Image Life Trails Based On Contrast Reduction Models For Face Counter-Spoofing

Balaji Rao Katika (k.balaji@iitg.ac.in)
Institute of Technology Guwahati https://orcid.org/0000-0002-7315-142X
Kannan Karthik
Institute of Technology Guwahati

Research

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

Posted Date: December 29th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-1161059/v1

RESEARCH

Image Life Trails Based On Contrast Reduction Models For Face Counter-Spoofing

Balaji Rao Katika[1*] and Kannan Karthik[1]

*Correspondence:
k.balaji@iitg.ac.in
[1]Department of Electronics and Electrical Engineering, Indian Institute of Technology Guwahati, Guwahati-781039 Assam , India
Full list of author information is available at the end of the article

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. On a client specific mode, when the subjects and faces are registered, secondary statistics which capture the prominence of self-shadow information, indicate that planar print-images tend to have highly suppressed self-shadows when compared with natural face images. An elaborate tuning procedure, based on a reduced set of training images was developed to first identify the optimal parameter set and then adapt the feature-vectors so that the error-rates were minimized for a specific dataset. Overall mean error rate for the calibration-set (reduced CASIA dataset) was found to be 0.267% and the error rates for other datasets such OULU-NPU and CASIA-SURF were 0.17% and 0.73% respectively.

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 smartphones, 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. 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, but do not have a counter-spoofing module inside [1]. 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;

(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 [1]. This is mainly because the customized design of a prosthetic tailored to mimic a particular individual's face (who owns the smartphone), is an extremely difficult scientific exercise. This problem is exacerbated by the fact that to prepare a 3D mask [2](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 [3][4]. 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 [5],[6], [7]; (ii) The reprinting process to synthesize a planar print brings about a progressive degradation in contrast [8], clarity, specularity [9], quality [10] or color-naturalness [11].

One type of statistical lens for detecting planar spoofing is a specularity check [12]. If the paper printing of the target's face is done on a glossy type of paper, this results in a dominant specular component [12][9] 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 [12]. In Emmanuel et al. [13], 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. [9]. 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, beards etc., this remains an non-robust primary feature.

Planar geometric constraints also impact the manner in which other parameters are influenced,such as contrast [8] or sharpness (or its opposite blur)[7], [5], [6].

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 [8]. 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 [8]. 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[5][14] [7] in the captured image. In the work of Kim et al. [5], 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. [5] to detect planar spoofing.

In another blur variability detection procedure [14], 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 [7] 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 [15], 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 [15]), 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 [15].
- 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" [3].

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 [11]. Thus, an image of a planar printed face may not exhibit on one hand all the true colours 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 [16]. 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 colour 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 [17] 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. [18], 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) [18] [16] [19], 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 [20]. Facial landmarks such as eye centers, nose tips etc., 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 [21]. 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. [22], statistics based on a mixed bag of features ranging from texture, colour 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, print colour quality remain unknowns; (ii) For tablet and other digitized presentations, the nature and extent of re-sampling noise [18], 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 [3] 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[16][19] tend to outshine the subject independent ones [8][10].

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 [23, 24]. 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" [3].
- 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 [11],[8], blur/sharpness [6], [7], specularity [9] 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 [3]. 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 [16][19].

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 [3][4].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 [3] from the centre 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 [3]. 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 [4], 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 [19] [16], 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 [25]. 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.

The proposed self-shadow formulation is built on another proposition called a image contrast reductionist life trail, which is facilitated by a non-linear iterated function mapping [25], discussed in detail in Section. 2. Feature extraction and secondary statistics which sit on top of the enhanced self-shadow images are discussed with preliminary classification results in Section. 3. Finally, positioning and error rates of the proposed architecture in relation to the state of the art are presented in Section. 4.

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 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, 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 Fig. 1 and Fig. 2. 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. 1.

The images captured with this arrangement for three different source locations are shown in Figs. 2(a,b,c). In Fig2(a), the light source has been positioned top-left-front of the person's face and beside the camera unit (north-west direction); In Fig 2(b), the source is positioned towards the left of the person and partly in front (west position), while in Fig. 2(c), 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).

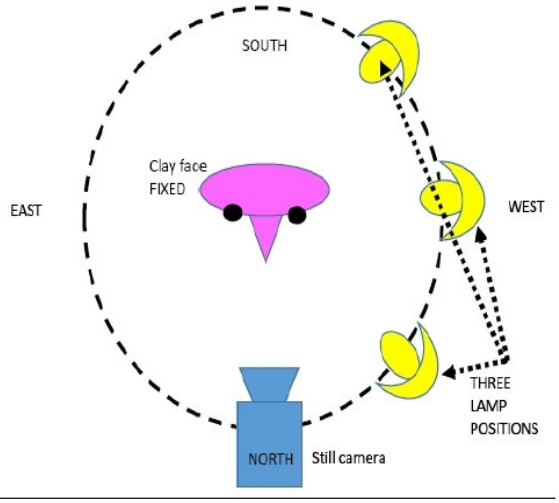

Figure 1 Experimental setup using a clay model and a fixed cell-phone camera for producing natural images with self-shadows.

