Identity Independent Face Anti-spoofing Based on Random Scan Patterns

Kannan Karthik(✉) and Balaji Rao Katika

Department of Electronics and Electrical Engineering,
Indian Institute of Technology Guwahati, Guwahati 781039, Assam, India
{k.karthik,k.balaji}@iitg.ac.in

Abstract. Conventional face anti-spoofing paradigms tend to operate on plain facial profiles and learn either the natural face space alone (one-class training problem) or both the natural face space as well as the spoof sample space (2-class training problem). However, this rigidity with respect to spatially constrained measurements, makes the base feature or statistic vulnerable to noise related to pose and camera perspective/orientational and scale changes. Noting that the sharpness profile computed on a natural face is largely independent of the pose and perspective change, it is imperative that the measurements be extracted in an identity independent setting by ignoring the pose/perspective variation. To facilitate this, we have deployed a 2-dimensional random walk for capturing lower order pixel correlation statistics from natural faces, with virtually no perceptual interference. The proposed identity independent frame has surpassed the state of the art with reference to a 3D mask dataset (image oriented, isolated frame setting), with an EER of 2.25% without auto-population and an EER of 0.45% with auto-population.

Keywords: Anti-spoofing · $3D$ mask · 2D random walk · Scan-patterns · Auto-population · Outlier detection · Prosthetic

1 Introduction

Since most facial recognition systems operate based on notion of perceptual similarity of images of human faces, a majority of them, cannot tell the difference between a spoofed version of a face, versus, a naturally captured facial image. It is therefore important to have a counter-spoofing layer that sits above the facial recognition system, which attacks the environment linked to the image acquisition process and attempts to anticipate any form of spoofing. Much of the literature, related to counter-spoofing, has been directed to a specific form of facial spoofing called planar spoofing [11], wherein the impersonator (X) tends to present planar images or printed photos of the target subject (Y), (who is being impersonated). This a type of geometrically constrained spoofing, in which, the identity of the targeted individual (Y), is embedded in the form of a planar intensity variation induced either by double printing or by capturing an image

B. Deka et al. (Eds.): PReMI 2019, LNCS 11941, pp. 3–12, 2019.
https://doi.org/10.1007/978-3-030-34869-4_1

of a natural photograph. This form of spoofing-model, albeit trivial, is enticing from a research perspective inviting solutions on multiple fronts, some related to sharpness reduction [6], some related to base-image quality degradation [4], some linked to model-specific, geometrically induced distortions [5,7] etc.

With the co-existence of a diverse entertainment industry, a rapid advancement in cosmetic technology and the inclusive growth and evolution of touch-up artists, facial spoofing frames have evolved considerably. Every individual tends to possess a distinct facial surface-contour, which, can be captured either overtly or surreptitiously. This surface contour, can be used to synthesize a prosthetic of that individual's face. This prosthetic can be designed either using paper-craft [2] or rigid plastic or some semi-elastic form of material. However, most prosthetics are customized according to the target individual who is being impersonated (viz. Y) and are usually de-linked from the identity of the individual who is the impersonator (X), to ensure that his/her identity is not revealed during the counter-spoofing or authentication process. This opens up the problem to the counter-spoofing community thanks to the following conjecture:

Claim-1: Given an impersonator X and a target subject Y, the prosthetic is designed to mimic the surface contour of Y and has very little to do with the surface contour of X. This is to ensure identity masking from the point of the view of the attacker (X). The only way this can be achieved, is by ensuring that this physical facial re-mapping or (physical face-morphing), is of a many-to-one type. Thus a single prosthetic designed to impersonate Y, can fit multiple individuals of the X-type. In other words, one mask is designed to fit many. This makes the prosthetic, presented as a synthetic surface contour of X, an over-smoothed approximation of X's facial profile with some depth information. One therefore anticipates a reduction in facial image sharpness as far as image of the prosthetic of X is concerned. This sharpness variation can be captured by performing a gradient based analysis.

