaches in which the researchers have focused their research work of face anti-spoofing countermeasures, and an introduction to the nuisance attribute projection. Then, the proposed method is presented, explaining in detail the core of the work presented in this paper, and focusing on the nuisance subspace computation and the different methods of further evaluation. After that, the experimental results are presented, presenting more in detail the two selected databases and the corresponding performance metrics and results. Finally, a short conclusion of the main contributions is shown, with the corresponding possible implications and improvements that could be covered in future work.

2. Related work

2.1. Face spoofing

Face recognition has been confirmed by the International Biometric Group (IBG) as the second largely deployed biometric in terms of market quota [36]. Nevertheless, it's also the biometric technology where most spoofing attacks and research have been detected, just after the fingerprint biometrics, leader of biometrics' market quota.

The use of masks to avoid being recognized has been present in almost all the well-known civilizations since centuries. Trends haven't changed so much, as today the use of plastic surgery [10] is becoming more and more popular, due to costs reduction and speed improvement. However, more elementary methods have also been used for attacking face recognition systems; for instance, using basic masks [58] or even only wearing regular make-up [23], face biometric systems can be spoofed.

Even if these techniques (*i.e.*, plastic surgery, face masks, make-up) have been traditionally used to hide the attacker's identity instead of trying to impersonate another user, it has been proven that they could even been used for direct attacks. For example, in Tabula Rasa's spoofing challenge conference [69], a woman succeeded to access in the place of a man into a face biometric systems, only by using some make-up.

Almost all the face spoofing techniques may be classified in two main groups based on the shape of the artefact used during the attack: 2D surfaces (*i.e.*, photos and video attacks) or 3D volumes (*i.e.*, mask attacks).

These artefacts have been used to perform three main types of attacks: ·

- **Photo Attacks:** Attacks performed presenting a picture of the genuine user to the biometric system. The attacker may have taken the picture or even have obtained it from the internet (*i.e.*, social networks [55]), and print it or even show it in a digital device [28] as

a mobile phone. A more advanced technique is the use of high-resolution printed masks where mouth and eyes have been cut out, making the attacker able to reproduce typical face movements such as eye blinking.

- **Video Attacks:** More complex attacks where the impostor replays a video of the genuine person using a digital device such as his mobile phone or laptop [64]. These attacks are more difficult to detect, as the dynamics of the face are also copied.

- **Mask Attacks:** The impostor uses a 3D mask of the genuine user's face, making more difficult its detection. In this case the complete 3D structure of the user face is replicated, making the use of depth cues (useful for the two previous types of attacks) inefficient.

These different types of attacks may be already found in the face spoofing databases available for researchers [5], [30], [84].

2.2. Face anti-spoofing

The creation of the Tabula Rasa european project [68] in 2010, focused on the study of the spoofing attacks to biometric systems, led to the revolution in anti-spoofing research. In addition, the distribution of several public face spoofing databases (mentioned before) [5], [19], [29], [30], [78], [84], was the other important factor that encouraged the development of anti-spoofing techniques, as the researcher could directly start focusing in the implementation of countermeasures.

In this section we will go through the works that have addressed countermeasures for biometric spoofing. It's difficult to decide which technique is better than the rest, as the performance depends of the type of attack and these methods loose accuracy when they are tested in different databases. Therefore, sometimes the best results may be obtained through the combination of several of these methods [30], [31].

We will cover the anti-spoofing techniques classifying them into four groups depending on their approaches:

- **Feature Level Dynamic Approaches**

It appeared as an anti-spoofing method against photo attacks which use printed faces. These methods are based in the detection of motion over a face video sequence. They mostly study the trajectory of face segments such as eye-blinking [38], [44], [54], [62], [80] or face and head gestures (*i.e.* smiling, nodding...). The latest being detected through face and gaze tracking [3], [11] or through optical flow estimation [4], [7], [43], [45]. The analysis of these trajectories makes possible the checking of face's liveness and, therefore, allows the discrimination between real faces and printed versions. However, these methods loose considerable accuracy when trying to detect video attacks where the face's movement is also replicated. Several research branches have appeared to try to overcome this problem such as obtaining 3D structure of the face by the analysis of 2D images with different poses [27], [81], using context-based analysis to also exploit the non-facial captured information [42], [63], [82], estimating the noise [22], obtaining temporal information [25], comparing face dynamics with other rigid objects dynamics (*i.e.* photos or masks) [47] or enhancing the motion in a video [8].

- **Feature Level Static Approaches**

This approach appeared due to the duration limitation of the dynamic approach. It's focused on the analysis of one single static image, making it faster than the previous method, sometimes at the cost of performance decrease. Most of these methods are based on the analysis of face texture using different image processing tools such as Fourier Spectrum [53]; multiple Difference of Gaussian (DoG) filters to extract frequency information [84] or even a combination of DoG with Lambertian Model [78]; providing good results also with bad illuminations [66]; partial least squares for low-level descriptors [73], its combination with high-level descriptors [83]; using Local Binary Patterns (LBP) [19], [49], its combination with shape information extracted using Histogram of Oriented Gradient (HOG) [57]; detecting paper's microtextures [6], [19], [56]; using video context (upper body location [48], pixel difference between consecutive frames [41]). Some of these techniques have been successfully combined at feature level showing improved accuracy [40], [44]. Also, comparative studies show that the fusion of static and dynamic techniques provides the best performance [30], [31].

- **Sensor Level Approaches**

Some approaches have been proposed using information outside the visual spectrum, such as infrared (IR) or near infrared (NIR) images, even able to distinguish between identical twins [65], [67]; or the use of LEDs and photodiodes to compare reflectance of real faces and fake materials [9], [85]. In addition, some personal authentication technologies could be potentially useful for anti-spoofing such as thermal imaging [15], [35]; facial vein pattern [14], [74]; or 3D face acquisition [50], [51], [52]. Finally, some multi-modality techniques emerged combining face and voice for detecting attacks [16], [17], [18], [46], [71] (*i.e.* lip movement).

