Universal Identity Independent Face Counter-Spoofing

Balaji Rao K

Universal Identity Independent Face Counter-Spoofing

A

Thesis submitted

for the award of the degree of

DOCTOR OF PHILOSOPHY

By

Balaji Rao K

DEPARTMENT OF ELECTRONICS AND ELECTRICAL ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI

GUWAHATI - 781 039, ASSAM, INDIA

June 2022

Certificate

This is to certify that the thesis entitled "**Universal Identity Independent Face Counter-Spoofing**", submitted by **Balaji Rao K** (126102032), a research scholar in the *Department of Electronics and Electrical Engineering, Indian Institute of Technology Guwahati*, for the award of the degree of **Doctor of Philosophy**, is a record of an original research work carried out by him under my supervision and guidance. The thesis has fulfilled all requirements as per the regulations of the institute and in my opinion has reached the standard needed for submission. The results embodied in this thesis have not been submitted to any other University or Institute for the award of any degree or diploma.

Dated:
Guwahati.

Dr. Kannan Karthik
Associate Professor
Dept. of Electronics and Electrical Engg.
Indian Institute of Technology Guwahati
Guwahati - 781 039, Assam, India.

To

LORD SRI SRI SRI. TIRUPATI BALAJI

for His blessings

My guide **Dr. Kannan Karthik**

for his guidance and inspiration

&

My Lovely Son Chi. Dakshith Raj Katika

Acknowledgements

I am obliged to GOD for His divine guidance and blessings. I solely dedicate my thesis in Lotus feet to Lord **SRI SRI SRI TIRUPATI BALAJI**.

This thesis would not have been possible without the immense help and support of several people in various measures. I would like to convey my acknowledgment to all of them.

First and foremost, I express my sincere gratitude to my Ph.D thesis supervisor (My Guru), Dr. Kannan Karthik for providing me an opportunity to work under his guidance. It is very difficult to describe my feelings in words to acknowledge my supervisor for his continuous guidance in all aspects, constant motivation and support throughout the doctoral studies. I am very much thankful to him for transforming me from an unstructured form to a structured form in every aspect of my life and showing me a different path of life. It would be completely impossible for me to bring the research as well as the thesis to this form without the immense facilities provided by him in the Signal Processing Research Laboratory and the freedom of work he has given to me.

I am thankful to my doctoral committee members Prof. Prabin Kumar Bora, Prof. M. K. Bhuyan, and Dr. P. Guha for their encouragement and valuable suggestions on my work. I would like to thank faculty members and the office staffs of the Department of Electronics and Electrical Engineering, IIT Guwahati, for their help in carrying out this research work. I am very much thankful to Mr. Sanjib Das for his kind support and help. With out Internet it would impossible to conduct research, I have express my sincere thanks to Mr. G Shakia Technical officer IIT Guwahati for providing extra slot to download standard databases to carryout research work.

I sincerely thank to Prof. Kannan S R for his help, when I was ill in my Ph.D journey. Prof. Kannan has provided valuable suggestions on my research, which had helped me in completing task in stipulated time.

I owe my deepest gratitude to my teacher Prof. S. R. M. Prasanna for giving valuable suggestions during most toughest period in my Ph.D journey. I would like to express my sincere thanks to Prof. Prasanna for giving me opportunity to work in his first class Signal Processing research labs.

I am very much thankful to my senior colleagues Dr. Haris, Dr. Deepak, Dr. Nagaraj, Dr. Pradhan, Mr. Ramesh sir Dr. Parveen who helped me whenever I need him. I am thankful to my friends Mr. Ramesh, Mr. Inala, Mr. Sarfraj, Mr. Das, and Dr. Kashi Dr. Sharma, Dr. Akilesh and

Dr. Rajesh, Dr. Ajay and Dr. Suman Dev for their assistance in writing my thesis.

Friends part of life would be incomplete in IIT Guwahati, if i had not spent a time with my friends Inala, Das, Kashi in discussing non technical matters in mess and coffee shops and watching a movies in late nights in weekends, eating in stalls enjoying pronight shows of Alcheringa cultural fest held in IIT Guwahati campus.

I also thank to my Junior colleagues/friends Mr. Shoubhik, Mr. Mrinmoy, Mr. Sandeep, Mrs. Deepika Ms. Protima for which I have spend couple of years in IIT Guwahati. Happy moments I still remember is chatting in coffee shops cracking jokes in Lab.

It is time to thank to my close friends Mr. Vikram, Mr. Vijay, Mr. Gangadhar and Mr. Sreenu, Mr. Nalla Karthik and others for giving encouragement during most difficult part of time I have faced.

It is time to thank to my wife Mrs. Sireesha Balaji, with out her patience and cooperation, this research work would have been next to impossible. I thank for her help care and support provided. I have to thank for my lovely son Chi. Dhakshith Raj Katika, as when i come out of working room he will smile and gets excited for play.

I attribute this achievement to my Mom and Dad Smt. Heerabai and Sri. Dharma Raj for their constant blessings, support, silent prayers for my success and moreover, making me stand in this position.

Finally this research would not have been possible without unlimited support of My Leader Prof. Kannan Karthik hence I sincerely dedicate this work to my Professor and thesis supervisor.

Balaji Rao Katika

Abstract

Face biometrics is a common contactless access control method used for applications such as entry into a secure work area or accessing devices like smart phone and laptops. When the access point is unmanned, this is subjected to spoofing attack with digital images, photo prints or prosthetics for illegitimate access. Most conventional face recognition techniques are not equipped to deal with face spoofing, since, as per that frame, spoof-facial presentations are also unfortunately treated as legitimate illumination, scale or pose distortions. Thus, a conventional face recognition system cannot tell the difference between a prosthetic and a geometric scaling operation. The face counter-spoofing layer is therefore an independent algorithmic layer that sits on top of the face recognition system to detect any form of artificiality in the face presented to the camera. Generating a prosthetic or a 3D mask, either from several natural scattered images or video sequences surreptitiously, is a computationally intensive and expensive affair. Hence, most attackers use a planar form of spoofing either via digital images or via printed versions of the target individual's face. Most of the existing counter-spoofing solutions are distortion type and phenomenon specific, meaning that every unique observation related to the manifestation of a certain form of distortion in the artificial image captured from the planar presentation, invokes a unique gamut of solutions. While there had been attempts to pool together the features, there was no single monolithic solution to this planar spoofing phenomenon.

Furthermore, almost all the existing solutions demanded samples from the spoof space to facilitate a 2-sided learning procedure with a view to capture the additional distortions picked up during the execution of spoofing operations. This arrangement selected to eliminate the bias associated with a common illumination environment, made these two sided solutions inflexible, demanding a heavy re-tuning when the acquisition environment was altered. Cross validation across multiple data sets was not very effective. Another major drawback with the state of the art feature sets was the manner in which spatial measure-

ments were taken. They were usually taken in a rigid and regularised fashion via a gridding procedure and secondary statistics were built on top of them. This gathering procedure made the feature vulnerable and sensitive to local content variability, particularly when the learning process was made client or subject independent. Thus, there was a need for an approach which was subject agnostic, yet that remained robust enough to trap and detect the distortions due to artificial presentations alone.

The thesis has two primary contributions encompassing the demand for universality of the solution under different spoofing and acquisition environments and at same time providing a robust spoof-detection procedure in a subject-agnostic frame, using a virtually contiguous random scan algorithm: (i) Design and development of an outlier detection algorithm by first characterizing the natural face space as the inlier space and picking up spoof-presentations as deviations; (ii) Deployment of a virtually contiguous random scan (inspired by the Space filling curves (SPCs) used for encrypting compressed videos), towards an entrapment of the pixel-correlation profile in natural faces across subjects, making the feature extraction subject- agnostic and content-agnostic;

Apart from these two contributions, an analytically heavy model specific solution specifically to detect print-spoofing, is presented as a secondary contribution. Here, an iterative function system in the form of a logistic map, is used to generate a contrast reductionist life trail sequence of images. Each image in this life-trail sequence, because of this image-intensity-swarm drift, progressively moves towards a zero contrast image. It was observed that the first first order difference, in this sequence, leaks significant information regarding the self-shadows. Thus, on a client/subject specific mode, secondary statistics derived from this self-shadow feature, was used to build a 2-class SVM, to isolate test cases involving print-spoofing. Accuracies were comparable to the state of the art CNN based methods.

Keywords: Face counter-spoofing, Countermeasures, Outlier detection, Specularity, Sharpness profiles, Random Scan, Life trails, Self Shadows.

Contents

List of Figures

List of Tables

List of Acronyms

APG	Accelerated Proximal Gradient
ANN	Artificial Neural Network
ASM	Antispoofing model
AUC	Area Under curve
AU	Action units
BU-3DFE	Binghamton University 3D Facial Expression
BC	Bhattacharyya Coefficient
CON	Constrast Score
CASIA	Chinese Antispoofing Dataset
CCA	Canonical Correlation Analysis
CDF	Cumulative Distribution Function
CASIA-SURF	CASIA-SURF Database
CRF	Conditional Random Fields
DET	Detection Error Tradeoff
DFT	Discrete Fourier Transform
EER	Equal Error Rate
EM	Expectation Maximization
EPC	Expected Performance Curve
FAR	False Acceptance Rate
FDD	Frequency Dynamic Descriptor
FR	Face Recongition
FRR	False Rejection Rate
GLCM	Grey level co-occurrence matrices
HFD	High Frequency Descriptor

HTER	Half Time Error Rate
HoG	Histogram of Gaussian
HCI	Human computer interaction
IA	Identification Accuracy
IDTFT	Inverse Discrete Time Fourier Transform
IR	Identification Rate
IQA	Image Quality Assessment
JFA	Joint Factor Analysis
LEV	Log eigen value
LDA	Linear Discriminant Analysis
LDP	Local Derivative Pattern
LP	Linear Prediction
LBP	Local Binary Patterns
LEV	Log eigen Value plot
LBP TOP	LBP three orthogonal planes
MAP	Maximum *a Posteriori* Adaptation
MLP	Multilayer Perceptron
MSE	Mean Square Error
MMSE	Minimum Mean Square Error
MR	Miss Rate
MSLBP	Multi Scale Local Binary Patterns
MSU-MFSD	MSU mobile face spoofing database
MSU-USSA	MSU Unconstrained Smartphone Spoof Attack Database
NN	Neural Network
OANDF	Outlier or Anomaly detection with respect to Natural faces
OULU-NPU	OULU-NPU Database
PDF	Probability Density Function
PCA	Principle Component analysis
PS	Person specific settings
PSD	Power Spectral Density

PCB	Principle Component Pursuit
RFID	Radio Frequency Identification
RBF	Radial Basics Function
ROC	Receiver Operating Curve
RBF	Radial basics function
SVM	Support Vector Machine
SFAR	Spoof false acceptance rate
TABULA RASA	Trusted biometrics under spoofing attacks
T-norm	Test Score Normalization
TPR	True Positive Rate
TNR	True Negative Rate
UVAD	Unicamp Video attack database
UBM	Universal Background Model
VQ	Vector Quantization
WCCN	Within Class Covariance Normalization
ZJU	ZJU eyeblink database

List of Symbols

$\boldsymbol{v}$	Arbitrary vector
E_{Avg}	Average energy threshold value
β	softmax threshold
CON	Contrast Score
d	Distance
V_{D_j}	Eigen Vector
Λ	Eigen Value
ϵ	energy of vector
$X(n,k)$	n^{th} feature frame in k^{th} feature space
X	feature vectors
F	Feature vector
$F(u,v)$	Fourier Transform
$\|\|X\|\|_F$	Frobineus norm
$H(I)$	Histogram of Image
H	entropy
$I(x,y)$	Intensity of image at x,y
$\psi_{u,v}(z)$	Gaussian Kernel
$K(x_i,x_j)$	kernel
LBP^{u2}	Uniform LBP
LBP^{ms}	Multiscale LBP
$LDP(Z_0)$	Local Derivative Pattern at pixel Z_0
l_1	l_1 Norm
$L(X,Y)$	Lagrange function
E_{MSE}	Mean Square Error

μ_X	Mean of Vector X
A_*	Nuclear Norm of matrix A
$\boldsymbol{\Gamma}$	Nonlinear transformation
$\boldsymbol{S}_n$	Normalized score
$\boldsymbol{S}$	Original score
$f(y)$	PDF of standard normal distribution
ρ	Probability Density
I_q	query image
r	Rank of a feature
EER_R	Relative improvement in EER
R_{th}	Rank Threshold
$\rho(x, y)$	reflection coefficient at x,y
$S_\epsilon[x]$	Soft threshold
$f(:,:)$	Thresholding function
$\sigma^2(X)$	Variance of feature vector X
$v(u, v)$	Visual Rhythm
σ^2	Variance
λ	Wavelength of light incident

1

Face counter-spoofing: Motivation and Scope

Contents

1.1 Need for a Face counter-spoofing module

Among all biometrics, the 'face of an individual', has a unique significance, mainly owing to its social positioning. In this digital age, virtually every person intends to create a digital-house for himself/herself, via a Twitter, Facebook, Instagram, Linked-in, Speaking-tree etc., account, to form a niche network (small or big). The motive for this form of positioning in this digital space could be to create some form of channel for marketing in-house products and performances related to art, music and cooking. Since, multiple instances of a particular digital facial identity can be found in several places on the web, a skillful attacker can use these myriad pose and scale variants of this target individual to generate a 3-dimensional version of the same target-face, in the form of a prosthetic [15], [8]. As such, for low-level smart-phone type attacks, particularly directed towards the older models, which do not have a counter-spoofing module, a high quality print copy of the targeted face can be deployed [2]. The access control coupled with recognition engine, which does not have a counter-spoofing module cannot tell the difference between a natural facial presentation and a synthetic face-like object presented to the camera. There is a reason why the recognition paradigm is very different from the counter-spoofing frame [16].

- Facial recognition systems, particularly deployed in the authentication mode, are designed to dissolve variability in the natural space corresponding to pose changes, scale changes and geometric distortions incurred during a legitimate face presentation to the camera. In totality and cumulatively this natural deviation, from a full and clear frontal face presentation translates to some form of geometric-noise which has an un-predictable yet structured pattern. The best methods, which are interest point based and may deploy graphs to string together multiple interest points [[17] [18] Interest point based algorithms, which also involve graphs], are designed to first detect and then characterize these interest points reliably (and dissolve this structured noise) before stringing them together. In the process of looking for something similar, while matching interest points, such systems often tend to overlook other coarse format changes in the facial pattern presented to the camera. Thus, a planar printed photo of person-X (PP(X)), despite its variability in terms of depth-loss, contrast-degradation, natural color deviation during printing and re-imaging, will seep through the recognition system, as the recognition engine will tend to treat this new form of noise as regular geometric-warp related noise. Since it is not noise that is profiled but the image, such engines cannot be used INCLUSIVELY as counter-spoofing

modules.

If person Y intends to masquerade as person X, he/she will require a synthetic and real facial identity corresponding to person X. Let MASK[Y as X] be this synthetic representation. This synthetic presentation can be of two types:

1.1.1 Planar spoofing

Easily available and reproducible either digitized or print versions of person X, when presented to the camera, passes through the access control system, if the recognition engine does not have a counter-spoofing module.These two sub-modalities may be represented as

- MASK[PP, Y as X]: where PP stands for planar print of X's face.
- MASK[PDIG, Y as X]: where PDIG stands for the digitized planar version of X's face.