In this paper, we focus on literature connected with prosthetic based facial spoofing. A 3D-mask dataset was developed using paper craft models in Erdogmus et al. [2], wherein the prosthetics were customized to target different subjects. Some examples of this are shown in Fig. 1. The natural faces of the subjects are shown in Fig. 1(b)(row-2), while their corresponding spoofed versions with prosthetics are shown in Fig. 1(a)(row-1). It is obvious that paper craft model has been cleverly designed to mimic the surface contours including the ocular and nasal profiles of each targeted subject. In Erdogmus et al. [2], the base feature used for recognition was the Local Binary Pattern (LBP) along with its variants. The 3D-Mask dataset was analyzed both as a sequence of static images, and also as a video sequence, by extending the LBP analysis to include both time and space differentials. A 2-class SVM was finally constructed by learning the prosthetic as well as the genuine face spaces, coupled with the decision boundary/surface. Spoof-detection was done by extracting the same features from a typical query test-face and checking its position with respect to the two reference clusters. In a video-based setting associated with the 3D mask database,

Fig. 1. (a) Examples of 3D mask faces for different subjects, taken from 3D mask database [2] (b) Examples of their corresponding real genuine samples.

more options exist, since it is possible to deploy space-time micro-feature analysis to search for liveliness in the facial profile, consistencies and naturalness in expression changes etc. Optical flow methods were used in Feng et al. [3] to detect differences in dynamism with respect to texture between an imposter and a genuine subject. A deep-learning network for attacking multi-biometric spoofing including facial spoofing was developed by Menotti et al. [9], but again the learning was two sided and assumed availability of samples related to the spoofing process. Wen et al. [10], proposed a mixed bag of features ranging from intensity and gradient all the way up to those which captured color and texture, with the objective of covering the complete gamut of statistics, which would help segregate the genuine face class from all forms of spoofing. However, once again, this arrangement demanded availability of spoof training samples, necessary for constructing a 2-class SVM. This existing frame had several issues:

- Very often the nature, texture and structure of the customized prosthetic may not be known. This implies that spoof-class training samples may not be available. Hence, it is important to shift and restrict the training process to the genuine face sample set, where the acquisition procedure, naturally captured facial profile coupled with the local statistics remains predictable.
- Since LBP features are highly localized in space and are registered, pose deviations and scale changes because of facial migrations with respect to the camera will lead to a contortion of measurements. This will interfere with the counter-spoofing procedure. We term this form of interference as perceptual interference, which arise when the measurements are registered in space.

The first problem related to absence of a spoof model, can be addressed through an inlier space characterization procedure by learning the space spanned by genuine natural facial images from different subjects, for different poses and mild illumination variations. This inlier space characterization was done through a query feature ranking procedure, in relation to the genuine face feature set, to detect outliers in [7]. Genuine face space characterization coupled with anomaly detection in a much more general setting by constructing a one-class SVM was done in Arashloo et al. [1].

While this arrangement was designed to care of the first problem related to the absence of a proper spoofing model, they were applied to planar spoofing alone. In both these papers [1,7], the measurements were registered in space either by gridding the image or by computing statistics in specific spatial zones, whose locations were largely static. They thus proved to be ineffective, when confronted with 3D-spoof models, wherein the prosthetics attempted to mimic the depth profile in the imposter's face.

Attacking this 3D-spoofing problem, with a single sided training procedure, involving only genuine face space characterization was the main challenge. This led to the proposed architecture which was placed on an identity independent setting. The rest of the paper is organized as follows: In Sect. 2 we propose a new paradigm based on identity independent feature auto-population through random scan patterns. Section 3 validates the choice of randomly scanned feature and builds a one-class SVM to characterize the space of natural faces. Experimental results and comparison with the state of the art are in Sect. 4.

2 Proposed Paradigm and Architecture

The anti-spoofing problem is a typical frame wherein the nature of impersonation remains unknown in practice. By treating this problem as a form of planar image or printed photo spoofing, the problem becomes analytically tractable, mainly because of physical constraints. To make the analysis model independent, without compromising on the robustness of the detection process, it is important to change the paradigm or the manner in which the measurements are gathered.

Claim 2: We claim that most anti-spoofing systems work best in an identity independent setting, wherein the measurements or features extracted are taken in such a way that perceptual relevance is given the least importance. However, the residual correlation or some other statistics, which may be derived from this dissolved identity, carry necessary information regarding the environment or channel in which the information has been captured to perform anti-spoofing. This identity dissolution, in our case, is performed using a constrained shuffle of pixels in the spatial domain using a 2-dimensional random walk. This 2-D random walk has been inspired by Space Filling Curves [8], which was originally devised for retaining the compressibility of video signals after encryption.