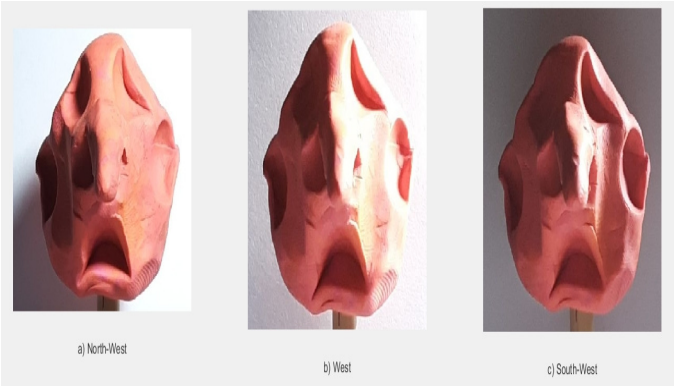

Figure 2 Images captured using the experimental setup (Fig. 1), for three different table lamp positions (north-west, west and south-west)

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 [25] 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 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:

$$\begin{aligned} SWARM(I_0) &= \{I_0(x, y), \\ &\quad s.t.(x, y) \in \{\{1, 2, ..., N\} \times 1, 2, ..., N\}\}\} \end{aligned}$$

This non-linear iterated function system is defined as [25],

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

with the initial value, $I_0 = I(n = 0) \in (0, 1)$ and I_n is the value at the $n^{th}, n \geq 0$ iteration with $I_n \in (0, 1)$. 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 images (Fig. 3). 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 Analytical proof for fixed point convergence

A simple proof has been provided for this convergence to this fixed point 0.5. Let $I_0(x, y)$ be the normalized intensity at spatial location (x, y). Note $0 \leq I_0(x, y) \leq 1$. The iterative mapping is, $I_{n+1} = 2I_n(1 - I_n)$. If $I_0 = 0$ then $I_1 = 0$ subsequently $I_n = 0 \forall_n \geq 1$. To add $I_0 \neq= 1$, This implies the drift will take place only if $0 < I_0 < 1$. The proof of convergence as follows:

$$I_{n+1} = 2I_n - I_n^2$$

Define the cost-function (or the check-point) which has to be minimized in iteration $n + 1$ as,

$$\begin{aligned}
C_{n+1} &= \frac{(I_{n+1} - 0.5)^2}{\frac{1}{4}} = (2I_{n+1} - 1)^2 \\
C_{n+1} &= (4I_n(1 - I_n) - 1)^2 = (1 + 4I_n^2 - 4I_n)^2 \\
&= 4(I_n^2 - I_n + \frac{1}{4})^2 \\
&= 4(I_n - 0.5)^4 \\
&= 4 \times \left[\frac{(I_n - 0.5)}{0.5}\right]^4 \times 0.5^4 \\
&= \left(\frac{1}{4}\right) \times C_n^2 \\
&= \left(\frac{1}{4}\right) \times \left[\frac{1}{4} \times C_{n-2}^2\right]^2 \\
&= \left(\frac{1}{4}\right)^{(1+2)} C_{n-2}^{2^2} \\
&= \left(\frac{1}{4}\right)^{(1+2)} \left[\frac{1}{4} \times C_{n-3}^2\right]^{2^2} \\
&= \left(\frac{1}{4}\right)^{(1+2+2^2)} C_{n-3}^{2^3} \\
&= \left(\frac{1}{4}\right)^{(1+2+2^2+\ldots+2^n)} \times C_0^{2^n}
\end{aligned}$$

with the initial cost, $C_{(n=0)} = C_0 = 4(I_0 - 0.5)^2 \in (0, 1)$, provided $0 < I_0 < 1$. This implies that, for n-moderate, this C_{n+1} has to converge and shrink to zero. When, $C_{n+1} \to 0$ as n becomes moderate depending on the initial condition, $I_{n+1} \to 0.5$ with probability '1'.

While waiting for a precise convergence of all points is not necessary, in an 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

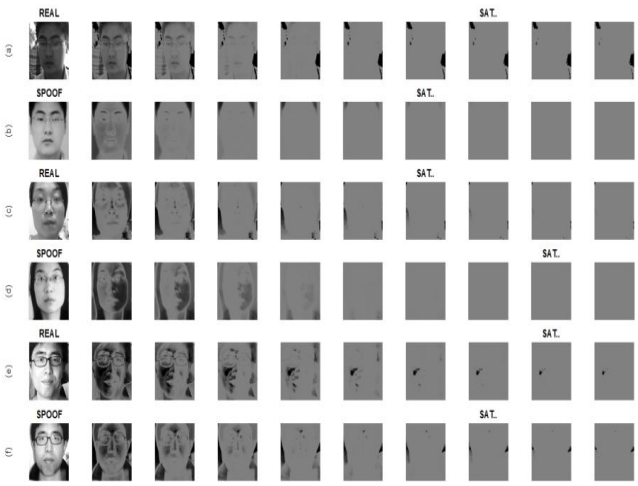

Figure 3 Contrast reductionist life trails for real and spoof image samples using the logistic map.

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,

$$\frac{|I_n - I_{n-1}|}{I_n} < \epsilon \tag{2}$$

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 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 Eqn. 2 individually. This swarm convergence trend has been tapped using a saturation curve based on a function $P(n)$ (Fig. 4), where n is the iteration number. Typical saturation curves for natural and spoof images are shown in Fig.4. Fig.3 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

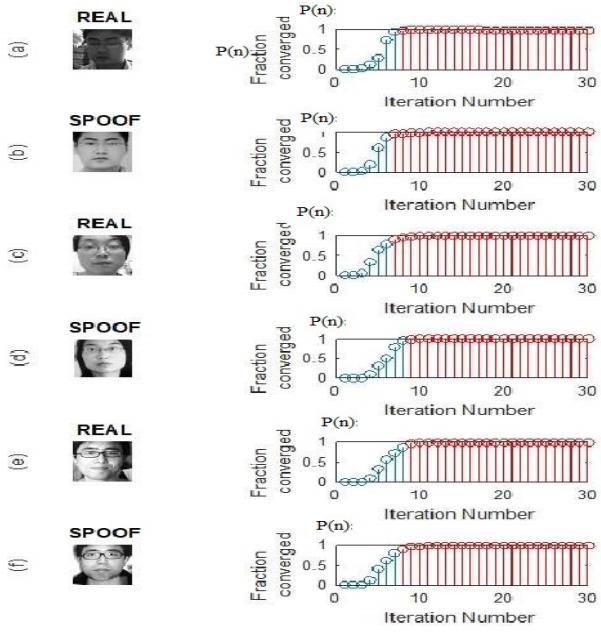

Figure 4 Saturation curves which bring out the trends linked to the rate at which the initial image samples (either natural or spoof) converge to a zero-contrast image in the life-trail.

planar print versions tend to usually be of a lower quality, typically lower contrast [8] and limited color [16] 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.3.