Figure 1.1: Examples of real access and attack in different scenario, In the top row samples from adverse scenario- in the bootm row samples from controlled scenario. Column from left to right show real access, printed photo, mobile phone and tablet attack faces [1].

Examples of both forms of planar spoofing: both print and digital image based are presented in Fig. 1.1. The difference between these two planar spoofing modes, is the analog printing process which ensues when the digitized version of person X's face is converted to a print form (either planar or warped/curved surface without any form of depth variation at different facial interest points). Both

these planar spoofing modalities have something distinctive about them, in terms of the distortion format, which can be used to pick out the attack type.

For instance, digital face images are backlit, while print versions have to be illuminated either from the front or from the sides. Both these presentations lack real depth information during the final still image capturing session, which can be detected with a precise and elaborate blur diversity check [7] [16]. Print versions have many more cumulative distortion patterns as compared to their digitized counterparts, among those, there are some which are exclusive to print versions (such as contrast degradation [3] and suppression of self-shadows (elaborated in Chapter 6, in our work)). There is one specific distortion pattern which is unique to digital planar presentations, which is the onset of re-sampling and interpolation noise and these patterns are called MOIRE' PATTERNS [19].

1.1.2 Prosthetic based spoofing

This is a more sophisticated form of spoofing wherein the facial parameters, except the eyes, nostrils and mouth of the individual are duplicated via some form of a synthetic 3D-mask built either out of paper craft [8] or rubber or clay [20]. An example of a prosthetic mask is presented in Fig. 1.2. This spoofing mode is represented as,

- MASK[PROSTHETIC, Y as X]

Since the face-shape and coarse apparent features are duplicated synthetically, this form of spoofing can fool an un-manned surveillance system even with a moderately complex counter-spoofing module. Counter-spoofing algorithms which work against planar spoofing will not be very effective against prosthetic masks [3].

- Person Y crawls through the web and gathers several face images of person X. Using these templates, he/she can re-create a 3D-version of person face [1], more along the lines of a facial hologram. Person Y would need only the facial topography of X (or rather frontal surface plot). Once obtained, person Y would then print out this facial surface of X onto some medium such as gelatin, rubber or clay and use that as a cover for his/her face with slits at the right places.

- Gaming consoles such as KINNECT [8] tend to generate digital holograms of a real player or use depth related information, to allow these players to enjoy immersive interactive experiences with several characters within the frame. Person Y can rig one such box and extract without

Figure 1.2: Example of a prosthetic [5].

the knowledge of the player, the player's 3D-facial profile and eventually print it out onto some medium.

- Either from a portrait sketch (with shading parameters) or from a well taken digital image with self-shadows, it is possible for the individual to use this self-shadow information to use existing 3D surface models to estimate the exact elevations at different points on the individual's face if the light source direction is known. Once this surface is generated, it can then be printed out onto some medium to form the mask for person-Y.

It is quite evident that the prosthetic-based spoofing is likely to be less appealing to most attackers since the 3D-recreation process (particularly at a clandestine level) is difficult and unless one has access to sophisticated "art labs or studios" where 3D-printing can be done, production of this rubber or clay mask is not possible. Hence, unless one is planning an upper end masquerading attack, such as an impersonation one in top-secret organization, prosthetics will not find a use. However, for lower end attack fronts such as smart phone unlocking [2] and others, one can enter systems using these planar spoofing modalities. This aspect has been highlighted in Table. 1.1.

Table 1.1: Application space for different spoofing modalities

Person Y as Person X	Application space	
Planar spoofing (digital image or print)	Smart phone unlocking	Impersonation (stealing facial digital IDs) such as PAN-cards etc., some of which do not have pseudonyms or hash-codes as virtual identifiers.
Using Prosthetics	Entry into top-secret organizations, particularly at points where un-manned surveillance is used.	Concealing identities at key points such as railway stations, airports etc., where surveillance cameras can be deployed.

Table 1.2: Parameters to look out for detecting different forms of spoofing.

Presentation Modalities				
Physical Parameter	Natural face	Planar spoofing		Prosthetics
		Print-spoofing	Digital-spoofing	
Contrast [3]	Rich and natural	Suppressed	Virtually same as original face due to backlighting	Same as normal face
Color quality and naturalness [21]	Natural and rich	Artificial in some cases with some distortion	Mild distortion	Same as original face
Self-shadow prominence (proposed, Chapter 6)	Normal and prominent	Suppressed	Normalcy retained due to backlighting	Similar to the natural face owing to the depth profile
Sharpness diversity (or its opposite BLUR variation) [22] [7]	Significant and diversity can be detected	More homogeneous with inherited diversity suppressed	More homogeneous	Sharpness diversity exists because of the depth profile. But due tthe "over-smoothing" constraint, it is less compared to a natural face.
Quality in general (assuming cumulative distortion in the case of spoofed versions) [23]	High Quality And Clarity	High cumulative distortion	Some distortion prevails	Virtually the same as a natural face
Moire' patterns (re-sampling and interpolation noise) [19]	NON-Existent	Low but may arise if camera zooming is done for the print submission	High because this resampling and interpolation is done twice before final acquisition	NON-Existent
Specularity [24] [25]	Virtually non-existent owing to depth profile	Depends on the nature of the paper on which the face is printed and the position of the local light source relative to the object and camera	Virtually non-existent owing to backlighting, unless some form of ghosting is observed	Virtually non-existent.

Table 1.3: The complete description of anti-spoofing database used to detect the various kind of spoof attack faces

Data set	No of real subjects (No of variations per subject)	No of spoof subjects (No of variations per subjects)	Variability in images of real subjects	Variability in images of spoof subjects	Remarks
CASIA	30(15)	30(15‘)	Recorded mostly with PC-mounted cameras: Poses are largely frontal with significant illumination, pose and scale change (even within the same subject).	Print quality variability and dominant specularity observed in most of the samples. Pose, Illumination and scale variability is less as compared to natural images. This is done to deliberately fool the surveillance system.	Ethnic base is mostly East Asian, largely dominated by people from China.
MSU-MFSD	35(30)	35(30)	Two kinds of cameras with different resolutions (720×480 and 640×480) were used to record the videos from the 35 individuals. Poses are frontal and there is minimal illumination variability but with scale change (subject moves towards and away from the camera).	Recorded with multiple cameras. Print quality variability and considerable specularity observed in most images. There is scale change as well.	Ethnicity is diverse and dataset includes people from Asia, Middle East Europe, America and Latin-America.
OULU-NPU	20(20)	20(20)	Six mobile devices camera such Samsung, HTC, OPPO, ASUS, MEIZU, Sony Xperia. All are frontal poses with minimal scale and illumination changes.	There is not much difference in the print-versions except for a reduction in contrast in relation to the natural versions.	Includes people from Europe and East Asia.
CASIA-SURF	50(35)	50(35)	Both RGB as well as Depth profile of real genuine faces have been recorded. Poses are frontal with minimal illumination change.	RGB and Depth profile recorded for Spoof-versions as well. Spoof presentations are with and without paper-cuttings exposing parts of the face such as EYES, NOSE etc.	Includes largely people from East Asia.
3D mask	17(50)	17(50)	Real done under controlled conditions, with frontal-view and neutral expression.	Prosthetic mask has designed with paper crafts.	Includes people from Europe and Middle-East.

1.2 Paradigms for Counter-spoofing

There are two primary elements which make up the counter-spoofing frame: (i) Modality or Modalities involved in the primary feature or feature set selection and (ii) Classifier model building strategy or counter-spoofing paradigms. When a spoofing operation takes place, the artificiality brings with it certain geometric constraints coupled with some form of a cumulative distortion in some cases. Generally, the choice of features is directed by a phenomenon or phenomena observed during the acquisition process, whether connected with natural faces or spoof ones.

Table 1.2, summarizes some perceptible as well as statistical differences between natural and spoof versions. It is quite obvious from the table that planar spoofing, particularly of the print type can be easily separated from natural face presentations on several grounds such as contrast [3], color distortion [21], specularity [24] [25], blur diversity vs homogeneity [22] [7] etc. On the other hand digital planar spoofing can be detected on grounds of blur homogeneity [22] [3] and re-sampling noise [19].

It is also obvious from the table that the hardest form of spoofing is the prosthetic version and the only way it can be detected is via a highly precise blur/sharpness analysis. There are several possibilities towards class-specific model building.

1.2.1 2-sided training with samples collected across subjects (subject independence)

, In this counter-spoofing frame one assumes both natural and spoof faces samples are available for building a robust 2-class model, provided the features chosen are discriminatory. These samples are gathered across subjects and most samples are assumed to be images of full frontal face-poses (example Fig. 1.3). The main drawback with this arrangement is that every individual has a distinct facial profile in terms of eye-positioning, eye-size and socket shape, cheek and jawline, nose shape and size, type of forehead and overall face outline. This coupled with scale, geometric and perspective distortions due to an imperfect presentation of the face to the camera, makes facial registration very difficult. The facial grid being rigid introduces what is known as “subject-content-noise” [16], which increases the intra-class variability.

Figure 1.3: Samples across subjects for 2-sided training from CASIA dataset [6]

1.2.2 2-sided training in client specific mode [1] [2]

It was observed while including faces from different subjects to build both natural and spoof class models, there was a large intra-class variability. To counteract this the client specific model was proposed in literature, wherein 2-class models were built keeping in mind an authentication arrangement wherein the identity of the subject presenting his/her face to the camera was known a priori to the counter-spoofing system. This allowed the system to build subject specific 2-class models and this suppressed the intra-class variability considerably so long as the poses were full frontal, with minimal scaling and rotation and had moderate lighting variations. There are several disadvantages with this arrangement:

- Very few samples are available on a subject specific basis, particularly because scale and pose-variants since the facial parameters need to be spatially registered. This leaves out a handful of frontal poses per subject for building the client specific class model (example Fig. 1.4).

- This also assumes spoof-examples are available, although the exact model and nature of spoofing operation is un-predictable. Due to this the spoof class is under-represented and in a comparative setting, this may not result in a robust 2-class client specific model. There could be subjects for which no spoof samples are available. In [2], this problem regarding limited data was addressed using what was called subject domain adaptation, where, synthetic features from other subjects could be used to derive synthetic features for subjects for which spoof samples were not available.

For these subjects, to facilitate the adaptation, the mapping was directed in the presence of their own natural samples.

- This also assumes spoof-examples are available, although the exact model and nature of spoofing operation is un-predictable. Due to this the spoof class is under-represented and in a comparative setting, this may not result in a robust 2-class client specific model. There could be subjects for which no spoof samples are available. In [2], this problem regarding limited data was addressed using what was called subject domain adaptation, where, synthetic features from other subjects could be used to derive synthetic features for subjects for which spoof samples were not available. For these subjects, to facilitate the adaptation, the mapping was directed in the presence of their own natural samples.

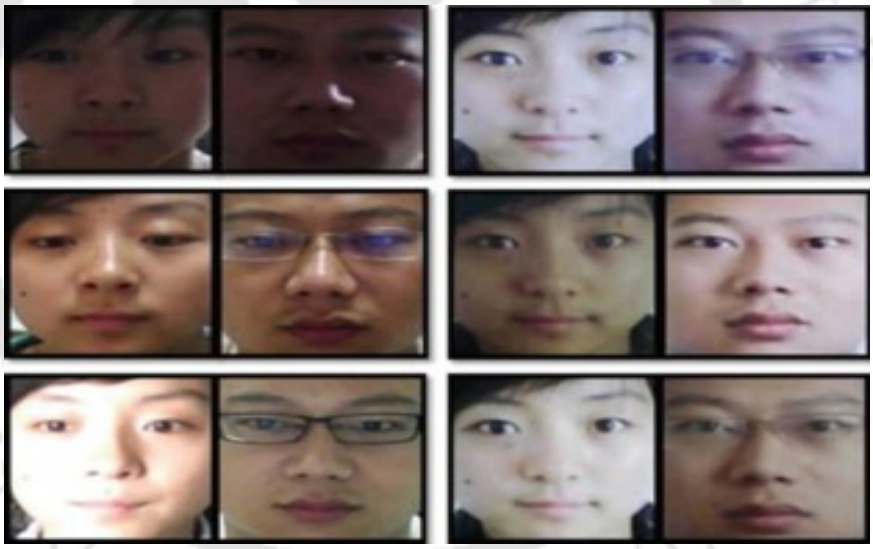

Figure 1.4: Client specific examples from CASIA dataset [1].

1.2.3 Natural space characterization and one-sided training [3] [4]

Since the spoof-class or modality remains un-predictable, it becomes necessary to focus on what is available in plenty and what well understood by the counter-spoofing system. This means the training has to be one sided and must have several natural examples across subjects. The raises a few questions:

- How does one choose the primary discriminatory feature?
- If the natural samples are gathered across subjects, how one mitigates feature variability within the natural face class?

- How does one construct a one-sided training model? Rather, on what basis does one establish where to draw the boundary or surface? Should this boundary be kept rigid to detect outliers or should the test-sample be positioned in relation to the data to generate some form of INLIER or OUTLIER-score? The latter is data adaptive and works when the number of training samples is small.

This calls for a UNIVERSAL solution in a subject independent setting [16] that is:

(i) **DISCRIMINATORY yet COMPACT and ROBUST as a feature**

(ii) **Subject content noise is suppressed despite lack of proper spatial registration.**

In our originally proposed work related to the Contrast parameter and the outlier detection frame [3], the outlier detection was done based on a simple contrast score ranking mechanism done in an inclusive fashion, wherein the scores of the natural training samples were fused with the score of the test sample and then the ordering was done. If the score of the test sample (test-point) fell well within the footprint of the natural scores (or natural points), then it was declared as an inlier, else an outlier.

At the same time, as an independent body of work, [4], used quality-assessment related features and statistics [23] to construct a 1-class Support Vector Machine (1-SVM). Both these methods had limited accuracies: (i) the contrast and outlier proposition [3], failed because contrast was a weak feature; (ii) The approach of Arashloo et al. [4] failed mainly for digital planar spoofing for the very reason that digital images are very different and have a higher quality as compared to the print-spoof counterparts [16]. It is unlikely to work against prosthetics.

1.2.4 Deep Learning approach for Face Counter-Spoofing

- DEEP-LEARNING (approach restricted to 2-class setting): Firstly, DEEP-LEARNING approaches become useful when there is a wealth of data available regarding both classes (not just one). It found that almost all deep-learning techniques, assumed a 2-class arrangement and the availability of data related to both the natural set (customized according to a certain environment) and also samples related to the spoof segment for the same subjects present in the repository for which the natural samples were also available.