Claim 3: By auto-populating the each facial image profile with these scans, it is possible to construct several variations of the same profile, which essentially carry the same pixel-correlation information, with minimal content interference. Thus a single facial profile is transformed into an ensemble of scans produced using independent 2-D random walk patterns. This ensemble carries significant information regarding the sharpness profile of the facial image and will have sufficient information to segregate 3D mask profiles from natural facial images.

2.1 Random Scan Algorithm

Between a complete shuffle and a raster scan, one can find a judicious trade-off between feature transparency and conservation versus the size of the auto-populated set based on the type of scan. With a perfect shuffle of pixel complete pixel correlation structure is lost, while in the case of a raster scan this correlation structure is preserved. However, in the case of the raster scan, the peeling is done in a regularized fashion and hence only one instance of the facial profile can be made available, imparting a significant rigidity to the generation of statistics. A trade in terms of conserving the correlation profile in pixels, while at the same time permitting multiple shuffle trajectories is via the randomized correlated scan.

Figure 2 shows two typical variations in the scan patterns, executed over the same image F_i. Eventually when the scan is completed, a 1-D vector $\bar{X}_i$ is generated as a function of the randomizer/key KEY_i:

$$\bar{X}_i = CSCAN(F_i, KEY_i) \tag{1}$$

where, $CSCAN(.,.)$, represents the randomized correlated scan algorithm based on a 2-dimensional random walk directed by a key sequence, KEY_i. While perceptual identity is lost in the unregistered feature vector $\bar{X}_i$ structure and format of the data captured is conserved. Gradient and sharpness features can now be computed on the top of this randomly scanned intensity feature. The key sequence carries information pertaining to the direction/trajectory of the random walk. In case, there is an abrupt termination of the walk, the key sequence also stores information related to the pixel jump. In a nutshell, the key sequence is a sequence of location pointers forming a linked list. To reduce scanning complexity, F_i is a down sampled version of the parent facial image. Let $\bar{X}_i = [x_{i,1}, x_{i,2}, ..., x_{i,n}]$. It is to be noted that not all the correlated scans are contiguous in terms of random walk.

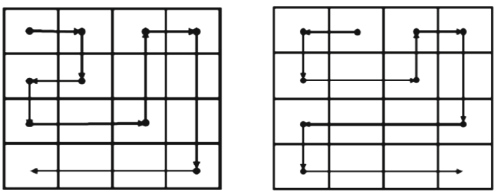

Fig. 2. Random but correlated scans for same facial image F_i (two different walks executed on the same facial image).

When the pointer in the 2-D random walk either walks into a corner or bumps into its own tail, it may encounter an abrupt termination. At this point, the pointer must hop to a new free cell within the same grid and resume the random walk. This process continues till all the pixels within the image grid are

traversed. If $N \times N$ is the size of the image, the length of the scanned vector $\bar{X}_i$ is $n = N^2$ and the path length is $n - 1$ units. The walk is rectangular in nature and diagonal transitions are not allowed. Because of this, the primary scanned vector $\bar{X}_i$ must be median filtered using a 1×3 ($w = 3$) window, to iron out singularities. The scanned vector becomes smoother with a larger window size w at the expense of a loss of detail and an un-necessary alteration of natural pixel correlation statistics. It is important that the median filter does interfere with the accuracy of the natural image statistics. Hence, the optimal choice for w is three. The pre-processed statistic is given by,

$$\bar{X}_{MED,i} = MEDIAN[\bar{X}_i, w] \tag{2}$$

with median filter window size, $w = 3$.

2.2 Final Differential Statistic

Based on earlier conjectures and observations, it is clear that the prosthetic arrangement is likely to have a smoother surface contour as compared to the natural face (partly owing to CLAIM-1 in Sect. 1). This is based on the one-mask fits many assumption, the mask designed to dissolve the identity of the imposter (X), while emulating the identity of the target (Y), who is being impersonated. Hence, a simple differential feature which captures the first or second order pixel derivative, will be sufficient to discriminate between a natural face as compared to one which has a prosthetic. The natural face is expected to have a greater roughness (culminating in a greater and more heterogeneous sharpness profile) as compared to that of the prosthetic. Let $\bar{D}_{X,i}$ be the differential statistic computed on the median filtered 1D sequence. If $\bar{D}_{X,i} = [d_{i,1}, d_{i,2}, ..., d_{i,n}]^T$ and $\bar{X}_{MED,i} = [x_{MED,i,1}, x_{MED,i,2}, ..., x_{MED,i,n}]$,

