Image life trails based on contrast reduction models for face counter-spoofing

Balaji Rao Katika* and Kannan Karthik

Abstract

Natural face images are both content and context-rich, in the sense that they carry significant immersive information via depth cues embedded in the form of self-shadows or a space varying blur. Images of planar face prints, on the other hand, tend to have lower contrast and also suppressed depth cues. In this work, a solution is proposed, to detect planar print spoofing by enhancing self-shadow patterns present in face images. This process is facilitated and siphoned via the application of a non-linear iterative functional map, which is used to produce a contrast reductionist image sequence, termed as an image life trail. Subsequent images in this trail tend to have lower contrast in relation to the previous iteration. Differences taken across this image sequence help in bringing out the self-shadows already present in the original image. The proposed solution has two fronts: (i) a calibration and customization heavy 2-class client specific model construction process, based on self-shadow statistics, in which the model has to be trained with respect to samples from the new environment, and (ii) a subject independent and virtually environment independent model building procedure using random scans and Fourier descriptors, which can be cross-ported and applied to new environments without prior training. For the first case, where calibration and customization is required, overall mean error rate for the calibration-set (reduced CASIA dataset) was found to be 0.3106%, and the error rates for other datasets such OULU-NPU and CASIA-SURF were 1.1928% and 2.2462% respectively. For the second case, which involved building a 1-class and 2-class model using CASIA alone and testing completely on OULU, the error rates were 5.86% and 2.34% respectively, comparable to the customized solution for OULU-NPU.

Keywords Face counter-spoofing, Self-shadows, Image life trail, Contrast reduction, Logistic maps, Iterated function

1 Introduction

Given the seamless integration of functionalities and technologies inside smart-phones, it is imperative to incorporate not only biometric access control features inside it, but also include algorithms and architectures, which can detect and protect the contents against any form of impersonation or biometric-spoofing [1]. The face as a biometric establishes an individual's identity in a social setting, and this entrenchment permits easy traceability both in the digital space, as well as across

surveillance networks. Phone models therefore tend to use the owner's face as a biometric unlocking feature [2]. It is practical to assume that the natural face capturing environment, which involves taking a single shot image of a person standing in front of a camera is well defined under somewhat constrained settings (of-course with some variability in lighting and pose). Spoofing operation however can be effected on multiple fronts: (i) presenting a planar printed photo as a mask, of the person who is being impersonated; (ii) replaying a video sequence from a tablet or another cell-phone of the target; and (iii) wearing a carefully designed prosthetic (with a certain texture and having appropriate slits) of the target individual.

There are many applications, particularly involving smart phones, where, prosthetic based spoofing is unlikely. This is mainly because the customized design

---

*Correspondence:
Balaji Rao Katika
k.balaji@iitg.ac.in
Department of Electronics and Electrical Engineering, Indian Institute of Technology Guwahati, Guwahati 781039, Assam, India

of a prosthetic tailored to mimic a particular individual's face (who owns the smart-phone) is an extremely difficult scientific exercise. This problem is exacerbated by the fact that to prepare a 3D mask [3] (flexible or rigid), tuned to a particular individual's most recent facial parameters, one needs to first prepare a cast of the person's face or derive some form of holographic representation of the individual's facial parameters surreptitiously. This is an extremely expensive and time consuming affair. Hence, much of the spoofing technology is likely to be directed towards planar spoofing, wherein low or high-resolution facial images of individuals are either downloaded from the web and either printed and presented or presented via tablets to a particular face authentication/identification engine. Since most authentication engines look for facial similarity, the modality in which the authentication is done tends to ignore formatting anomalies connected with spoofing operation. One of the reasons why an authentication engine gets fooled by a planar print is because, while from a machine vision perspective this engine is designed to be robust to pose and illumination variations, this robustness comes at a price of overlooking format changes associated in the manner in which facial parameters are presented to the camera [4, 5]. Hence, there is a need for a counter-spoofing algorithmic layer, which searches for some form of naturalness based on some statistical lens, with respect to the facial parameters presented to the camera.

1.1 Counter-spoofing based on physical models

When the spoof-type is planar with a high probability, the counter spoofing solution can be designed more effectively by picking that statistical or forensic lens which separates the natural face class from the planar spoofed version. Very often the selection of this lens is governed by the manner in which the planar print representation is viewed or analyzed. When a planar printed photo is presented to the camera, on physical grounds it is easy to see that there are multiple fronts on the basis of which the so called naturalness can be compromised: (i) a planar presentation does not have depth, hence, the blur-profile in the target image is largely homogeneous [6–8], and (ii) the reprinting process to synthesize a planar print brings about a progressive degradation in contrast [9], clarity, specularity [10], quality [11], or color-naturalness [12].

One type of statistical lens for detecting planar spoofing is a specularity check [13]. If the paper printing of the target's face is done on a glossy type of paper, this results in a dominant specular component [10, 13] in the trapped image. While the non-specular component is a function of the object's color reflectivity profile and texture/roughness, its specular component is a measure of the object surface geometry witnessed by the camera in relation to a

fixed light source. In the case of a natural face, on account of a natural depth variation, the magnitude of the specular component is likely to be highly heterogenous while it is largely homogeneous for planar-print presentations [13]. In Emmanuel et al. [14], primary low rank specular features were derived from training face-images belonging to both classes. However, a principal components analysis (PCA) model was built for the natural face space alone, in Balaji et al. [10]. The training samples were projected onto this natural eigenspace. Since the spoof projections were ideally expected to correspond to the null space in relation to this PCA model, they were observed to have much lower magnitudes as compared to natural specular samples. Since the natural variability associated with the specular component is a function of many factors such as ethnicity, facial profile, presence of cosmetics, and other facial elements such as glasses and beards, this remains an non-robust primary feature.

Planar geometric constraints also impact the manner in which other parameters are influenced,such as contrast [9] or sharpness (or its opposite blur) [6–8].

When natural photographs are either re-printed or re-imaged and re-presented to a still camera, there is a reduction in contrast which follows a power law drop [9]. This reduces the dynamic range in the intensity profile considerably, eventually resulting in a more homogeneous contrast profile throughout the image. This contrast homogeneity can be measured by fusing local contrast statistics, using a global variance measure [9]. One of the main issues with this choice of high-level feature is the lack of consistency when it comes to print re-production. There are high quality printers available for re-creating the original subject-face in virtually the exact same form before presenting it as a mask to the camera. Thus, this cannot be treated as a universal feature from the print of view of planar printing.

Alternatively, in literature, while examining the planar-spoofing problem, it was observed that in the case of closed cropped natural faces, the natural depth (or distance) variation with respect to the camera often had a tendency to reflect as a spatially varying blur [6, 8, 15] in the captured image. In the work of Kim et al. [6], two sets of images were taken of the same subject. In one case, the depth of field was narrowed deliberately to induce a significant blur deviation across the entire natural image. In case of a planar spoofing, the blur differential between the original and de-focused image is likely to be very small. This dis-similarity in the de-focus patterns was used by Kim et al. [6] to detect planar spoofing.

In another blur variability detection procedure [15], a camera with a variable focus was used in the experiment and was designed to focus manually at two different points on the person's natural face: (i) nose of

the individual which is closest to the camera and the (ii) the ear of the individual which is the farthest from the camera. In the manual search procedure, the focal length adjustment was done to ensure clarity of one of these two facial-entities (nose or ear). It was observed that in the case of the natural face, the number of iterations required for the two cases were very different. On the other hand for a planar spoof presentation, virtually the same number of iterations were required to produce either a clear nose or a clear ear image. This difference between convergence trends was used to detect planar spoofing.

In an isolated image analysis setting (without deploying multiple entrapments and variable focus cameras), a pin-hole camera model was presented in [8] to bring out the problem connected with this blur phenomenon. A simple sharpness profile analysis based on gradients and gradient-thresholding was done to generate a statistic which gave an approximate measure of the sharpness measure for the presented image. In the case of planar spoofing, since the referential plane of focus (or object plane) need not coincide precisely with the spoof-print presentation, a homogeneous blur is likely to be superimposed on top of the original natural blur trapped in the printed version. Because of this, the average sharpness of the planar print version is expected to be much lower as compared to mean sharpness computed from a natural face image. The statistic proved to be sub-optimal, particularly for cases where the plane of focus was close to the print-object plane for print-presentations. The other problem was that with regular cameras in which the depth of field covers the complete face, the blur deviation is likely to be subtle. Thus, this blur diversity cannot be easily trapped without deploying a highly precise single face image based depth map computation algorithm.

Entrapment of scene related immersive information particularly regarding the positioning of light sources [16] is possible in the case of natural faces. This is because for portions of the face which are smooth in nature such as the cheeks and the forehead, the surface normal directions, for fixed ethnic group of individuals can be reliably estimated based on 3D registration frames. This becomes a referential pattern available in the repository. Now, when the subject presents his/her face to camera, at precisely the same spatial locations, based on the apparent intensity gradient and the known source co-ordinates relative to the subject, the surface normal directions are re-estimated. When there is a similarity in direction at a majority of the points where the measurements are taken, then the presentation can be declared as a natural one. When the estimated surface normal directions deviate considerably from the test subject, then it is highly probable that this inconsistency is due to a planar

spoofing. While the approach is interesting there are some issues with this:

- Multiple light sources are required at the surveillance point (at least two as in [16]), so that the same subject's face presentation can be illuminated from multiple directions. The overall setup requires additional lights, timers and switches and the per-subject assessment time is significant. This makes this architecture quite infeasible in large scale public scanning environments.
- Intra-natural face class errors associated with the normal direction estimation tend to climb if there are pose, scale, and expression changes in the individual [16].
- Since the points at which the measurements are taken must be registered in space, in a subject independent setting, identification of these key-points becomes a noisy affair for an arbitrary pose and scale presentation. This presents itself as what can be called subject-mixing noise or registration noise [4].

Planar spoofing (both print and digitized presentations) tend to imbibe some form radiometric distortion which stems from the additional printing and re-imaging stages which are constrained and lossy in nature [12]. Thus, an image of a planar printed face may not exhibit on one hand all the true colors which were originally present in natural face image of the same subject. Given the availability of both natural and spoof samples, this radiometric model can be estimated at a generic level but confined to a subject/client specific analysis [17]. When a test image arrives, its affiliation with the subject-specific radiometric distortion model is done via some form of regression analysis to establish the trueness or naturalness of the image. There are several issues with this arrangement:

- To ensure that only the illumination and color profile confined to the facial-region of a particular subject is analyzed, the background is painted and cropped via a segmentation procedure. The close cropping is extreme to the extent that no part of the person's hair or lower neck/shoulders are included in the segmented region. When this close cropping is not done, then both the radiometric (real, planar) model-estimation, along with the detection procedure, becomes noisy and quite unreliable.
- When there is subtle pose change, considerable illumination variation and scale change in the training sets, the model learning procedure (even on a subject specific note) becomes highly unreliable. Because of

this lack of model reliability, the accuracy reported for difficult datasets such as CASIA [18] was found to be on the lower side.

1.2 Counter-spoofing based on image texture and quality analysis

It was proposed in Maatta et al. [19] that planar spoofing tends to bring about a change in texture and facial perspective (apparent or projected face) compared to real facial images. Local binary patterns (LBPs) [17, 19, 20], Gabor and Histogram of Gradients (HoG), can therefore be used to capture texture statistics linked to both the classes and build a 2-class SVM model. But without a crisp differential noise analysis, with respect to natural and planar spoof representations, features/statistics picked may not be robust enough.

In the same context of texture, facial micro-analysis via landmark identification can be used track faces across real-time surveillance videos [21]. Facial landmarks, such as eye centers and nose tips, once identified from a sequence of frames using standard face detection protocols, pixel information from their local neighborhoods can be collated to construct a statistical model for each landmark. These so called landmark-descriptors when stitched together in the form of a connected graph, can be tracked across videos. In a dynamic camera and still face arrangement, multiple collections of landmark-sets taken from a series of video frames can be used to recreate a generic 3D model of the person's face [22]. In the case of planar spoofings, these gathered measurements will result in the re-creation of face surfaces which are largely flat and lacking in depth information. There are several issues with this arrangement:

- Need for relative movement between the subject and the camera is must in this arrangement to re-create either a 3D-representation by aligning the landmark features from multiple frames or for establishing whether the presentation is planar in nature. This relative dynamism may not always be feasible at an un-manned surveillance point, particularly when the camera is expected to move relative to a static face.
- If too many landmark-points are identified, the graph structure is expected to become un-stable (leading to alignment problems) when there is a pose variation or an illumination profile change. Too few landmark points will result in an imprecise model in the context of 3D surface reconstruction. Under varying ethnic origins, this optimization

problem will turn subject specific and difficult to handle. Cross-porting a particular counter-spoofing architecture/arrangement tuned to one dataset may not be very effective on a dataset housing subjects from a different geographical region.

1.3 Mixed bag techniques

Apart from model based approaches, in Wen et al. [23], statistics based on a mixed bag of features ranging from texture, color diversity, degree of blurriness were deployed, assuming that the extended acquisition pipeline (in a spoof-environment), connected with a re-printing and re-imaging procedure, tends to alter and impose constraints on this bag of features on a multitude of fronts. There were several issues with this arrangement:

- In a diverse planar spoofing environment, there exist several uncertainties related to the spoofing-medium: (i) for paper-print-presentations, the nature of the paper (glossy/non-glossy), printing resolution, and print color quality remain unknowns; (ii) for tablet and other digitized presentations, the nature and extent of re-sampling noise [19], resolution, color re-transformation, and reproduction remain unknown. Thus, using a common and diverse statistical lens to segregate natural and planar-spoofings may not be very effective. What works for one type of spoofing may not work work for another.
- The other main problem in conducting the training in a subject independent fashion is the influx of content dependent noise connected with subject-type variability [4] which stems from differences in facial parameters such as eye structures, their separation, nose profiles, and cheek and jaw-bone patterns. This is where client/subject dependent models [17, 20] tend to outshine the subject independent ones [9, 11].

Texture analysis in a broader context can be visualized as a quality assessment measure, wherein in most cases natural images are expected to possess a higher quality and clarity as compared to spoofed images [24, 25]. This blind quality assessment is brought about via a differential analysis wherein differential information between the original and its low pass filtered version is analyzed. Natural faces tend to exhibit a greater noise differential as compared to planar prints. Statistics such as pixel difference, correlation, and edge based measures were used to quantify the differential noise parameters and subsequently the overall quality score. There were several issues with this arrangement:

- Since-edge related statistics are heavily dependent on the subject facial profiles, the measures were not subject-agnostic, inviting subject-specific content interference or "subject mixing noise" [4].
- There was no scientific basis or analytical justification for choosing such a potpourri of statistics for performing this noise analysis. Hence, these features/statistics were not all that precise.
- The differential noise and image quality analysis was done in a 2-class setting (real versus spoof), and assuming prior availability of sample training images from the spoof-segment, which is impractical.

1.4 Subject mixing noise

Overall, in the approaches discussed so far, features connected with intensity, contrast [9, 12], blur/sharpness [7, 8], specularity [10]. and differential statistics such as localized binary patterns (LBPs) and its variants collected in regular fashion are pooled together to generate a 2-class model assuming that spoof-print samples are available. The problem with this paradigm is that in this frame one cannot avoid what can be called "subject mixing noise," as subject-related perceptual content tends to interfere with the regularized measurements. This "mixing" problem stems from a lack of proper face registration due to pose and face-scale changes [4]. This problem can be mitigated to some extent in a client-authentication rather than a client-identification setting by restricting the analytical and decision space to specific subjects/clients [17, 20].

Since the facial parameters such as eye-type and relative positioning, nose (size and shape), mouth, and cheek bones are distinct but largely fixed for a given individual, registered measurements taken in a certain order for a natural image can be weighed against those taken from a print-spoof image without worrying about "subject-mixing noise." There are many more choices as far as feature selections are concerned in a client specific arrangement as opposed to a client agnostic one. While lack of portability and customization of the detection algorithm is a drawback of this architecture, a big advantage is the higher accuracy one can achieve, since the "subject mixing noise" is nullified provided, pose variation and scale change is minimal.

1.5 Identity independent counter-spoofing via random scans

This so called subject-mixing noise can be combated in a subject agnostic setting by noting that short-term pixel intensity correlation profiles carry significant immersive information regarding both the type of object presented to the camera and also the lighting environment [4, 5].

Thus, by trapping this short-term correlation profile without inviting content dependent texture-noise, one can detect natural presentations. The first, second or third order pixel correlation profiles can be trapped by executing a simple random walk [4] from the center of the image. Multiple realizations of this random walk phenomenon can be used to auto-populate the features associated with a natural image. By ignoring the macro-structure in the face image, only the format differences are extracted via first order differential scan statistics [4]. This allows this random walk based counter-spoofing algorithm to transcend a variety of planar-spoof-media, lending itself as a monolithic yet universal solution. While such a random walk approach can tell the difference between a over-smoothed prosthetic and a natural face [5], with albeit a reduced degree of reliability, it has a tendency to hit an error-rate ceiling when the acquisition format or scene variability in the inlier/natural face space class is on the higher side. The error rates reported for CASIA-CASIA are therefore likely to saturate at EER = 1.89% and 2.16% for printed and digital planar spoof-sets respectively. This may not even decrease, even if one drifts to a client/subject specific frame.