- The UNIVERSAL method have eventually proposed is RANDOM SCANS with DIFFERENTIAL STATISTICS for extracting CONTENT AGNOSTIC features linked to the NATURAL

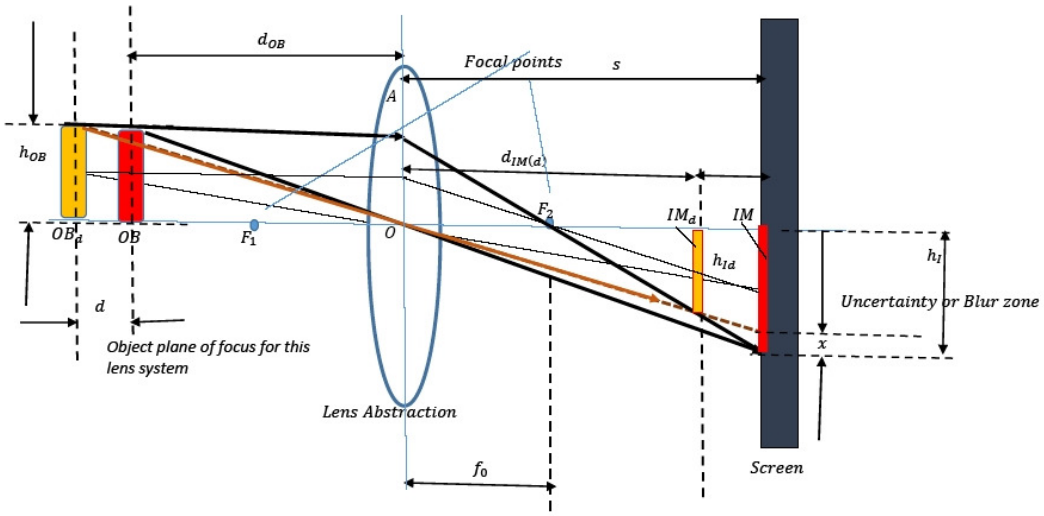

Figure 1.5: PINHOLE camera model highlighting the blur phenomenon [7].

FACE SPACE ALONE. The model building here is ONE-SIDED and INTRINSIC. This means, one can BUILD the MODEL on NATURAL-FACE-CASIA ALONE and test it on both NATURAL and SPOOF MSU-MFSD. Cross-validation is effective in this arrangement. This form of generalizability is not possible for the DEEP-LEARNING models as they are 2-sided and environment dependent. There are however CNN-linked MULTI-MODAL fusion based papers which have attempted this aspect connected to "generalizability" [13].

Chapter 5 (RANDOM-SCANS with INTRINSIC-learning) and Chapter-6 (LIFE-TRAILS) are our main final contributions. While bringing in the comparisons related to deep-learning we bring to your notice, tables in Chapters 5 and 6.

1.3 MAIN Proposition: Natural space characterization in a content agnostic way

The lessons learnt from Sub-sections 1.2.1, 1.2.2 and 1.2.3 led to the following demands of an effective counter-spoofing system:

(i) Natural face space characterization is a must.

(ii) To ensure a large training ground, the natural samples must be taken across subjects, i.e. must be made subject-independent.

(iii) Subject content variability owing to the spatial de-registration problem must be suppressed.

(iv) The type of feature or statistic, designed or chosen, must work universally to detect all spoofing modalities, including prosthetics. This means two things, "availability or detectability across all modalities" and "statistical distinctiveness natural versus spoof".

Thus, in this light we propose a UNIVERSAL IDENTITY INDEPENDENT COUNTER-SPOOFING SYSTEM based on CONTIGUOUS RANDOM SCANS [16]. In Table 1.2, it was seen that, the only phenomenon that leaves a distortion in both the natural and spoof classes (including prosthetics), while at the same time exhibits a distinct statistical pattern is the BLUR-phenomenon, which stems from the PINHOLE-camera model.

1.3.1 Blur quantification and analysis because of the PINHOLE camera effect

When a snapshot is taken of a natural face, the plane of focus is a vertical plane parallel to the camera screen, somewhere, depending on the position of the vertical slice on which the camera is focussed. All points on this plane of focus appear very clearly. However, all points either in front of it or behind it will appear blurred depending on the distance of that plane from this plane of focus. In this exemplar figure below Fig. 1.8, the optical axis of the camera passes through the tip of the nose and is normal to the object plane indicated in red. In the front and side view, belonging to the same subject- instance, we highlight the points of interest which we wish to track in the imaging process.

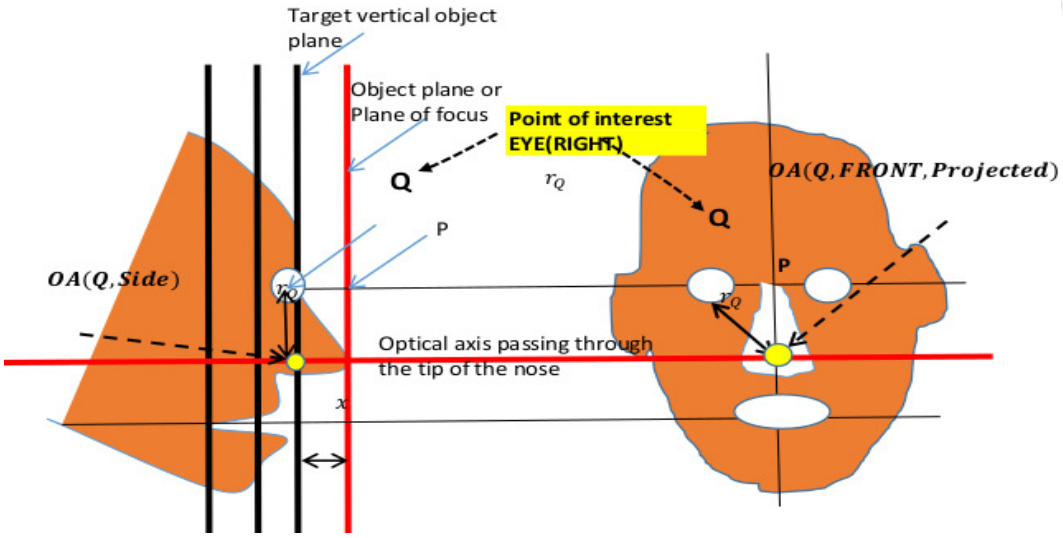

Figure 1.6: SIDE views of a particular subject's face. The target vertical object plane has been laid out to bring out the radial distance of the point of interest Q.

Note that in the conventional lens-equation system, when one restricts the imaging to points on the plane of focus or the object plane,

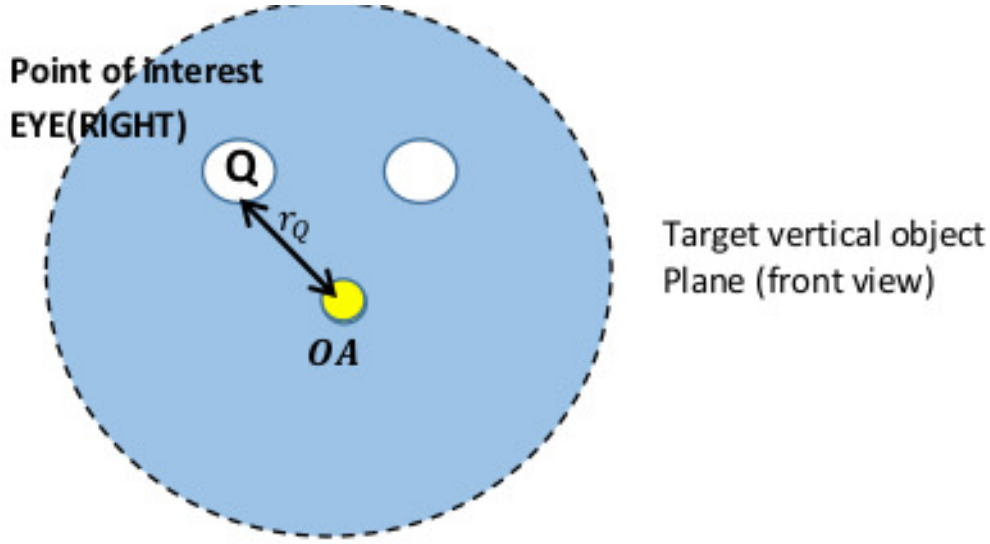

Figure 1.7: Front views of a particular subject's face. The target vertical object plane has been laid out to bring out the radial distance of the point of interest Q..

$$\frac{1}{F} = \frac{1}{u} + \frac{1}{v} \tag{1.1}$$

Where v is screen distance from equivalent optic centre and u distance of plane of focus of object plane. Hence $d_{OB} = \frac{SF}{S-F}$.

This changes here, as the point of interest Q is not on the plane of focus. This point is in fact at a horizontal distance $d_Q = x$ behind the plane of focus, and thus at a distance $u_Q = d_{OB} + x$ from the optical –centre with respect to the lens-system as shown in Fig. 1.8. The vertical plane containing the point Q intersects the optical axis at point OA. Thus the radial distance of the point Q from this point of intersection OA is r_Q. which is its "effective-height" when the plane containing OA and Q and the optical-centre is rotated and positioned vertically. The transformed arrangement to simplify the geometric analysis is presented in Fig 1.8. The intermediate 3D-arrangement is shown in Fig. 1.8.

In Fig. 1.8 now consider a point P suspended in space. This point will appear clearly on the screen. There will be absolutely no region of uncertainty about the image of this point. This result will hold for all the points on the object plane, irrespective of the radial distance. Now consider the target vertical plane at a distance x behind the object plane and a point Q on it at a radial distance r_Q coinciding with the eye of the person. This point is not a suspended point but is a point on the person's face. The radial distance seen from the front view is r_Q. This point Q will appear blurred

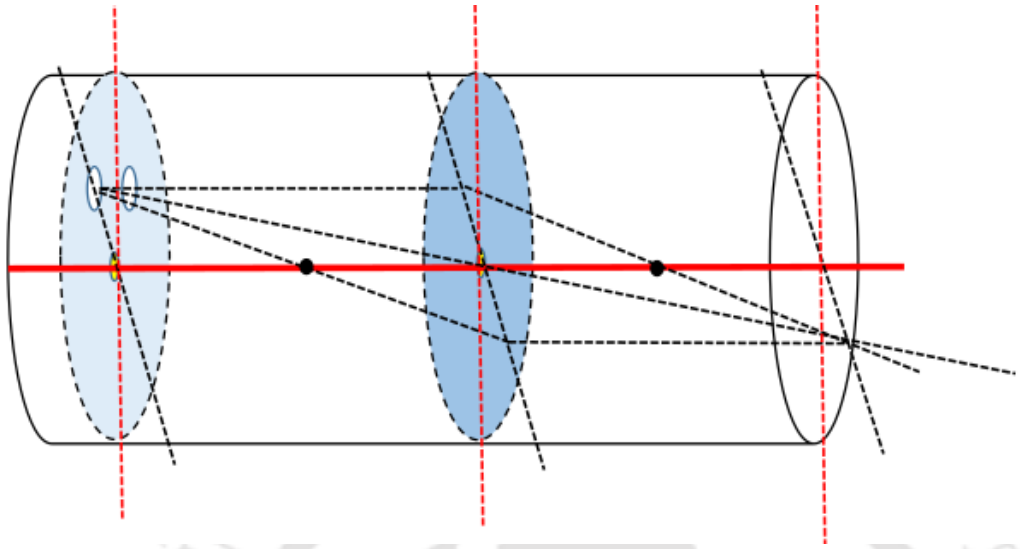

Figure 1.8: Front views of a particular subject's face. The target vertical object plane has been laid out to bring out the radial distance of the point of interest Q..

on the screen. This is because this point Q will not focus on the camera screen but rather on a plane in front of it. The lens equation for this is,

$$\frac{1}{F} = \frac{1}{d_{OB} + x} + \frac{1}{v_Q} \tag{1.2}$$

Where, v_Q horizontal distance of the image of Q on the camera side with respect to the optical centre. Thus the image of point Q which is outside the screen culminates in the cone of uncertainty or a footprint of uncertainty on the screen. Deploying the similar triangle relation based on the PINHOLE model, it follows that,

$$\frac{r_Q}{d_{OB} + x} = \frac{r_{Q_{IMG}}}{v_Q} \tag{1.3}$$

Where, r_Q, is the radial distance of the image of Q from the optical axis. Finally $r_{Q_{IMG}}$ can be written as

$$r_{Q_{IMG}} = \frac{F r_Q}{d_{OB} + x} \tag{1.4}$$

Since the imaging plane is in front of the screen as seen in Fig. 1.9, there will be a line of uncertainty extending outwards and passing through the geometric centre of the screen. This line of uncertainty has a bound and this bound defines the spread of the image of point Q on the screen in the form of a line trace. This trace has a width of Δ_Q units. Focussing on the similar triangles pivoted about the

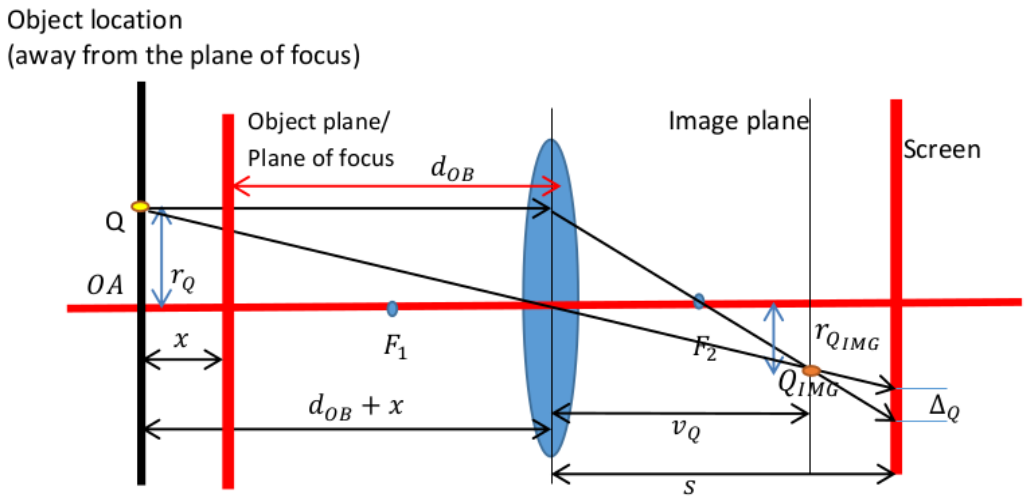

Figure 1.9: Tracking the images of the points of interest when captured by the lens system. Note here the plane containing the points Q, OA and the OPTIC-CENTRE has been rotated to make it vertical along the page of this sheet to simplify the geometric interpretation and analysis. The transitional 3D-figure is shown in Fig. 1.8.

image of Q (viz. Q_{IMG}), it follows that,

$$\frac{r_Q}{\Delta_Q} = \frac{V_Q}{S - V_Q} \tag{1.5}$$

$$\Delta = \frac{r_Q(S - V_Q)}{V_Q} = r_Q(\frac{S}{V_Q} - 1) = r_Q(\frac{S}{[\frac{F(d_{OB}+x)}{d_{OB}+x-F}]} - 1)$$

$$\Delta_Q = r_Q(\frac{S}{F}[1 - \frac{F}{d_{OB}+x}] - 1) \tag{1.6}$$

Using Equation 1.2, with respect to the point P on the object plane, and substituting for d_{OB}.

$$\Delta_Q = r_Q(\frac{S}{F}[1 - \frac{F}{\frac{SF}{S-F}+x}] - 1) = [\frac{r_Q x(\frac{S}{F}-1)^2}{S + x(\frac{S}{F}-1)}] \tag{1.7}$$

1.3.2 Depth diversity and Indirect Depth Profiling

A natural face has a curvature profile, a surface topographical variation which unique for every subject. When this face is presented to a still camera, depending on the positioning of the object plane (or plane of focus, OB as shown in Fig. 1.5) PINHOLE MODEL [3] everything behind and in front of this plane of focus, OB, will appear blurred. The extent of blur has been shown to be function of the relative distance or depth with respect to the plane OB and the curvature and type