$$d_{i,r} = x_{MED,i,r} - x_{MED,i,(r-1)} \tag{3}$$

for $r \in \{1, 2, ..., n\}$ and with initial conditions, $x_{MED,i,(0)} = 0$. The vector, $\bar{D}_{X,i}$ is the final feature vector, extracted from the natural face image class alone, is fed to a one-class SVM [1] for characterizing the inlier space [7] (or the natural face space).

3 Feature Validation and Training the One-Class SVM

Feature validation is done by splitting the 3DMAD dataset [2] (composition given in Table 1), into natural faces and prosthetic based images. The base feature used for this comparison is the norm of the final differential vector, $\bar{D}_{X,i}$, which is given by,

$$E_i = ||\bar{D}_{X,i}||_2 = \sqrt{d_{i,1}^2 + d_{i,2}^2 + ... + d_{i,n}^2} \tag{4}$$

The conditional distributions, $f_{E/NATURAL}(e)$ and $f_{E/SPOOF}(e)$ are computed on the same scale in Fig. 3, for the 3D-MASK dataset (these are essentially

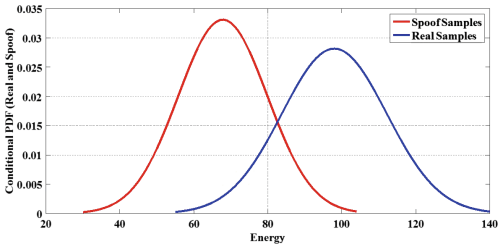

Fig. 3. Conditional distributions of the differential energy feature for natural and spoof samples, computed from the 3DMAD dataset. (Color figure online)

conditional histograms which have been interpolated to impart smoothness to the functions). In Fig. 3, the conditional distribution shown in blue corresponds to the genuine face space energy profile, while that shown in red corresponds to the energy profile generated from the prosthetic samples. As expected, the differential statistics produced from the natural face space have a larger mean and larger variance (because of the increased roughness and intensity diversity), while that of the prosthetic shows a smaller mean and variance (owing to over-smoothing stemming from the one-mask fits all claim). While the conditional distributions demonstrate the feature separability and ability of the random scans to conserve the lower order correlation statistics present in the image, the impact of the of the random scan in obscuring the identity of the individual subjects is demonstrated in Fig. 4. Notice that the scanned versions presented for simplicity as a 2-D shuffled version in Fig. 4(b, d), have no resemblance to their corresponding un-scanned counterparts (Fig. 4(a, c)). Thus, the processing and feature extraction is done in truly an identity independent setting.

Table 1. Description and composition of 3D mask database [2]

3D mask [2]	No of subjects	No of poses/subject
Faces with 3D masks	17	50
Natural faces	17	50

The set of natural faces from the 3DMAD database is split into a one-class training set for characterizing the inlier space and a test set which comprises of both natural faces as well as spoofed faces using the prosthetic. Given final differential base feature vectors $\bar{D}_{X,i}$, $i \in \{1, 2, ..., N_{GEN,T}\}$, where, $N_{GEN,T}$ is the number of training images from the genuine and natural face space. A one-class SVM [2], is constructed by building a hyper-sphere around the genuine multi-dimensional base differential features vector set corresponding to genuine face images, with an α-trim outlier fraction set to 10%.

4 Experimental Results

The 3DMAD database [2], whose composition is presented in Table 1, is split into three sets: (i) Genuine face set for training, $x\%$ of the total genuine face space; (ii) Genuine face set for testing, remaining $(100 - x)\%$ of the remaining genuine face space; (iii) Spoof samples ONLY for testing, from the paper-craft based prosthetic arrangement, $y = 100\%$ of the spoof set.