1.6 Motivation and problem statement

In this work, as opposed to a universal one, a spoof model directed approach on client-specific grounds has been proposed wherein the spoofing frame is considered as a planar print presentation. This streamlining permits the design and deployment of a much more precise solution with a higher detection accuracy as compared to the universal case. As discussed earlier, this client specific weighing (in the image analysis domain, natural versus spoof) allows a mitigation of "subject mixing noise." The counter-spoofing system here knows the identity of the face presented to the camera and can access stored samples related to that "presented-subject" from the repository, with a client/subject-dependent [17, 20], 2-class support vector machine (SVM) model and use that prior data to perform the classification of this new test image sample. The main contributions in this work are:

- Proposition of a new contrast reductionist frame for planar print counter-spoofing, by deploying a discrete logistic map at the pixel level [26]. This has been termed as an image life trail wherein the contrast of the original test image (real or spoof) drops with each iteration and eventually reaches a virtually zero contrast state (saturation point).
- A self-shadow enhancement procedure which feeds on this life trail to make the self-shadows trapped in natural images much more prominent. It has been observed that planar-print spoof images tend to

have suppressed self-shadows as compared to natural ones, which serves as a discriminatory feature for segregating the two classes.

- A simple statistical model based on the dynamic range associated with intensity distributions connected with real and spoof/print classes has been used to justify the choice of first, first difference ratio statistic for enhancing self-shadow information and also arrive at the optimal choice of the exponent α^* via a calibration process and shape the final feature used to build the subject-specific 2-class model.

The proposed overall architecture has been split into two segments/blocks: (i) feature extraction, based on contrast reductionist image life trails leading to the extraction of critical information pertaining to self-shadows found in natural face-images (Fig. 1), and (ii) the training, subject-specific model building and final testing procedure shown in Fig. 2.

The section-specific organization is as follows: the proposed self-shadow formulation, i.e., base for the work in this paper where contrast reduced life trails are generated using logistic maps [26], is discussed in Section 2. The analytical frame and model in which the image is abstracted as random variable has been used to validate some of the claims made particularly linked to the life trails and the convergence rates of real and print images in Section 3. The self-shadow image statistic which is derived from the image life trail and further enhancements have been supported with an analytical justification in Section 4. Once the primary statistics have been finalized, it is known that every new illumination environment will demand a re-calibration and training for its own subjects. A method for arriving at the operating point for every new dataset is discussed in Section 5. Database description is given in Table 1 and the experimental results are presented in Section 8. Finally, to impart a certain flexibility a path has been proposed in which cross-porting can be done with a random scan front followed by a Fourier descriptor, to build subject agnostic models in Section 9.

2 Motivation and formulation for extracting self-shadows

Natural faces taken under constrained lighting conditions, with a frontal camera view and the light source positioned at an incline related to the face tend to exhibit what are known as self-shadows. A self shadow is formed mainly because of the following reasons: (i) the natural face which is exposed to a particular lighting
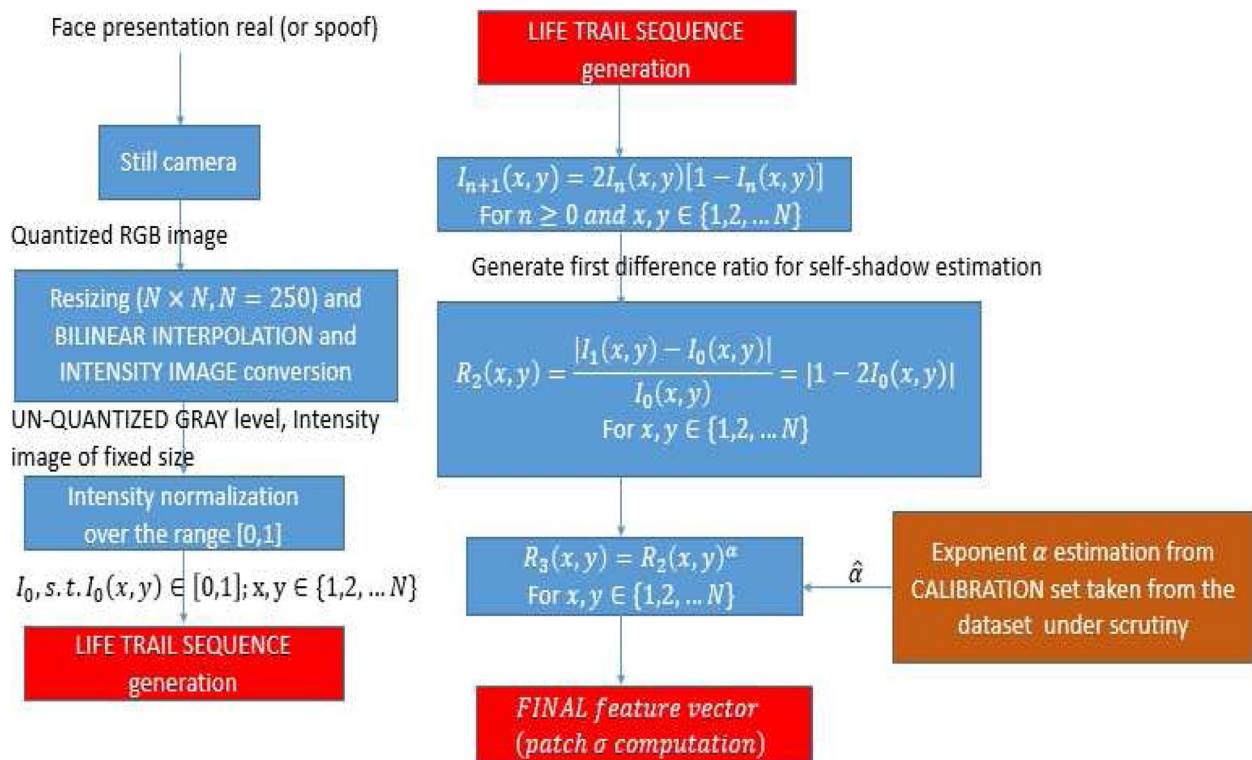
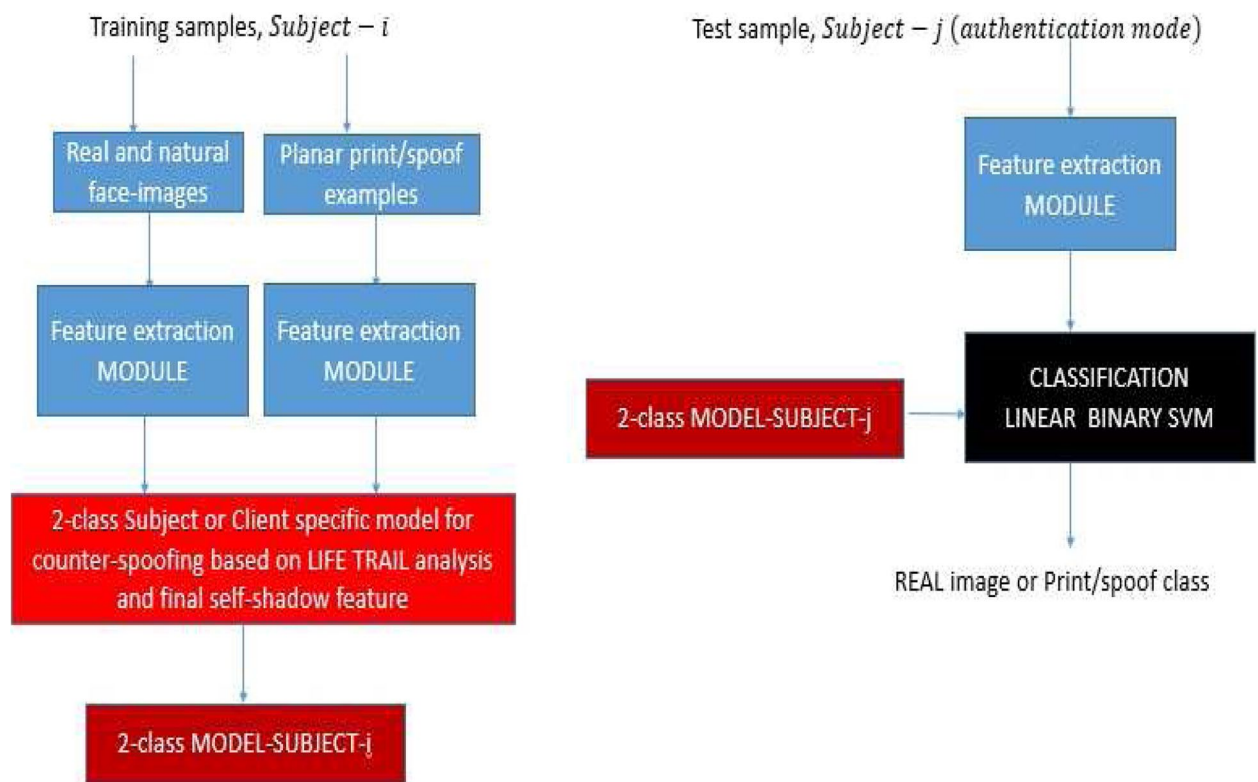

Fig. 1 Block-diagram of feature extraction procedure

Fig. 2 Block-diagram of training and classification/detection module

Table 1 Selective face anti-spoofing datasets and related parameters

Datasets	Year	No. of subjects	Camera	Image representation	Spoof-type	Race
CASIA [18]	2012	50	VIS	RGB	Print photo, cut photo	Chinese
MSU-MFSD [23]	2015	35	Phone Laptop	RGB	Printed photo	Asia, Middle east
						Hispanic, Europeans
						Latin Americans
Oulu NPU [27]	2018	20	VIS	RGB	Printed photo	Europeans and Middle east
CASIA SURF [28]	2018	1000	Real sense	RGB/depth	Print and cut photos	Almost all the races

environment has an irregular 3-dimensional surface contour, depending on the facial features of the individual. (ii) When light is projected onto one side of the face, the elevated parts of the face, such as the nose, high cheek bones, and facial curvature on either side of the cheeks tend to serve as occlusions to the projected light, leaving behind a self-shadow or a partial shadow on the other side. An example of this has been illustrated via a clay model as shown in Figs. 3 and 4. The camera positioned in front of the individual can be marked as the referential northern direction, relative to the person's face (which is in the southern direction). This camera (viz. an attached and aligned cell-phone camera unit) coupled with the clay-face itself is kept

fixed for the entire experiment. There are three light source orientations relative to the clay-face model indicated in a yellow-shade in Fig. 3.

The images captured with this arrangement for three different source locations are shown in Fig. 4a–c. In Fig4a, the light source has been positioned top-left-front of the person's face and beside the camera unit (north-west direction); in Fig 4b, the source is positioned towards the left of the person and partly in front (west position), while in Fig. 4c, the source is positioned behind the person in the south-west position. Self-shadows are evident in all the three images but minimal in the case of the north-west position and maximum when the light source is behind the clay-face (south-west position).
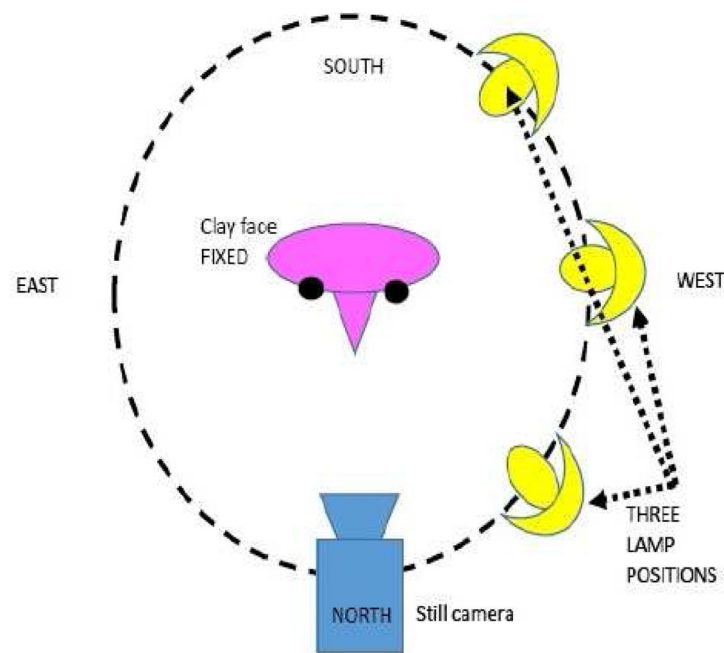
Fig. 3 Experimental setup using a clay model and a fixed cell-phone camera for producing natural images with self-shadows
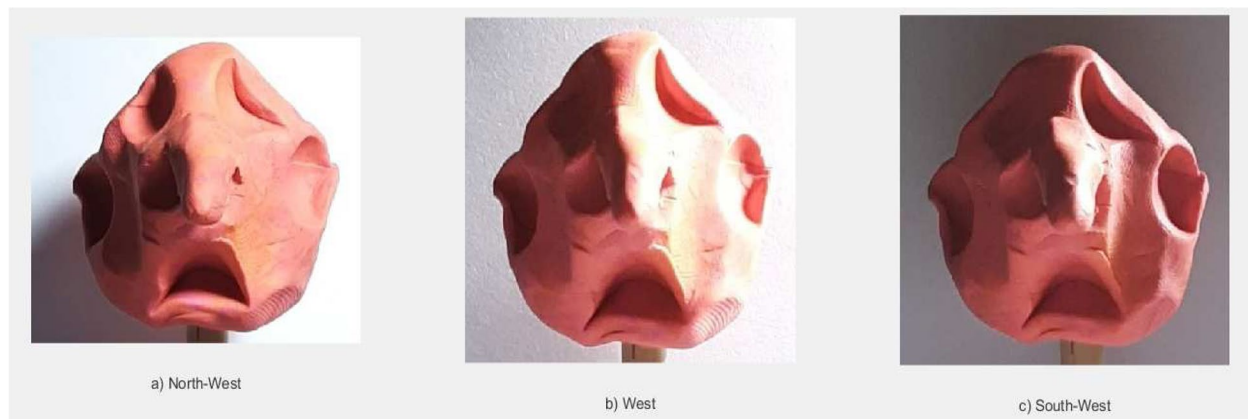

a) North-West   b) West   c) South-West

Fig. 4 Images captured using the experimental setup (Fig. 3), for three different table lamp positions (north-west, west and south-west)

Claim 1 *The first claim is that these self-shadows can be enhanced by first deploying an iterative contrast reducing procedure using a non-linear logistic map and then taking a relative difference ratio with the parent image. This difference image carries precious information related to the self-shadows.*

Claim 2 *The second claim is that in the case of a camera imaging of a planar print of a particular subject's face, these self shadows remain in a suppressed state. The original self-shadows which were trapped in the planar print of a natural*

facial image, are no longer fully visible, mainly owing to the secondary lighting environment, which leads to the formation of a much more uniformly illuminated image.

To facilitate an enhancement of this self-shadow pattern in the natural image, a non-linear logistic mapping [26] is deployed. This is an iterated function system that operates on an initial scalar value repeatedly and eventually converges to a "fixed point." One of the advantages of this logistic map is that on an average the convergence rate is quite fast and the fixed point is
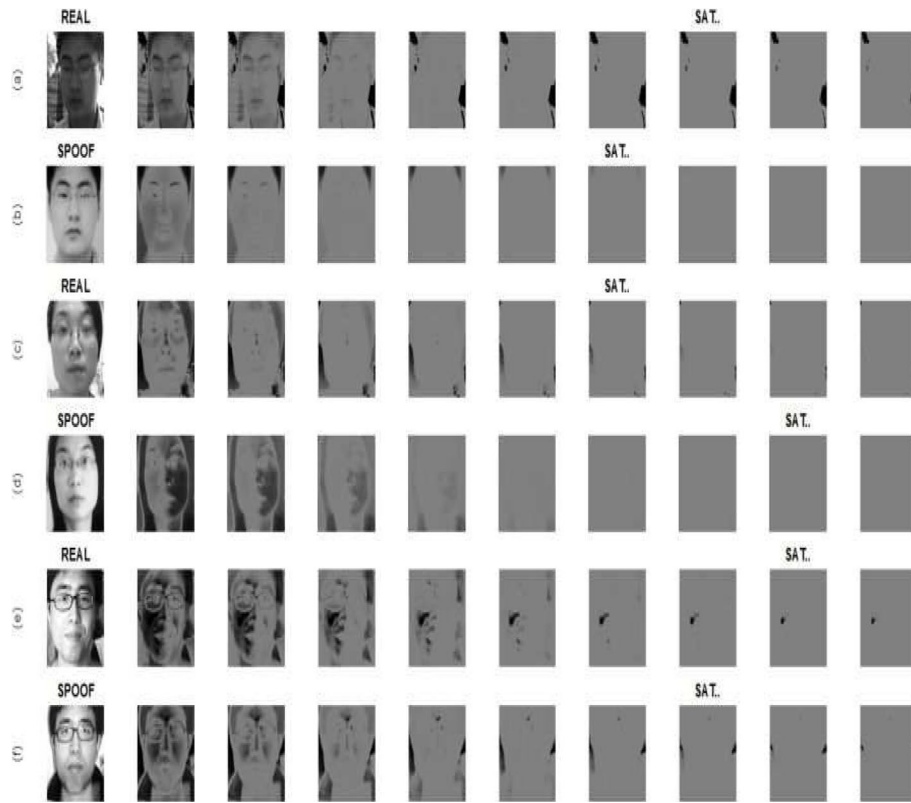
Fig. 5 Contrast reductionist life trails for real and spoof image samples using the logistic map

reached quickly, irrespective of the initial state (on an average).