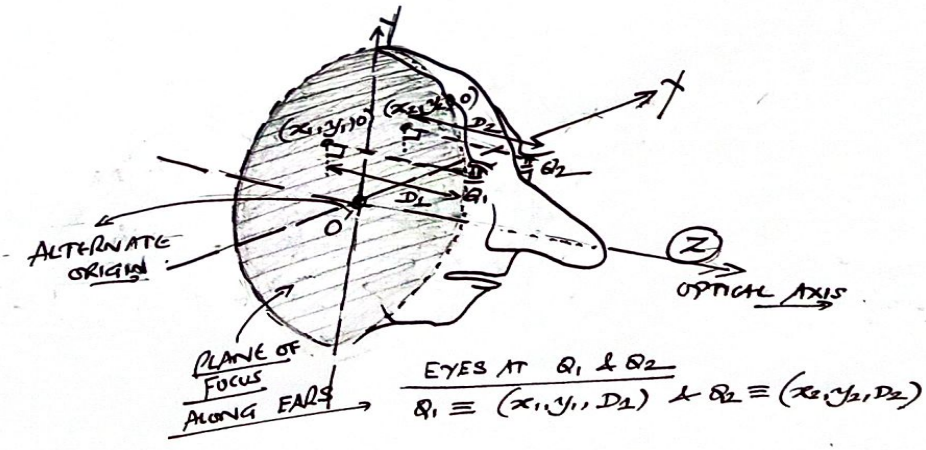

Figure 1.10: The plane of focus is represented by the shaded portion which is along the X-Y plane. The origin O' of this alternate co-ordinate system is formed at the intersection of this plane of focus and the OPTICAL AXIS. The two referential points are the two eye-locations denoted by Q1 and Q2. The co-ordinates of these points are Here, represent the perpendicular distances of points Q1 and Q2 with respect to the plane of focus (shaded portion). In general represents the face-surface map.

of lens (normal or fish-eye etc.) and focal length selection [7]. There is a body of work connected with this line of reasoning. In KIM et al. [22], it was observed that planar print presentations lack depth, hence, when a real natural facial presentation is compared with a print-version, there will be a difference in the degree of sharpness and in particular the sharpness diversity seen in the images trapped in these two presentation modes. By ensuring a narrow depth of field during still image photography it was possible to generate differential sharpness profiles for natural images and print spoof versions. With the natural curvature or change in the surface topography associated with a person's face, depending on the position of the plane of focus in relation to the optical center of the lens, a certain zone in the face will appear clearer as compared to the rest. Given point on the facial surface, the linked to its image produced on the screen will be a function of its vertical distance from the plane of focus. This was illustrated in the PINHOLE model Fig. 1.5. Thus, each person should ideally produce a distinctive depth map based on the surface topography as can be seen in the Fig. 1.11 THREE-SIDE-POSES below. When one compares the facial topography of a natural face of person X versus two synthetic arrangements: (i) Planar print version in which the printed image of X is presented to the camera; (ii) Prosthetic mask tailormade according to X's facial profile by Y and presented to the camera, the depth variation with respect to the plane of focus (PF) or object

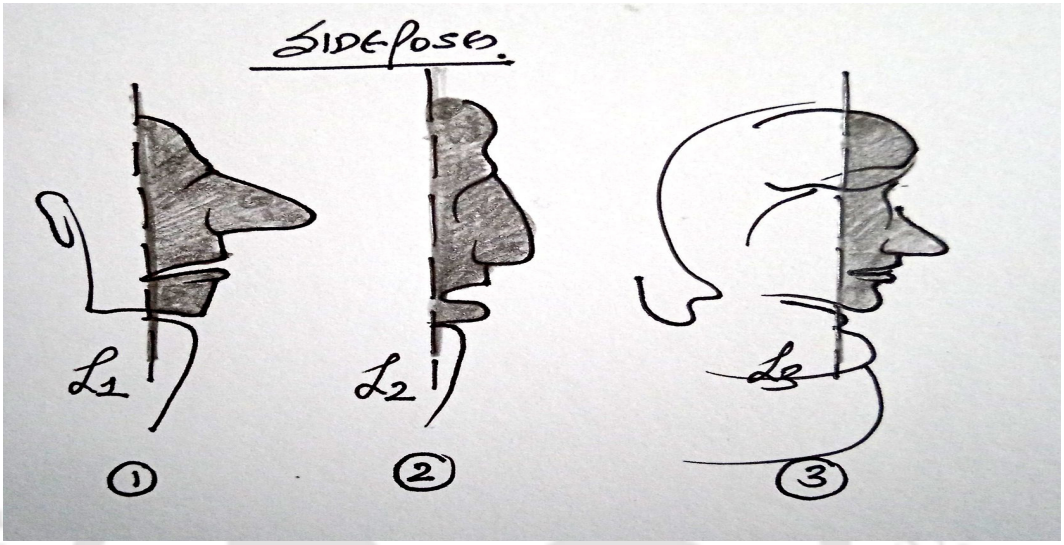

Figure 1.11: Three side-poses with different surface topographies should generate distinct depth maps.

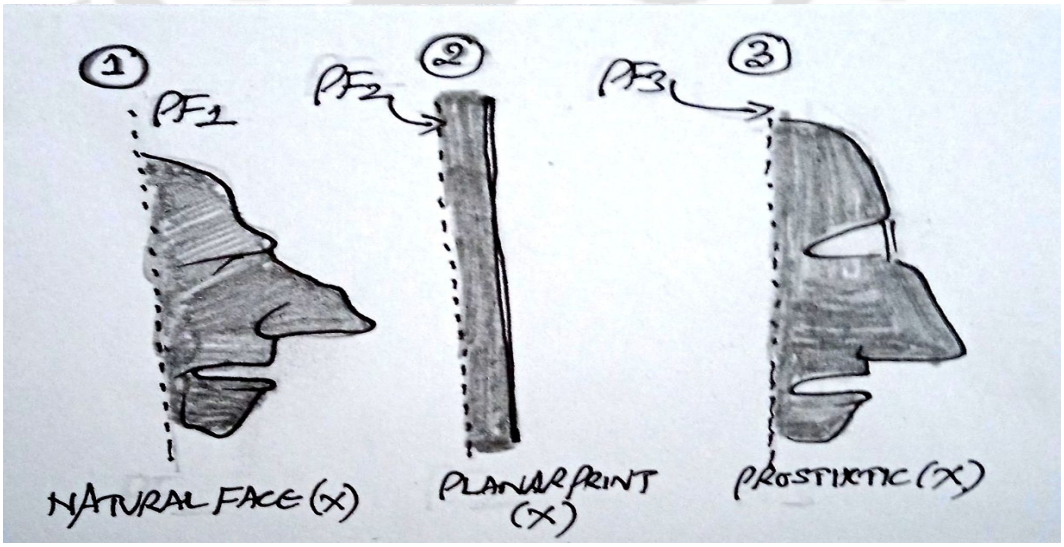

Figure 1.12: Side-poses, natural face, planar print and prosthetic of same subject.

plane (OB) (one and the same), is shown in Fig. 1.12 Let $\sigma(x_0, y_0)$, represent the approximated blur standard deviation (assuming isometry but a space varying blur as a function of relative distance), at a specific interest point location such as one of the eye centers (left or right eye). The σ, profile, over the entire image, for different presentation modes corresponding to subject X's face can be discussed as follows:

- Since this blur is a function of the relative depth or distance with respect to the object plane, one can express one as a monotonic function of another:

$$B_{NAT}(x, y) \approx f_{MONOTONE}(D(x, y)) \tag{1.8}$$

  Where, $D(x, y)$ happens to be the distance from the object plane or plane of focus. This relative depth variation appears in the form of a heterogeneous blur deviation over the entire image. Estimating the degree of blur at a given point in space is an indirect way to quantify the depth profile trapped at that location. This depth profile $D_{NAT}(x, y)$ has two components.

- Since natural faces have a rugged topography, the surface relative to the object plane (i.e. the relative distance profile) can be modelled as,

$$D_{NAT}(x, y) = D_{SMOOTH}(x, y) + N_{NAT}(x, y) \tag{1.9}$$

  Where, $D_{SMOOTH}(x, y)$ is the over-smoothed surface topographical profile associated with the subject's face, while, $N_{NAT}(x, y)$ is the immersive micro-acquisition noise connected with many natural imaging factors such as contrast diversity, presence of localized self-shadows, micro-blur variation since the surface topography associated with a particular subject's face varies from point to point on the surface. Collectively and cumulatively this noise pattern $N(x, y)$ carries crucial information regarding the degree of naturalness linked to the imaging process.

$$D_{NAT}(x, y) = \lambda_1 D_{NAT}(x, y) * G_\sigma(x, y) + \lambda_2[D_{NAT}(x, y) - D_{NAT}(x, y) * G_\sigma(x, y)] \tag{1.10}$$

The sketching and caricatures were included to bring out the impact of the exaggerated facial parameters firstly on the facial topography and subsequently the relative depth profile presented to the camera system. The LINES L1, L2 and L3 in Fig. 1.11 indicate the PLANE OF FOCUS or the OBJECT PLANE. The greater the perpendicular distance of a typical point on the

surface of the face from this plane of focus, greater will be the degree of blur. Since the surface topographies are different for different people, the blur diversity will vary but in terms of a quantum will be on the higher side. The second figure Fig. 1.11, is used to illustrate what happens when a FACIAL-PLATE (second figure in Fig. 1.7 below) is presented to the camera-system. The resultant blur is weighted combination of a homogenous blur with an inherited diverse blur stemming from the natural face which was originally imaged. Finally, in the third figure in Fig. 1.7, which involves a prosthetic, the "over-smoothing constraint" emerging from the "one-mask fits all problem", is illustrated. This over-smoothing is exactly what gives away the prosthetic mask.

Since it is hard to find faces and annotate them to suit our problem-presentation, we prefer to retain our own hand-drawn figures.

In Chapter. 5, we demonstrate that this subtle overriding natural noise can be trapped more effectively using a contiguous random walk process [16] [26] and then computing the differential statistics related to this randomly scanned vector to profile the noise pattern. We show that when a random scan is deployed, the noise profiling becomes content agnostic, as opposed to a DOG operator which tends to leave behind edge artefacts (crease lines involving wrinkles, eyebrows, edges of the nose, outline of the mouth and prominent singular points such as the eyes etc.).

- In the case of a planar print, if the plane of focus does not coincide with the plane of the paper-print presented to the camera, there will be a uniform homogeneous blur superimposed on top of the original blur diversity trapped in the image which was printed. Thus the apparent depth profile trapped from the image of the planar print can be written as,

$$D_{PP-APPARENT}(x,y) = \alpha_0 D_{NAT}(x,y) * G_{\sigma_p}(x,y) \tag{1.11}$$

Where, the FIXED Gaussian parameter, σ_p is a function of the relative distance between the plane of focus and the plane of the printed photo and $\alpha_0 \in [0,1]$ is a scaling parameter which in a crude way indicates a reduction in the dynamic intensity range during the printing process and '*' represents a 2D-discrete space convolution.

- Notice the with the prosthetic, the depth profile is real and not apparent and in our claim

in Chapter 5/6 [RANDOM SCANS], it is brought out why a prosthetic is an artificially over-smoothened version of a particular individual X. In a nutshell, the MASK(Y as X) has to fit any one among many individuals which have a facial structure similar to Y, as the mask is not exactly designed keeping Y in mind, but X. Thus, the depth profile for a prosthetic can be written as,

$$D_{PROSTHETIC}(x, y) = D_{SMOOTH}(x, y) \tag{1.12}$$

Note here that in the smoothened version, there is a spatial blur variation unlike the planar print version.

Claim-CH1.1: Since the acquisition frame for the natural face image of any subject, at the micro-level has a noise profile that imparts a certain ruggedness to the image, we claim that this is lost for both the planar print and the prosthetic. We claim that this can be trapped with the help of a CONTIGUOUS RANDOM SCAN algorithm which preserves first order and second order pixel intensity correlation properties [16] [26]. Our work (technological GUN transfer), was inspired by a connected paper related to SPACE FILLING CURVES (SPCs), invented by Matias and Shamir (1987) [27], to compress encrypted videos. Owing to the contiguity of these curves (SPCs), correlation properties were preserved, which allowed the communication system to apply entropy codes to facilitate some form of compression after encryption.

In our work, once we deploy a simpler version of this contiguous random walk, the first order, second order and to a small extent third order correlation properties are conserved, but beyond a certain order, information tends to get mixed due to divergence of the walk. Once several scanned vectors of same image patch (or image) are generated, one can compute the first and second order differential statistics (e.g. simply the differential energies, as discussed in Section of this Chapter). This is illustrated in the following sub-section 1.1.

1.3.3 Randoms scans to trap acquisition noise

In Chapter-5, we show that while prosthetics no matter how good, have depth attributes, there is the "over-smoothing" problem, which gives away the synthetic presentation provided the acquisition noise profiling algorithm is good enough. This has been achieved via a contiguous random process [16], which traps the correlation profile in images very effectively. This proposed contiguous random walk process for acquisition noise profiling and face-feature auto-population [26], is essentially inspired by

the work from a paper regarding Space Filling Curves [SPCs] [27], which were originally designed to compress encrypted videos. These SPCs, a form of controlled shuffling, preserved the correlation profile of the video data and thus allowed the encoder to perform entropy coding even after the encryption process was completed.