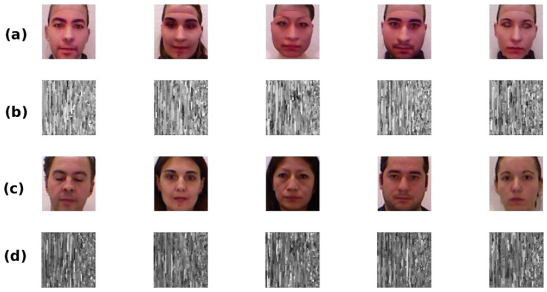

Fig. 4. (a) Samples of 3D mask faces (b) Random walk features extracted for mask faces (c) Samples of real genuine face (d) corresponding random walk features.

Table 2. Error rates for different trim factors α and database splits with 3D mask set. Note that best results are obtained for $\alpha = 10\%$.

Inlier $(x(\%))$/Outlier samples $(y(\%))$	EER@ $\alpha = 5\%$	EER@ $\alpha = 10\%$	EER@ $\alpha = 15\%$	EER@ $\alpha = 20\%$
$x = 30\%, (100 - x) = 70\%, y = 100\%$	5.9291	5.0115	8.4520	10.4653
$x = 40\%, (100 - x) = 60\%, y = 100\%$	5.8956	3.4048	5.2192	8.0694
$x = 50\%, (100 - x) = 50\%, y = 100\%$	3.2860	2.2594	4.2459	5.8806
$x = 60\%, (100 - x) = 40\%, y = 100\%$	2.5087	1.3078	3.3742	4.1446
$x = 70\%, (100 - x) = 30\%, y = 100\%$	1.2161	0.1852	3.0428	3.0760

Selection of the trim factor in the one-class SVM α, is a careful tradeoff between extent of generalization of the natural face space versus weeding out spoof samples which are likely to be close in structure with respect to the natural space. With limited training samples, the need for generalization calls for an expansion of the hyper-sphere (or a reduction of α), while the urge to weed out almost all spoof samples as outliers, demands a compaction or a contraction of the hyper-sphere (or an increase in α). Either way there will be mis-classifications either in the form of false positives or in the form of false negatives. Somewhere in between there is compromise and this optimal trimming factor was found to be $\alpha = 10\%$, as the outlier fraction. This is visible in Table 2, wherein best results are obtained for a trim factor of $\alpha = 10\%$. For a specific inlier (genuine space)

training fraction $x\%$ (viz. along a specific row in Table 2), the Equal error rate (EER) decreases and then increases when α is varied from 5% to 20%, with the minima hovering around $\alpha = 10\%$. Note that in this table no auto-population is done using the random scans. The EER therefore is slightly on the higher side for $x = 50\%$, 50% natural face samples for training, wherein the minimum EER (corresponding to $\alpha = 10\%$) was found to be 2.25%.

4.1 Auto-Population Results and Comparisons

It is natural to deploy the proposed random scan tool to derive statistically equivalent but identity independent representations of the same natural facial profile. Thus, every natural face training image is converted into an ensemble of scans which carry equivalent statistical information pertaining to the lower order pixel correlation profile. Since the walk is randomized each realization of the parent image is distinct and provides a unique perspective. Results are therefore expected to improve considerably with this form of auto-population. The effective number of training samples is magnified by a significant amount, viz. by a scale factor N_{SCAN}. Impact of different ensemble sizes (or scale factor N_{SCAN}) is shown in Table 3. Note that $N_{SCAN} = 1$, corresponds to results without auto-population and the EER numbers are expected to drop from left to right along a specific row. Saturation is expected beyond a certain point as the additional scans carry no new information for characterizing the inlier space. For $x = 50\%$, 50% face training, the lowest EER is obtained for $N_{SCAN} = 20$, highlighted in bold in Table 3, with a percentage of $EER = 0.43\%$, which is way below the number obtained in the same row corresponding to $N_{SCAN} = 1$, which is, $EER = 2.25\%$.

A fair comparison is possible only when the state of the art algorithms are compared on an image analysis front (with or without implicit auto-population) but applied to the 3DMAD database. It is unfair to compare video processing algorithms which attempt to detect liveliness in faces by examining wrinkle and crease line dynamics to track consistency in emotional transitions of subjects. The only paper that fits this constraint is the original work by Erdogmus et al. [2]. Both the random scan versions of the proposed algorithm with and without auto-population out-perform the state of the art. This validates the identity independent paradigm (Table 4).

Table 3. Performance with optimal trim factor, $\alpha = 10\%$ and auto-population using the proposed random scan algorithm. EER results saturate beyond a certain point.