- **Score Level Approaches**

This approaches are more recent and they focus on the topic of score-level anti-spoofing strategies for 2D face recognition systems. For instance, one of these approaches consists on the study of the impact of anti-spoofing techniques on the performance of face recognition systems, by analysing different score fusion techniques [20], and the

combination of anti-spoofing modules, as the static and dynamic fusions mentioned above [55], [72], in order to improve the performance when we change of dataset [26].

As we mentioned before, almost all these techniques demonstrated really good performance while working in the database that was originally used during the research. Nevertheless, only some of these techniques have been tested in more than one dataset, and even then, the performance gets really worst when the approach is tested in a different database. In the next section we will introduce the Nuisance Attribute Projection (NAP), that will be later studied as a possible cross-database stable countermeasure technique for biometric spoofing.

2.3. Nuisance attribute projection (NAP)

NAP is an important technique originally used in the field of speaker recognition for compensation of channel effects regardless of its source, the main problem of automated speaker recognition [75], [76]. The channel effects compensated by NAP are assumed to lie a low dimensional variability subspace. In the field of biometrics in unconstrained environment, the variability sources are mixed and unknown.

Previous works proposed a normalization scheme based in NAP to remove the illumination artifacts [77], by projecting away multiple dimensions of a low dimensional illumination space. Then, based on this work, another research took place, applying NAP this time to compensate for any kind of variability factors that affect the face recognition performance [79]. Both researches open the possibility of trying to use NAP for detecting attacks into face recognition systems, as we will cover in this paper.

In this paper we will compute the NAP subspace and evaluate if the nuisance factors of a face contain relevant information for face spoofing detection.

3. The proposed method: NAP subspace and normalization

In this section we present our proposed procedure, which consists on the computation of the NAP subspace and the normalization of the extracted information for further evaluation. To do it, we have first used the training data to obtain the projection matrix to the NAP subspace. Then we have used this matrix to normalize all the dataset's images in this NAP subspace, to finally obtain the variability attributes that will be used during the evaluation to classify genuine and spoofing attempts in face recognition systems.

3.1. NAP compensation scheme

One of the most exploited variants of the NAP technique projects away multiple dimensions of a specifically designed subspace, called NAP (or nuisance) subspace, with the goal of reducing the channel induced variability. In our case, the variability attributes ('channel' effects) would be all the effects induced to the image due to the face spoofing attack (print or video replay attacks), such as reflections, shadows... In fact, these variability attributes are what we finally plan to obtain use to distinguish between genuine and spoofing attempts during the evaluation.

Consider a set of n images (in vector form) of size $N = a \times b$ pixels arranged into a $n \times N$ column data matrix $\mathbf{X} = [x_1, x_2, ..., x_n]$. NAP tries to remove any additive distortion in the images as follows:

$$\mathbf{X'} = \mathbf{P}(\mathbf{X} - \mathbf{M}), \tag{1}$$

where **X'** denotes the new data matrix with compensated 'channel' effects, **M** denotes a matrix of the same size as **X** containing in each of its columns the global mean of the images in **X** and **P** stands for the $n \times n$ projection matrix:

$$\mathbf{P} = \mathbf{I} - \sum_{i=1}^{d'} w_i w_i^T. \tag{2}$$

Here, **I** denotes the $n \times n$ identity matrix, w_i represents the i-th NAP direction, d' stands for the number of NAP directions.

Now, when looking at Eq. (2) we may ask ourselves how we determine the d', which corresponds to the number of NAP directions needed for the compensation scheme. To answer this question lets assume that we have n_{C_j} sample images from the j-th class, where $C_1, C_2, ..., C_r$ represents the class labels of the images in **X**.

For each of these images we can write:

$$x_{C_{j,k}} = \hat{x}_{C_{j,k}} + c_k \tag{3}$$

Here, we modified some of our notations, the symbols used in Eq. (3) denoting: C_j - the class label of the image, k - index of the image in the j-th class, $\hat{x}_{C_{j,k}}$ - the channel-free part of $x_{C_{j,k}}$, c_k - the vector encoding the channel effects for the k-th image of the j-th class.

Lets assume that the channel effect vector c_k represents a random variable drawn from the standardized normal distribution $N(0, 1)$. Then the class-conditional sample mean is defined:

$$\mu_{C_j} = \frac{1}{n_{C_j}}\left(\sum_{k=1}^{n_{C_j}} \hat{x}_{C_{j,k}} + \sum_{k=1}^{n_{C_j}} c_k\right) = \frac{1}{n_{C_j}} \sum_{k=1}^{n_{C_j}} \hat{x}_{C_{j,k}}. \tag{4}$$

The above expression suggests that the mean value of each of the r classes μ_{C_j} $(j = 1, 2, ..., r)$ represents a channel-free estimate of an image from the j-th class (if $n_{C_j} >> 1$). Thus, by removing the corresponding class means from each image in the data matrix **X** we arrive at a new data

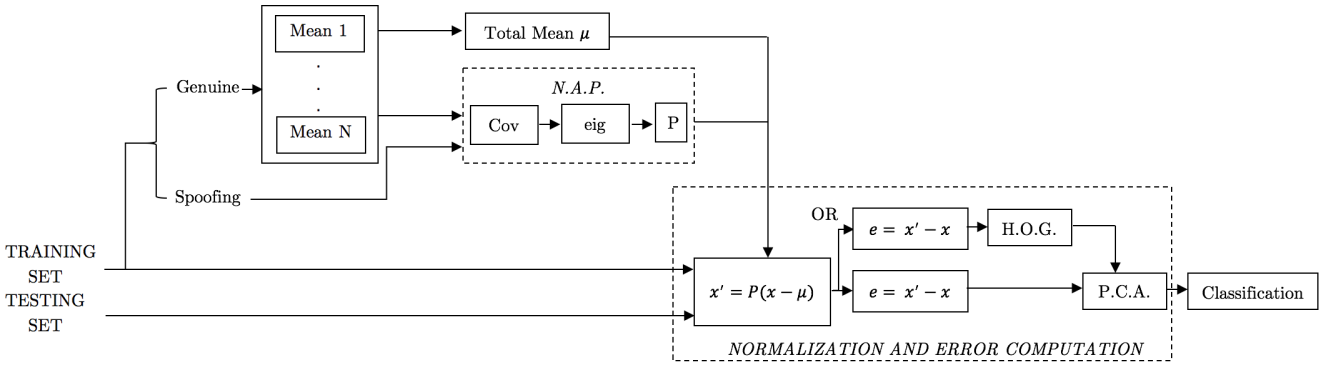

Figure 2. Overview of the proposed method. The procedure first computes the means of the genuine attempts of the different training subjects, to then calculate the NAP subspace P projection matrix by calculating the covariance and extracting the eigenvector of the spoofing training attempts. This matrix is then used, as well as the total mean, to normalize the images of the dataset and compute the 'error' image. This 'error' is then, or used to extract the HOG descriptors, or directly used, to compute the Principal Component Analysis (PCA) and finally classify the testing images.