2.1 Logistic maps and image life trails

Assume, $I_0(x, y)$ to be the normalized intensity value at particular spatial location (x, y) in an $N \times N$ face image of a particular subject, such that $I_0(x, y) \in [0, 1]$ and $I_0(x, y) = 0$ represents the completely black; $I_0(x, y) = 1$ represents the completely white pixel. The logistic map is a contrast reducing mapping which when applied to a "swarm" of image pixels independently, eventually after a few iterations the entire image reduces to a zero contrast image. We define an image "swarm" as the communion of all the intensity states of N^2 pixels undergoing this non-linear transformation. The length of this contrast-reductionist trail has been termed as an "image life trail." The life-line here refers to the number of iterations required for the parent image to reach a virtually zero contrast image or reach a point wherein almost all the pixels in this image swarm have come close to the fixed point value. To begin with, this pixel swarm is defined as follows:

$$SWARM(I_0) = \{I_0(x, y), s.t. x, y \in \{1, 2, ..., N\}\}$$

This non-linear iterated function system is defined as [26],

$$I_{n+1}(x, y) = 2I_n(x, y)(1 - I_n(x, y)) \tag{1}$$

with the initial value, $I_0(x, y) \in (0, 1)$ and $I_n(x, y)$ is the value at the nth, $n > 0$ iteration with $I_n(x, y) \in (0, 1)$. Irrespective of the initial value the Logistic map directs the value towards what is well known as a fixed point which in this case happens to be 0.5. By design with every iteration this value drifts closer and closer to the fixed point.

When such a map is applied to the swarm on a pixel by pixel basis, the entire swarm undergoes a transformation with each iteration, eventually producing what can be called a sequence of low contrast image (Fig. 5). Finally, the swarm results in a zero contrast image when almost all the pixels have converged to a value close to the fixed point 0.5 (which corresponds to gray level value 128).

2.2 Dynamic ranges of real and print face-images

At this point with respect to the life trail analysis, it is important to draw a distinction between the trails of a natural and spoof/print image. Any pixel having a

particular normalized intensity in the range (0, 1) will converge to the fixed point 0.5 eventually, upon repeated application of the logistic map. However, the trail dynamics when considering the pixel swarm or rather the collective convergence will depend on the slowest among the myriad pixel convergence trails (over the image), as a function of the intensity value spread (or rather the dynamic intensity-range). Smaller the dynamic range, faster will be the convergence. Hence, trails of low-contrast spoof images are likely to converge much faster as compared to natural face images.

NATURAL VERSION decays much SLOWER and It was surmised in [9] that given two registered face images (belonging to the same subject), the original normalized intensity version can be linked to the planar printed version via a power law relation,

$$I_{pp}(x, y) \approx I_{ORIG}(x, y)^{\gamma}$$

where *gamma* > 1 and and subsequent images of planar prints can be represented by the relation,

$$I_{pp[m]}(x, y) \approx I_{pp[m-1]}(x, y)^{\gamma}$$

where $m \geq 2$ with $I_{ORIG}(x, y) \in [0, 1]$ and $I_{pp[m]}(x, y) \in [0, 1]$. This implies that with subsequent printing, the moderately dark zones become darker and the lighter zones become darker. Eventually, as the planar printing is iterated, the entire image becomes completely dark. Hence, a planar printing procedure via a gamma power law is also a contrast reductionist transformation, wherein the transformed image has a lower intensity dynamic range as compared to the original image. The other thing that comes out of this is that a planar print version will always have a lower contrast as compared to that of the parent original image.

Consider the generation and deployment of a contrast score metric for measuring the dynamic range and score generated for eight subjects from the CASIA dataset (both real and spoof) [29]. Based on the metric used the scores produced for the natural faces are higher as compared to the spoof/print versions of the same subjects. Since all images have been resized to $N \times N$, let the normalized intensity value at position (x, y) be represented/mapped as:

$$I((x - 1)N + y) = I_0(x, y) \in [0, 1]$$

with $(x, y) \in 1, 2, ...N$. Pull out the non-trivial intensity values and let $I_{NZ}(k), k \in 1, 2, ..., M$ ($M \leq N^2$) be given by,

$$I_{NZ}(k) = I_0(x, y); \text{ provided } I_0(x, y) > \epsilon_1 \tag{2}$$

Using these non-zero intensity values, compute the mean and standard deviation over the entire image,

$$\mu_G = \frac{1}{M} \sum_{k=1}^{M} I_{NZ}(k)$$

$$\sigma_G = \sqrt{\frac{1}{M} \sum_{k=1}^{M} (I_{NZ}(k) - \mu_G)^2} \tag{3}$$

The final contrast score can be computed as [9], with a slight modification to account for images with very dark foregrounds:

$$CON = \frac{\sigma_G}{\mu_G} \tag{4}$$

To check the validity of this contrast metric from a perceptual view point the scores produced for real and print versions are shown in Fig. 6. Print versions tend to have a lower contrast scores as compared to natural faces.

To link up this apparent contrast degradation seen in print images with the exponential gamma law presented earlier in this section and also in [9], the same dynamic range numbers have been computed using the standard deviations σ (over the intensity profiles), on synthetically produced images via an application of this gamma-exponentiation on a natural faces of subjects. For all the intensity values in the set derived from a natural image, the exponential law is applied as,

$$I_{synthetic}(i, \gamma) = I(i)^{\gamma}$$

where $\gamma > 1$ and $I(i) \in SET_0$. The dynamic range scores for $\gamma = 1$ (i.e., no transformation), and then for $\gamma = 1.5, 3, 5$, for natural face images of four subjects are shown in Fig. 7. A simple statistical model is used to understand the differences between natural and print versions and as to how the contrast reductionist life trails evolve in both these cases.

3 Analytical frame for validation

The motive for this section is to abstract the image (real or print) as a random variable and bring out various elements linked to the problem connected with image trail and at the same time in-part validate some of the results analytically. Two facial images of the same subject (one original and one print-version) are expected to have intensity distributions which are similar to a scale factor (in terms of shape). However, the planar print version is expected to exhibit a lower dynamic range with respect to the intensity distribution. The following aspects are evaluated in the subsequent sub-sections:

- Statistical model and convergence to a fixed point and subsequent proof given in Appendix A.
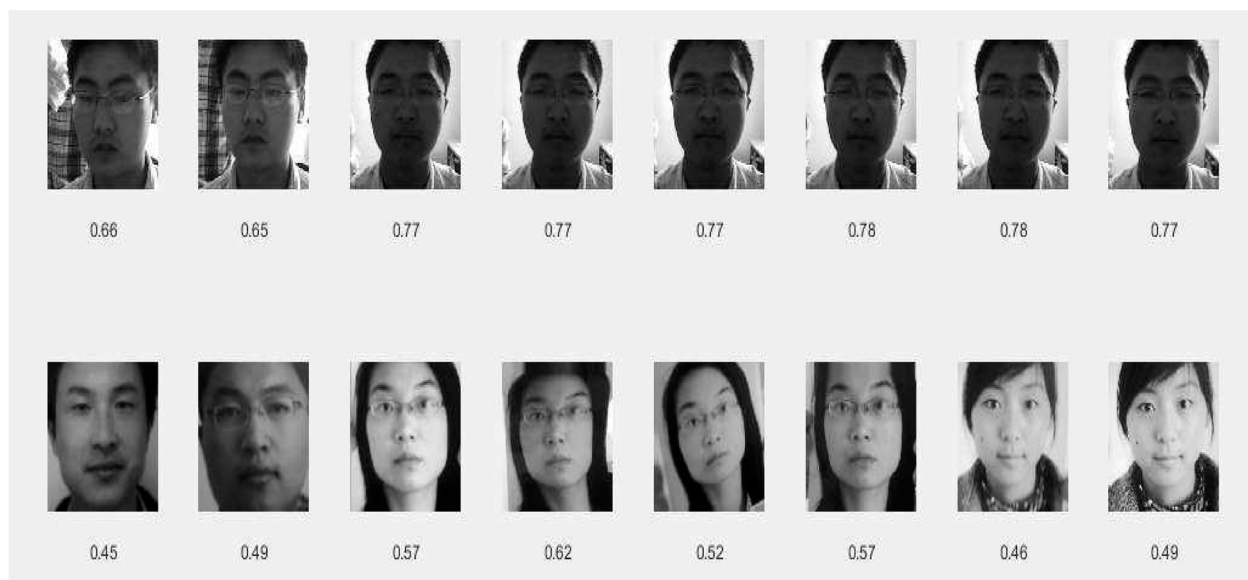
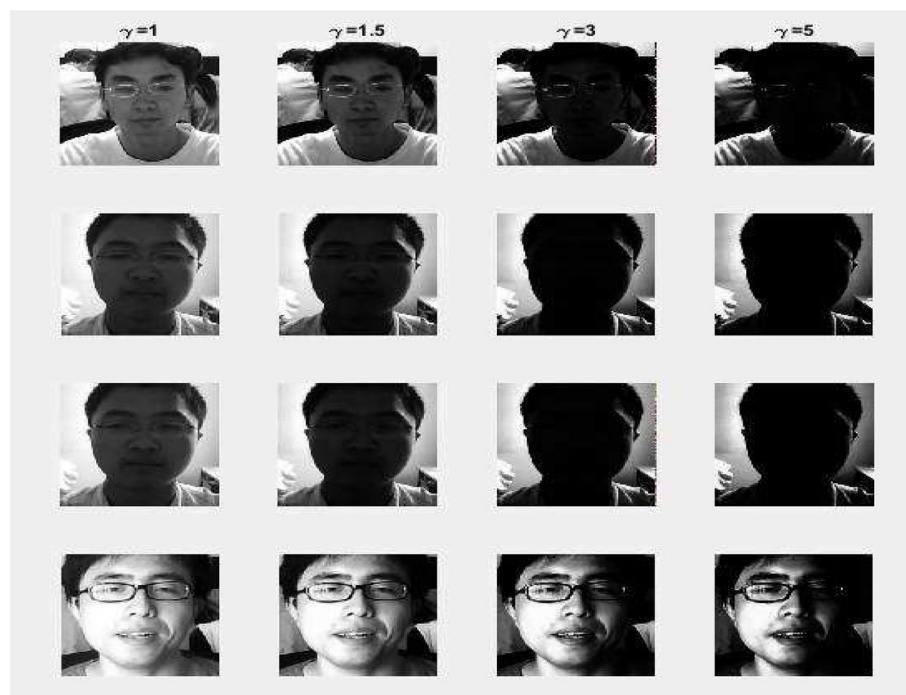**Fig. 6** Real and planar prints with contrast scores as per Eqn. (4)

Fig. 7 Impact of the Gamma power law on the degradation of the contrast profile of the original image (in the corresponding synthetic versions). Results are shown for γ = 1 (no transformation) and for 1.5, 3, 5

- Life trail dynamics discussing the rate of convergence of the real and print-abstractions as random sequences with proof details in Appendix B.

3.1 Fixed point and convergence analysis based on a simple statistical model

In this section, a simple statistical model is presented, to reflect the difference in dynamic range of natural and print

images. The original image is modeled as a random variable X_0 with a uniform distribution over the range (0, 1), while the print version is mapped to a uniform random variable Y_0, with reduced range (0, 1/a), where $a > 1$.

Impact of the iterative map function map on these two types of random variables is examined and some of the proofs are elaborated in Appendix A.

Let $f_x(x)$; $x \in (0, 1)$ represent the referential probability density function (PDF) of a normal face image corresponding to the global pixel-intensity distribution. In a crude way, its low contrast version after planar printing is defined based on the functional mapping based on the exponential law discussed in the earlier section,

$$Y_0 = X_0^\gamma \tag{5}$$

and this is expected to have a PDF,

$$f_{Y_0}(y) = a \times f_{X_0}(ay)$$

with $a > 1$, a shrinking of the referential density function is created, without compromising on the overall structure of the intensity probability density function (the number of inflection points and their relative positioning would remain the same). Note that $y \in [0, \frac{1}{a}]$; $a > 1$ with, $a = e^{1/\gamma}$ with $\gamma > 1$. Upon the application of the logistic map [26] to both these random variables and its planar-printed and low contrast counter-part, $Y_0 = X_0^\gamma$, secondary random variables (after one iteration) X_1 and Y_1 are formed, $X_1 = 2X(1 - X)$ and $Y_1 = 2Y(1 - Y)$.

It can shown that if $f_{X_0}(x)$ $UNIFORM[0, 1]$, then over successive iterations of this logistic map,

$$X_n = 2X_{n-1}(1 - X_{n-1}); n \geq 1 \tag{6}$$

the PDF of the transformed natural random variable, X_n, via this logistic map in the nth, $n \geq 1$ iteration is,

$$f_{X_n}(x) = \left(\frac{1}{2}\right)^{n-1}\left[\frac{1}{\sqrt{1-2x}}\right]^n \tag{7}$$

with $x \in [0, 0.5]$, which implies that once the logistic map is applied, for all the following iterations the points stay on the left side of $x = 0.5$ and approach the fixed point from the left. As n becomes large, it can be shown that $f_{X_n}(x) \approx \delta(x - \frac{1}{2})$, i.e.,

$$X_n \to 0.5 \text{ with prob. '1' for large } n \tag{8}$$

Similarly starting off with Y_0 $UNIFORM[0, (1/a)]$; $a > 1$ (uniformly distributed but reduced dynamic range) and applying the logistic map several times, one can manipulate the equations to obtain the result:

$$Y_n \longrightarrow 0.5 \text{ with probability '1' for } n >> 1 \tag{9}$$

This is illustrated in Appendix A.

3.2 Life trail dynamics

The intention here is to demonstrate when an image having a higher dynamic range in terms of intensity is subjected to the same logistic mapping, the convergence rate towards the fixed point is slower. For images with smaller dynamic ranges, the convergence is faster. The iterative functional mappings for both the natural (modeled by random variable X and print abstractions (modeled as random variable Y are:

$$X_n = 2X_{n-1}[1 - X_{n-1}]$$

with $n > 0$, and $X_0 = X\tilde{U}NIFORM[0, 1]$.

$$Y_n = 2Y_{n-1}[1 - Y_{n-1}]$$

with $n > 0$, and $Y_0 = Y\tilde{U}NIFORM[0, 1/a]$ such that $a > 1$. To monitor and track the fixed point convergence, the normalized first order difference metric is defined as,

$$G_n = \frac{X_n - X_{n-1}}{X_{n-1}} = 1 - 2X_{n-1}; n \geq 1$$

$$H_n = \frac{Y_n - Y_{n-1}}{Y_{n-1}} = 1 - 2Y_{n-1}; n \geq 1$$

It is shown in Appendix B, that the print-abstraction error sequence sequence, H_n converges faster in comparison with its counterpart, G_n, the real-image-abstraction error-sequence. Thus, it follows that the parent Y_n print-sequence because of a reduced dynamic range converges faster than the corresponding parent real image sequence X_n. In other words, life trails of low-contrast print images are shorter than the trails of real images.

3.3 Actual image life trails

While waiting for a precise convergence of all points is not necessary, in a practical image analysis setting, this convergence is approximate and designed to meet perceptual grounds with respect to a zero contrast image.