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. 3(e,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 overall 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. The authors claim that 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.

2.3 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 [25] (one iteration only), a contrast reduced image is obtained, $I_1(x, y)$ such that,

$$I_1(x, y) = 2I_0(x, y)[1 - I_0(x, y)] \tag{3}$$

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

$$R_1(x, y) = |I_1(x, y) - I_0(x, y)| = |I_0(x, y) - 2\left[I_0(x, y)\right]^2| \tag{4}$$

$$R_2(x, y) = \left[\frac{|I_1(x, y) - I_0(x, y)|}{I_0(x, y)}\right] = |1 - 2I_0(x, y)| \tag{5}$$

$$R_3(x, y) = [R_2(x, y)]^{\alpha} \tag{6}$$

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.

2.4 Selecting the exponential parameter α

The TWIN-image [26] in Fig. 5, 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 same pose is best for some intermediate value of α. The right-twin image represents a spoofed

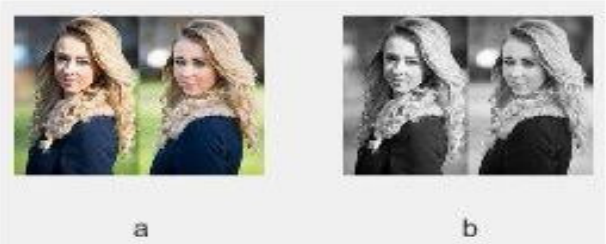

Figure 5 TWIN-image where one version is taken under normal outdoor lighting and the other one under diffused sunlight.

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. The diversity scores are computed for the left and right images independently for different values of α, shown in rows 2 and 3 . respectively.

The raw mean diversity scores (discussed later using Eqn. 7), were computed for all the cases and shown below the respective sub-figures in Fig. 6, with $\alpha = (1, 2, 2.5, 4, 6, 8)$. The separation is maximum around $\alpha = 4$ and drops beyond that.

$$S_{DIFF}(\alpha) = |STAT_{RAW}(TOP\ IMG)(\alpha) - STAT_{RAW}(BOTTOM\ IMG) \tag{7}$$

Where, the diversity computation (un-normalized, $STAT_{RAW}$) has been discussed later in Sub-section. 3.2 Step-5 (corresponding to the TYPE-1 setting).

Table 1 Differential scores for different values of α (top and bottom ratio images generated from the TWIN-figure, Fig. 6)

α	$STAT_{RAW}(TOP)$	$STAT_{RAW}(BOTTOM)$	S_{DIFF}(Eqn. 7)
1	0.188	0.185	0.03
2	0.168	0.16	0.08
2.5	0.155	0.146	0.09
4	**0.123**	**0.12**	**0.11**
6	0.094	0.084	0.10
8	0.075	0.066	0.09

One of the reasons for stopping with the first iteration is because the swarm convergence rate of this non-linear logistic map is quite fast as can be seen from the analytical proof (Sub-section. 2.2). This turns out to be less than 10 iterations for natural images and subsequent ratio images have been observed to contain less prominent and less enhanced self-shadows. Once this ratio image is generated, this is subjected 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 we 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. Thus $\alpha = 4$ was a compromise based on observations with respect to the TWIN-image (see Table. 1 in relation to Fig. 6).

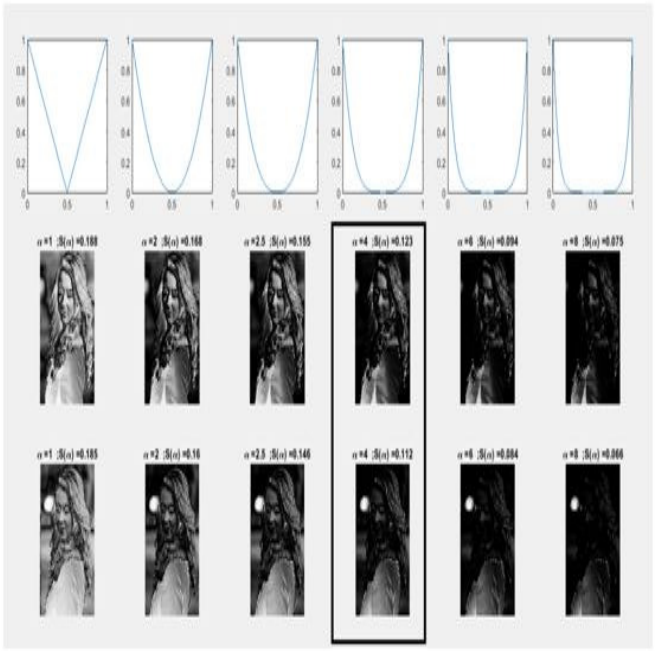

Figure 6 Impact of changes in the exponential parameter α on both the versions from the TWIN-image set [26]. As the exponent increases, the self-shadows become much more discernible for the version where the lighting is normal. Beyond a certain point the ratio images corresponding to both the normal version and the diffused version become dark.