matrix containing only information about the channel effects in the original data, that's exactly what we are looking for. If we wanted to remove these effects from the input data, we would have to estimate the NAP directions w_i $(i = 1, 2, ..., d')$ that correspond to the principal axes of the scatter matrix Σ_w:

$$\Sigma_w = \sum_{j=1}^{r} \sum_{k=1}^{n_{C_j}} (x_{C_j} - \mu_{C_j})(x_{C_j} - \mu_{C_j})^T, \qquad (5)$$

where the axes are computed as the leading eigenvectors of the following eigenproblem:

$$\Sigma_w w_i = \lambda_i w_i, i = 1, 2, ..., d' \leq n - r \qquad (6)$$

3.2. Normalization and 'error' computation

Starting from the Eq. (1), we notice that the channel effects can be removed from the facial image using the presented NAP compensation scheme, by estimating the NAP directions corresponding to the artefacts. Any input image x (in vector form) can easily then be normalized with respect to this artefacts by projecting away a number of directions in the NAP subspace. The normalization procedure can be written as:

$$x' = \mathbf{P}(x - \mu), \qquad (7)$$

where μ represents the global mean of the images in $\mathbf{X}$.

In our case, as we want to use the information present in the variability attributes, we just have to subtract this normalized version of the testing images to the corresponding original testing images, obtaining the 'error' representing the artefacts resulting of the NAP subspace computation:

$$e = x - x'. \qquad (8)$$

In the experiments made in the next section, we used this 'error' images and also extracted the HoG descriptors from it, to evaluate how this information could be useful for a classifier in order to identify the spoofing attacks.

An overview schema of the whole proposed method explained in this section is shown in Figure 2.

4. Experimental results

In this section we first introduce both datasets, then explain briefly the pre-processing applied to the training and testing sets of these datasets and finally present the results obtained after feature extraction and classification.

4.1. Datasets

To study the effectiveness of the use of NAP as a countermeasure for anti-spoofing, two popular face anti-spoofing databases were chosen, named CASIA face anti-spoofing database [84], and, OULU-NPU mobile face presentation database with real-world variations [13].

The first, the CASIA face anti-spoofing database, contains 600 video clips, 12 videos of each of the 50 genuine subjects. This dataset covers three different type of attacks: warped photo attacks, cut photo attacks and video playback attacks; with three different image qualities (named low, normal and high quality). In fact, each subject contains 12 videos (3 genuine and 9 fake), one for each category and quality. The training set consists on 20 of these subjects and the resting 30 subjects are leaved for testing. In Figure 3 the overall data for one subject is shown.

The second, the OULU-NPU mobile face presentation database with real-world variations, contains 4950 video clips that where recorded with 6 different smartphones, divided between 55 subjects. To consider the real-world variations, the real videos and attack materials were collected in 3 sessions with different acquisition conditions. The attack types considered in the OULU-NPU database are print (two different printers) and video-replay (two different dis-

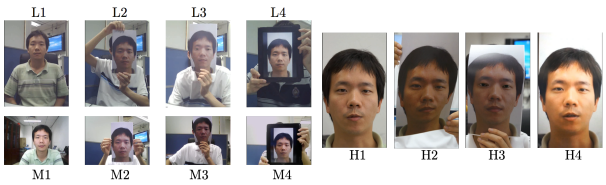

Figure 3. One complete video set of CASIA dataset for a subject. The left top four images represent the low quality videos (L1-L4), the left bottom are the normal quality videos (N1-N4), and the right are the high quality videos (H1-H4). For each quality, from left to right are genuine, warped photo attack, cut photo attack and video attack.

plays). Due to the big amount of video clips present in this dataset, we have just used the clips from one of the different smartphones (considering the 3 capture sessions and different attacks). Both testing and training parts consist on 20 different subjects. In Figure 4 the overall data captured for one subject with one smarpthone in one of the capture sessions is shown.

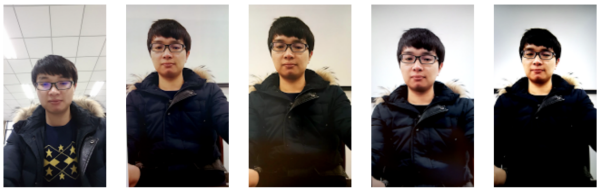

Figure 4. One complete video set of OULU-NPU dataset for a subject captured with one smartphone in one of the captured sessions. From left to right, the images represent the real genuine attempt, the two print attacks and the two video replay attacks.

4.2. Pre-processing: training and testing

As the final purpose of this experimental section was to test if the nuisance subspace contains relevant information for spoofing attacks detection, and this will be tested using some classifier, we first need to pre-process the datasets' videos.

For each of the videos (genuine and spoofing) from each of the sets (training and testing) of both datasets (CASIA and OULU-NPU), we extracted the frames and extracted the grayscale cropped region corresponding to the face detected on it using Viola-Jones algorithm. Finally, to speed up the computation, we decided to resize the cropped faces to 100 x 100 pixels.

Finally, for CASIA's dataset, we got a total of 45400 images for training (12980 genuine and 32420 spoofing) and 79680 images for testing (27060 genuine and 52620 spoofing). Then, for OULU-NPU dataset, we got a total of 22715 images for training (2843 genuine and 19872 spoofing) and 25272 images for testing (4582 genuine and 20690 spoofing). An example of a pre-processed genuine and spoofing images is shown in the Figure 5.