For a particular pixel positioned at location, (x, y), which is subjected to this non-linear mapping, the pixel is considered active if the value in the next iteration is significantly different from the earlier value. When two or more successive values are close, then the pixel in an approximate sense has assumed to have reached a saturation point and close enough to the fixed point. If I_n is the intensity level at iteration n, the pixel is considered to have converged and reached a saturation point if,

$$\frac{|I_n - I_{n-1}|}{I_n} < \epsilon \tag{10}$$

All the pixels with a non-zero intensity state are expected to drift towards the fixed point, which is 0.5 eventually. Note that the convergence rates are non-uniform and a
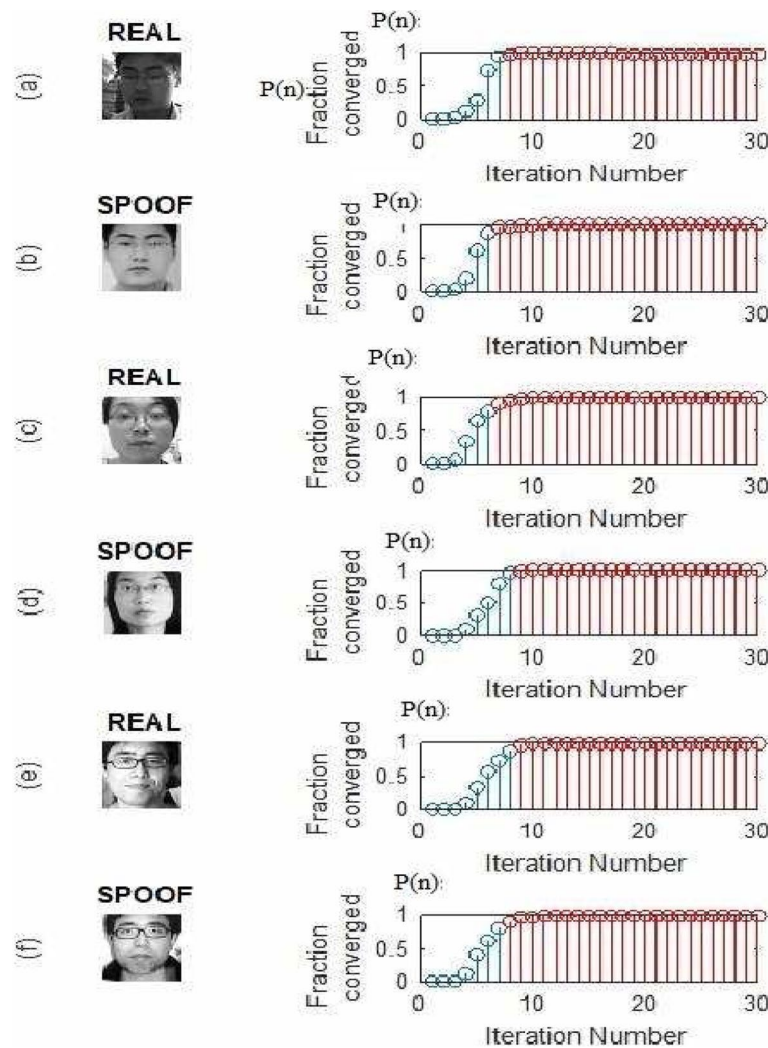
Fig. 8 Saturation curves which bring out the trends linked to the rate at which the initial image samples (either natural or spoof) converge to a zero-contrast image in the life trail

function of the initial value (or intensity state) of a particular pixel within the swarm. Hence, greater the spread of intensity levels (or diversity in the intensity profile), slower will be the swarm convergence. The entire swarm $SWARM(I_0)$ is said to have converged at iteration $n = s$, where s is the approximated saturation point of the complete image swarm if more than γ percent of the N^2 pixels ($\gamma \geq 0.9$) have met the convergence constraint given in Eq. 10 individually. This swarm convergence trend has been tapped using a saturation curve based on a function $P(n)$ (Fig. 8), where n is the iteration number. Typical saturation curves for natural and spoof images are shown in Fig. 8.

Figure 5 shows the contrast life trails of both natural and spoof images along with the termination points/saturation points. The overall swarm will converge only if almost all the pixels have converged and now the final image

saturation time to some extent depends on the MAXIMUM over all possible saturation timings across individual pixels. It is obvious that the more diverse the intensity profile, the greater the spread of intensity values, slower will be the swarm convergence. Natural face images tend to exhibit a higher dynamic range with respect to intensity in comparison with their planar print counter parts. The planar print versions tend to usually be of a lower quality, typically lower contrast [9], and limited color [17] as compared to the natural face images. Subsequently, on a subject specific note, these planar print images tend to have a shorter overall swarm life trail as compared to natural images. This can be seen in Fig. 5.

In the CASIA data-set, it was observed that there were some cases where the print versions had a very high quality and good clarity. Such cases turn out to be anomalies when examined from a life trail perspective. An example

of this is CASIA subject-11 shown in Fig. 5e, f, wherein the print quality almost matches the natural face quality.

Images with scale changes also tend to exhibit some form of anomalous behavior. Certain subjects tend to present their faces much more closer to the camera compared to others. A scale increase in a face turns out to be tantamount to a contrast reduction as the amount of detail in the image is reduced because of this zoom-in effect.

The swarm activity trails can be captured in the form of a global-image saturation level spotted at each iteration. These saturation graphs can be termed as S-graphs which tend to reflect an inverse trend in some cases. Hence, under scale variations and printing quality differences, the spoof detection may not prove to be fully effective. To attack this lack of universality with respect to the life trail lengths or S-curve trends, the focus is shifted to self-shadows. These self-shadow enhanced versions can be siphoned and generated from the same Image life trail when the original image swarm is passed through this logistic map.

4 Enhancing the self-shadows

One trend that is universal and remains independent of scale change in natural images and printing quality variations is the notion of perceptible self-shadows. These self-shadows are less prominent in spoof-print images, where they remain in a suppressed mode mainly owing to printing limitations and the superposition of secondary frontal lighting during the re-imaging process. Particularly, in the case of planar printing, the same natural image originally gathered from some unknown route is printed and presented again to an unmanned camera unit with a view to overcome the counter-spoofing system. Typically, such presentations are designed for low-end systems such as smart-phones which rely on their local mobile cameras for performing facial recognition to grant access to legitimate cell-users. Since in the case of planar spoofing the attacker must ensure a full face presentation with proper

uniform illumination to guarantee him/her access to a phone unit which belongs to another individual, a part of the originally trapped self-shadow information present in the printed photo tends to get suppressed by this secondary lighting. It is precisely this difference that this body of work picks out by extracting and enhancing the self-shadows.

This type of analysis is viable in indoor lighting and capture scenarios where invariably the sources are positioned towards one side of the individual's face creating in some cases a partial self-shadow. Given the original intensity normalized image $I_0(x, y)$, when this is passed through the logistic map [26] (one iteration only), a contrast reduced image is obtained, $I_1(x, y)$ such that,

$$I_1(x, y) = 2I_0(x, y)[1 - I_0(x, y)] \qquad (11)$$

A differential image can be generated from the life trail in one of the following ways,

$$R_1(x, y) = |I_1(x, y) - I_0(x, y)| = |I_0(x, y) - 2[I_0(x, y)]^2| \qquad (12)$$

$$R_2(x, y) = \left[\frac{|I_1(x, y) - I_0(x, y)|}{I_0(x, y)}\right] = |1 - 2I_0(x, y)| \qquad (13)$$

$$R_3(x, y) = [R_2(x, y)]^\alpha \qquad (14)$$

where, $\alpha \geq 1$. Since all these ratios can be exclusively expressed as a function of the original intensity pattern: $I_0(x, y)$, this can be treated as an intensity transformation.

The TWIN-image [30] in Fig. 9 has been used to illustrate the impact of the exponent α under two different illumination conditions: diffused lighting (right image) and virtually no self-shadows and regular outdoor lighting (left image) with the facial image showing prominent self-shadows. The main objective was to illustrate that when this exponent α is increase from "1" to a larger number, visually, the separation between the two images (RIGHT vs LEFT) with virtually the

Fig. 9 TWIN-image where one version is taken under normal outdoor lighting and the other one under diffused sunlight

same pose is best for some intermediate value of α. The right-twin image represents a spoofed low contrast image with virtually no self-shadows while the left-twin image mimics a natural image with prominent self-shadows further enhanced by the introduction of the exponential parameter α.

This exponentiation leads to an intensity transformation, which, makes the penumbral zones darker (zones where there are partial self-shadows). The part where there is no penumbra is made lighter. This is precisely why a power-law arrangement of the form $y = x^2$ or $y = x^\alpha$, where $\alpha > 1$ was deployed. Thus, the final enhanced image-statistic was, $E_\alpha(x, y) = R_{n=1}(x, y)^\alpha$.

For most natural images, it was found that when this α was increased beyond a certain point, even the non-penumbral zones were darkened. On the other hand, too small a value of α did not have much of an impact on the original self-shadows. This process of arriving at the optimal α can be done more reliably with an analytical twist using the same probability model discussed earlier.

4.1 Justification for first, first-order difference ratio

Analytical proof as to why the first, first-order difference provides maximum information related to the self-shadows is provided in this segment. Given the normalized error term for the natural image abstraction, $G_n = (1 - 2X_0)^{2n}$ for $n \geq 2$ and $G_1 = (1 - 2X_0)$, where, X_0 has a uniform PDF over the interval $[0, 1]$.

For $n \geq 2$, the PDF of G_n can be derived using the classical random variable transformation analysis [31] as,

$$f_{G_n}(g) = \frac{1}{2n}\left(g^{\left(\frac{1}{2n} - 1\right)} \right) \tag{15}$$

where $g \in [0, 1]$. The continuous/differential entropy ([32]) of G_n can be evaluated as,

$$H[G_n] = -E_{G_n}\left[\ln\left(f_{G_n}(G) \right) \right] \tag{16}$$

where the expectation is with respect to $G_n = G$.

$$H[G_n] = -\int_{g=0}^{1} f_{G_n}(g) \ln[f_{G_n}(g)] dg$$

Can show that this evaluates to,

$$H[G_n] = \ln(2n) - 2n + 1$$

which is a decreasing function of n, with the value obtained for $n = 2$ as, $H[G_2] = 2 \times 0.693 - 3 = -1.6137$. For $n = 1$, since the same random variable evaluated at $n = 1, i.e., G_1 = 1 - 2X_0$ is uniform over the interval $[-1, 1]$, the entropy $H[G_1] = \ln(2) = 0.693$ is MAXIMUM and is greater than the entropies evaluated for $n \geq 2$. This is a decaying trend with respect to entropy.

This implies that the self-shadow statistic provides maximal information when $G_n = 1$ is used as the normalized ratio statistic. All other differences larger than $n = 1$, provide less information than the information contained in the first difference ratio. Since, the distribution for G_1 is uniform in a larger sense this can serve as a SUFFICIENT STATISTIC for trapping maximal self-shadow information.

4.2 Connection of the exponential parameter with the statistical model

The first difference normalized ratio as seen in the earlier section, traps the self-shadow pattern to a certain degree of statistical sufficiency. Thus, it is enough to use this ratio statistic to derive the final feature vector for building a subject-specific 2-class SVM model. From the point of view of model building there were two motives for choosing this additional parameter and not just feeding on the ratio statistic:

- While the conditional ratio statistics, $G_1 = 1 - 2X_0$ and $H_1 = 1 - 2Y_0$ where $X_0 : UNIF[0, 1]$ and $Y_0 : UNIF[0, \frac{1}{a}]$ carry sufficient information to trap self-shadow information, one factor which is of prime importance is the class separation with respect to real and spoof. It may be possible to post process these stats in such a way that the self-shadow profiles associated with real and spoof images are pushed further apart. This has been attempted via an exponentiation procedure as the exponentiation is likely to modify the dynamic ranges of both ratios.
- Let $R_{REAL} = [G_1]^\alpha$ and $R_{SPOOF} = [H_1]^\alpha$ with $\alpha > 0$. Define $\Delta H(\alpha) = H[R_{REAL}] - H[R_{SPOOF}]$, as the difference between the information contained in the self-shadow profiles of the real and spoof versions, where $H[R_{REAL}] = -E_R[\ln f_{REAL}(R)]$. Selection of α must be done to ensure $-\Delta H(\alpha)$ is as small as possible.
- On the other hand, the absolute information contained in the self-shadow profile of the natural face image, i.e., $H[R_{REAL}] = E_{REAL}(\alpha)$. should not be reduced significantly as this would impede the detection procedure.

Claim 3 *The selection of the exponent α is based on judicious tradeoff between maximizing the self-shadow information present in natural faces while at same time increasing the class-separation between the self-shadow distributions of the real and spoof classes. These two requirements are slightly conflicting.*

Thus, the choice of the exponential parameter must be done to ensure $-\Delta H(\alpha)$, is lowered as much as possible

without compromising on the information contained in the absolute entropy of modified ratio statistic corresponding to the real face image, i.e., $E_{REAL}(\alpha)$ must be as large as possible.

It can be shown that,

$$|G_1| : UNIF[0, 1]$$

and

$$|H_1| : UNIF[0, \frac{1}{a}]$$

with $a > 1$. Using the random variable transformation formulation from Papoulis et al. [31],

$$f_{R_{REAL}}(r) = (\frac{1}{\alpha}) r^{\frac{1}{\alpha} - 1} \qquad (17)$$

where $r \in [0, 1]$ and

$$f_{R_{SPOOF}}(r) = (\frac{a^{\alpha}}{\alpha}) r^{\frac{1}{\alpha} - 1} \qquad (18)$$

where $r \in [0, \frac{1}{a}^{\alpha}]$. Subsequently,

$$H[R_{REAL}] = ln[\alpha] + 1 - \alpha \qquad (19)$$

$$H[R_{SPOOF}] = ln(\frac{\alpha}{a}) + (1 - \alpha)(1 + ln(a)) \qquad (20)$$

for $a > 1$ and $\alpha > 0$. This gives two important metrics, (i) connected with the difference between real and spoof self-shadow entropies,

$$-\Delta H(\alpha) = H[R_{SPOOF}] - H[R_{REAL}] \qquad (21)$$

$$= -\alpha \times ln(a) \qquad (22)$$

and (ii) absolute entropy of the natural face self-shadow statistic as,

$$E_{REAL}(\alpha) = ln(\alpha) + 1 - \alpha \qquad (23)$$

When the dynamic range parameter a is known or is estimated from the real and spoof versions corresponding to a particular calibration set, the operating point is decided by the point of intersection of the two constraints for the measured $\hat{a}$. This is illustrated in Fig. 11. For different values of a different sets of contraints are obtained out of which one has to be picked based on the computation. Keeping in mind that the attacker will ensure a reasonable quality associated with planar prints, one need not expect a to go above 2-units. A value of $a = 2$ would correspond to a 50% drop in the dynamic range of the print version in relation to the natural intensity profile (Fig. 10).

5 Operating point and initial calibration

The right choice of exponent α to strike a balance between the quantum of self-shadow information obtained from the differential ratio statistic taken from the life trail of natural faces and the differential entropy statistic is decided by a calibration process. The family of curves (seen in Fig. 11) is dependent on
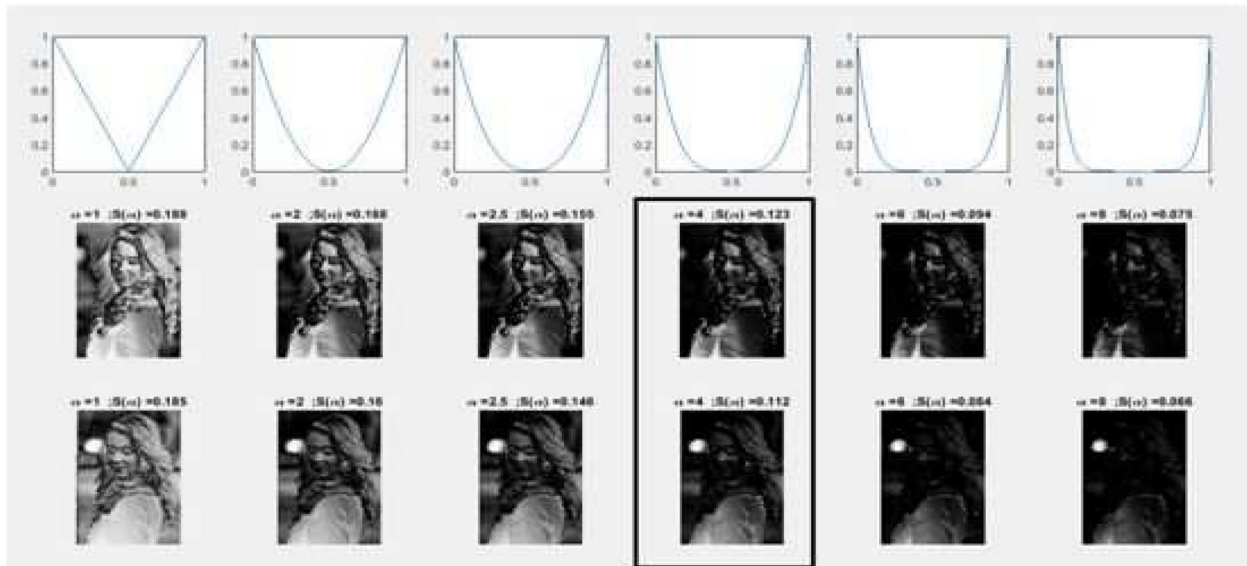