3 Final feature extraction procedure and Client Specific Classification

3.1 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 $N \times \lfloor \rho N \rfloor$ pixels, with $N = 250$ and the ratio, $\rho > 0$. 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:

$$E_\alpha(x, y) = R_3(x, y) = \left[\frac{|I_1(x, y) - I_0(x, y)|}{I_0(x, y)} \right]^\alpha \tag{8}$$

This self shadow enhanced image with parameter $\alpha > 1$, was placed in a rectangular grid and intensity standard deviations were 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,

$$S = \begin{pmatrix} \sigma_{1,1} & \sigma_{1,2} & \dots & \sigma_{1,n} \\ \sigma_{2,1} & \sigma_{2,2} & \dots & \sigma_{2,n} \\ \dots & \dots & \dots & \dots \\ \sigma_{n,1} & \sigma_{n,2} & \dots & \sigma_{n,n} \end{pmatrix} \tag{9}$$

with,

$$\sigma_{i,j} = \sqrt{\frac{1}{W^2}\sum \sum_{(x,y)\in PATCH(i,j)} (R_3(x,y) - \mu_{i,j})^2} \tag{10}$$

where,

$$\mu_{i,j} = \frac{1}{W^2}\sum \sum_{(x,y)\in PATCH(i,j)} R_3(x,y) \tag{11}$$

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

3.2 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 $N_C = 250 \times N_c$ where, $N_c = 250 \times \left\lfloor \frac{N_2}{N_1} \right\rfloor$.

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

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

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

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 $n_{TYPICAL} = 30$.

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

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

$$R(x,y) = \frac{(|I_1(x,y) - I_0(x,y)|)}{I_0(x,y)}$$
$$E_\alpha(x,y) = R(x,y)^\alpha$$

Step-4: Computing the patch-wise intensity diversity statistic Let $\beta(0,1)$ 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. $250 \times N_c$. Set $\beta = \beta^* \in (0,1)$ ($\beta \in \{2\%, 5\%, 10\%, 20\%\}$ of mean image length $N_{avg} = (250 + N_c)/2$), based on simulation experiments conducted and the tuning procedure related to a specific dataset. Let the patch size be $W \times W$ with $W = \lfloor \beta \times N_{avg} \rfloor$. 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

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

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,

$$STAT_{RAW}(I_0) = \frac{1}{N_{patches}} \sum_{\forall patches} \sigma_p \textbf{ TYPE1}$$
$$LSTAT_{NORM}(I_0) = \frac{2}{N_{patches}} \sum_{\forall patches} \left[|ln(\frac{\sigma_p}{\sqrt{W}})| \right] \textbf{ TYPE2}$$

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

3.3 2-class SVM Models for each Client/Subject

The original CASIA set [17] 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. 7 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. 7(a), 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. 7(a), third and fourth images).
- In Subject-11, the problem was very different and existed in the spoofing segment (Fig. 7(b), 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

(a) Some natural variabilities of subjects 4, 6 and 11 respectively

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

Figure 7 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.

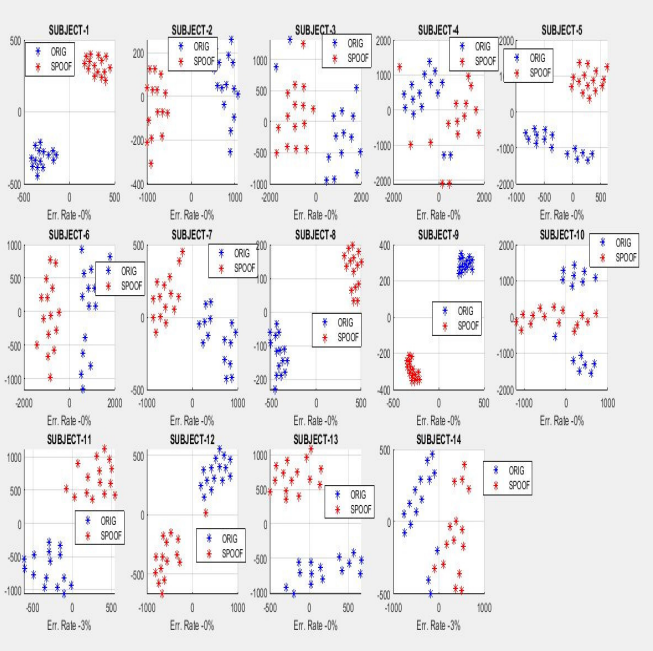

Figure 8 Cluster separation (subject-wise) in a 2-class setting, for the reduced CASIA-dataset comprising of 14-subjects (out of a total of 50) in which 50% of the variations per subject, were used for testing.

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 [16] [19]. The remain 50% of the samples from both the natural and spoof segments were used for testing. The t-SNE maps [29] of the reduced CASIA set test set on a subject specific basis are shown in Fig 8 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 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 [29] (a stochastic map which presents a fairly realistic lower dimensional representation of higher dimensional data) in Fig. 8. In all the subject specific subplots of the test-data, Fig. 8(a-n), the cluster separation was found to be excellent, attesting and reinforcing CLAIMS 1 and 2.

4 Experimental results and comparisons

4.1 Description of Databases

A summary of the datasets used for final round testing of the proposed life-trail algorithm is provided in Table. 2. The original CASIA face dataset [17] shown in Fig. 9 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.

Table 2 Selective face anti-spoofing datasets and related parameters.

Datasets	Year	No of Subjects	Camera	Modal types	Spoof attack	Race
CASIA [17]	2012	50	VIS	RGB	Print photo, cut photo	Chinese
MSU-MFSD[22]	2015	35	Phone Laptop	RGB	printed photo	Asia, Middle east Hispanic, Europeans Latan 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 phots	Almost all the races

The OULU-NPU dataset[27] shown in Fig. 10, 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. 11, 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.