When applied towards face-format profiling, 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 $d-footstepsaway$ (or walk units away). This has been illustrated in Fig. 1.13: 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 the Fig. 1.13 below. Number of distinct paths is, $N_{paths} = 4 \times 3 \times 2$ This three-point scan be represented by a directed graph linking the

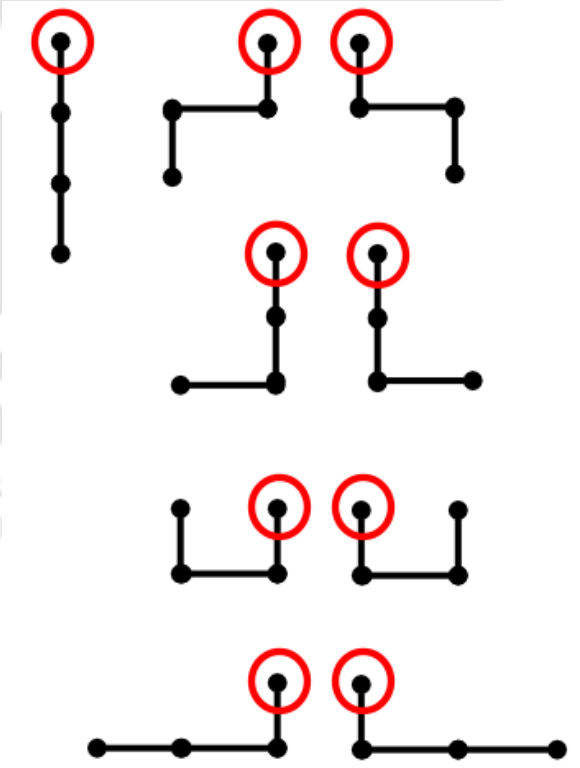

Figure 1.13: CONTIGUOUS RANDOM WALK: Destination pixel marked in RED and last-mile entry is from the bottom pixel (i.e. pixel located below the final destination pixel).

pixel intensities in a fixed destination pixel and two of pixel intensities from the past two locations: $I(P_2) \rightarrow I(P_1) \rightarrow I(T)$. T denotes the position of the destination in the image and and the past two

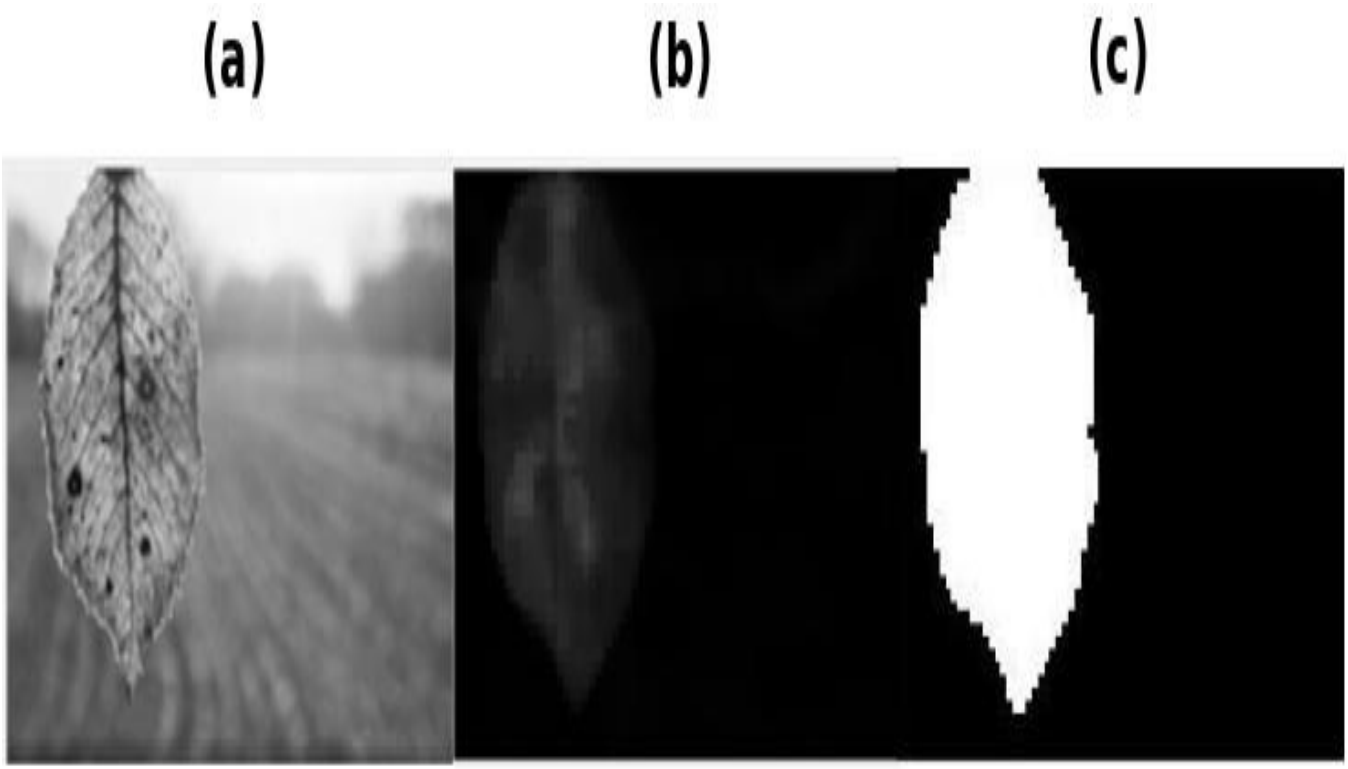

Figure 1.14: Leaf image and the corresponding differential energy profile revealing the depth map.

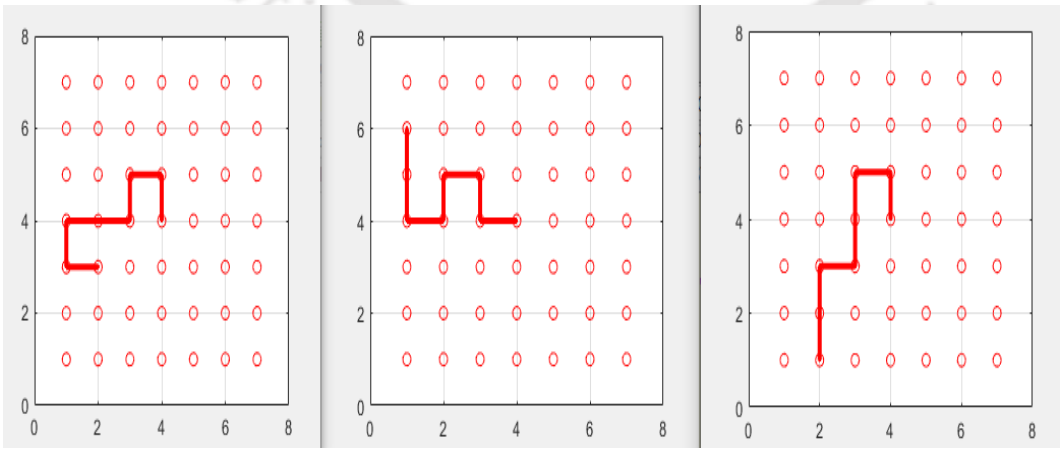

Figure 1.15: Some exemplar random walk patterns leading to the central pixel. Walk length is $d = 7$ units and the size of ensemble (or extent of auto-population) was 15-scans.

locations. The corresponding pixel intensity values are indicated by the function: $I(T)$, $I(P_1)$ and $I(P_2)$ respectively. The corresponding scanned vector is: $S = [I(P_2), I(P_1), I(T)]$.

For each entry from left, right, top or bottom, there are NINE possibilities. For one such case (out of four) where the entry is from the bottom, is illustrated in Fig. 1.13.

This three-point scan be represented by a directed graph linking the pixel intensities in a fixed destination pixel and two of pixel intensities from the past two locations: $I(P_2) \rightarrow I(P_1) \rightarrow I(T)$. T denotes the position of the destination in the image and and the past two locations. The corresponding pixel intensity values are indicated by the function: $I(T)$, $I(P_1)$ and $I(P_2)$ respectively. The corresponding scanned vector is: $S = [I(P_2), I(P_1), I(T)]$. When applied to face image regions, these scans tend to possess the following characteristics

(i) Contiguity guarantees lower order pixel intensity correlation [16] (at least up to the second or third order), beyond which information divergence takes place and the source and destination "pixel-values" become virtually independent, i.e. $I(T)$ and $I(P_{d-1})$ where $d > 3$ is the walking distance become nearly independent.

(ii) Content agnostic statistical profiling possible] By taking the first order difference of the scanned vector, S, the vector itself and a shifted version of it (by one unit to the right), we get the DIFFERENTIAL VECTOR: $DV = [d_0, d_1, d_2]$. Where, $d_0 = I(P_2)$, $d_1 = I(P_1) - I(P_2)$ and $d_2 = I(T) - I(P_1)$. Consider the energy of this differential vector averaged over the $4 \times 3 \times 3 = 36$ possible realizations, one is guaranteed to get a realistic estimate of the first order correlation pixel correlation profile. This differential energy analysis can be done at a spatially local level as well as at a global level.

$$D_{ENERGY}(ENSEMBLE) = \frac{1}{36} \sum_{P_1, P_2 \in ENSEMBLE} I(P_2)^2 + [I(P_2) - I(P_1)]^2 + [I(T) - I(P_1)]^2 \tag{1.13}$$

One application of this form of scan auto-population is towards deriving depth maps from naturally and deferentially blurred images, particularly those taken over a narrow depth of field. Shown in Fig. 1.14 is a leaf image where the entire leaf on the left hand side is in focus, but the background is blurred. The leaf has a diverse texture. However, because of this ensemble effect, the space varying blur due to the PINHOLE-camera phenomenon shown in [7] can be estimated reliably using these differential statistics. The patch size was chosen as $W \times W$ with $W = 7$ and

Table 1.4: COMPARISON OF RANDOM SCAN STATISTICS for REAL and SPOOF VERSIONS

Statistic	Natural face	Planar print or digital image (spoof)	Prosthetic (spoof)
Mean random scan differential energy (first order)	HIGH	VERY LOW	MODERATE (close to the natural version)
Mean blur estimate	MODERATE	LOW/MODERATE	LOW
Standard deviation blur estimate	HIGH	VERY LOW	LOW

the walk length leading to the central pixel of the patch was chosen as, $d = 7$ Some exemplar walk patterns are shown in Fig. 1.15. The number of instances or realizations for producing the ensemble of scans was chosen as 15. Thus, with very little data, a fairly good estimate of the depth map (or sharpness deviation) can be obtained as can seen from the quality of the result in Fig. 1.14.

1.3.4 Application of Random scans and derived statistics towards various face presentation modalities

This first and second order differential energy profile when conditioned on the nature of the face presentation format carries the following information:

(i) Clearly owing to the rugged nature of a natural face, the mean differential energy corresponding to a natural face is expected to be higher than its corresponding spoof counter-parts (the planar version and the prosthetic). The blur estimate, which is the inverse of the differential energy, should have a moderate mean for natural faces and moderate/low mean for planar spoof versions (see Table. 1.4). However the standard deviation of the blur profile should be highest for the natural face and should register low values for planar spoofing and prosthetics. Since the conditional density function related to the blur profile are distinct for different modalities (natural vs planar-spoof vs prosthetic), segregation on the basis of the blur feature is possible.

To begin with examples of multiple instances of random scans of a particular 7×7 patch are shown in Fig. 1.16. The walk length is set as $d = 3 \times W = 21$, where, $W \times W$ is the patch size. A complete scan of the patch (if small) is not required as the amount of new information captured at some point becomes irrelevant. As long as the locality or neighborhood is sufficiently mapped and represented, the partial scan can do same the trick as the full scan.

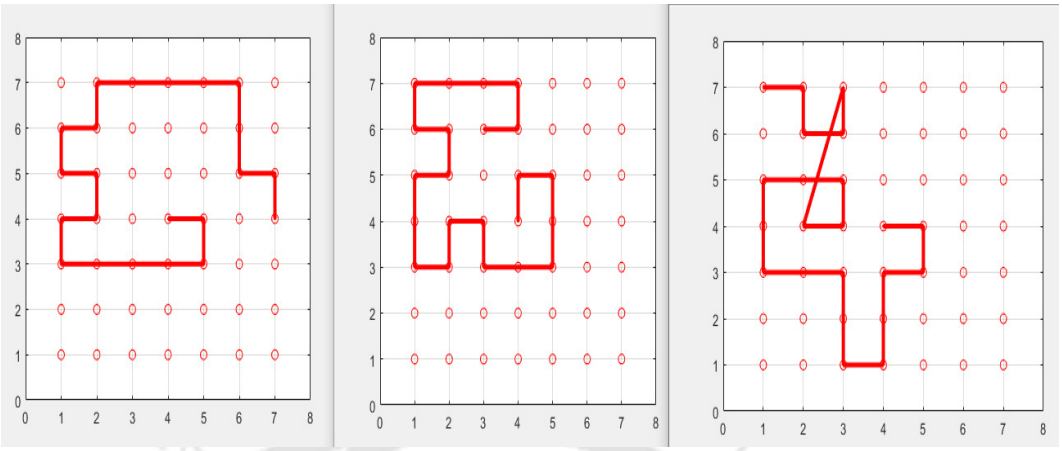

Figure 1.16: Contiguous random scan examples with $w = 7$ and and path length $d = 21$.

To bring out the comparison between the real and spoof versions, via scan statistics, consider the following three modes of presentation of a particular subject's face taken from the database [2]. The print version which was not present in the original dataset, was re-created by printing the natural face onto regular printer paper. The same scan parameter values were chosen for the analysis: $W = 7$ (patch size 7×7) and walk length from the central pixel as, $d = 3 \times W = 21$. The first order differential energy scores were generated over multiple instances ($N_S = 15$) or multiple random scans over the same patch and these scores were averaged as discussed in Section 1.1. For a particular patch P_0, the first order difference of the randomly scanned vector corresponding to instance, the first order difference of the randomly scanned vector corresponding to instance $k \in [1, 2, ...N_S]$, is represented as, $D(P_0, k)$ This has a length of $L = 3 \times W = 21 units$. And the corresponding differential energy scalar value is, $ED(P_0, k)$, The final patch statistic which gives the blur estimate is, $BS(P_0) = \frac{1}{\sum_{k=1}^{N_S}} ED(P_0, k)$. A BLUR-score HISTOGRAM IS CONSTRUCTED from the all $BS(patch)$ values taken from all the patches in the rectangular grid. The normalized blur histograms for three modalities: one real and two spoof, corresponding to the face images of the same target-subject (depicted in Fig. 1.17), are shown in Fig. 1.18. Since the roughness/sharpness profile is much better for the natural and real face, the micro-noise energy trapped via differential scan statistics registers a much higher mean as compared to the print version. Subsequently the blur diversity is greater for the natural image as compared to the spoof versions. Furthermore because of the depth homogeneity in the case of the planar print version, the variance of the blur profile is less as compared to the real version. This holds even for the prosthetic.

Figure 1.17: Presentation modes: Natural, Planar print and Prosthetic

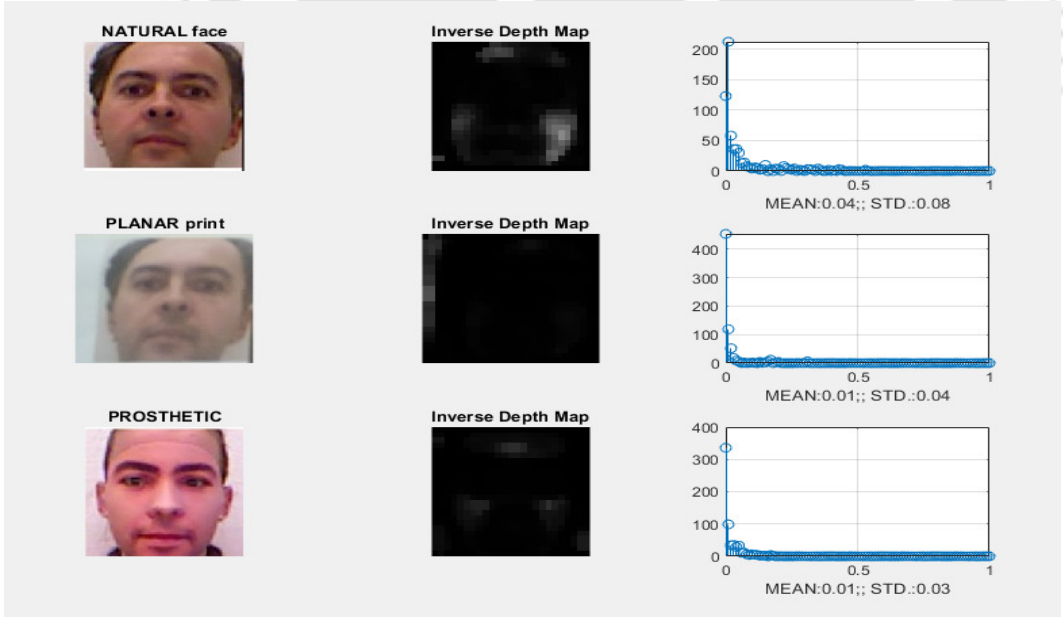

Figure 1.18: Conditional histograms of the blur scores for natural, planar-print and prosthetic. Observe the higher blur-score mean and standard deviation for the natural face as compared to the planar-print and the prosthetic.