Fig. 10 Impact of changes in the exponential parameter α on both the versions from the TWIN-image set [28]. As the exponent increases, the self-shadows become much more discernible for the version where the lighting is normal. Beyond a certain point the ratio images corresponding to both the normal version and the diffused version become dark
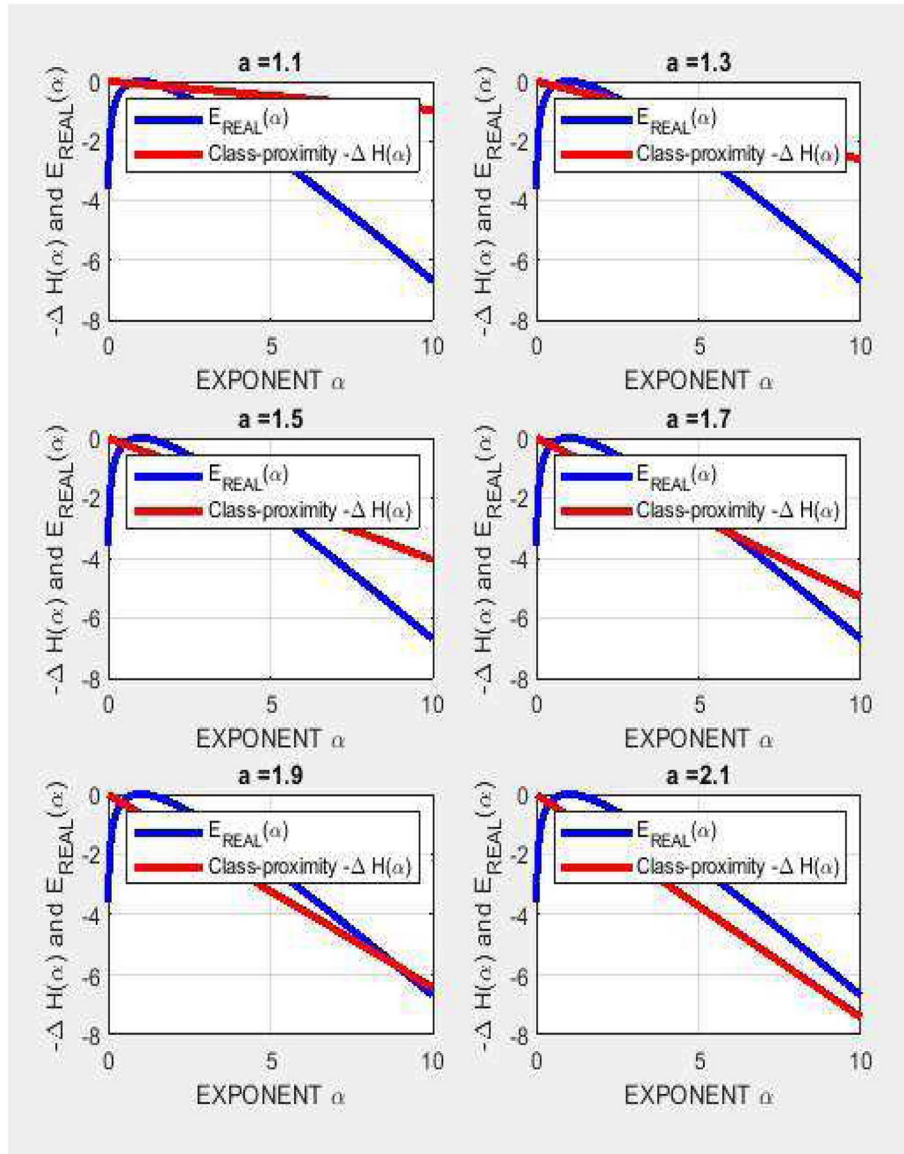
Fig. 11 The selection of the operating point, as the point of intersection towards the right of the RED and BLUE curves, to maximize class separation (not the one on the left) is shown for different values of α

the knowledge of the dynamic range parameter $\hat{a}$, connected with the print-spoof image intensity profile. It is therefore imperative that there be an elaborate procedure for estimating this parameter $\hat{a}$, on both relativistic as well as approximate grounds, via measurements taken over the real and spoof image sets derived from calibration data. This calibration procedure for α is designed as follows,

- Take 5 subjects with a total of 75-samples from both real and spoof classes, from the the dataset being scrutinized.

- For a particular image sample in the real-class, generate the global contrast score [9], (obtained from Eq. (4)).

$$CR_i = \sigma_i / \mu_i \tag{24}$$

- The mean contrast score for natural faces is,

$$CON_{REAL(E)} = \frac{1}{N_{CALIBREAL}} \sum_{i \in SET_{CALIBREAL}} CR_i \tag{25}$$

where $N_{CAL_{REAL}}$ is the number of real subject face samples and $SET_{CALIBREAL}$ is the set of indices of real images deployed towards calibration.

- Similarly, for the spoof/print segment from the calibration set,

$$CON_{SPOOF(E)} = \frac{1}{N_{CALIBSPOOF}} \sum_{i \in SET_{CALIBSPOOF}} CS_i \qquad (26)$$

where $N_{CAL_{SPOOF}}$ is the number of spoof/print subject face samples and $SET_{CALIBSPOOF}$ is the set of indices of spoof images deployed towards calibration.

- To cross-reference this measurement profile against the analytical model and the curves shown in Fig. 11, the mean contrast score of the real-calibration set is referenced against the spoof set taking a ratio of the two:

$$\hat{a}_F = \frac{CON_{REAL(E)}}{CON_{SPOOF(E)}} \qquad (27)$$

Note that if this relativistic normalized dynamic range parameter, $\hat{a}_F$, is close to UNITY or is smaller than unity, then the counter-spoofing system based on contrast reductionist life trails will not be very effective. However, because of the physical acquisition process, the spoof print version will always have a lower contrast as the corresponding original version. This will induce a high likelihood towards the EVENT, $\hat{a}_F > 1$, from the measurements taken over the calibration set. This also explains why this method may not work on backlit planar images produced by tablets and laptops.

Use the family of curves from Fig. 11 (or an elaborate lookup table) and pick out the optimal value of α for that dataset based on the corresponding quantum-value associated with $\hat{a}_F \in [1.1, 1.3, 1.5, 1.7, 1.9, 2.1]$. For the CASIA-dataset, 5 subjects, with 75 samples per class, the parameters estimated were $CON_{REAL(E)} = 0.5889$; $CON_{SPOOF(E)} = 0.4716$; and $\hat{a}_F = 1.2487$. This quantum corresponds to $a = 1.2487$ pointing to an operating point of $\alpha_{CASIA} = 2.7$.

6 Final feature extraction procedure and client-specific classification

Block diagrams of the feature extraction procedure following by the classification and testing are shown in Figs. 1 and 2 respectively.

6.1 Secondary statistics

To derive the feature sets and statistics for every image I_0, a size normalization was done and all images were resized to $N \times N$ pixels, with $N = 250$. The enhanced self-shadow image $R(x, y)$, is constructed by passing this swarm $SWARM(I_0)$, through a logistic map, to produce contrast reduced image represented by $SWARM(I_1)$ in the life trail. A secondary differential ratio image as discussed earlier was generated:

$$E_\alpha(x,y) = R_3(x,y) = \left[\frac{|I_1(x,y) - I_0(x,y)|}{I_0(x,y)} \right]^{\hat{\alpha}} \qquad (28)$$

where $\hat{\alpha}$ can be obtained via a calibration process discussed in the previous section. This self shadow enhanced image with parameter $\hat{\alpha}$ is placed in a rectangular grid and intensity standard deviations are computed for every patch. The patch size was chosen as 10% of the image size for this initial simulation setup. The secondary statistics matrix can be written as,

$$S = \begin{pmatrix} \sigma_{1,1} & \sigma_{1,2} & ... & \sigma_{1,n} \\ \sigma_{2,1} & \sigma_{2,2} & ... & \sigma_{2,n} \\ ... & ... & ... & ... \\ \sigma_{n,1} & \sigma_{n,2} & ... & \sigma_{n,n} \end{pmatrix} \qquad (29)$$

with,

$$\sigma_{i,j} = \sqrt{\frac{1}{W^2} \sum_{(x,y) \in PATCH(i,j)} (R_3(x,y) - \mu_{i,j})^2} \qquad (30)$$

where

$$\mu_{i,j} = \frac{1}{W^2} \sum_{(x,y) \in PATCH(i,j)} R_3(x,y) \qquad (31)$$

The complete algorithm from the image to the final feature and scalar statistics (both normalized and un-normalized) is discussed below :

6.2 Complete algorithm: generating self-shadow statistics from images

Step 0: Image size normalization while preserving the aspect ratio

Resizing the original $N_1 \times N_2$ image to $N \times N$, with $N = 250$

Step 1: Formation of swarm/collection of pixel intensity values over the entire image

$$DOMAIN_0 = \left\{ \begin{array}{c} (x,y) \ s.t. \ x \in \{1, 2, ...250\} \ and \\ y \in \{1, 2, ...N_c\} \end{array} \right\}$$

$$SWARM_0 = \{I_0(x,y) : \ s.t. \ (x,y) \in DOMAIN_0\}$$

where $I_0(x,y) \in [0,1]$ is the normalized luminance intensity level in the facial image.

Step 2: Application of the non-linear mapping to the entire swarm individually. Evaluate this iteratively for the entire *SWARM* for $n = 1, n = 2, \ldots, n = n_{TYPICAL}$ where $n_{TYPICAL} = 30$.

$$\forall (x,y) \in DOMAIN_0, \ I_n(x,y) = 2I_{n-1}(x,y) \big[1 - I_{n-1}(x,y)\big]$$

Based on observations across subjects picked from the CASIA dataset, typical convergence timing, in terms of number of iterations for natural images, is around 10 and for spoof images is around 8. To ensure complete convergence as far as the life trail is concerned, the maximum number of iterations has been set to $n_{TYPICAL} >> MAX(N_{TYP-NAT}, N_{TYP-SPOOF})$.

Step 3: Self-shadow enhancement via first-order differences as one traverses the LIFE trail

Stop with the first iteration: $I_{(n=1)}(x,y) : (x,y) \in DOMAIN_0$ Define

$$R(x,y) = \frac{\left(|I_1(x,y) - I_0(x,y)|\right)}{I_0(x,y)}$$

$$E_\alpha(x,y) = R(x,y)^{\hat{\alpha}}$$

Step 4: Computing the patch-wise intensity diversity statistic. Let $\beta \in (0,1)$ be the fractional patch size with respect to the ratio image ($E_\alpha(x,y) = R(x,y)^\alpha$), which is of the same size as the original image, i.e., 250×250. Set $\beta = \beta^* \in (0,1)$ ($\beta \in \{2\%, 5\%, 10\%, 20\%\}$ of $N = 250$, based on simulation experiments conducted and the tuning procedure related to a specific dataset. Let the patch size be $W \times W$ with $W = \lfloor \beta \times N \rfloor$. Let (x_p, y_p), be the top-left corner of the patch within the RATIO image statistic: i.e., $E_\alpha(x,y)$.

$$DOMAIN_{Patch(p)} \left\{ \begin{array}{c} (x,y) : s.t \\ \in x \in \{x_p, ..., (x_p + W - 1)\} \\ y \in \{y_p, ..., (y_p + W - 1)\} \end{array} \right\}$$

$\forall (x,y) DOMAIN_{Patch(p)}$ Compute

$$\mu_p = \frac{1}{W^2} \sum_{(x,y) \in DOMAIN_{Patch(p)}} E_\alpha(x,y)$$

$$\sigma_p = \sqrt{\frac{1}{W^2} \sum_{(x,y) \in DOMAIN_{Patch(p)}} \left[E_\alpha(x,y) - \mu_p\right]^2}$$

Step 5: Statistics for analysis. Two types of statistics were computed. $TYPE - 1$: Pure variances from the ratio-image patches and their mean as the scalar statistic. This arrangement suffered from a statistical aperture effect with respect to patch size fractional increase (i.e., due to an increase in β). Hence, a normalized version was developed as $TYPE - 2$. The latter, i.e., TYPE-2 was deployed in the final test, while TYPE-1 was used in the calibration segment with respect to the trimmed version of the CASIA dataset (14 subjects). The scalar feature parameter can be chosen for the given image as, the mean diversity from the ratio image,

$$STAT_{RAW}(I_0) = \frac{1}{N_{patches}} \sum_{\forall patches} \sigma_p \textbf{ TYPE1} \quad (32)$$

$$LSTAT_{NORM}(I_0) = \frac{2}{N_{patches}} \sum_{\forall patches} \left[|ln(\sigma_p)|\right] \textbf{ TYPE2} \quad (33)$$

The vector feature is a simple raster scan of all the σ parameters.

6.3 2-class SVM models for each client/subject

The original CASIA set [18] was deployed in the final testing round (50 subjects, 3×30 variations per subject at three different quality levels: low, medium, and high). From the original CASIA set, a reduced version was used as a calibration set from the point of view of algorithm refinement, final feature selection, keeping difficult subjects, and their variations in the backdrop. Final round test databases chosen for unbiased evaluation were OULU-NPU [27] and CASIA-SURF [28].

The reduced CASIA set had 14 subjects with 30 variations per subject covering both natural and print-spoof images. Thus, there were a total of 420 images across 14 subjects for natural and 420 images covering 14 subjects for print-spoofing. Out these 14 subjects, subjects 4, 6, and 11 have been identified as the anomalous and difficult ones (Fig. 12) keeping in mind various factors:

- From the point of view of subject 4, there was a significant scale change/increase since the subject was closer to the camera than normal. This reduced the dynamic range in the intensity space leading to shorter life trails for natural faces as compared to the spoof ones (Fig. 12a, first and second images).
- From the point of view of subject 6, there were cases where the light source was present in front but above the subject. This suppressed the self-shadow profile considerably for some natural images (Fig. 12a, third and fourth images).
- In subject 11, the problem was very different and existed in the spoofing segment (Fig. 12b, fifth and sixth images), wherein the printing and re-imaging quality was very high and comparable to that of a natural face image.

Thus, the life trail lengths turned out to be similar for natural and spoof faces for these anomalous cases.

To check the precision of the proposed algorithm, the CASIA set was segregated subject-wise (across both natural and spoof segments) and 50% of the variations per natural or print-version was used to build a 2-class-subject-specific SVM model [17, 20]. The remain 50% of the samples from both the natural and spoof segments were used for testing. The t-SNE maps [33] of the reduced CASIA set test set on a subject specific basis are shown in Fig. 13. The corresponding error rates for the test samples are shown alongside. The overall error mean equal error rate (EER) across all subjects for this reduced calibration CASIA dataset is 0.48% for the ratio-mapping parameter $\alpha = 2.5$. The error rates climb for
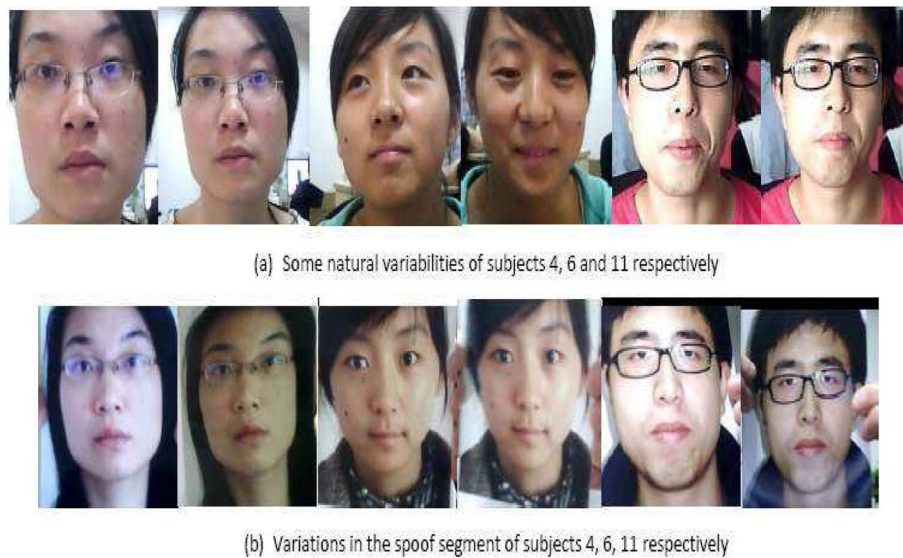
(a) Some natural variabilities of subjects 4, 6 and 11 respectively

(b) Variations in the spoof segment of subjects 4, 6, 11 respectively

Fig. 12 Anomalous cases in CASIA which have a tendency to induce mis-classifications (Subjects 4, 6 and 11); (**a**) Some natural variations; (**b**) Spoof variations; Ordering is Subject 4, 6 and then 11

values less than $\alpha = 2.5$ and larger than $\alpha = 3.5$. The client/subject specific cluster separations have been generated using t-SNE mappings [33] (a stochastic map which presents a fairly realistic lower dimensional representation of higher dimensional data) in Fig. 13. In all the subject specific subplots of the test-data, Fig. 13a–n, the cluster separation was found to be excellent, attesting and reinforcing CLAIMS 1 and 2.

7 Database description

In this section, a description of three different datasets, CASIA [18], OULU [27], and CASIA-SURF [28] is provided, then in the second phase of the calibration protocol in which the parameter β^* is decided based on a parameter sweep for database-specific values of α^* obtained using the calibration protocol discussed earlier. Based on these optimized parameters, subject-specific model building, testing, and comparisons form the last few subsections.

A summary of the datasets used for final round testing of the proposed life trail algorithm is provided in Table 1. The original CASIA face dataset [18] shown in Fig. 14 which was created from Chinese individuals showed significant variability on both the natural face front as well as the planar spoofing front. The variability as far as the natural faces were concerned encompassed minor pose variations, significant light source positional variations, scale changes, etc. The variability as far as print-spoofing was concerned stemmed from color variations and minor scale variations depending on the manner in which the printing was done. The CASIA print set had 50 subjects and images were captured under different image acquisition resolutions (low, medium, and high). Each resolution level had 30 variations

per subject for both natural and print classes. The OULU-NPU dataset [27] shown in Fig. 15, on the other hand, contained spoof samples related to print-photo and video attacks, along with natural face samples. The face presentation attack sub-database consisted of 4950 real access and attack videos that were recorded using front facing cameras of six different smartphones over a varied price range. The print attack was created using two printers (printer 1 and printer 2) and two display devices (display 1 and display 2) out of which 20 subjects were publicly available. The enrolled users were mostly Europeans and people from the middle east. Pose and scale changes were minimal here.