Figure 5. Example of a pre-processed training genuine (left) and spoofing images (right) from CASIA dataset.

Once we had the training and testing sets ready, we computed the NAP subspace as described in the previous section. During these experiments, we tried to see how the different classifiers (that will be introduced later) will perform depending on the number of NAP directions. After testing different NAP directions values as *d'=25,50,75,100,200,500* and classifying the resulting NAP 'errors' we concluded that the results were optimal for a NAP directions value of 50, so we took *d'=50* as default value for the rest of experiments.

4.3. Feature extraction

Now it was time to check if this 'error' information, contained in the NAP subspace, has relevant information for spoofing detection. However, not only the whole 'error' images were used in the evaluation of classifiers' performance. In fact, different feature extraction approaches where considered, such as the extraction of the LBP (Local Binary Patterns) [1] or of SIFT (Scale-Invariant Feature Transform) [33] or HOG (Histogram of Oriented Gradients) [2] descriptors.

After testing these different approaches, we decided to keep going the classification step with the HOG features and the original 'error' images from the NAP subspace, as both approaches provided considerably better results compared to the rest of feature extraction methods.

4.4. Principal Component Analysis (PCA)

As we are working with images of 100x100 pixels, the 'error' images will have the same dimension (10000 dimensional vector). On the other hand, for HOG we found that the best results without comprising computational problems were obtained for a cellsize of 4 during the HOG descriptors extraction, resulting in 19375 dimensional vectors as HOG descriptors of one image. This could definitely cause problems during classification due to the hardware constraints of the computer used during the experiments, which lacks of GPU.

To solve this issue, and to try to uncorrelate the different variables of the feature vectors, PCA (Principal Component Analysis) was performed in top of these vectors. To obtain the optimal dimensionality of the PCA, we checked for each database, the number of eigenvalues of the diagonal

eigenvalues matrix, presented in the Eq. (6), that must be summed up in order to satisfy the following relation:

$$\frac{\lambda_1 + \lambda_2 + ...}{\sum_{i=1}^{n} \lambda_i} > 0.90 \quad (9)$$

where the number of eigenvalues needed in the numerator of the fraction to satisfy this relation, represents the dimensionality of PCA. After checking this for both datasets, we decide to use a PCA dimensionality of 50 for OULU-NPU's dataset and of 25 for CASIA's one.

4.5. Classification and scoring

The performance of different classifiers recognizing spoofing attacks has been tested, such as the Naïve Bayes classifier, the Random Forest classifier, the Multilayer Perceptron (MLP) classifier and, finally, the Support Vector Machine (SVM) classifier. The results of the classification of the NAP 'error' images of the testing images and the corresponding HOG's descriptors obtained from them are reflected in the Table 1.

CLASSIFIER	TP RATE	FP RATE	F-MEASURE	ROC AREA
a.1) CASIA DATASET – FULL ERROR IMAGES				
NAÏVE BAYES	65.2 %	34.3 %	66.1 %	67.5 %
RANDOM FOREST	68.3 %	34.5 %	68.9 %	74.5 %
M.L.P.	69.5 %	28.8 %	70.3 %	74.3 %
S.V.M.	68.2 %	32.1 %	68.9 %	73.0 %
a.2) CASIA DATASET – H.O.G. FEATURES				
NAÏVE BAYES	63.6 %	35.0 %	64.6 %	66.3 %
RANDOM FOREST	67.9 %	33.5 %	68.6 %	73.1 %
M.L.P.	68.0 %	28.7 %	68.9 %	73.5 %
S.V.M.	68.5 %	32.2 %	69.2 %	74.4 %
b.1) OULU-NPU DATASET – FULL ERROR IMAGES				
NAÏVE BAYES	53.3 %	13.6 %	57.4 %	81.2 %
RANDOM FOREST	87.3 %	48.4 %	85.6 %	86.6 %
M.L.P.	72.2 %	25.9 %	75.2 %	82.0 %
S.V.M.	62.6 %	25.3 %	66.9 %	81.1 %
b.2) OULU-NPU DATASET – H.O.G. FEATURES				
NAÏVE BAYES	74.9 %	27.5 %	77.3 %	79.3 %
RANDOM FOREST	88.3 %	42.7 %	87.1 %	88.2 %
M.L.P.	81.1 %	21.5 %	82.6 %	88.5 %
S.V.M.	72.9 %	17.9 %	75.9 %	88.1 %

Table 1. Performance results of 'error' images and HOG features classification with Naïve Bayes, Random Forest, MLP and SVM classifiers, for CASIA and OULU-NPU datasets.

Different metrics have been used to compare the performance of the different classifiers:

- True Positive Rate (TPR): it measures the quantity of positives that are correctly identified as such.
- False Positive Rate (FPR): it measures the quantity of negatives that are wrongly identified as positives.
- F-Measure: the harmonic mean of the precision (fraction of relevant instances among the retrieved instances) and recall (fraction of relevant instances over the total amount of relevant instances). This metric is better than the two metrics that it comprises, as it make easier the performance comparison of different classifiers.
- ROC Area: is the area under the receiver operating characteristic curve (ROC). As ROC is the most used metric for analysing the performance of classifiers and reflects the TPR against the FPR, its area is really representative of the classifier performance.

The ROC curves of the different experimental tests described above can be observed in the Figure 6, where *a.1)* and *b.1)* correspond to the performance of full 'error' images for CASIA's and OULU-NPU's datasets respectively. Then, *a.2)* and *b.2)* curves represent respectively CASIA's and OULU-NPU's performance while using HOG descriptors as features.

Taking a closer look to the Table 1, we can observe that Random Forest classifier and MLP classifier are the ones that perform better globally, obtaining TPR up to 88.3% for OULU-NPU dataset and of 69.5% for CASIA dataset. The rest of metrics present in the table make possible to select one classifier among the previous two. For example, as it can be seen, Random Forest classifier usually performs similar or better considering TPR, F-measure or the area under the ROC curve. However, taking a look into the FPR, we see that it's considerably smaller for the MLP classifier, what makes it even more suitable than the Random Forest one. On the other hand, a big performance difference is observed between both datasets, probably due to the difference between both databases, as OULU-NPU's one is the most recent and controlled one.