Figure 9 Examples (both natural and spoof-print versions) from the original CASIA dataset [17].

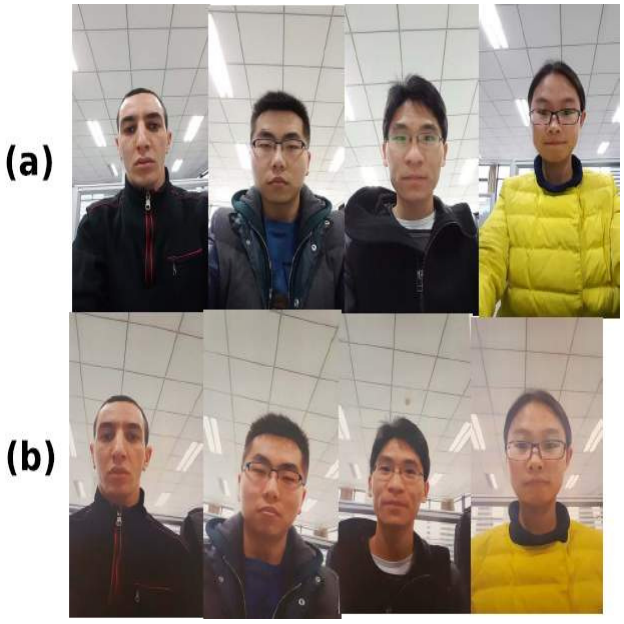

Figure 10 Examples (both natural and spoof-print versions) from the original OULU-NPU dataset [27]

Figure 11 Examples (both natural and spoof-print versions) from the original CASIA-SURF dataset [28].

4.2 Parameter Tuning

While the optimal choice of α is function of the indoor lighting arrangement on a mean scale across subjects, the optimal choice of parameter β is slightly tricky. The selection of β is related to the fractional space occupied by the person's face in relation to the background. 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). In images such as the OULU, the face is small part of the background (here the ratio of face to whole rectangular area drops to 1:4). In such cases the optimal patch fraction (β) is expected to increase. To shortlist the optimal parameter set 5-subjects from the training set with 15-variations per subject were chosen and used to generate the diversity scores from the ratio images (for both the natural and spoof training segments). To compensate for the statistical aperture effect stemming from the patch size increase, a normalizing factor 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. 3.2(Step-5).

If σ_p is the patch standard deviation

$$LS_{NORM(p)} = |ln\left(\epsilon + \frac{\sigma_p}{\sqrt{(W)}}\right)| \tag{12}$$

Where, ϵ is a small positive number. From this the mean normalized log score, the average self-shadow diversity score for a given image, is computed as,

$$LSTAT_{NORM} = t_{IMG} = \frac{1}{N_{patches}} \sum_{p=1}^{N_{patches}} LS_{NORM(p)} \tag{13}$$

$$\begin{aligned} G_{OIG}(\alpha) &= \frac{(\sigma_{OIG}^2 + \sum_{t \in CLUSTER_{OIG}} (t - \mu_{SF})^2)}{\sigma_{OIG}^2} \\ &= 1 + \frac{1}{\sigma_{OIG}^2} \sum_{t \in CLUSTER_{OIG}} (t - \mu_{SF})^2) \end{aligned} \tag{14}$$

$$\begin{aligned} G_{SF}(\alpha) &= \frac{(\sigma_{SF}^2 + \sum_{t \in CLUSTER_{SF}} (t - \mu_{OIG})^2)}{\sigma_{SF}^2} \\ &= 1 + \frac{1}{\sigma_{SF}^2} \sum_{t \in CLUSTER_{SF}} (t - \mu_{OIG})^2) \end{aligned} \tag{15}$$

$$S_{Check}(\alpha) = G_{OIG}(\alpha) + G_{SF}(\alpha) \tag{16}$$

Let $t_1, t_2, ...$ be the mean scores computed for a particular image-class (real or spoof across subjects) and let μ_{ORIG} and μ_{SPOOF} be the mean normalized-log scores for respective classes and σ_{ORIG}^2 and σ_{SPOOF}^2 the corresponding variances.

Table 3 Performance of proposed features with CASIA Database

Parameters	CASIA Dataset Reduced (5 Subjects 15 Variations per subjects)								
β, α	1	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5
0.02	3.952	9.349	9.687	8.389	7.732	7.450	7.355	7.359	7.419
0.05	3.298	7.522	8.704	8.119	7.724	7.563	7.545	7.612	7.729
0.1	2.212	5.762	8.184	8.131	7.935	7.898	7.972	8.105	8.267
0.2	1.848	5.040	8.693	9.668	**9.859**	**9.998**	**10.162**	**10.333**	**10.496**
0.3	2.901	5.804	8.310	9.700	10.453	10.934	11.293	11.583	11.826
0.4	4.127	8.478	11.988	14.141	15.426	16.205	16.661	16.895	16.970

Table 4 Performance of proposed features with OULU Database

Parameters	OULU-NPU Dataset Reduced (5 Subjects 15 Variations per subjects)								
β, α	1	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5
0.02	0.048	0.049	0.051	0.053	0.056	0.058	0.061	0.062	0.064
0.05	0.063	0.061	0.061	0.063	0.066	0.070	0.074	0.078	0.082
0.1	0.039	0.035	0.034	0.036	0.041	0.048	0.059	0.072	0.087
0.2	0.017	0.018	0.022	0.029	0.041	0.060	0.086	**0.120**	**0.160**
0.3	0.009	0.011	0.011	0.010	0.008	0.007	0.005	0.004	0.004
0.4	0.003	0.004	0.006	0.008	0.010	0.014	0.020	0.032	0.050

Table 5 Performance of proposed features with CASIA-SURF Database

Parameters	CASIA-SURF Dataset Reduced (5 Subjects 15 Variations per subjects)								
β, α	1	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5
0.02	0.816	0.834	0.913	1.027	1.150	1.263	1.357	1.433	1.492
0.05	1.049	0.925	0.837	0.809	0.843	0.922	1.020	1.117	1.201
0.1	**1.355**	**1.264**	**1.187**	1.131	1.103	1.100	1.118	1.150	1.186
0.2	1.190	1.034	0.917	0.839	0.799	0.790	0.804	0.832	0.866
0.3	0.381	0.243	0.166	0.145	0.163	0.205	0.258	0.314	0.368
0.4	0.334	0.246	0.186	0.149	0.132	0.131	0.144	0.166	0.197

Table 6 Database and optimal parameter values for various databases, based on the tuning procedure.