The CASIA-SURF [28] shown in Fig. 16 is a wide dataset with real and spoof samples along with depth profiles. This dataset contained samples of 1000 Chinese individuals from 21000 videos across three modalities (RGB, Depth, IR). There were six scenarios under which the print-photo attacks were implemented:

- Attack 1: Person holding his/her flat face photo with the eye-region cut
- Attack 2: Person holding his/her curved face photo with eye-region cut
- Attack 3: Person holding his/her flat face photo with eye and nose regions cut
- Attack 4: Person holding his/her curved face photo with eye and nose regions cut
- Attack 5: Person holding his/her flat face photo when eye, nose and mouth regions are cut
- Attack 6: Person holding his/her curved face photo when eye, nose and mouth regions are cut
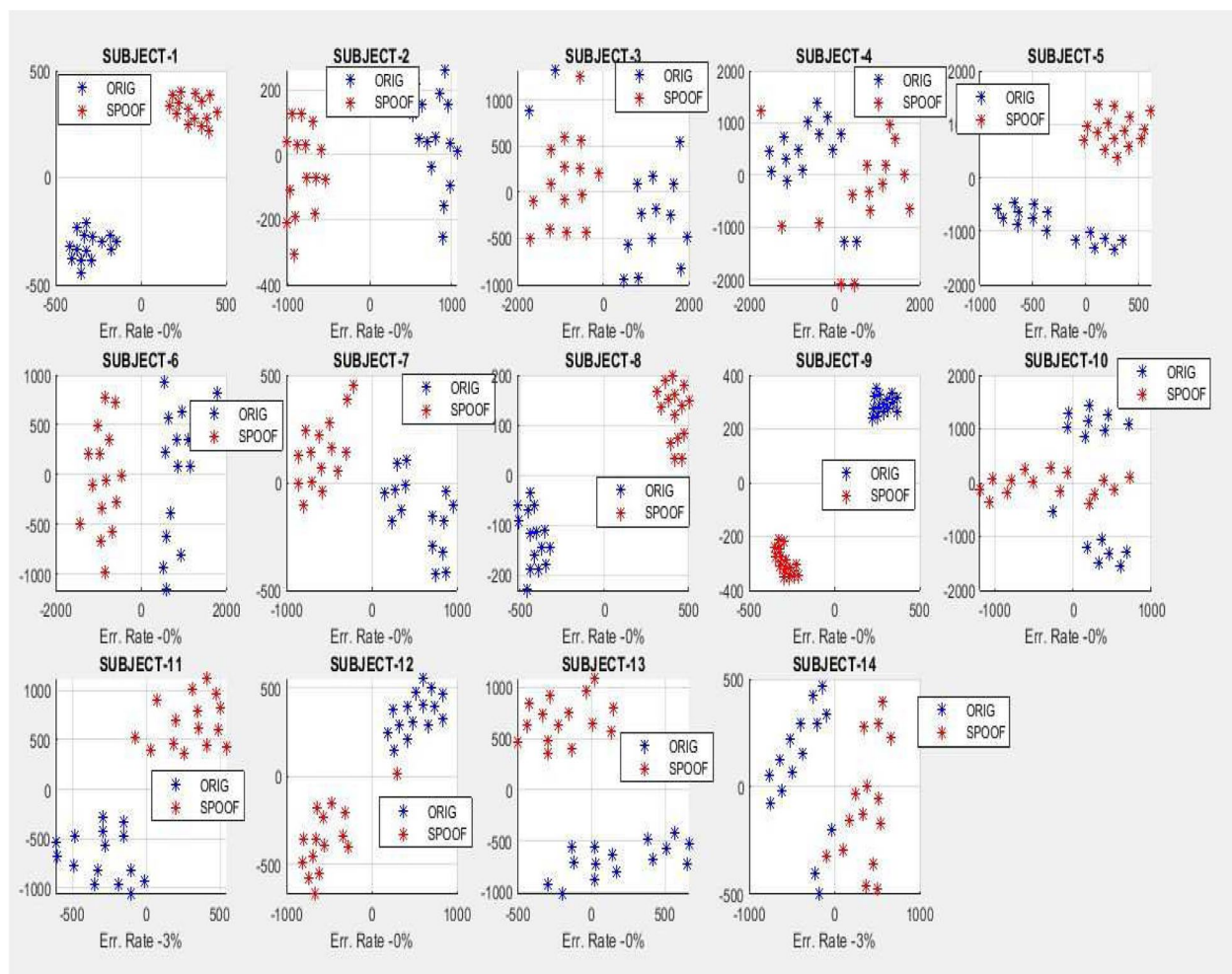
Fig. 13 Cluster separation (subject-wise) in a 2-class setting, for the reduced CASIA-dataset comprising of 14-subjects (out of a total of 50) in which 50% of the variations per subject, were used for testing

Fig. 14 Examples (both natural and spoof-print versions) from the original CASIA dataset [18]
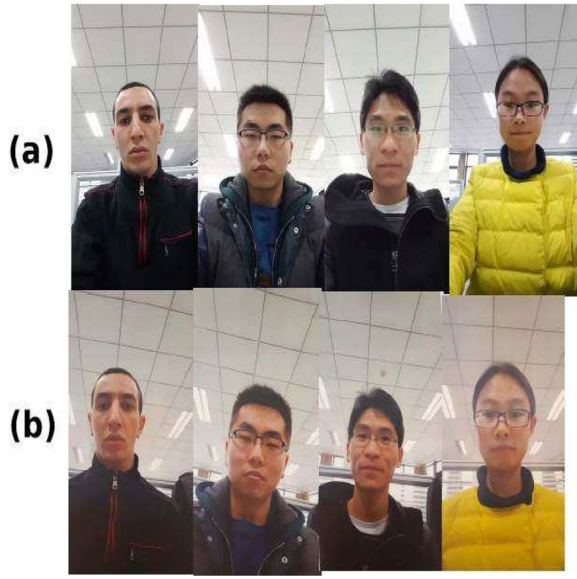
Fig. 15 Examples (both natural and spoof-print versions) from the original OULU-NPU dataset [31]

8 Final customized calibration and testing on different datasets

There are two parameters which are a function of the acquisition process and the environment in which the face images are generated. These are the exponent α, which is associated with the first, normalized first difference ratio statistic which captures the self-shadow information with a certain degree of sufficiency and other happens to be the patch size-fraction $\beta \in [0, 1]$ which decides the dimensionality of the feature space.

In close cropped images from datasets such as CASIA and CASIA-SURF, the face is virtually fully inscribed inside the "image-rectangle" (we take this as the referential 1:1 scenario). Here, the patch fraction β is expected to be around 10% to 20%. However, in datasets such as OULU, where the face is small part of a bigger background (here the ratio of face to whole rectangular area

inversely proportional to the square root of the size of the patch was introduced (this is mentioned as the TYPE-2 statistic in the scalar abstraction in the Algo. 6.2(Step 5).

If σ_p is the patch standard deviation, the quantum of self-shadow information present in it can be approximately represented as,

$$L_p = |ln(\epsilon + \sigma_p)| \tag{34}$$

where ϵ is a small positive number. The average self-shadow information for a given image can then be computed as,

$$LSTAT = \frac{1}{N_{patches}} \sum_{p=1}^{N_{patches}} L_p \tag{35}$$

Let $u_1, u_2, ..., u_r$ be the LSTAT-scores computed from the natural face calibration set and let $v_1, v_2, ... v_r$ ($r = 75$) be the LSTAT-scores produced from the spoof-set. From these conditional LSTAT-scores, two conditional means and two conditional variances are computed:

$$
\begin{aligned}
\mu_{REAL} &= \frac{1}{r} \sum_{k=1}^{r} u_k \\
\mu_{SPOOF} &= \frac{1}{r} \sum_{k=1}^{r} v_k \\
\sigma_{REAL}^2 &= \frac{1}{r} \sum_{k=1}^{r} (u_k - \mu_{REAL})^2 \\
\sigma_{SPOOF}^2 &= \frac{1}{r} \sum_{k=1}^{r} (v_k - \mu_{SPOOF})^2
\end{aligned}
\tag{36}
$$

The separation between the two clusters as function of the parameter β for a particular calibrated α^* can be determined based on the symmetric version of the Kullback-Liebler (KL) divergence [34], under a conditional Gaussian assumption for the two classes: real and spoof. This metric based on KL-divergence for two univariate Gaussian distributions can be computed as:

$$SEPARATION_{KLD} = \left(\frac{\sigma_{REAL}^2}{\sigma_{SPOOF}^2} + \frac{\sigma_{SPOOF}^2}{\sigma_{REAL}^2} \right) + (\mu_{REAL} - \mu_{SPOOF})^2 \left(\frac{1}{\sigma_{REAL}^2} + \frac{1}{\sigma_{SPOOF}^2} \right) \tag{37}$$

drops to 1:4), the optimal patch fraction (β) is expected to decrease, keeping the volume of perceptual information connected with self-shadow details the same.

To shortlist the optimal parameter for each dataset, 5 subjects with a total of 75 samples from each class were chosen and used to generate the class separation scores. To compensate for the statistical aperture effect stemming from the patch size increase, a normalizing factor

The impact of a β parameter sweep for specific values of α, i.e., obtained via the initial exponential parameter calibration procedure is shown in Table. 2. For a specific database, when β is varied for a fixed α, the separation scores show a clear maximum for some $\beta = \beta^*$. It was observed that for the CASIA-SURF dataset, where the dynamic ranges of both the natural and spoof/print faces were close, optimal $\beta_{CASIA-SURF} = 0.15$ corresponding to an $\alpha_{CASIA-SURF} = 1.7$. On the other hand

Table 2 Separation scores for all three datasets: CASIA, OULU, and CASIA-SURF

Parameters	$\longleftarrow \beta \longrightarrow$					
	0.05	0.1	0.15	0.2	0.25	0.3
$\alpha_{CASIA} = 2.7$	1.75	1.79	1.97	3.43	**4.80**	2.43
$\alpha_{OULU} = 3.4$	42.26	**95.27**	64.80	48.44	40.70	56.74
$\alpha_{CASIA-SURF} = 1.7$	4.75	3.05	**6.90**	1.09	1.20	1.14

for the standard CASIA dataset, such fine grained scrutiny of the self-shadow image was not required and the optimal $\beta_{CASIA} = 0.25$ for an $\alpha_{CASIA} = 2.7$. For OULU, however, since the face information was a small part of a larger background, it was natural to expect the optimal β^* to drop to $\beta_{OULU} = 0.1$ for an $\alpha_{OULU} = 3.4$. The final parameters from the two-stage calibration procedure have been captured in Table 3.

8.1 Testing: Experimental results and comparison with literature

There are two primary paradigms designed to suit two different types of applications: (i) the subject identity not known a priori, i.e., a face is presented to the camera and the counter-spoofing system must decide whether face-presentation is natural [4, 15, 18, 23, 35], and (ii) the subject identity is known to the counter-spoofing system (more like an authentication environment) [17, 20].

The proposed image trail architecture was evaluated over a client specific frame (i.e., Type-(ii), subject ID known). Since client specific architectures effectively suppress subject-mixing noise or registration noise, the error scores are much lower here (Table 4) as compared to the subject-independent error scores (Table 5). The best among them is the random walk/scan-based algorithm [4, 5] which uses short-stepped random walks to not just trap the short-term spatial correlation statistics but also to generate several equivalent randomly scanned realizations of the same parent face-image to transform an image feature to blob (or an ensemble), which can be used highly reliably to capture the natural immersive environment in a truly subject agnostic fashion. Error rates for the print-presentation attack (CASIA) for the random scan algorithm were reported as 3.5122% (without auto-population) and 1.8920 % (with auto-population). To begin with, this became one of the benchmark error measures

Table 3 Database and optimal parameter values for various databases, based on the tuning procedure

Database and Optimal parameters	$\alpha*$	$\beta*$
CASIA	2.7	0.25
OULU	3.4	0.1
CASIA-SURF	1.7	0.15

against which the proposed life trail based approach in a client-specific setting needed to be compared.

For the complete CASIA print dataset (50 subjects, 3×30 variations per subject for three different quality levels), the proposed life trail algorithm showed a comparable error rate of 0.310% Table 4. With respect to state of the art client-specific face counter-spoofing architectures, the proposed life trail algorithm performed better than most on the planar-printing front.

The error rates of the proposed algorithm observed for the OULU-NPU dataset [27] was 1.192% and that for the CASIA-SURF [28] was found to be 2.246%. These numbers were comparable with the convolutional neural network (CNN)-based solutions shown in Table. 4. Notice that in the case of CASIA-SURF, the CNN-based solutions available in [43], depth map information was augmented with RGB information to support the learning process. With pure RGB information, these error numbers will be higher.

9 Random scan extension to facilitate cross-validation

Random scans [4, 5] were developed to capture acquisition noise statistics while suppressing both content and subject-content interference. Contiguous random scans in the form of space filling curves (SPCs) [44], were originally designed for communications applications to facilitate compression of videos after shuffling. These contiguous random scans when deployed towards face counter-spoofing have a few interesting properties:

- The scans preserve the first, second and third order pixel-intensity correlation statistics in a particular image.
- By executing the same scan multiple times on the same image or patch, one can auto-populate the features or statistics derived from a typical scan, at an ensemble level. An illustration of a short contiguous scan in given in Fig. 17.
- Secondary differential statistics can be computed over the scanned vectors of the first, second, and third order to trap the mean acquisition noise energy over the entire image. Thus, every image can be abstracted as a 3-dimensional feature vec-

Table 4 State of the art methods within a client specific frame.C_{sp}: represents the subject-specific or client specific mode of training and testing; in*Protocol − I*below used on the OULU set, the sameC_{sp}mode has been deployed

Method	Classifier	Train data	Test data	Threshold	EER
LBP [36]	GMMUBM	CASIA	CASIA	C_{sp}	10.09
	Two-class SVM	CASIA	CASIA	C_{sp}	9.87
LBP-TOP [36]	GMM UBM	CASIA	CASIA	C_{sp}	6.36
	Two-class SVM	CASIA	CASIA	C_{sp}	3.95
Motion [36]	GMMUBM	CASIA	CASIA	C_{sp}	9.66
	Two -class SVM	CASIA	CASIA	C_{sp}	11.27
MSLBP [20]	Two-class SVM	CASIA	CASIA	PS-iFAS Test-S	5.60
	Two-class SVM	CASIA	CASIA	PS-iFAS Test-T	2.26
	Two-class SVM	CASIA	CASIA	PS-iFAS Test	3.59
HOG [20]	Two-class SVM	CASIA	CASIA	PS-iFAS Test-S	0.82
	Two-class SVM	CASIA	CASIA	PS-iFAS Test-T	5.045
	Two-class SVM	CASIA	CASIA	PS-iFAS Test	3.35
CNN [37]	Deep CNN	CASIA	CASIA	$C_{sp}, \alpha, \beta, \gamma$	1.85
Radiometric distortion [12]	Two-class SVM RBF	CASIA	CASIA	C_{sp}	0.00
CPqDN [38]	CNN	OULU-NPU	OULU-NPU	*Protocoll* [39]	6.9
GRADIANT [38]	CNN	OULU-NPU	OULU-NPU	*Protocoll* [39]	6.9
STASN [40]	CNN	OULU-NPU	OULU-NPU	*Protocoll* [39]	1.9
FaceDs [41]	CNN	OULU-NPU	OULU-NPU	*Protocoll* [39]	1.5
STPM [42]	CNN	OULU-NPU	OULU-NPU	C_{sp}, RGB+Depth	1.0
NHF [28]	CNN	CASIA-SURF	CASIA-SURF	C_{sp}, RGB+Depth	4.7
Single-scale SEF [28]	CNN	CASIA-SURF	CASIA-SURF	C_{sp}, RGB+Depth	2.4
Multi-scale SEF [28]	CNN	CASIA-SURF	CASIA-SURF	C_{sp}, RGB+Depth	0.8
PSMM-Net [43]	CNN	CASIA-SURF	CASIA-SURF	C_{sp}, RGB+Depth	0.4
PSMM-Net(CeFA) [43]	CNN	CASIA-SURF	CASIA-SURF	C_{sp}, RGB+Depth	0.2
Proposed image life trail	Two-class SVM linear	CASIA$\alpha = 2.7, \beta = 0.25$	CASIA$\alpha = 2.7, \beta = 0.25$	C_{sp}	**0.3106**
Proposed image life trail	Two-class SVM linear	Oulu-NPU$\alpha = 3.4, \beta = 0.1$	Oulu-NPU$\alpha = 3.4, \beta = 0.1$	C_{sp}	**1.1928**
Proposed image life trail	Two-class SVM linear	CASIA-SURF$\alpha = 1.7, \beta = 0.15$	CASIA-SURF$\alpha = 1.7, \beta = 0.15$	C_{sp}	2.2462

tor, which may contain crucial information regarding a certain phenomenon such as, BLUR-diversity (due to a PINHOLE LENS-effect [8]) or self-shadow prominence (in this paper).