Finally, taking another look at the ROC curves, we can confirm the hypothesis mentioned in the previous paragraph observing the shape of the different curves. Both MLP and Random Forest classifier curves are the ones that look better in all cases. Nevertheless, the observation of these graphs makes us possible to conclude that it looks like the system works better with the HOG features than with the whole 'error' images, as the receiver response curves look considerably more stable in HOG's case (with less oscillations).

Therefore, we can conclude that our method performs the best using HOG descriptors as features and MLP as classifier in OULU-NPU's dataset.

5. Conclusions

In this paper we have presented the first, to the best of our knowledge, use of the Nuisance Attribute Projection (NAP) for detecting spoofing attacks to face recognition systems. We have proven that NAP, in combination with Histogram of Oriented Gradients (HOG) features, could be considered as a countermeasure for biometric spoofing, alone or in combination with other techniques mentioned in the state-of-the-art section. We've also shown that a countermeasure approach can be easily tested in more than one dataset, even if our performance results vary more than expected from one dataset to another.

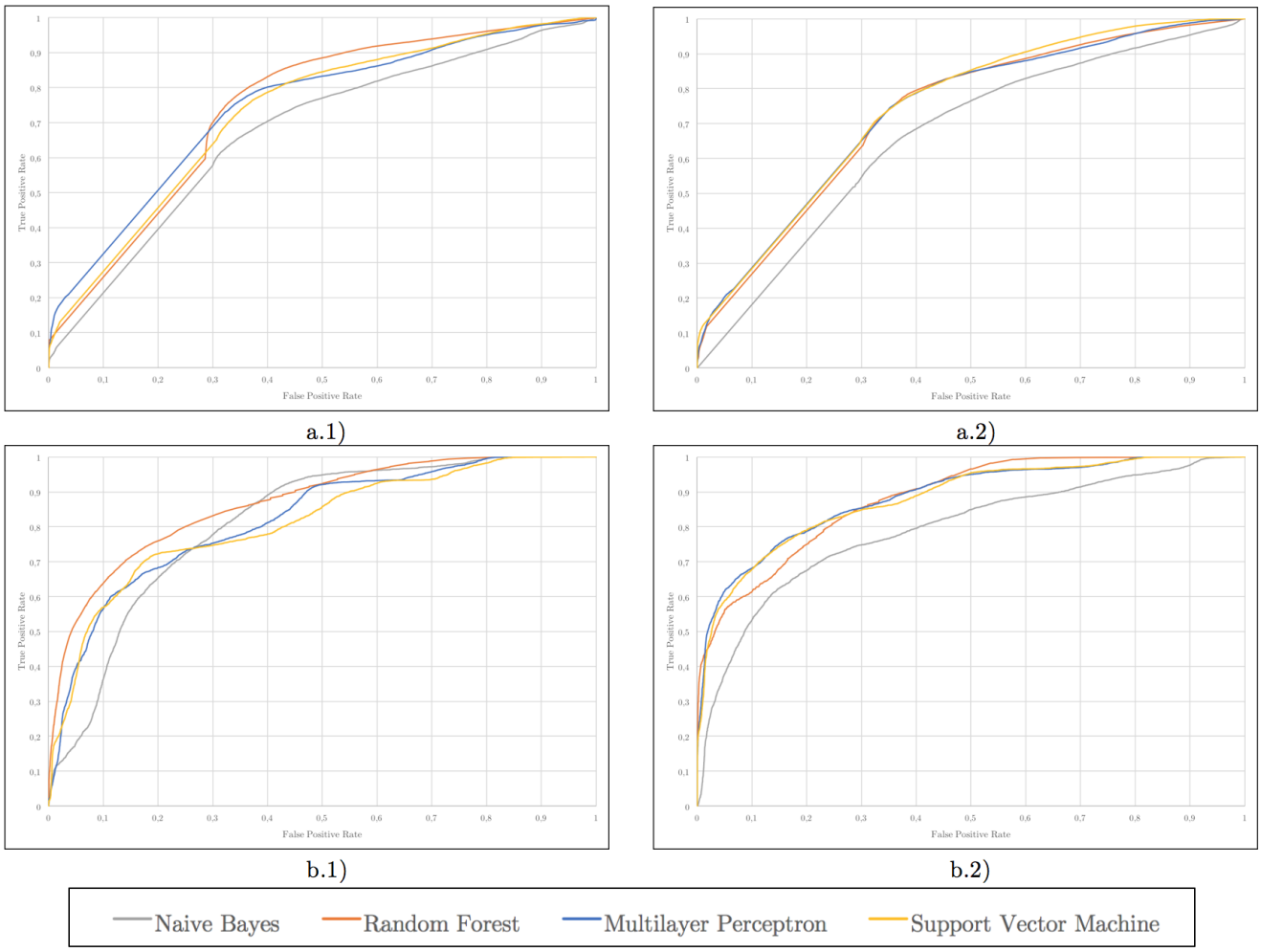

Figure 6. ROC curves of the four considered classifiers: Naïve Bayes classifier, Random Forest classifier, MLP classifier and SVM classifier. The curves a.1) and b.1) correspond to CASIA's and OULU-NPU's full 'error' images evaluation, respectively. On the other hand, curves a.2) and b.2) correspond respectively to CASIA's and OULU-NPU's HOG features evaluation.

Future work may try to improve the performance results obtained in this paper, varying parameters such as NAP's dimensionality or the feature extraction method. Moreover, the use of RGB frames could also reveal more artefacts after NAP computation that may improve also our results.

References

[1] T. Ahonen, A. Hadid, and M. Pietikäinen. Face recognition with local binary patterns, 2004. ECCV, European Conference on Computer Vision, pp 469-481.

[2] A. Albiol, D. Monzo, A. Martin, J. Sastre, and A. Albiol. Face recognition using hog–ebgm, 2007. I-TEAM, Universidad Politecnica de Valencia, Spain.

[3] A. Ali, F. Deravi, and S. Hoque. Liveness detection using gaze collinearity, 2012. in Proc. IEEE Int. Conf. Emerg. Secur. Technol. (ICEST), pp. 62–65.