1.3.5 Distortions due to lightning/pose in Spoof faces

Given the differential scan statistics from the auto-populated scans, what form of information do these scans and derived differential statistics capture regarding the facial-object or face presented to the camera?

These scans conserve the first order, second order and third order pixel-intensity correlation profile.

This pixel intensity correlation profile, when attached to natural face presentations in indoor illumination environments, carry information on the following fronts:

- Firstly, they carry precious information regarding the heterogeneous BLUR-variation induced by the CAMERA-LENS, in a typical PIN-HOLE setting.
- Natural contrast variation associated with the self-shadows which are produced over person's face when the light-source is positioned on one side.

While the latter linked to "contrast" may not surface (self-shadow patterns are spatially slow varying, as the surface topographies in most faces are largely smooth), the first parameter, i.e. blur change (diversity) over the entire face can be profiled using these differential scan statistics. When auto-populated and multiple STAT-vectors are produced over the same image, the model becomes much more robust and helps towards profiling natural faces primarily on grounds of BLUR-heterogeneity.

Blur-heterogeneity, we believe is largely immune to pose variations and scale changes as the depth maps would be different but the blur-diversity statistics would exhibit the same pattern as the full-frontal natural face-pose.

Since the final stats are differential scan based they are:

- Subject agnostic.
- Illumination environment and light source positioning agnostic (for the same reason these differential scan statistics may not pick up the self-shadow profiles when directly applied on the face-image).

Thus, when we cross-tested the original natural model trained on CASIA on other datasets (without performing subject level calibration), results were excellent and superior to the state of the art.

1.3.6 Generalization to unknown attacks

There are two main contributions which gel together well with some adaptation:

- Outlier detection frame which was proposed for a contrast feature oriented planar spoofing model (print-type spoofing).
- The RANDOM SCAN based differential statistics which carry precious information regarding the first and second order pixel correlation profile.

The UNION of the Outlier based frame with the RANDOM SCAN algorithm leads to a model which is useful for characterizing the NATURAL FACE SPACE ALONE. The scalar version of the Outlier model which we had proposed in 2017 has extended to a multi-dimensional setting using the 1-class SVM arrangement of Arashloo-2017.

We have demonstrated in some of the chapters that the random scan algorithm when applied locally to several image patches can be used to trap the depth and roughness profile.

Thus, what is NATURAL to a face gets defined in terms of the face surface topographical statistics (which involve depth or "depth diversity" and roughness).

So long as we restrict the differentials to first and second order, the statistics become independent of the local illumination arrangement and even pose variations.

Furthermore, and most importantly, since the walk is random, the statistic is subject-agnostic. So the counter-spoofing may not just detect the SPOOFING of a REGULAR SUBJECT, it may also detect the PROSTHETIC of an UNKNOWN PERSON.

Thus, with NATURAL SPACE CHARACTERIZATION via INTRINSIC ONE-SIDED learning with RANDOM SCAN statistics, we a SPOOF-TYPE-independent and a SUBJECT-type independent and an IMAGING-environment independent solution.

1.4 Contributions of this thesis

(i) **Outlier detection frame and deployment of Random scans for trapping acquisition noise statistics:** The thesis has two primary contributions encompassing the demand for universality of the solution under different spoofing and acquisition environments and at same time providing a robust spoof-detection procedure in a subject-agnostic frame, using a virtually contiguous random scan algorithm: (i) Design and development of an outlier detection algorithm by first characterizing the natural face space as the inlier space and picking up spoof-presentations as deviations; (ii) Deployment of a virtually contiguous random scan (inspired by the Space filling curves (SPCs) used for encrypting compressed videos), towards an entrapment of the

pixel-correlation profile in natural faces across subjects, making the feature extraction subject-agnostic and content-agnostic.

(ii) **Contrast reductionist Life trails for trapping self-shadows in images:** part from this a secondary yet important contribution is the deployment of an iterative functional mapping for reducing a particular face image to a zero contrast one. This contrast reductionist so called life trail (Chapter 6), carries precious information in the form of self-shadows. This self-shadow profile can be extracted by processing the first few images within this life-trail sequence. Since planar prints have reduced self-shadows, this method when used in a client specific mode gives excellent accuracies towards planar print detection.

1.5 Organization of Thesis Chapters

Chapter 2, discusses one of our primary contributions, which involves the rank-ordering based OUTLIER DETECTION FRAME with CONTRAST as the base feature. This is the place we introduce our case for NATURAL SPACE CHARACTERIZATION and illustrate how even with a simple statistic such as contrast, if the base feature is robust and discriminative, a DATA-oriented MODEL can be built for the natural face class alone. We also discuss briefly, with the exponential Gamma-law, as to how printed version of a person's face (denoted by the intensity profile I_{pp} exhibits a lower contrast as compared to natural face (denoted by the intensity profile I_{ORIG}.

$$I_{pp} = (I_{ORIG})^{\gamma} \tag{1.14}$$

Where, the intensity profiles are assumed to be normalized over the range zero to one and the parameter $\gamma > 1$. Finally, the contrast score or statistic is generated in conjunction with Weber's law [28] [WEBER-contrast or relative change], which talks about relative change in intensity about the mean local intensity. The contrast score presented here is the inverse of the Weber's equation: $CON(patch) = \frac{\mu}{\sigma}$. Where, μ is the mean patch intensity and σ^2 is the variance in intensity over the patch.

Chapter 3, extends the model-specific work connected with the specular-phenomenon spotted in most printed photos albeit with a natural space angle. While the learning here is not truly one-sided, the EIGENSPACE model building is natural space directed. The eigen-space projection vectors of the natural and planar spoof samples over the NATURAL EIGEN-SPACE, form the 2-class model.

Chapter 4, llustrates to begin with as to why BLUR-analysis in images is essential to detect planar spoofing using a very simple PINHOLE-camera model. Later in Chapter 5 (Random scans) we provide an argument as to why prosthetics tend to exhibit an over-smoothing as far as the surface topography is concerned and thus have a lower blur-diversity as compared to natural images. A sharpness-based analysis, using a texture filter called GSM (Gradient Significance Map) [29] gives away the natural face versions versus the printed presentations. Since, the texture-based sharpness statistic, turns out to be a weak feature, we had to deploy two-sided training to build a robust model.

Chapter 5 contains our PRIMARY contribution which seeks to attack all three spoofing modalities: Print spoofing, Digital image spoofing and Prosthetics using RANDOM SCANS. It surmises here that the best solution is that which ignores the content and focuses on the format of the data. The acquisition format here relates to the imaged face-topography and the subsequent micro-noise which is retained in the image, can be captured using contiguous random scans, which preserve the first, second and third order pixel intensity correlation statistics. The inspiration for this work came from another application where such scans in form of a more sophisticated SPACE FILLING CURVE [27], were used for compressing encrypted/shuffled videos. Random scans in our work, now applied to trap acquisition noise can not only be used for depth-profiling (which was demonstrated in Chapter 1, Section 1.3), but can also be used to AUTO-POPULATE the data very effectively. When used in conjunction with natural space characterization and model building, excellent spoof-detection accuracies were obtained.

Chapter 6 contains our secondary, yet highly significant contribution connected with a novel form of analysis related to iterated function systems. We observed that when a face image was subjected to a logistic map of the form, $I_n(x,y) = 2I_{n-1}(x,y)[1 - I_{n-1}(x,y)]$ Where, $I_0(x,y) = \Im$ (the original image/template presented) and $n > 1$ is the iteration number, a sequence of contrast reductionist images are produced and eventually the original image is reduced to a zero contrast image (or a fixed point). In this contrast reductionistic sequence, the first, first order difference between original image and its next image in sequence contains maximal self-shadow information. This self-shadow information can be used to segregate planar-print versions from natural face images. With a suitable calibration procedure, it is possible for one to arrive at the right operating point for a given training set with a handful of samples. A 2-class, client specific frame was chosen as the model, with self-shadow statistics forming the base feature. Excellent accuracies were obtained with respect to print-spoof

detection.

Chapter 7 covers the main highlights of this thesis and opens up work for the future.

2

Image Quality Assessment Using Contrast score and outlier detection

Contents

Objective *Planar spoofing is a well researched problem, wherein a high quality planar photograph can be replayed in front of a still camera as a substitute for another individual's face. Most modern day face recognition systems can be fooled by this process, as the perceptual information contained in a photo-of-a-photo, is virtually the same as that of a natural photograph of an individual. Current solutions attempt to detect this form of planar-spoofing through an extrinsic training process wherein both planar samples as well as regular photos are included as separate training sets. To avoid this form of explicit discriminant model-learning, we propose a single class training procedure for establishing and quantifying the quality of natural photographs taken under different lighting conditions, in terms of their CONTRAST PROFILE. Once this distribution is learnt, a suitable threshold is set based on the mean and standard deviation to pick up outliers. In this chapter, we show that with just single poses of subjects, it is possible to achieve a low Equal Error Rate (EER) of 21.56% on the CASIA dataset and a rate of 8.57% upon cross-validation with a trimmed and shortened version of the MSU dataset.*

2.1 Introduction

While face recognition is a viable and active research problem, one main assumption made in these systems, is that the face presented to the camera module and the subsequent recognition-engine, is an authentic and real one. Given the widespread design and use of art models, prosthetics and high-resolution printing, there is a concern whether face presented to the camera is really authentic. If a person X claims to be Y, the face detection system is forced to ask the question, *"Is this person who claims to be Y, really Y?"* or *"Is he another individual who is impersonating Y?"*. In the case of still image captures, there are numerous ways in which a person can emulate the face of another individual:

- By covering his own face and presenting a high resolution printed photograph of the person, who is to be impersonated as a mask [30] [31] [32].
- By wearing a prosthetic mask of the face of the individual who is to be impersonated [33].

This chapter, restricts the class of attacks to 2D printed photograph based operations, wherein spoofing is attempted by presenting either warped or un-warped versions of a photo-of-a-photo of some subject. Early approaches have tried to derive simple features from facial image profiles which mimic or capture

the extent of diversity in photographs. It was observed in Jiangwei et al. [34], that the Fourier spectrum of a spoofed image is much more compact as compared to that of a natural image. In other words, a natural image will have more significant high frequency components as compared to that of a spoofed image. In one of their segments, they computed a high frequency descriptor (HFD) to quantify the fraction of energy in the higher frequency bands. This was compared against a threshold for preliminary classification. There are several issues with this implementation:

- Images including faces are non-stationary in nature, viz. two different realizations of the same subject photo, due to a pose variation or an illumination setting change, will induce very different Fourier spectral profiles. It is therefore in-appropriate and difficult to define a so called cut-off frequency (at the cusp of what can be termed low or high frequency). It boils down to asking the question: "How high is high enough?" Arguably, if a bandwidth/cut-off frequency threshold is fixed as a certain number, this number will not account for local environmental changes, pose variations and also skin color and texture changes. By holding the premise regarding spectral richness of natural photographs, does not really help unless there is a more concrete and robust measure for quantifying this form of richness.

- Furthermore, they used a fixed threshold for the High Frequency Descriptor (HFD) for detecting spoofed images. The main issue with this is that this threshold assumes a controlled setting in which the experimentation is carried out and also assumes that the faces are somewhat cropped and registered. No cross validation with other databases was done.

- When faces are presented with elaborate backgrounds containing objects such as room furniture and curtains, the diversity of the image increases considerably, making it pointless to use the image Fourier spectrum as a global feature.

Alternative approaches, have been devised to deploy features to quantify the clarity of edges, such as Local Binary Patterns (LBPs) [35] and Difference of Gaussians (DoGs) [31] to train a classifier for segregating real faces from spoofed ones. However, these approaches are forced to learn the conditional distributions for both the classes, through significant training, before forming the decision boundary. This is purely based on the assumption that there is adequate representation for both the classes (both natural as well as spoofed). While it is reasonable to assume that the former (viz. a set of natural photographs of subjects) are available, it is highly improbable to find widespread samples from the

latter class (i.e. the spoofed one).

Attempts have been made to arrive at a heuristic distortion model for image spoofing to segregate natural images from those of photos of photos. In recent work by Wen et al. [32], the intensity profile of a spoofed image was expressed as a function of the intensity profile of the parent natural image. If $I_p(\bar{x})$ is the natural image, the intensity profile of the photo of this natural photograph was expressed as of a diffusion component and a specular reflection component,

$$I_{pp}(\bar{x}) = f\left[I_p(\bar{x})\right] + E(\bar{x}) \tag{2.1}$$

The diffusion component carries information regarding the image content, but the specular component $E(\bar{x})$, is largely a function of the light source orientation relative to the object and camera and thus serves as pure noise. This diffusion component can be further decomposed as histogram operation over a blurred version of the parent image. The premise for this was that in case of a photo of a photo, there is expected to be a degradation in contrast. This evidence supports the argument presented in our chapter. In our independent work, it present a much simpler model for contrast degradation based on power laws. The main disadvantage with the approach of Wen. et al. [32], is the lack of simplicity towards the detection and classification process, as they tend to accumulate texture, color and intensity based local and global features for training their classifier.

There were many more issues with this implementation:

- What works for one spoofing modality may not work for another and might in fact impede the reliability of spoof-detection. Some features may not even be present in a particular spoofing mode. Examples for these include:
  - Contrast as a feature is not a discriminating parameter with respect to planar digital image spoofing but is indeed a valuable feature with respect to print based spoofing.
  - Quality in general of digital planar (tablet-based) images is better as compared to print versions, hence may of the quality linked features will not work on planar digital spoofing.
- The most important issue with this arrangement was paradigm itself, wherein features were gathered assuming some form of facial registration (in a similar manner as one would accumulate features to perform a face authentication).

Both these problems have been addressed in the proposed work in different chapters:

– In Chapter-2 (this chapter), we attack problem-1 by making the analysis model-specific with suitable justifications. Furthermore it also recognize natural images have a high contrast. Thus instead of forming a model for the spoof-segment based on the contrast statistic, it turn inwards and characterize the natural face space based on the same contrast parameter. This is achieved via an Outlier detection algorithm [3] discussed towards the end of Chapter-2.

– The second problem, which is more intricate, is attacked in Chapter-5 (our main contribution), via a random scan process to content independent yet correlated measurements (several sets of those from the same image region) to trap the natural depth profile, coupled with micro-sharpness details present in natural images [26] [16]. We show that this works well against both digital and print planar spoofing and also against prosthetics (which can be detected as mentioned in Chapter-1 due to the over-smoothing problem).

It was observed in Galbally et al. [23] that planar print images in general suffer from several quality degradation. Thus by estimating the quality of the images in a blind fashion it could be possible to separate natural face images from spoofed version.

Since full quality assessment is not possible without image registration, the secondary image is derived by filtering the parent image with a Gaussian kernel. Their premise was that when the edges in an image are enhanced (or the texture is enhanced) by taking a differential between the parent and a blurred version it, the degradation in quality (or richness) is much more in the case of a natural photograph as compared to that for a photo of a photograph. A series of image quality measures were used to form the feature sets for classification. Finally a two-class classifier was trained using both the original and spoofed data-sets and tested on two different databases.