- The features and statistics are content and subject agnostic.

One has to note that these contiguous random walks tend to diverge considerably beyond a certain number of steps. Rather, when viewed conversely, given a walk length of d units, one can construct a graph from the destination pixel to one of its myriad origins $d − foot − steps − away$ (or walk units away). This has been illustrated in Fig. 17: CONTIGUOUS RANDOM WALK, where the final destination is flagged by a RED-CIRCLE and length of the walk has been chosen as $d = 3$ units. The original source pixel from which the 3-unit distance walk had originated and the distinct paths traversed are shown in Fig. 17, where the final mile entry is from the bottom. The entry can similarly be from the left or right or above. Thus, the number of distinct

paths is, $N_{paths} = 9 \times 4 = 36$ for $d = 3$-walk-length-units. Some exemplar generated walk patterns are shown in Fig. 18. Since the random scan can be fed with any target image or image-like-statistic, for this application which is concerned with life trails and self-shadows, it is fed with the following first, first difference ratio image,

$$G_1(x, y) = |X_1(x, y) − X_0(x, y)| / X_0(x, y) \qquad (38)$$

where

$$X_1(x, y) = 2X_0(x, y)[1 − X_0(x, y)] \qquad (39)$$

with $X_0(x, y) = IM_{norm}(x, y) \in (0, 1)$ representing the original normalized real and natural intensity image. An enhanced version of this is created by raising $G_1(x, y)$ to the power $\alpha = 2.5$ (fixed) to ensure that the self-shadows in the natural face image are brought out much more clearly. This enhanced version is given by,

$$G_{1E}(x, y) = G(x, y)^{\alpha} \qquad (40)$$

Table 5 State of the art methods which assume a client/subject independent frame and the corresponding error rates

Method	Classifier	Train data	Test data	Threshold	EER
LBP [36]	GMMUBM	CASIA	CASIA	C_{gb}	21.69
	Two-class SVM	CASIA	CASIA	C_{gb}	15.42
LBP-TOP [36]	GMM UBM	CASIA	CASIA	C_{gb}	12.65
	Two-class SVM	CASIA	CASIA	C_{gb}	8.53
Motion [36]	GMMUBM	CASIA	CASIA	C_{gb}	12.52
	Two-class SVM	CASIA	CASIA	C_{gb}	11.53
IMQ [11]	One-class SVM	CASIA	CASIA	-	23.07
BSIF [11]	One-class SVM	CASIA	CASIA	-	36.06
LPQ [11]	One-class SVM	CASIA	CASIA	-	35.19
LBP [11]	One-clas SVM	CASIA	CASIA	-	25.06
Random scan [4]	One-class SVM	CASIA	CASIA	$N_S = 1$	3.5122
Random scan [4]	One-class SVM	CASIA	CASIA	$N_S = 20$	1.8920

Fig. 16 Examples (both natural and spoof-print versions) from the original CASIA-SURF dataset [32]

Fig. 17 CONTIGUOUS RANDOM WALK: Destination pixel marked in RED and last-mile entry is from the bottom pixel (i.e. pixel located below the final destination pixel)
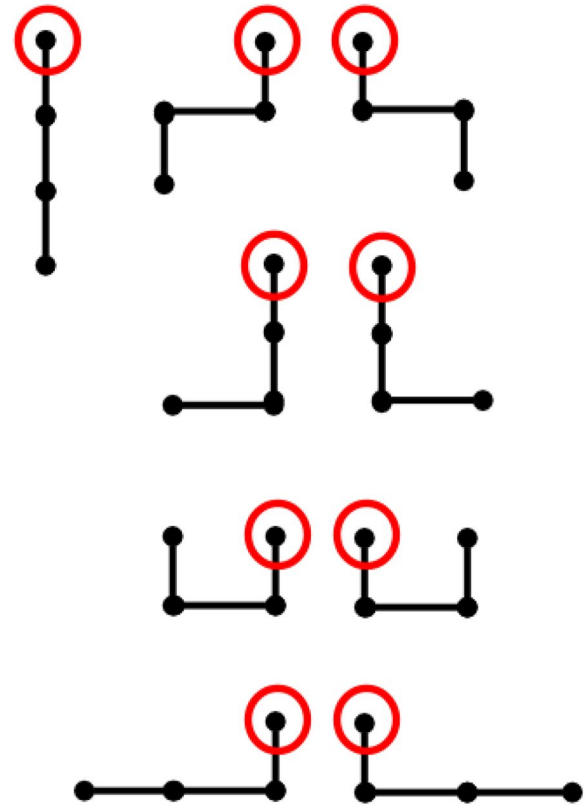
This enhanced first first difference ratio G_{1E} is then fed to the random scan algorithm. Let the scanned self-shadow intensity vector of length L-walk units for a particular instance $k \in 1, 2, .. N_S$ be:

$$\bar{S}(G_{1E}, k) = [s_{k,1}, s_{k,2}, ..., s_{k,L}] \tag{41}$$

where N_S is the size of the ensemble of scans or number of differently scanned vectors from the same self-shadow image/statistic.

9.1 Need for a secondary Fourier descriptor

Note that the motive in this segment is to detect the presence of the self-shadow, irrespective of its size, shape and positions. This size, shape, self-shadow prominence, and location is a function of an interplay between the light source orientation relative to the face-surface topography which is being photographed by a frontal camera. Most poses are assumed to be full-frontal, but mild scale changes and pose variations are allowed and expected. Thus, under natural lighting conditions there are clear bright and dark zones, the only issue being that the fraction of the zone that is dark and constitutes the self-shadow remains uncertain. When an image of a planar print is analyzed, in relation to the diffused lighting analogy of the TWIN-image of Fig. 9, the difference between the two cases is in the presence

of darker zones for natural images versus suppressed umbral-penumbral zones for planar print images. Thus, a spectral analysis eventually leading to a computation
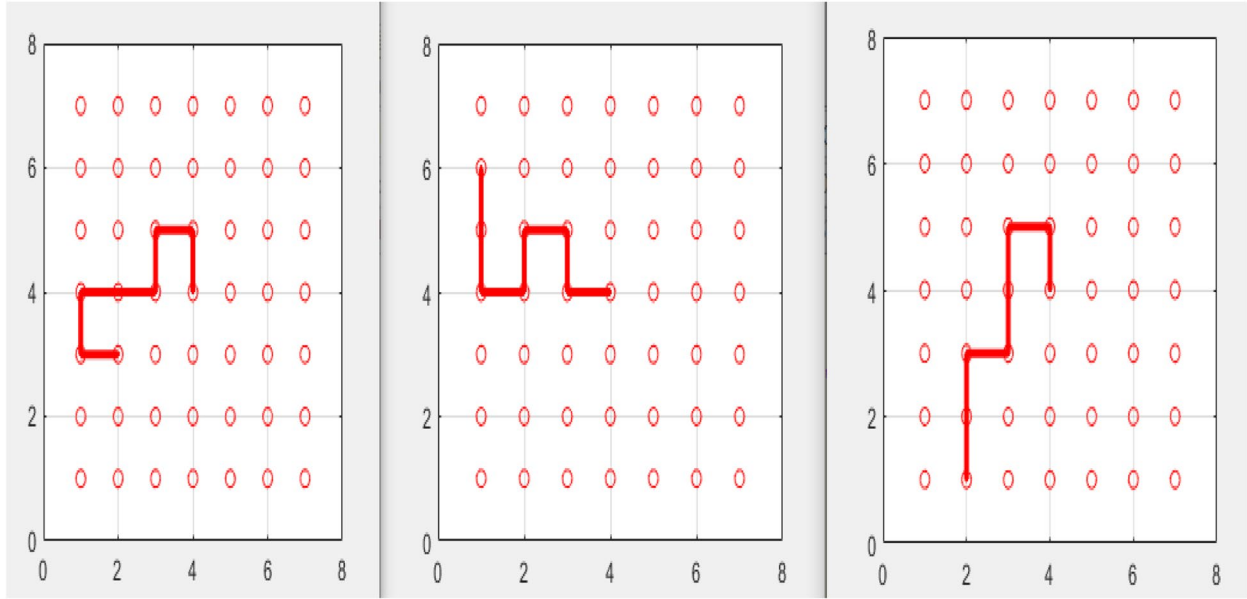
Fig. 18 Some exemplar random walk patterns leading to the central pixel. Walk length is d = 7 units and the size of ensemble (or extent of auto-population) was 15-scans

of parameters such as spread of power over the discrete frequency space should be able to segregate natural spectra derived from self-shadow images from print spectra.

Claim 4 *In this section, it is claimed that the bandwidth of the first, first difference ratio statistic G_1, carries enough discriminatory information to distinguish natural face images from print spoof version via a self-shadow spectral analysis. Furthermore, it is also claimed that a contiguous random walk starting from the center of the image preserves the correlation statistics and subsequently some of the 2D-spectral parameters in the self-shadow image-statistic. Thus, by executing the random walk on the self-shadow image statistic G_1 and then analyzing its magnitude spectrum, one can construct a robust Fourier descriptor for the natural face image.*

The 1D-discrete Fourier transform (DFT) of the scanned vector $\bar{S}(G_{1E}, k)$, corresponding to instance or walk realization k is given by,

$$FS_{G_{1E},k}(r) = \sum_{n=0}^{L-1} s_{k,n} W_L^{nr} \qquad (42)$$

where W_L is the Twiddle factor, given by $W_L = e^{-(j*2*pi)/L}$, where $j = \sqrt{-1}$. The magnitude spectrum is given by,

$$MFS_{G_{1E},k}(r) = \sqrt{FS_{G_{1E},k}(r) \times FS_{G_{1E},k}^*(r)} \qquad (43)$$

Assuming L to be an even integer, the following BAND-related, spectral cumulative statistics are computed:

$$
\begin{aligned}
A(1) &= \frac{4}{L} \sum_{r=0}^{L/4-1} MFS_{G_{1E},k}(r) \\
A(2) &= \frac{4}{L} \sum_{r=L/4}^{L/2-1} MFS_{G_{1E},k}(r) \\
A(3) &= \frac{4}{L} \sum_{r=L/2}^{3L/4-1} MFS_{G_{1E},k}(r) \\
A(4) &= \frac{4}{L} \sum_{r=3L/4}^{L-1} MFS_{G_{1E},k}(r)
\end{aligned}
\qquad (44)
$$

To ensure robustness to self-shadow variations, shape and size mainly, another set of spectral statistics are derived from the above set:

$$
\begin{aligned}
B(1) &= A(1) + A(2) \\
B(2) &= A(2) + A(3) \\
B(3) &= A(3) + A(4) \\
B(4) &= A(2) + A(3) + A(4)
\end{aligned}
\qquad (45)
$$

The final feature or descriptor which can now be deployed in the cross-validation experiment is now:

$$FD_{CRV,k} = [A(1), A(2), A(3), A(4), B(1), B(2), B(3), B(4)] \qquad (46)$$
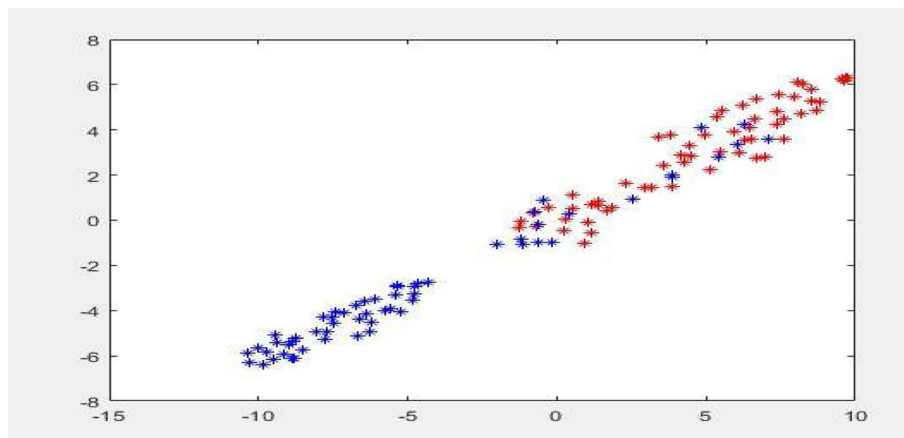
for scan-instance k.

Fig. 19 Clusters from 60-natural spectral descriptors and 60 print spectral descriptors. N S = 1 (which means only one random scan was generated per image statistic) and scan parameters were: Image-statistic or patch size W × W , W = 21 and walk length (complete) covering full image-statistic, L = W²

9.2 Calibration with CASIA dataset

Sixty images across 10 subjects from the CASIA set (real and spoof) were used for testing the spectral frame and also for calibrating the parameters. All images were resized to 100×100, and then the self-shadow image-statistic was computed. The exponent α was fixed as 2.5, and what was fed to the random walk process was the enhanced self-shadow image-statistic:

$$G_{1E}(\alpha) = G_1^\alpha \qquad (47)$$

The final feature descriptors were produced for each image. Fig. 19, shows the t-SNE plot of the real descriptors versus the print descriptors. A good separation with a small overlap can be seen in Fig. 19.

9.3 Cross-validation with OULU dataset

To check whether the model developed using subjects and images from the CASIA set can be applied to other datasets, one can take one of two pathways:

- Single sided training: Characterize the natural space class alone [9, 11], via the self-shadow feature and its secondary statistics using the random scans and Fourier descriptor. The training model is confined to the CASIA dataset and is built across subjects (this subject-agnostic training is helped in part by the contiguous random scan which does not require feature registration [4]). Once the 1-class SVM [11] model is built, this is then tested on another dataset, OULU [27]. The complete image set, natural and print versions from OULU are used for testing. No part of OULU is used in model building.
- Two-class training: Here, the training model is built with natural and spoof samples from CASIA.

Table 6 Error rates when features are trained on CASIA and tested on OULU exclusively

Method	Train dataset	Test dataset	EER (%)
Proposed features + 2 class SVM	CASIA	OULU	2.3475
Proposed features + 1 class SVM	CASIA	OULU	5.8676

Testing is done in the same way as described earlier over OULU.

The parameters for the 1-class and 2-class model building were as follows:

- *1-class model*: 14 subjects from a reduced CASIA dataset with 15 variations per subject for the natural face class alone was used to form the 1-SVM model with the final random scan-induced Fourier descriptors, discussed in the early of this section. All images were resized to 100×100, and the ratio statistic G_1 was computed first and then raised to an exponent $\alpha = 2.5$ (fixed). These enhanced self-shadow statistics G_1^α were then resized to 21×21 and subjected to a random scan followed by a Fourier analysis to generate final descriptors.
- *2-class model*: The only difference here is that 14 subjects with 15 variations across subjects from BOTH CASIA classes were used to form the 2-class SVM model. All other parameters remained the same.

Table 6 shows the error rates on the OULU set when the proposed random scan based with Fourier descriptor, self-shadow model (learnt on the CASIA dataset) was applied.

Table 7 EEE for optimal values of $\alpha*$ and *beta** for

Database	$\alpha*$	$\beta*$	EER (%)
CASIA	2.7	0.25	0.3106
Oulu	3.4	0.10	1.1928
CASIA-SURF	1.7	0.15	2.2462

The error rates for both one- and two- class SVMs were on the lower side (5.86% and 2.34% respectively) and comparable with the results obtained when customized calibration was done for the OULU-dataset (1.19%). This modified approach de-links the data-set training from the data-set testing and makes it more general (Table 7).

10 Summary and conclusions

In this paper, a novel contrast reductionist life trail based image sequence is generated using a non-linear logistic map, in such a way that successive images down the pipeline tend to have a progressively lower contrast when compared with previous iterations. Eventually, the sequence converges to a zero contrast image. A simple statistical model was used to show not just the proof of convergence but also arrive at fact that the first, first difference ratio statistic from the life trail carried sufficient and maximum information pertaining to self-shadows. This corroborated with the observations from the TWIN-image life trail analysis. The model also provided an insight into the selection of the optimal parameter $\alpha*$ based on an intersection between two constraints: (i) absolute self-shadow entropy from the natural face ratio-statistic after exponentiation and (ii) class separation parameter $-\Delta H(\alpha)$, leading to the crystallization of the operating point $\alpha*$ if the dynamic range parameter $\hat{(a)}_F = \hat{a}$ can be extracted via measuremen (Fig. 19).

For each dataset which was being tested, a small fraction of samples (both classes) were set aside for calibration which was done in two phases, and this was done in a subject agnostic fashion: (i) estimation of $\alpha*$ based on measurements and the two constraints and (ii) varying the patch fraction β to trap the localized entropy score related to the self-shadow statistic and checking the separation between the real and spoof conditional distributions. The β which corresponding to the highest separation value was chosen as the optimal $\beta*$.

When tested on three datasets, error rates for the proposed algorithm when applied to CASIA (the calibration database) and OULU-NPU and CASIA-SURF were found to be 0.3106% ($\alpha* = 2.7, \beta* = 0.25$), 1.1928% ($\alpha* = 3.4, \beta* = 0.1$), and 2.2462% ($\alpha* = 1.7, \beta* = 0.15$) respectively for planar-print-type spoofing operations.