Database and Optimal prameters	$\alpha*$	$\beta*$	Justification for $\alpha*$, $\beta*$
CASIA	3	0.2	Prominent self-shadows in most images already present.
OULU	5	0.2	Diffused indoor lighting leading to less prominent self-shadows first up demanding a larger value of α.
CASIA-SURF	1	0.1	Poor quality spoof class images with very low contrast. Hence no change is probably required in the ratio statistic (i.e. exponent α is '1').

Table 7 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[30]	GMMUBM	CASIA	CASIA	C_{gb}	21.69
	Two Class SVM	CASIA	CASIA	C_{gb}	15.42
LBP-TOP[30]	GMM UBM	CASIA	CASIA	C_{gb}	12.65
	Two Class SVM	CASIA	CASIA	C_{gb}	8.53
Motion[30]	GMMUBM	CASIA	CASIA	C_{gb}	12.52
	Two Class SVM	CASIA	CASIA	C_{gb}	11.53
IMQ[10]	One class SVM	CASIA	CASIA	-	23.07
BSIF [10]	One class SVM	CASIA	CASIA	-	36.06
LPQ [10]	One class SVM	CASIA	CASIA	-	35.19
LBP[10]	One clas SVM	CASIA	CASIA	-	25.06
Random Scan[3]	One class SVM	CASIA	CASIA	$N_S = 1$	3.5122
Random Scan [3]	One class SVM	CASIA	CASIA	$N_S = 20$	1.8920

Table 8 State of the art methods within a client specific frame.

Method	Classifier	Train Data	Test Data	Threshold	EER
LBP[30]	GMMUBM	CASIA	CASIA	C_{sp}	10.09
	Two Class SVM	CASIA	CASIA	C_{sp}	9.87
LBP-TOP[30]	GMM UBM	CASIA	CASIA	C_{sp}	6.36
	Two Class SVM	CASIA	CASIA	C_{sp}	3.95
Motion[30]	GMMUBM	CASIA	CASIA	C_{sp}	9.66
	Two Class SVM	CASIA	CASIA	C_{sp}	11.27
MSLBP[19]	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[19]	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 [31]	Deep CNN	CASIA	CASIA	α, β, γ	1.85
Radiometric Distortion [11]	Two class SVM RBF	CASIA	CASIA	$C = 1$	0.00
CPqDN[32]	CNN	OULU-NPU	OULU-NPU	$ProtocolI$	6.9
GRADIANT[32]	CNN	OULU-NPU	OULU-NPU	$ProtocolI$	6.9
STASN[33]	CNN	OULU-NPU	OULU-NPU	$ProtocolI$	1.9
FaceDs[34]	CNN	OULU-NPU	OULU-NPU	$ProtocolI$	1.5
STPM[35]	CNN	OULU-NPU	OULU-NPU	$ProtocolI$	1.0
NHF[28]	CNN	CASIA-SURF	CASIA-SURF	RGB+Depth	4.7
Single-scale SEF[28]	CNN	CASIA-SURF	CASIA-SURF	RGB+Depth	2.4
Multi-scale SEF[28]	CNN	CASIA-SURF	CASIA-SURF	RGB+Depth	0.8
PSMM-Net[36]	CNN	CASIA-SURF	CASIA-SURF	RGB Depth	0.4
PSMM-Net(CeFA)[36]	CNN	CASIA-SURF	CASIA-SURF	RGB+Depth	0.2
Proposed Image life trail	Two class SVM Linear	CASIA $\alpha = 3$, $\beta = 0.2$	CASIA $\alpha = 3$, $\beta = 0.2$	$C = 1$	**0.267**
Proposed Image life trail	Two class SVM Linear	Oulu-NPU$\alpha = 5$, $\beta = 0.2$	Oulu-NPU$\alpha = 5$, $\beta = 0.2$	$C = 1$	**0.17**
Proposed Image life trail	Two class SVM Linear	CASIA-SURF$\alpha = 1$, $\beta = 0.1$	CASIA-SURF $\alpha = 1$, $\beta = 0.1$	$C = 1$	0.73

The values of α and β, which induce a good separation between distributions have been selected as the optimal parameter set. This was done via a cluster separation protocol similar to the usage of the Mahalanobis distance [37]. The impact of a parameter sweep for specific values of $\alpha \geq 1$ and specific values of patch fraction, β, $0 < \beta < 1$ on the misclassification error rates is shown in Tables, Table. 3 (for the CASIA dataset), Table. 4 (for the OULU dataset) and Table. 5 (for the CASIA-SURF dataset).

Instead of simply settling with the parameter-pair (α^*, β^*) which yields the highest separation score for a specific database, the shortlisting procedure entailed first the selection of the column which gave the best score for primary parameter, α for values of *beta* only up to 0.2 (keeping in mind the statistical aperture effect due to large patch sizes). The idea was to fix β and identify the range of α for which the top few scores had registered. The row corresponding to that value of β for which

highest scores were generated (other than the one corresponding to the last row related to $\beta = 0.3$), was shortlisted and the highest scores flagged in bold font.