[4] A. Anjos, M. M. Chakka, and S. Marcel. Motion-based countermeasures to photo attacks in face recognition, 2014. IET Biometrics, vol. 3, no. 3, pp. 147–158.

[5] A. Anjos and S. Marcel. Counter-measures to photo attacks in face recognition: A public database and a baseline, 2011. in Proc. IEEE Int. Joint Conf. Biometrics (IJCB), Oct. 2011, pp. 1–7.

[6] J. Bai, T. T. Ng, X. Gao, and Y. Q. Shi. Is physics-based liveness detec- tion truly possible with a single image?, 2010. in Proc. IEEE Int. Symp. Circuits Syst. (ISCAS), pp. 3425–3428.

[7] W. Bao, H. Li, N. Li, and W. Jiang. A liveness detection method for face recognition based on optical flow field, 2009. in Proc. Int. Conf. Image Anal. Signal Process. (ICIASP), pp. 233–236.

[8] S. Bharadwaj, T. I. Dhamecha, M. Vatsa, and R. Singh. Computationally efficient face spoofing detection with motion magnification, 2013. in Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW), pp. 105–110.

[9] S. Bharadwaj, T. I. Dhamecha, M. Vatsa, and R. Singh. Computationally efficient face spoofing detection with motion

magnification, 2013. in Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW), pp. 105–110.

[10] H. S. Bhatt, S. Bharadwaj, R. Singh, and M. Vatsa. Recognizing surgi- cally altered face images using multiobjective evolutionary algorithm, 2013. IEEE Trans. Inf. Forensics Security, vol. 8, no. 1, pp. 89–100.

[11] J. Bigun, H. Fronthaler, and K. Kollreider. Assuring liveness in biometric identity authentication by real-time face tracking, 2004. in Proc. IEEE Int. Conf. Comput. Intell. Homeland Secur. Pers. Safety (CIHSPS), pp. 104–111.

[12] W. W. Bledsoe. The model method in facial recognition, 1964. Panoramic Res., Inc., Palo Alto, CA, USA, Tech. Rep. PRI:15.

[13] Z. Boulkenafet, J. Komulainen, L. Li, X. Feng, and A. Hadid. Oulu npu: A mobile face presentation attack database with real world variations, 2017. 2017 12th IEEE International Conference on Automatic Face and Gesture Recognition, Washington, DC, pp. 612-618.

[14] K. W. Bowyer, K. Chang, and P. Flynn. A survey of approaches and challenges in 3d and multi-modal 3d+2d face recognition, 2006. Comput. Vis. Image Understand., vol. 101, no. 1, pp. 1–15.

[15] P. Buddharaju, I. T. Pavlidis, P. Tsiamyrtzis, and M. Bazakos. Physiology-based face recognition in the thermal infrared spectrum, 2007. IEEE Trans. Pattern Anal. Mach. Intell., vol. 29, no. 4, pp. 613–626.

[16] G. Chetty and M. Wagner. Liveness verification in audio-video authentication, 2004. in Proc. 8th Int. Conf. Spoken Lang. Process. (ICSLP), pp. 2509–2512.

[17] G. Chetty and M. Wagner. Liveness detection using cross-modal cor- relations in face-voice person authentication, 2005. in Proc. Annu. Conf. Int. Speech Commun. Assoc. (INTERSPEECH), pp. 2181–2184.

[18] C. C. Chibelushi, F. Deravi, and J. S. D. Mason. A review of speech- based bimodal recognition, 2002. IEEE Trans. Multimedia, vol. 4, no. 1, pp. 23–37.

[19] I. Chingovska, A. Anjos, and S. Marcel. On the effectiveness of local binary patterns in face anti-spoofing, 2012. in Proc. IEEE Int. Conf. Biometrics Special Interest Group (BIOSIG), pp. 1–7.

[20] I. Chingovska, A. Anjos, and S.Marcel. Anti-spoofing in action: Joint operation with a verification system, 2013. in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW), pp. 98–104.

[21] T. Crunch. Woman uses tape to trick biometric airport fingerprint scan, 2009. [Online]. Available: http://techcrunch.com/2009/01/02/woman-uses-tape-to-trick-biometric-airport-fingerprint-scan.

[22] A. da Silva Pinto, H. Pedrini, W. Schwartz, , and A. Rocha. Video-based face spoofing detection through visual rhythm analysis, 2012. in Proc. 25th Conf. Graph., Patterns Images (SIBGRAPI), pp. 221–228.

[23] A. Dantcheva, C. Chen, and A. Ross. Can facial cosmetics affect the matching accuracy of face recognition systems?, 2013. in Proc. IEEE 5th Int. Conf. Biometrics, Theory, Appl. Syst. (BTAS), pp. 391–398.

[24] K. H. Davis, R. Biddulph, and S. Balashek. Automatic recognition of spoken digits, 1952. J. Acoust. Soc. Amer., vol. 24, no. 6, pp. 637–642.

[25] T. de Freitas Pereira, A. Anjos, J. M. de Martino, and S. Marcel. Lbp-top based countermeasure against face spoofing attacks, 2012. in Proc. Int. Workshop Comput. Vis. Local Binary Pattern Variants (ACCV), pp. 1–12.

[26] T. de Freitas Pereira, A. Anjos, J. M. D. Martino, and S. Marcel. Can face anti-spoofing countermeasures work in a real world scenario?, 2013. in Proc. IEEE Int. Conf. Biometrics (ICB), pp. 1–8.

[27] M. de Marsico, M. Nappi, D. Riccio, and J. Dugelay. Moving face spoofing detection via 3d projective invariants, 2012. in Proc. IEEE Int. Conf Biometrics (ICB), pp. 73–78.

[28] N. M. Duc and B. Q. Minh. Your face is not your password. face authentication bypassing lenovo—asus—toshiba, 2009. San Francisco, CA, USA: Black Hat.

[29] N. Erdogmus and S. Marcel. Spoofing in 2d face recognition with 3d masks and anti-spoofing with kinect, 2013. in Proc. IEEE Biometrics, Theory, Appl. Syst. (BTAS), pp. 1–6.