We draw a distinction of the proposed approach from that of Galbally et al. [23] and Wen et al. [32] on the following fronts:

- The basic idea is to examine and qualify an image based on its "clarity" or "discernibility" of the face captured. Without focussing on the edge profile or the richness in texture, we adopt an indirect indication in terms of the image "contrast". Contrast here is defined as the extent of clarity perceived in a given image. This "clarity" can be quantified and extracted from any given image WITHOUT ANY FORM of reference. Hence our approach virtually becomes a blind image quality assessment procedure. We achieve this by an extremely simple metric called

as the contrast score which is simply the ratio of the mean of an image (or image block) to its standard deviation. In our opinion, this represents the extent of fuzziness or contrast in a given scene or image. A cumulation of these scores over the entire image quantifies the extent of clarity. A score close to 0 indicates high clarity, while a large score indicates extreme fuzzyness and low contrast.

$$CON_{score} = \mu_{block}/\sigma_{block} \tag{2.2}$$

In a natural photograph, there is a direct reflection from the objects in the scene which induces a diversified intensity profile. When the same scene abstraction is presented as a photograph to the camera, a fraction of the reflected light is absorbed in the photograph itself based on the patch-intensity profile presented by the photograph to the light source. The effect is non-linear in the intensity space: viz. Gray regions become darker and Lighter regions drift towards a Gray shade. Thus, the dynamic range of intensity values in the image of a photograph drops in comparison with a natural photo. Subsequently SPOOFED images register larger scores as compared to NATURAL image because of a reduction in contrast.

- About the time our work connected with spoof-linked outlier detection was published based on rank-ordering and contrast, there was an independent body of work connected with the same counter-spoofing paradigm by [4], [3], but here the features used were mixed bag based and quality related and a higher dimensional 1-class SVM was deployed for classification.

  The database against which the query is compared comprises of only natural photographs and the classification is done by collectively ordering and ranking the query "contrast-score" along with the natural database images. If the query score ends up in the tail of this ordered set, it is treated as an OUTLIER and then classified as a SPOOFED image. This outlier boundary is TUNED using the CASIA database and then cross-validated using a trimmed version of the MSU-dataset. Since there is no elaborate training procedure, complexity associated with outlier detection is minimal.

The rest of the chapter is organized as follows: In Section. 2.2, we devise a metric for quantifying contrast in images and provide substantial proof of this new but simple statistic towards mimicking human definition of clarity (or quality). We then set up this statistic for segregating natural faces from photos of faces purely based on an outlier detection procedure discussed in Section. 2.3. Finally

we cross-validate the results with another dataset in Section. 2.4.

2.2 Quantifying Contrast in Images

For a specific gray-scale image block or image, contrast may be defined as the average change in intensity value over the entire block with respect to the mean intensity value. For an $N \times N$ image block, comprising of intensity levels $X(i,j) : i,j \in \{0,1,2,...,N-1\}$ with $X(i,j) \in 0,1,...,255$, this contrast can be quantified as,

$$CON_X = \mu_X / \sigma_X \tag{2.3}$$

where,

$$\mu_X = \frac{1}{N^2} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} X(i,j) \tag{2.4}$$

$$\sigma_X = \sqrt{\frac{1}{N^2} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} [X(i,j) - \mu_X]^2} \tag{2.5}$$

Lower the contrast score registered, larger is the deviation per mean value and subsequently richer is the image. Contrast definition for a gray-scale image may seem simple, however, in the color space for a color image, this definition can become slightly more complex. Quantifying contrast in a color rich image by understanding the color spaces, is an extremely difficult problem connected with the way we perceive relative color settings in a complex natural image. This is beyond the scope of this chapter. However, it is possible to generate an abstraction of the color rich image in terms of its luminance profile and then define contrast for this gray scale setting. This is the path we have adopted in this chapter. It was observed by first examining and then analyzing natural photographs of faces and photos of natural facial-photographs, that natural photos appeared to exhibit a higher contrast as compared to their synthetic counterparts (photo of photos). We propose the following theory:

Theory for the RELATIVE CONTRAST SHIFT (REAL vs SPOOF): *When there is a direct illumination of a specific object exhibiting a certain curvature in 3-D space, depending on the position of the light source relative to the object, the intensity of the reflected light that is captured by the camera would vary as a function of this relative positioning and object "shape". This in turn would influence the way the image of the object appears. When a photograph of the above image itself (of that of the object), is presented to the light source, this becomes a case of SECONDARY REFLECTION, wherein, there is a reduction in the intensity of the captured image commensurate to the extent of "darkness"*

in a specific patch. A darker patch tends to reflect lesser light as compared to a lighter patch, hence the same scene appears a little fuzzier as compared to the natural picture of the object. Therefore, the contrast affiliated with natural images of objects is likely to be on the higher side as compared to the images of natural images (on a majority). We therefore claim that the intensity profile of a natural photograph I_p of a face can be associated with that of a photo of the same photograph I_{pp} as,

$$I_{pp} = (I_p)^{\gamma} \tag{2.6}$$

where, I_p is normalized in the range $[0, 1]$ representing the natural photograph, where $\gamma > 1$ is a intensity shrinkage parameter and I_{pp} is the associated intensity value of the "photo of this photograph". On an iterative setting, the r^{th} photo of a photo can be linked in terms of intensity with the original photograph as,

$$I_{pp(r)} = \left(I_{pp(r-1)}\right)^{\gamma} = (I_p)^{(\gamma)^r} \tag{2.7}$$

The waning of the contrast setting in the r^{th} photo-of a photo is very steep. This γ parameter is unique to the camera setting, local illumination profile, the texture of the printed paper etc. This theory was tested on real and spoof photographs from the CASIA database [31]. The contrast scores were computed for 10 sample images from each set, in Fig. 2.1. The contrast scores were observed to be distinctly lower for real images as compared to spoof images, indicative that the two classes are clearly separable. The γ values were estimated by taking a ratio of the mean intensity of the spoofed image to that of the natural photograph of the same subject (both of which were resized to 64×64). Note that the photo of the natural photo was not registered with the natural photograph of the subject. This induces deviations in the γ values. The gamma values for the 10 subjects were: 1.5771, 1.7407, 1.4662, 0.9729, 0.5463, 1.3655, 1.5267, 0.9885, 1.6045 and 1.6621 respectively, a majority of them significantly larger than one.

2.3 ANTI-SPOOFING by OUTLIER TUNING with the CASIA Dataset

Unlike existing approaches [31], wherein both real and spoofed images were used for training and testing, here we deploy ONLY real natural photographs for the TUNING process. Once the space for natural photographs is understood, any query image which does not fall within this space will be treated as an outlier. The process is as follows:

Figure 2.1: Two classes of images: Set-1: Photos of natural photographs of faces; Set-2: Natural photographs of faces; Observe the distinct separation between the two classes. Natural photographs tend to register smaller contrast scores as compared to photos of natural photographs. The γ values for the 10 subjects were: 1.5771, 1.7407, 1.4662, 0.9729, 0.5463, 1.3655, 1.5267, 0.9885, 1.6045 and 1.6621 respectively.

- Contrast scores are computed for all N_D natural images stored in the database. Let these scores be $Scores_D = \{CON_1, CON_2, ..., CON_{N_D}\}$.

- These scores are then sorted according to increasing magnitude (or in ascending order): $Scores_{sorted(D)} = \{CON_{s(1)}, CON_{s(2)}, ..., CON_{s(N_D)}\}$, with,

$$CON_{s(i)} \leq CON_{s(i+1)} \tag{2.8}$$

- When a query or test image is presented, its contrast score is computed as, CON_Q. This score is appended with the set $Scores_{sorted(D)}$ from the database and is then re-sorted in the ascending order. The location of the query score in this new sorted list is noted.

- If the query score is ranked in the bottom $\alpha\%$, the query image is declared as an outlier (or a spoofed image). Otherwise, it is declared as a natural photograph. Based on experimentation with the CASIA database we have chosen this OUTLIER threshold parameter as $\alpha = 25\%$. The tail of any conditional density function is always unreliable. When we form a distribution of contrast scores for natural images, the distribution tails to zero for larger contrast values. If the threshold is shifted closer to the MEDIAN of the DATABASE scores, this would reject most of the outliers, but then the inlier space would be constrained leading to an increase in the number

of false rejections.

$$S_{MED} = MEDIAN(Scores_D) \tag{2.9}$$

One the other hand if the threshold is moved away from the mean, then we would enter the tail of the distribution of natural contrast scores, inviting interference from the tail of the other conditional corresponding to the outliers. This is likely to increase the false acceptance rate. Creating a perfect separation between the natural and spoof scores is an extremely difficult task as this would depend on the quality and stability of the feature (in our case it happens to be a contrast score). The rank threshold is set as,

$$R_{TH} = (N_D + 1) - \lfloor \alpha \times (N_D + 1) \rfloor \tag{2.10}$$

where, $\lfloor . \rfloor$ corresponds to the FLOOR function and when the query value upon sorting with the rest of the database, acquires a RANK R_Q,

$$R_Q > R_{TH} \tag{2.11}$$

It is classified as an OUTLIER or a SPOOFED IMAGE.

An example of this outlier detection process is given in Fig. 2.2 with $N_D = 10$ and $\alpha = 25\%$. The rank threshold for this setting is, $R_{TH} = 9$. The query contrast value is denoted by a RED DOT amidst the sorted database values represented by BLUE DOTS. In the first two cases Fig. 2.2(b,c) the rank of the query is $R_Q = 8$ and $R_Q = 9$ respectively. Hence the query will be treated as an inlier (or a natural photograph). On the other hand in Fig. 2.2(d,e) [cases 3 and 4], the rank of the query is $R_Q = 10$ and $R_Q = 12$ respectively. Hence this query image will be treated as an outlier (or a spoofer image). The database was arranged as follows:

- Number of subjects: $N_{SUB} = 15$.
- Number of natural illumination and subtle pose variations/expression changes of the same subject: $N_{VAR-NAT} = 30$ (natural photographs/subject).
- Number of spoofed and warped photo of photos per subject: $N_{VAR-SPOOF} = 30$.
- Number of natural photographs of subjects used for TUNING: $N_D = 10$.
- Total number of images for testing: $N_{TOT} = 15 \times 30 \times 2 = 900$.

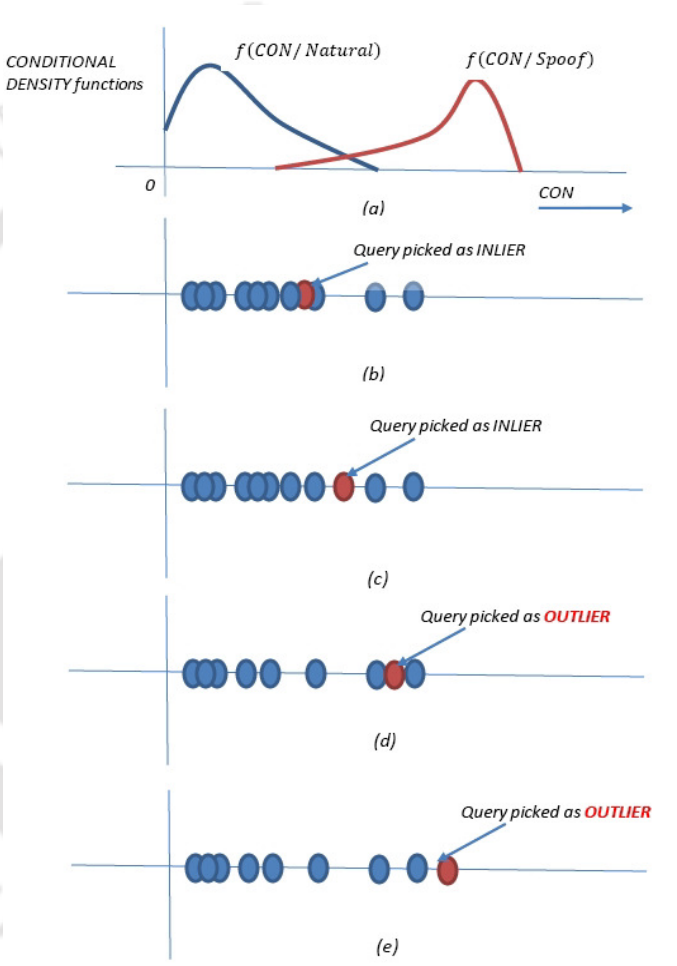

Figure 2.2: An example for the outlier detection procedure. Number of images in the database: $N_D = 10$. Outlier rank threshold $= \alpha = 25\%$ or $R_{TH} = (N_D + 1) - \lfloor \alpha \times (N_D + 1) \rfloor$; $R_{TH} = 9$. Rank of the query larger than this rank-threshold would result in this being classified as a SPOOFED IMAGE. (a) Typical conditionals for the contrast scores (natural and spoofed images); (b) Rank of the query is $R_Q = 8$, which indicates that it will be picked up as an INLIER; (c) $R_Q = 9$ borderline and declared as an inlier; (d) $R_Q = 10$, borderline and declared as an outlier; (e) Clearly declared as an outlier.

Fig. 2.3 shows the approximated conditional density functions for the two classes by accumulating the contrast scores for real and spoofed images. The figure shows that natural photographs tend to register lower contrast scores as compared to spoofed images. The class separation is distinct and provides adequate experimental proof that segregation of real from spoofed images is certainly possible on grounds of contrast. To compute the efficiency of the system and determine the quality of the

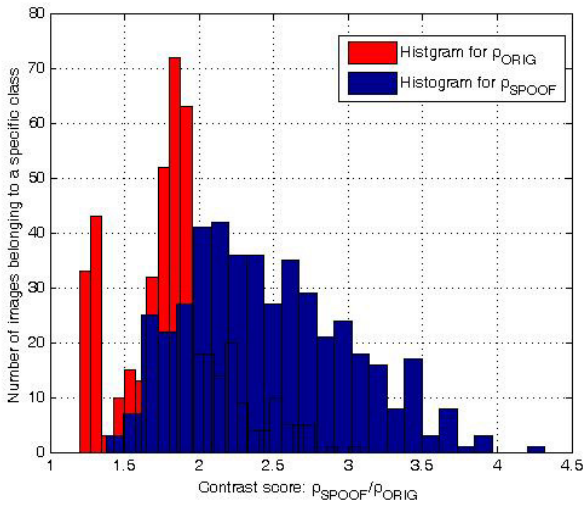

Figure 2.3: Histogram of contrast scores for natural photos of faces and photos of natural photos.

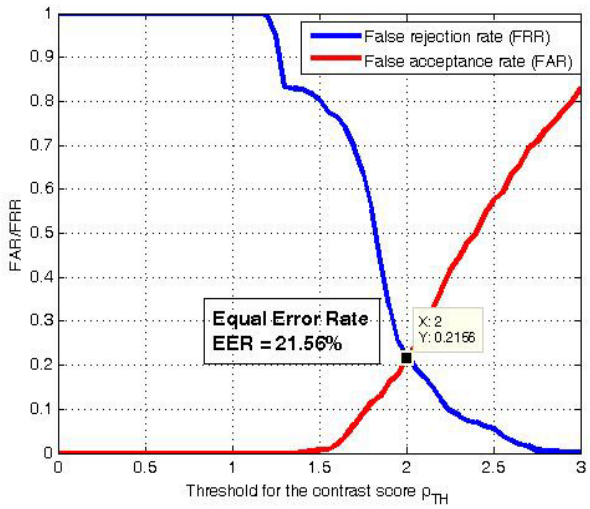

Figure 2.4: System operating point with optimal error rate obtained from the intersection of the FAR and FRR curves. The Equal Error rate (EER) was found to be 21.56% with the corresponding threshold being $\rho_{TH} = 2$.