Inlier ($x(\%)$)/ Outlier samples ($y(\%)$)	EER@ $\alpha = 10\%$				
	$N_{Scan} = 1$	$N_{Scan} = 5$	$N_{Scan} = 10$	$N_{Scan} = 20$	$N_{Scan} = 30$
$x = 30\%, (1-x) = 70\%, y = 100\%$	5.0115	2.0163	2.1085	2.1059	2.1426
$x = 40\%, (1-x) = 60\%, y = 100\%$	3.4048	1.2387	0.4527	0.4560	0.3486
$x = 50\%, (1-x) = 50\%, y = 100\%$	2.2594	**0.9489**	**0.6657**	**0.4310**	**0.4510**
$x = 60\%, (1-x) = 40\%, y = 100\%$	1.3078	0.5489	0.2626	0.2441	0.2605
$x = 70\%, (1-x) = 30\%, y = 100\%$	0.1852	0.1131	0.0871	0.0776	0.0731

Table 4. Comparison with the state of the art, which has used the 3DMAD dataset as a sequence of images. Training fraction, 50% from the natural face space.

Algorithm	Classifier	EER %
Ergodomus et al. [2]	SVM	4.92
Proposed random scan (NO auto-population $N_{SCAN} = 1$)	SVM	**2.25**
Proposed random scan (with auto-population $N_{SCAN} = 20$)	SVM	**0.4310**

5 Conclusions

This paper proposes an identity independent paradigm for facial anti-spoofing based by deploying 2-D random walks, to preserve the lower order pixel correlation in images, while dissolving the identity of subjects. With the suppression of perceptual interference, stemming from this form of constrained shuffle of pixels, results have improved significantly, in relation to the state of the art techniques. The EER rates with and without auto-population for this identity independent frame have been found to be 2.25% and 0.45% respectively.

References

1. Arashloo, S.R., Kittler, J., Christmas, W.: An anomaly detection approach to face spoofing detection: a new formulation and evaluation protocol. IEEE Access **5**, 13868–13882 (2017)
2. Erdogmus, N., Marcel, S.: Spoofing face recognition with 3D masks. IEEE Trans. Inf. Forensics Secur. **9**(7), 1084–1097 (2014)
3. Feng, L., et al.: Integration of image quality and motion cues for face anti-spoofing: a neural network approach. J. Vis. Commun. Image Represent. **38**, 451–460 (2016). http://www.sciencedirect.com/science/article/pii/S1047320316300244
4. Galbally, J., Marcel, S.: Face anti-spoofing based on general image quality assessment. In: 2014 22nd International Conference on Pattern Recognition (ICPR), pp. 1173–1178. IEEE (2014)
5. Garcia, D.C., de Queiroz, R.L.: Face-spoofing 2D-detection based on moiré-pattern analysis. IEEE Trans. Inf. Forensics Secur. **10**(4), 778–786 (2015)
6. Karthik, K., Katika, B.R.: Face anti-spoofing based on sharpness profiles. In: 2017 IEEE International Conference on Industrial and Information Systems (ICIIS), pp. 1–6. IEEE (2017)
7. Karthik, K., Katika, B.R.: Image quality assessment based outlier detection for face anti-spoofing. In: 2017 International Conference on Communication Systems, Computing and IT Applications (CSCITA), pp. 72–77. IEEE (2017)
8. Matias, Y., Shamir, A.: A video scrambling technique based on space filling curves (Extended Abstract). In: Pomerance, C. (ed.) CRYPTO 1987. LNCS, vol. 293, pp. 398–417. Springer, Heidelberg (1988). https://doi.org/10.1007/3-540-48184-2_35
9. Menotti, D., et al.: Deep representations for iris, face, and fingerprint spoofing detection. IEEE Trans. Inf. Forensics Secur. **10**(4), 864–879 (2015)
10. Wen, D., Han, H., Jain, A.K.: Face spoof detection with image distortion analysis. IEEE Trans. Inf. Forensics Secur. **10**(4), 746–761 (2015)
11. Zhang, Z., Yan, J., Liu, S., Lei, Z., Yi, D., Li, S.Z.: A face antispoofing database with diverse attacks. In: IEEE International Conference on Biometrics (ICB), pp. 26–31 (2012)