Note for the CASIA set (Table. 3), the row corresponding to $\beta = 0.2$, yielded high separation scores, which tend to saturate for $\alpha \geq 3$.

As far as the OULU Table. 4 is concerned, the row corresponding to $\beta = 0.2$ (again for $\beta \leq 0.2$) give high scores for $\alpha = 4.5 and 5$ (marked in BOLD FONT).

For the CASIA-SURF Table. 5, the patch fractional score which resulted in high scores (of course ignoring rows corresponding to $\beta > 0.2$) was $\beta = 0.1$ (marked in bold font for different α for a fixed β). Since the $\beta - values$ have now been fixed, the customized values of α for CASIA, OULU and CASIA-SURF were picked as $\alpha_{CASIA} = 3$, $\alpha_{OULU} = 5$ and $\alpha_{SURF} = 1$ respectively. The spoof images used in CASIA-SURF had a very low contrast, due to which virtually no power-law intervention was required, which explained why, $\alpha_{SURF} = 1$ (i.e. the original ratio image could be used as it is). On the other hand, the OULU images both real and spoof were captured in a research lab-indoor environment (i.e. virtually diffused lighting and partial self-shadows for the natural images). As a result the ratio images had to be enhanced considerably before generating the secondary statistics (thus $\alpha_{OULU} = 5$). This justification is given in Table. 6.

4.3 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 apriori, i.e. a face is presented to the camera and the counter-spoofing system must decide whether face-presentation is natural [38] [17] [14] [22] [3]; (ii) The subject identity is known to the counter-spoofing system (more like an authentication environment) [16] [19];

The proposed Image-trail architecture has been 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. 8)as compared to the subject-independent error scores (Table. 7). The best among them is the random walk/scan based algorithm [3][4] 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 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 0.213% Table 8. 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 0.17% and that for the CASIA-SURF [28] was found to be 0.73%. These numbers were comparable with the Convolutional Neural Network (CNN) based solutions shown in Table. 8.

5 Conclusion

Unlike subject-agnostic counter-spoofing solutions, client specific ones tend to offer a higher precision towards the detection of facial spoofing operations mainly because the spatial grids are registered for a particular subject. Within this client or subject specific frame, 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. It was observed that the transition from the first image to the second, carried significant information regarding self-shadows and that a differential ratio image could be created which was found to enhance the self-shadows present in natural closeup images of objects photographed in indoor lighting.

The algorithm was function of two primary parameters: (i) Exponent α deployed to generate the modified ratio-statistic based on a power law to enhance the self-shadow patterns and (ii) Patch fraction β from the secondary statistic computed over the ratio-image, which generates local patch diversity numbers. A tuning procedure was designed based on a reduced number of training-subjects for every new test-dataset to arrive at the optimal parameter set.

Error rates for the proposed algorithm when applied to CASIA (the calibration database) and OULU-NPU and CASIA-SURF were found to be: 0.267%, 0.17% and 0.73% respectively for planar-print-type spoofing operations.

Acknowledgements
Not applicable

Funding
There is no funding

Abbreviations
SVM: Support Vector Machines; CNN: Convolution Neural Netweok

Availability of data and materials
CASIA [17], MSU-MFSD database [22]

Ethics approval and consent to participate
Not applicable

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

Authors' contributions
Client specific Face Counter-Spoofing Model

Author details
[1]Department of Electronics and Electrical Engineering, Indian Institute of Technology Guwahati, Guwahati-781039 Assam , India. [2]Department of Electronics and Electrical Engineering, Indian Institute of Technology Guwahati, Guwahati-781039 Assam , India.

References

1. Patel, K., Han, H., Jain, A.K.: Secure Face Unlock: Spoof Detection on Smartphones. IEEE Trans. Information Forensic and Security **11**(10), 2268–2283 (2016)
2. Nesli Erdogmus, S.M.: Spoofing in 2d face recognition with 3d masks and antispoofing with kinect. In: IEEE BATS, USA (2013)
3. Katika, B.R., Karthik, K.: Face anti-spoofing by identity masking using random walk patterns and outlier detection. Pattern Analysis and Applications, 1–20 (2020)
4. Karthik, K., Katika, B.R.: Identity independent face anti-spoofing based on random scan patterns. In: 2019 8th PREMI International Conference on Pattern Recognition and Machine Intelligence (PREMI), India (2019). Springer
5. Kim, S., Yu, S., Kim, K., Ban, Y., Lee, S.: Face liveness detection using variable focusing. In: Biometrics (ICB), 2013 International Conference On, pp. 1–6 (2013). IEEE
6. Y. Kim, S.Y.J.L. Jaekenun: Masked fake face detection using radiance measurments. In: Conference on Optical Society OF America, vol. 26, pp. 1054–1060 (2009)