[30] I. C. et al. The 2nd competition on counter measures to 2d face spoofing attacks, 2013. in Proc. IAPR Int. Conf. Biometrics (ICB), pp. 1–6.

[31] M. M. C. et al. Competition on counter measures to 2-d facial spoofing attacks, 2011. in Proc. IEEE Int. Joint Conf. Biometrics (IJCB), pp. 1–6.

[32] J. Galbally, S. Marcel, and J. Fierrez. Biometric antispoofing methods: A survey in face recognition, 2014. in IEEE Access, vol. 2, pp. 1530-1552.

[33] C. Geng and X. Jiang. Face recognition using sift features, 2009. 16th IEEE International Conference on Image Processing (ICIP), Cairo, pp. 3313-3316.

[34] T. Guardian. iphone 5s fingerprint sensor hacked by germany's chaos computer club, 2013. [Online]. Available: http://www.theguardian.com/technology/2013/sep/22/apple-iphone- fingerprint-scanner-hacked.

[35] G. Hermosilla, J. R. del Solar, R. Verschae, and M. Correa. A comparative study of thermal face recognition methods in unconstrained environments, 2012. Pattern Recognit., vol. 45, no. 7, pp. 2445–2459.

[36] IBG. Biometrics: Market shares, strategies, and forecasts, worlwide, 2015-2021, 2015. International Biometrics Group, Virginia, USA, Tech. Rep.

[37] A. K. Jain, R. Bolle, and S. Pankanti. Biometrics: Personal identification in a networked society, 1999. Kluwer Academic Publishers.

[38] H. K. Jee, S. U. Jung, and J. H. Yoo. Liveness detection for embedded face recognition system, 2005. Int. J. Biol. Life Sci., vol. 1, no. 4, pp. 235–238.

[39] M. D. Kelly. Visual identification of people by computer, 1970. Stanford AI Project, Stanford, CA, USA, Tech. Rep. AI-130.

[40] G. Kim, S. Eum, J. K. Suhr, D. I. Kim, K. R. Park, and J. Kim. Face liveness detection based on texture and frequency analyses, 2012. in Proc. 5th IAPR Int. Conf. Biometrics (ICB), pp. 62–72.

[41] S. Kim, S. Yu, K. Kim, Y. Ban, and S. Lee. Face liveness detection using variable focusing, 2013. in Proc. IEEE/IAPR Int. Conf. Biometrics (ICB), pp. 1–6.

[42] Y. Kim, J. H. Yoo, and K. Choi. A motion and similarity-based fake detection method for biometric face recognition systems, 2011. in Proc. IEEE Int. Conf. Consum. Electron. (ICCE), pp. 171–172.

[43] K. Kollreider, H. Fronthaler, and J. Bigun. Evaluating liveness by face images and the structure tensor, 2005. in Proc. IEEE Workshop Autom. Identificat. Adv. Technol. (AutoID), Oct. 2005, pp. 75–80.

[44] K. Kollreider, H. Fronthaler, and J. Bigun. Verifying liveness by multiple experts in face biometrics, 2008. in Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit. (CVPR), pp. 1–6.

[45] K. Kollreider, H. Fronthaler, and J. Bigun. Non-intrusive liveness detection by face images, 2009. Image Vis. Comput., vol. 27, no. 3, pp. 233–244.

[46] K. Kollreider, H. Fronthaler, M. I. Faraj, and J. Bigun. Real-time face detection and motion analysis with application in 'liveness' assess- ment, 2007. IEEE Trans. Inf. Forensics Security, vol. 2, no. 3, pp. 548–558.

[47] J. Komulainen, A. Hadid, and M. Pietikainen. Face spoofing detec- tion using dynamic texture, 2012. in Proc. Asian Conf. Comput. Vis. Workshops (ACCV-W), vol. 7728., pp. 146–157.

[48] J. Komulainen, A. Hadid, and M. Pietikainen. Context based face anti-spoofing, 2013. in Proc. IEEE 6th Int. Conf. Biometrics, Theory, Appl. Syst. (BTAS), pp. 1–8.

[49] N. Kose and J. Dugelay. Classification of captured and recaptured images to detect photograph spoofing, 2012. in Proc. IEEE Int. Conf Inform., Electron. Vis. (ICIEV), pp. 1027–1032.

[50] N. Kose and J. L. Dugelay. Countermeasure for the protection of face recognition systems against mask attacks, 2013. in Proc. 10th IEEE Int. Conf. Workshops Autom. Face Gesture Recognit. (FG), pp. 1–6.

[51] N. Kose and J. L. Dugelay. Reflectanceanalysisbasedcountermeasure technique to detect face mask attacks, 2013. in Proc. IEEE Int. Conf. Digit. Signal Process. (DSP), pp. 1–6.

[52] N. Kose and J. L. Dugelay. Shape and texture based countermeasure to protect face recognition systems against mask attacks, 2013. in Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW), pp. 111–116.

[53] J. Li, Y. W. T. Tan, and A. K. Jain. Live face detection based on the analysis of fourier spectra, 2004. in Proc. SPIE, Biometric Technol. Human Identification (BTHI), vol. 5404, pp. 296–303.

[54] J. W. Li. Eye blink detection based on multiple gabor response waves, 2008. in Proc. IEEE Int. Conf. Mach. Learn. Cybern. (ICMLC), pp. 2852–2856.

[55] Y. Li, K. Xu, Q. Yan, Y. Li, and R. H. Deng. Understanding osn-based facial disclosure against face authentication systems, 2014. in Proc. ACM Asia Symp. Inf., Comput. Commun. Security (ASIACCS), 2014, pp. 413–424.

[56] J. Maatta, A. Hadid, and M. Pietikainen. Face spoofing detection from single images using micro-texture analysis, 2011. in Proc. IEEE Int. Joint Conf. Biometrics (IJCB), pp. 1–7.

[57] J. Maatta, A. Hadid, and M. Pietikainen. Face spoofing detection from single images using texture and local shape analysis, 2012. IET Biometrics, vol. 1, no. 1, pp. 3–10.

[58] T. D. Mail. The man in the latex mask, 2012. [Online]. Available: http://www.dailymail.co.uk/news/article-2153346/Black-armed-robber-disguised-white-man-using-latex-mask.html.