To impart a certain degree of flexibility in the solution and avoid repeated calibration and tuning each time the acquisition and illumination environment is changed, a model was built on the basic enhanced self-shadow statistic using random scans, to make the information gathering subject agnostic. The basic idea was to focus on detecting only the presence of self-shadows and not in profiling the shape, position, and prominence of this self-shadow present. The moment the focus shifted from profiling to detection, this called for a Fourier analysis, particularly because self-shadow statistics from natural images tend to have dominant higher frequencies and exhibit higher bandwidths as compared to their print counterparts. Based on this random scan-induced Fourier descriptor, the proposed model which was trained on the CASIA set alone, was found to very effective when cross-ported to OULU.

The proposed algorithm and pipeline has other distinct advantages:

- Since the main computation involves a swarm of parallel pixel-wise intensity manipulations using the logistic map, the model building is very simple and fast. Note that interestingly the computation is so simple that not even a simple image filtering operation is done. In a way, this trail building process demands a certain purity in the acquired image. While resizing introduces some quantum of interpolation noise, self-shadow profiles are not compromised. Thus, owing to its simplicity, it can be used a quick-check in most counter-spoofing applications.
- A high accuracy was obtained with the proposed frame, both with calibration and customization and also while cross-porting (with a random scan inclusion and Fourier-descriptor subject-agnostic twist) to other datasets and environments such as OULU-NPU.

We however note that while the proposed solutions (including cross-validation) are precise enough to detect print-planar spoofing, it may not be effective against digital planar image presentation cases based on tablets and laptops. This is so because the back-lighting tends to enhance the self-shadow profiles present even in digitally spoofed segments.

Appendix A: Proof of convergence of the print-image-mapped γ-sequence

Problem: Starting off with Y_0 *UNIFORM*$[0, (1/a)]$; $a > 1$ (uniformly distributed but reduced dynamic range) and applying the logistic map several times prove that:

$$Y_n \longrightarrow 0.5 \text{ with probability '1' for } n \gg 1 \tag{48}$$

Proof

The iterative function map with respect to the Y-sequence is,

$$Y_n = 2Y_{n-1}(1 - Y_{n-1}); n \geq 1 \tag{49}$$

Thus, subtracting Y_n from 0.5, we get,

$$\begin{aligned}
\frac{1}{2} - Y_n &= \left(\frac{1}{2} - 2Y_{n-1} + 2Y_{n-1}^2 \right) \\
&= 2 \times \left(\frac{1}{2} - Y_{n-1} \right)^2 \\
&= 2 \times 2^2 \times \left(\frac{1}{2} - Y_{n-2} \right)^4
\end{aligned} \tag{50}$$

Multiplying both sides by factor of 2, the above equation can be re-written as,

$$\begin{aligned}
1 - 2Y_n &= 2^4 \times \left(\frac{1}{2} - Y_{n-2} \right)^4 \\
&= (1 - 2Y_{n-2})^4 \\
&= (1 - 2Y_0)^2 n
\end{aligned}$$

Let,

$$Z = (1 - 2Y_0)^2 \tag{51}$$

It can be shown that the positive power of the random variable Z, i.e.,

$$Q_n = Z^n; n \gg 1 \tag{52}$$

will approach a deterministic zero with probability '1' as n becomes very large, i.e.,

$$f_{Q_n(z)} \longrightarrow \delta(z) \text{ for } n \gg 1 \tag{53}$$

where, $\delta(.)$ is the DIRAC-DELTA function. This leads to the result that for large n,

$$Prob(Z^n \longrightarrow 0) \longrightarrow 1; n \gg 1 \tag{54}$$

This implies that based on Eq. (51) and then Eqs. (51), (52) and (53),

$$Prob(1 - 2Y_n \longrightarrow 0) \longrightarrow 1; n \gg 1 \tag{55}$$

Since, $Y_n \in [0, 0.5]; n \geq 1$, it follows that,

$$Y_n \longrightarrow 0.5 \text{ with probability '1' for } n \gg 1 \tag{56}$$

Thus, the proof.

Appendix B: Convergence rates of real and print life trail sequences

Problem: It is to be shown that the print-abstraction related sequence Y_n, converges faster as compared the real-image-abstraction sequence X_n. It suffices to show that the dynamics associated with the error sequence H_n is greater as compared to the original error sequence G_n. This means the change and drift to a zero contrast image is faster for a print version as compared to a natural one.

Proof

To monitor and track the convergence rates of the two trails, the normalized first order difference (or error) metric is defined as,

$$G_n = \frac{X_n - X_{n-1}}{X_{n-1}} = 1 - 2X_{n-1}; n \geq 1$$

$$H_n = \frac{Y_n - Y_{n-1}}{Y_{n-1}} = 1 - 2Y_{n-1}; n \geq 1$$

Furthermore, it can be shown that,

$$\begin{aligned}
G_n - G_{n-1} &= 2[X_{n-2} - X_{n-1}] \\
&= 2 \left[\frac{X_{n-2} - X_{n-1}}{X_{n-2}} \right] X_{n-2}
\end{aligned} \tag{57}$$

$$= -2G_{n-1}X_{n-2}; n \geq 2 \tag{58}$$

Can show that,

$$G_n = G_{n-1}(1 - 2X_{n-2}) = G_{n-1}^2; n \geq 2 \tag{59}$$

It follows that,

$$G_n = (G_0)^{2n}; n \geq 2 \tag{60}$$

with $G_1 = 1 - 2X_0$; Similarly,

$$H_n = (H_0)^{2n} = (2Y_0 - 1)^{2n}; n \geq 2 \tag{61}$$

where, $H_1 = 1 - 2Y_0$; Now, let the expected value, $E[H_n] = \mu_H$. Given $E[H_n] = \mu_H$ and

$$E[H_n] = a \int_{u=0}^{u=1/a} (2u - 1)^{2n} du \tag{62}$$

$$= \frac{a}{2(2n+1)} \left[1 + \left(\frac{2}{a} - 1 \right)^{2n+1} \right] \tag{63}$$

Taking the limit as $a \longrightarrow 1^+$, since a is a number larger than '1',

$$E[H_n] \longrightarrow E[G_n] \tag{64}$$

$$= \frac{1}{2(2n+1)}; \text{for } n \geq 2 \tag{65}$$

It can be shown and verified analytically/numerically that for three different values of the parameter $a \in 1.1, 1.25, 1.5$, the function,

$$E[H_n] > E[G_n] \text{for } n > 1 \tag{66}$$

This result is in fact valid for all $a > 1$. This means that the error sequence associated with the print version has a greater magnitude as compared to the original error sequence. Two observations can be drawn from this:

- The original (natural image related) sequence X_n decays much slower as compared to the print (i.e. spoof image related) sequence Y_n. This happens because $E[H_n] > E[G_n]$ for all $n > 1$ and for $a > 1$.
- The separation between two errors $E[H_n] - E[G_n]$, is maximum for $n = 1$, which implies that the class separation is maximum for the first, first order difference.

Thus, the proof.

Abbreviations
SVM   Support Vector Machines
CNN   Convolution Neural Netweok

Authors' contributions
Both authors read and approved the final manuscript.

Availability of data and materials
CASIA [18], MSU-MFSD database [23].

Declarations

Ethics approval and consent to participate
Not applicable

Competing interests
The authors declare that they have no competing interests.

References
1. L.A.M. Pereira, A. Pinto, F.A. Andaló, A.M. Ferreira, B. Lavi, A. Soriano-Vargas, M.V.M. Cirne, A. Rocha, The rise of data-driven models in presentation attack detection, in *Deep Biometrics. Unsupervised and Semi-Supervised Learning*. ed. by R. Jiang, C.-T. Li, D. Crookes, W. Meng, C. Rosenberger (Springer, Cham, 2020), pp.289–311. https://doi.org/10.1007/978-3-030-32583-1_13
2. K. Patel, H. Han, A.K. Jain, Secure face unlock: spoof detection on smartphones. IEEE Trans. Inf. Forensic Secur. **11**(10), 2268–2283 (2016)
3. S.M. Nesli Erdogmus, *Spoofing in 2d face recognition with 3d masks and antispoofing with kinect* (IEEE BATS, USA, 2013)
4. B.R. Katika, K. Karthik, Face anti-spoofing by identity masking using random walk patterns and outlier detection. Patt. Anal. Appl. 1–20 (2020)
5. K.. Karthik, B.R. Katika, Identity independent face anti-spoofing based on random scan patterns, in *2019 8th PREMI International Conference on Pattern Recognition and Machine Intelligence (PREMI)*. (Springer, India, 2019)
6. S. Kim, S. Yu, K. Kim, Y. Ban, S. Lee, in *Biometrics (ICB), 2013 International Conference On*. Face liveness detection using variable focusing. (IEEE, 2013). pp. 1–6
7. Y. Kim, S.Y.J.L. Jaekenun, in *Conference on Optical Society OF America*, vol. 26. Masked fake face detection using radiance measurments. (2009) pp. 1054–1060
8. K. Karthik, B.R. Katika, in *Industrial and Information Systems (ICIIS), 2017 IEEE International Conference On*. Face anti-spoofing based on sharpness profiles (IEEE, 2017) pp. 1–6
9. K. Karthik, B.R. Katika, in *Communication Systems, Computing and IT Applications (CSCITA), 2017 2nd International Conference On*. Image quality assessment based outlier detection for face anti-spoofing. (IEEE, 2017), pp. 72–77
10. B.R. Katika, K. Karthik, Face anti-spoofing based on specular feature projections, in *Proceedings of 3rd International Conference on Computer Vision and Image Processing*. ed. by M. Chaudhuri, M. Nakagawa, P. Khanna, S. Kumar (Springer, Singapore, 2020), pp.145–155
11. S.R. Arashloo, J. Kittler, W. Christmas, An anomaly detection approach to face spoofing detection: a new formulation and evaluation protocol. IEEE Access **5**, 13868–13882 (2017)
12. T. Edmunds, A. Caplier, Face spoofing detection based on colour distortions. IET Biom. **7**(1), 27–38 (2017)
13. X. Gao, T.-T. Ng, B. Qiu, S.-F. Chang, in *2010 IEEE International Conference on Multimedia and Expo*. Single-view recaptured image detection based on physics-based features. (2010). pp. 1469–1474. https://doi.org/10.1109/ICME.2010.5583280
14. E.J. Candès, X. Li, Y. Ma, J. Wright, Robust principal component analysis? CoRR abs/0912.3599 (2009)
15. S. Kim, Y. Ban, S. Lee, Face liveness detection using defocus. Sensors. **15**(1), 1537–1563 (2015). https://doi.org/10.3390/s150101537
16. X. Zhang, X. Hu, M. Ma, C. Chen, S. Peng, in *2016 23rd International Conference on Pattern Recognition (ICPR)*. Face spoofing detection based on 3d lighting environment analysis of image pair. (2016). pp. 2995–3000. https://doi.org/10.1109/ICPR.2016.7900093
17. I. Chingovska, A.R. Dos Anjos, S. Marcel, Biometrics evaluation under spoofing attacks. IEEE Trans. Inf. Forensics Secur. **9**(12), 2264–2276 (2014)
18. Z. Zhang, J. Yan, S. Liu, Z. Lei, D. Yi, S.Z. Li, in *Biometrics (ICB), 2012 5th IAPR International Conference On*. A face antispoofing database with diverse attacks. (IEEE, 2012). pp. 26–31
19. J. Määttä, A. Hadid, M. Pietikäinen, Face spoofing detection from single images using texture and local shape analysis. IET Biom. **1**(1), 3–10 (2012)
20. J. Yang, Z. Lei, D. Yi, S.Z. Li, Person-specific face antispoofing with subject domain adaptation. IEEE Trans. Inf. Forensics Secur. **10**(4), 797–809 (2015)
21. J.M. Saragih, S. Lucey, J.F. Cohn, Deformable model fitting by regularized landmark mean-shift. Int. J. Comput. Vis. **91**(2), 200–215 (2011)
22. T. Wang, J. Yang, Z. Lei, S. Liao, S.Z. Li, in *2013 International Conference on Biometrics (ICB)*. Face liveness detection using 3D structure recovered from a single camera. (IEEE, 2013). pp. 1–6
23. D. Wen, H. Han, A.K. Jain, Face spoof detection with image distortion analysis. IEEE Trans. Inf. Forensics Secur. **10**(4), 746–761 (2015)
24. J. Galbally, S. Marcel, in *Pattern Recognition (ICPR), 2014 22nd International Conference On*. Face anti-spoofing based on general image quality assessment. (IEEE, 2014), pp. 1173–1178

25. J. Galbally, S. Marcel, J. Fierrez, Image quality assessment for fake biometric detection: application to iris, fingerprint, and face recognition. IEEE Trans. Image Process. **23**(2), 710–724 (2014)
26. E.W. Weisstein, Logistic map
27. Z. Boulkenafet, J. Komulainen, L. Li, X. Feng, A. Hadid, in *2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)*. Oulu-npu: a mobile face presentation attack database with real-world variations. (IEEE, 2017), pp. 612–618
28. S. Zhang, X. Wang, A. Liu, C. Zhao, J. Wan, S. Escalera, H. Shi, Z. Wang, S.Z. Li, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. A dataset and benchmark for large-scale multi-modal face anti-spoofing. (2019), pp. 919–928
29. Z. Zhang, D. Yi, Z. Lei, S.Z. Li, in *Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference On*. Face liveness detection by learning multispectral reflectance distributions. (IEEE, 2011), pp. 436–441
30. C. Bell, Bright sunlight (2015)
31. A. Papoulis, *Probability, random variables, and stochastic processes*. (McGraw-Hill, 1991)
32. T.M. Cover, J.A. Thomas, *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)* (Wiley-Interscience, USA, 2006)
33. L. Van der Maaten, G. Hinton, Visualizing data using t-sne. J. Mach. Learn. Res. **9**(11), (2008)
34. Kullback-leibler divergence
35. J. Galbally, S. Marcel, J. Fierrez, Biometric antispoofing methods: a survey in face recognition. IEEE Access. **2**, 1530–1552 (2014)
36. I. Chingovska, A.R. Dos Anjos, On the use of client identity information for face antispoofing. IEEE Trans. Inf. Forensics Secur. **10**(4), 787–796 (2015)
37. Y. Sun, H. Xiong, S.M. Yiu, Understanding deep face anti-spoofing: from the perspective of data. Vis. Comput. **37**, 1015–1028 (2021)
38. Z. Boulkenafet, J. Komulainen, Z. Akhtar, A. Benlamoudi, D. Samai, S.E. Bekhouche, A. Ouafi, F. Dornaika, A. Taleb-Ahmed, L. Qin, et al. in *2017 IEEE International Joint Conference on Biometrics (IJCB)*. A competition on generalized software-based face presentation attack detection in mobile scenarios. (IEEE, 2017), pp. 688–696
39. Z. Boulkenafet, J. Komulainen, Z. Akhtar, A. Benlamoudi, D. Samai, S.E. Bekhouche, A. Ouafi, F. Dornaika, A. Taleb-Ahmed, L. Qin, F. Peng, L.B. Zhang, M. Long, S. Bhilare, V. Kanhangad, A. Costa-Pazo, E. Vazquez-Fernandez, D. Perez-Cabo, J.J. Moreira-Perez, D. Gonzalez-Jimenez, A. Mohammadi, S. Bhattacharjee, S. Marcel, S. Volkova, Y. Tang, N. Abe, L. Li, X. Feng, Z. Xia, X. Jiang, S. Liu, R. Shao, P.C. Yuen, W.R. Almeida, F. Andalo, R. Padilha, G. Bertocco, W. Dias, J. Wainer, R. Torres, A. Rocha, M.A. Angeloni, G. Folego, A. Godoy, A. Hadid, in *2017 IEEE International Joint Conference on Biometrics (IJCB)*. A competition on generalized software-based face presentation attack detection in mobile scenarios. (2017), pp. 688–696. https://doi.org/10.1109/BTAS.2017.8272758
40. X. Yang, W. Luo, L. Y. Bao, D. Gong, S. Zheng, Z. Li, W. Liu, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Face anti-spoofing: model matters, so does data. (2019), pp. 3507–3516
41. A. Jourabloo, Y. Liu, X. Liu, in *Proceedings of the European Conference on Computer Vision (ECCV)*. Face de-spoofing: anti-spoofing via noise modeling. (2018), pp. 290–306
42. Z. Wang, Z. Yu, C. Zhao, X. Zhu, Y. Qin, Q. Zhou, F. Zhou, Z. Lei, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Deep spatial gradient and temporal depth learning for face anti-spoofing. (2020), pp. 5042–5051
43. A. Liu, Z. Tan, J. Wan, S. Escalera, G. Guo, S.Z. Li, in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. Casia-surf: A benchmark for multi-modal cross-ethnicity face anti-spoofing. (2021) pp. 1179–1187
44. Y. Matias, A. Shamir, in *Conference on the Theory and Application of Cryptographic Techniques*. A video scrambling technique based on space filling curves. (Springer, 1987), pp. 398–417

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.