7. Karthik, K., Katika, B.R.: Face anti-spoofing based on sharpness profiles. In: Industrial and Information Systems (ICIIS), 2017 IEEE International Conference On, pp. 1–6 (2017). IEEE
8. Karthik, K., Katika, B.R.: Image quality assessment based outlier detection for face anti-spoofing. In: Communication Systems, Computing and IT Applications (CSCITA), 2017 2nd International Conference On, pp. 72–77 (2017). IEEE
9. Katika, B.R., Karthik, K.: Face anti-spoofing based on specular feature projections. In: Chaudhuri, B.B., Nakagawa, M., Khanna, P., Kumar, S. (eds.) Proceedings of 3rd International Conference on Computer Vision and Image Processing, pp. 145–155. Springer, Singapore (2020)
10. Arashloo, S.R., Kittler, J., Christmas, W.: An anomaly detection approach to face spoofing detection: A new formulation and evaluation protocol. IEEE Access **5**, 13868–13882 (2017)
11. Edmunds, T., Caplier, A.: Face spoofing detection based on colour distortions. IET biometrics **7**(1), 27–38 (2017)
12. Gao, X., Ng, T.-T., Qiu, B., Chang, S.-F.: Single-view recaptured image detection based on physics-based features. In: 2010 IEEE International Conference on Multimedia and Expo, pp. 1469–1474 (2010). doi:10.1109/ICME.2010.5583280
13. Candès, E.J., Li, X., Ma, Y., Wright, J.: Robust principal component analysis? CoRR **abs/0912.3599** (2009). 0912.3599
14. Kim, S., Ban, Y., Lee, S.: Face liveness detection using defocus. Sensors **15**(1), 1537–1563 (2015). doi:10.3390/s150101537
15. Zhang, X., Hu, X., Ma, M., Chen, C., Peng, S.: Face spoofing detection based on 3d lighting environment analysis of image pair. In: 2016 23rd International Conference on Pattern Recognition (ICPR), pp. 2995–3000 (2016). doi:10.1109/ICPR.2016.7900093
16. Chingovska, I., Dos Anjos, A.R., Marcel, S.: Biometrics evaluation under spoofing attacks. IEEE Transactions on Information Forensics and Security **9**(12), 2264–2276 (2014)
17. Zhang, Z., Yan, J., Liu, S., Lei, Z., Yi, D., Li, S.Z.: A face antispoofing database with diverse attacks. In: Biometrics (ICB), 2012 5th IAPR International Conference On, pp. 26–31 (2012). IEEE
18. Määttä, J., Hadid, A., Pietikäinen, M.: Face spoofing detection from single images using texture and local shape analysis. IET biometrics **1**(1), 3–10 (2012)
19. Yang, J., Lei, Z., Yi, D., Li, S.Z.: Person-specific face antispoofing with subject domain adaptation. IEEE Transactions on Information Forensics and Security **10**(4), 797–809 (2015)
20. Saragih, J.M., Lucey, S., Cohn, J.F.: Deformable model fitting by regularized landmark mean-shift. International Journal of Computer Vision **91**(2), 200–215 (2011)
21. Wang, T., Yang, J., Lei, Z., Liao, S., Li, S.Z.: Face liveness detection using 3d structure recovered from a single camera. In: 2013 International Conference on Biometrics (ICB), pp. 1–6 (2013). IEEE
22. Wen, D., Han, H., Jain, A.K.: Face spoof detection with image distortion analysis. IEEE Transactions on Information Forensics and Security **10**(4), 746–761 (2015)
23. Galbally, J., Marcel, S.: Face anti-spoofing based on general image quality assessment. In: Pattern Recognition (ICPR), 2014 22nd International Conference On, pp. 1173–1178 (2014). IEEE
24. Galbally, J., Marcel, S., Fierrez, J.: Image quality assessment for fake biometric detection: Application to iris, fingerprint, and face recognition. Image Processing, IEEE Transactions on **23**(2), 710–724 (2014)
25. Weisstein, E.W.: Logistic map
26. Bell, C.: Bright sunlight (2015)
27. Boulkenafet, Z., Komulainen, J., Li, L., Feng, X., Hadid, A.: Oulu-npu: A mobile face presentation attack database with real-world variations. In: 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017), pp. 612–618 (2017). IEEE
28. Zhang, S., Wang, X., Liu, A., Zhao, C., Wan, J., Escalera, S., Shi, H., Wang, Z., Li, S.Z.: A dataset and benchmark for large-scale multi-modal face anti-spoofing. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 919–928 (2019)
29. Van der Maaten, L., Hinton, G.: Visualizing data using t-sne. Journal of machine learning research **9**(11) (2008)
30. Chingovska, I., Dos Anjos, A.R.: On the use of client identity information for face antispoofing. IEEE Transactions on Information Forensics and Security **10**(4), 787–796 (2015)
31. Sun, Y., Xiong, H., Yiu, S.M.: Understanding deep face anti-spoofing: from the perspective of data. The Visual Computer **37**, 1015–1028 (2021)
32. Boulkenafet, Z., Komulainen, J., Akhtar, Z., Benlamoudi, A., Samai, D., Bekhouche, S.E., Ouafi, A., Dornaika, F., Taleb-Ahmed, A., Qin, L., *et al.*: A competition on generalized software-based face presentation attack detection in mobile scenarios. In: 2017 IEEE International Joint Conference on Biometrics (IJCB), pp. 688–696 (2017). IEEE
33. Yang, X., Luo, W., Bao, L., Gao, Y., Gong, D., Zheng, S., Li, Z., Liu, W.: Face anti-spoofing: Model matters, so does data. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 3507–3516 (2019)
34. Jourabloo, A., Liu, Y., Liu, X.: Face de-spoofing: Anti-spoofing via noise modeling. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 290–306 (2018)
35. Wang, Z., Yu, Z., Zhao, C., Zhu, X., Qin, Y., Zhou, Q., Zhou, F., Lei, Z.: Deep spatial gradient and temporal depth learning for face anti-spoofing. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 5042–5051 (2020)
36. Liu, A., Tan, Z., Wan, J., Escalera, S., Guo, G., Li, S.Z.: Casia-surf: A benchmark for multi-modal cross-ethnicity face anti-spoofing. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 1179–1187 (2021)
37. Mahalanobis Distance, pp. 325–326. Springer, New York, NY (2008). doi:10.1007/978-0-387-32833-$1_2 40$.
38. Galbally, J., Marcel, S., Fierrez, J.: Biometric antispoofing methods: A survey in face recognition. IEEE Access **2**, 1530–1552 (2014)