[59] B. News. Malaysia car thieves steal finger, 2005. [Online]. Available: http://news.bbc.co.uk/2/hi/asia-pacific/4396831.stm.

[60] S. News. Fake fingers fool hospital clock-in scanner, 2013. [Online]. Available: http://news.sky.com/story/1063956/fake-fingers-fool-hospital-clock-in-scanner.

[61] S. News. Hackers 'fool' iphone x face id with a simple mask, 2017. [Online]. Available: https://news.sky.com/story/hackers-fool-iphone-x-face-id-with-simple-mask-11124604.

[62] G. Pan, L. Sun, Z. Wu, and S. Lao. Eyeblink-based anti-spoofing in face recognition from a generic webcamera, 2007. in Proc. IEEE 11th Int. Conf. Comput. Vis. (ICCV), pp. 1–8.

[63] G. Pan, L. Sun, Z. Wu, and Y. Wang. Monocular camera-based face live- ness detection by combining eyeblink and scene context, 2011. Telecommun. Syst., vol. 47, nos. 3–4, pp. 215–225.

[64] K. Patel, H. Han, A. K. Jain, and G. Ott. Live face video vs. spoof face video: Use of moiré patterns to detect replay video attacks, 2015. 2015 International Conference on Biometrics (ICB), Phuket, pp. 98-105.

[65] I. Pavlidis and P. Symosek. The imaging issue in an automatic face/disguise detection system, 2000. in Proc. IEEE Workshop Comput. Vis. Beyond Vis. Spectr., Methods Appl., pp. 15–24.

[66] B. Peixoto, C. Michelassi, and A. Rocha. Face liveness detection under bad illumination conditions, 2011. in Proc. IEEE Int. Conf. Image Process. (ICIP), pp. 3557–3560.

[67] F. J. Prokoski and R. B. Biel. Infrared identification of faces and body parts, 1999. in Biometrics: Personal Identification in Networked Society. Boston, MA, USA: Kluwer, pp. 191–212.

[68] T. Rasa. Trusted biometrics under spoofing attacks, 2010. [Online]. Available: http://www.tabularasa-euproject.org/.

[69] T. Rasa. Tabula rasa spoofing challenge, 2013. [Online]. Available: http://www.tabularasa-euproject.org/evaluations/tabula-rasa- spoofing-challenge-2013.

[70] T. Register. Get your german interior minister's fingerprint here, 2008. [Online]. Available: http://www.theregister.co.uk/2008/03/30/german_interior_minister_fingerprint_appropriated.

[71] E. A. Rua, H. Bredin, C. G. Mateo, G. Chollet, and D. G. Jimenez. Audio-visual speech asynchrony detection using co-inertia analysis and coupled hidden markov models, 2009. Pattern Anal. Appl., vol. 12, no. 3, pp. 271–284.

[72] B. Schneier. Inside risks: The uses and abuses of biometrics, 1999. Commun. ACM, vol. 48, no. 8, p. 136.

[73] W. R. Schwartz, A. Rocha, and H. Pedrini. Face spoofing detection through partial least squares and low-level descriptors, 2011. in Proc. IEEE Int. Joint Conf. Biometrics (IJCB), pp. 1–8.

[74] A. Seal, S. Ganguly, D. Bhattacharjee, M. Nasipuri, and D. K. Basu. Automated thermal face recognition based on minutiae extraction, 2013. Int. J. Comput. Intell. Stud., vol. 2, no. 2, pp. 133–156.

[75] A. Solomonoff, W. M. Campbell, and C. Quillen. Nuisance attribute projection, 2007. in Speech Communication. Elsevier Science BV, Amsterdam.

[76] A. Solomonoff, C. Quillen, and W. Campbell. Channel compensation for svm speaker recognition, 2004. in: Proc. of Odyssey, pp. 5762.

[77] V. Struc, B. Vesnicer, F. Mihelic, and N. Pavesic. Removing illumination artifacts from face images using the nuisance attribute projection, 2010. in Proceedings of the IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP 2010), Dallas, Texas, USA, pp. 846–849.

[78] X. Tan, Y. Li, J. Liu, and L. Jiang. Face liveness detection from a single image with sparse low rank bilinear discriminative model, 2010. in Proc. Eur. Conf. Comput. Vis. (ECCV), vol. LNCS 6316, pp. 504–517.

[79] P. Tome, R. Vera-Rodriguez, J. Fierrez, and J. Ortega-Garcia. Variability compensation using nap for unconstrained face recognition, 2012. in Distributed Computing and Artificial Intelligence AISC 151, Springer-Verlag Berlin Heidelberg, pp. 129-139.

[80] L. Wang, X. Ding, and C. Fang. Face live detection method based on physiological motion analysis, 2009. Tsinghua Sci. Technol., vol. 14, no. 6, pp. 685–690.

[81] T. Wang, J. Yang, Z. Lei, S. Liao, and S. Z. Li. Face liveness detection using 3d structure recovered from a single camera, 2013. in Proc. IEEE/IAPR Int. Conf. Biometrics (ICB), pp. 1–6.

[82] J. Yan, Z. Zhang, Z. Lei, D. Yi, and S. Z. Li. Face liveness detection by exploring multiple scenic clues, 2012. in Proc. 12th Int. Conf. Control, Autom., Robot. Vis. (ICARCV), pp. 188–193.

[83] J. Yang, Z. Lei, S. Liao, and S. Z. Li. Face liveness detection with component dependent descriptor, 2013. in Proc. IEEE/IAPR Int. Conf. Biometrics (ICB), pp. 1–6.

[84] Z. Zhang, J. Yan, S. Liu, Z. Lei, D. Yi, and S. Z. Li. A face antispoofing database with diverse attacks, 2012. in Proc. IAPR Int. Conf. Biometrics (ICB), pp. 26–31.

[85] Z. Zhang, D. Yi, Z. Lei, and S. Z. Li. Face liveness detection by learning multispectral reflectance distributions, 2011. in Proc. IEEE Int. Conf. Autom. Face Gesture Recognit. (AFGR), pp. 436–441.