CONTRAST STATISTIC in question, all images were first converted to gray-scale and then resized to 64×64 before computing the contrast statistic. The threshold for classification was varied from $CON_{TH} = 0$ all the way up to $CON_{TH} = 3$. If $\beta_{NAT-MIS}$ represents that fraction of authentic and

natural images which are misclassified as outliers and $\beta_{SPOOF-MIS}$ represents the fraction of spoofed image misclassified as INLIERS (or natural images), the false rejection and false acceptance rates can be expressed as:

$$FRR(\%) \quad = \quad \beta_{NAT-MIS} \times 100 = \frac{N_{NAT-MIS}}{N_{NAT}} \times 100$$

$$FAR(\%) \quad = \quad \beta_{SPOOF-MIS} \times 100 = \frac{N_{SPOOF-MIS}}{N_{SPOOF}} \times 100$$

The intersection of these two curves is the Equal error rate (EER), which determines the overall efficiency of the system. Fig. 2.4 shows the FAR and FRR curves over a range of thresholds, with the intersection point being $EER = 21.56\%$ (operating threshold $CON_{TH} = 2$).

2.4 CROSS VALIDATION

While the first CASIA database was used to TUNE the parameters and set the rank threshold to $\alpha = 0.25$, the second dataset (trimmed version of the MSU spoof database [32]) was used to test the algorithm on the same grounds. The second dataset is distinctly different from the first one on many fronts:

- The faces are not cropped, opening up the background region containing objects such as lab furniture, screens, curtains etc. This makes the contrast estimation process over the entire image a little tricky.
- The scales associated with the face frames are non-uniform. Some faces have been captured at a distance, while others have been acquired at close proximity.
- Some of the subjects exhibit pose variations: a slight tilt or head rotation with respect to the camera. Most of the photos of photos correspond to full frontal poses. This is not the case for the regular and direct photographs, since these two sets were formed at different times. The natural photographs are therefore not registered with respect to the photos of natural photographs.

All these variations can be witnessed in Fig. 2.5. The scores of the spoofed images corresponding to the same subjects are in Fig. 2.6. Note that the contrast scores of the spoofed images are much higher than those for the natural images.

To account for the background variations, the mean contrast score for a particular image was computed as follows:

- All the images were converted to gray-scale and then re-sized to 256×256 pixels.
- This resized image was split into smaller blocks of size 32×32.
- Localized contrast scores were computed for each of these blocks and then averaged over the entire image.
- To account for constant or very slowly varying patches, the following policy was adopted: The inverse ratio of patch standard deviation to the mean patch intensity was computed and compared against a threshold of $\delta = 0.1$. This was tantamount to comparing the inverse of the contrast-score for a patch with the threshold.

$$\frac{1}{CON_{p,q}} > \delta \tag{2.12}$$

where, $p, q \in 1, 2, ...N_B$ with N_B being the number of blocks. Only blocks which satisfied the condition, were included in the cumulative score computational process. Since the contrast scores decrease as the patches become richer in texture and color, it is only natural that its inverse will assume a larger number as the patch becomes contrast rich. This weeds out virtually constant patches.

If N_D is the number of reference natural images from the database used for detecting outliers, the FAR and FRR rates for 35 original and 35 spoofed images as a function of N_D is shown in Table. 2.1. Table. 2.1, provides rather interesting results. While it clearly indicates that the mis-classification

Table 2.1: FAR and FRR rates for different database sizes.

N_D	FRR (%)	FAR (%)
5	2.9	11.4
15	8.5	8.5
25	8.5	8.5
35	11.4	8.5

rates are low, the trend with an increase in the number of original database images N_D is rather interesting. While one may expect the FAR and FRR rates to drop further with an increase in N_D, the opposite takes place. The FRR rate increases which implies that the fraction of natural images/samples entering the tail of the conditional distribution associated with the contrast scores

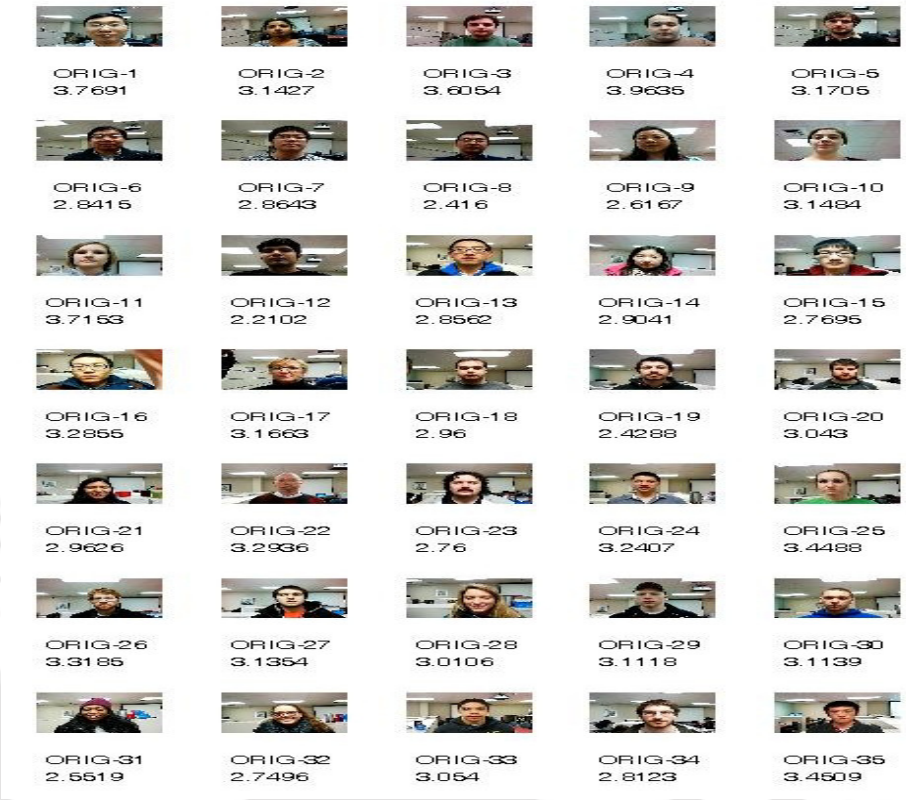

Figure 2.5: Contrast scores for selective original images from the MSU database.

of natural images $f(C_{CON}/Natural)$ increases. The real reason for this is the enhancement in the diversity of the database because of extensive pose variations, scale changes and background changes. This increase in diversity pushes some of the contrast scores to the tail of the conditional density function. Because of this one sided tuning or database construction with respect to natural images the same trend is not observed in the case of the FAR scores. They tend to drop with an increase in N_D.

2.5 Conclusion

In this chapter, we have proposed a virtually blind image quality assessment procedure for detecting spoofed images based on their contrast profile. The reference images constitute natural images of subjects from the database. The contrast scores from these natural reference images are compared with the contrast score derived from the query image to detect OUTLIERS. The base statistic for quantifying contrast was as simple as a ratio of the mean value over an image/image block to the standard deviation of the image. This base statistic was deployed in face anti-spoofing, through a

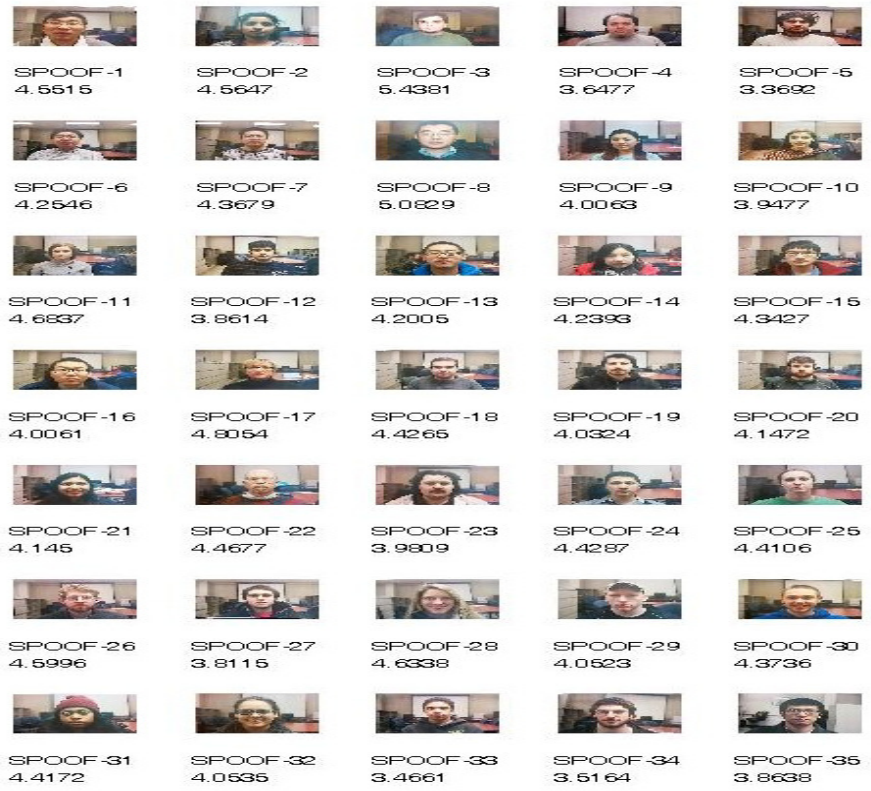

Figure 2.6: Contrast scores for SPOOFED images of the same subjects from the MSU database.

one-class in-house TUNING procedure for defining only the inlier population comprising of only these reference images from the database. The EER for the CASIA database was 21.56%. Cross-validation was done for a much shorter version of the MSU database and results were found to be promising with an FRR of 8.5% and FAR of 8.5% corresponding to a database size of 15 images.

3

Proposed pipeline for effective use of specular features

Contents

Objective

The need for facial anti-spoofing has emerged to counter the usage of facial-prosthetics and other forms of spoofing at un-manned surveillance stations. While some part of literature recognizes the difference in texture associated with a prosthetic in comparison with a genuine face, the solutions presented are largely prosthetic-model specific and rely on two sided calibration and training. In this paper we focus on the specular component associated with genuine faces and claim that on account of the natural depth variation, its feature diversity is expected to be much larger as compared to prosthetics or even printed photo impersonations. In our work concerning one sided calibration, we first characterize the specular feature space corresponding to genuine images and learn the projections of genuine and spoof data onto this basis. The trained SVM corresponding to genuine projections, 3D mask projections and printed photo projections is then used as an anti-spoofing model for detecting impersonations.

3.1 Introduction

The first thing to recognize in this problem space is that while the class of genuine facial images is available at our disposal, the spoof-class is likely to have diverse modalities and formats. In a typical impersonation attack, some individual (either a legitimate part of an organization or an outsider), passes on as another individual who is from the same organization. Because of this genuine concern, that one legitimate individual should not pass on as another genuine individual, this anti-spoofing problem is an *Identity Independent Detection problem* [7]. In this frame the anti-spoofing algorithm does not concern itself with the actual biometric feature associated with the individual, but rather focusses on a layer which can be be used to establish the genuineness of the feature. In many ways, this anti-spoofing problem draws several parallels from ideas connected with blind image quality assessment algorithms [36] [3]. In the present literature survey, it restricts our analysis and scope to static facial-images as opposed to videos.

Much of the earlier was driven towards the development of spoof-model specific anti-spoofing solutions, wherein images were captured on both fronts: genuine/natural and also with masks/prosthetics. These images were used to train suitable classifiers, with the objective of minimizing the mis-classification probability [2] [8]. Some papers [8], have assumed even prior knowledge of the imposter set, to account for mask contortions coming from the mismatch in facial profiles, between the imposter and the person being impersonated.

When the spoof-model is regularized and it becomes possible to recreate the spoof-settings to a certain degree of precision, then the corresponding anti-spoofing solutions can be optimized and diversified. This is largely true for spoofing through printed photographs [2] [9], wherein the line of anti-spoofing solutions encompass physical and imaging model based solutions involving physical

constraints. In Gao et al. [37], it was surmised that photographs of planar images tend to have a more homogeneous specular illumination component as compared to natural faces.

The same printed photograph was treated as spatial re-sampling problem by Garcia et al. [19], leading to the introduction of a few artifacts in the image domain, termed as Moire patterns, which can be treated as noise. Since this noise is wideband, its power can be estimated by filtering the query images and then subtracting the original image from the filtered image. The drawback of this frame is that this Moire model assumes a zooming in effect with respect to the printed photo, otherwise the model would change to an aliasing one and the same analysis cannot be repeated.

In Karthik et al. [7], the recapturing phenomenon has been shown physically to be imperfect registration process if the original object distance is unknown. This results in a cumulative blur, which supersedes the natural blur resulting from depth variation, lowering the mean patch sharpness profile in the printed photo. If the original settings can be recreated, then there would be no cumulative blur, but this is unlikely. This frame works for both zooming in and zooming out operations while recapturing the photo.

In KIM et al. [22], it was observed that planar print presentations lack depth, hence, when a real natural facial presentation is compared with a print-version, there will be a difference in the degree of sharpness and in particular the sharpness diversity seen in the images trapped in these two presentation modes. By ensuring a narrow depth of field during still image photography it was possible to generate differential sharpness profiles for natural images and print spoof versions.

Further more with a planar spoofing model, there is a lack of depth information, which is normally present in natural scenes. Light field cameras [38], tend to accumulate information from different angles and formulate a depth map of the scene or the facial frame. This aspect has been used to detect planar spoofing in Ji et al. [38].

We position our work as a model specific solution which hinges on certain assumptions or physical constraints. Since 3D masks and printed photos are largely smooth in nature, they tend to have a larger specular component, albeit more homogeneously distributed in the case of the printed photo [37]. However, in the case of natural facial photos, the specular component is expected to be more diverse, mainly because of the depth variations and the self shadowing effects. Two printed photos belonging to two different individuals are expected to show some coherence in terms of their specular profiles, however the specular profiles from two natural faces from two different individuals, are unlikely to

exhibit much similarity, because of the depth variation associated with the facial profiles. Since the specular component distribution is a function of the surface geometry, the association between two natural photos is expected to be more complex diverse in relation to the association between a natural photograph and the printed photo or between two printed photos. This philosophy has led to the following contributions:

- One sided characterization of the specular feature eigenspace corresponding to natural photos.
- Learning of the conditional density functions associated with the specular projections of natural photos onto natural photos and projections of spoofed images onto natural photos. Using a Maximal Likelihood inferencing frame by building an SVM to classify the query images into natural and spoof classes.
- Of less significance is the application of specular component isolation algorithm by Candes et al. [39] to extract features from both natural and spoofed faces.

3.1.1 Novelty of the Eigen space features

Firstly, it states this this form of specular analysis is likely to work only if a PRINTED VERSION of a face is re-imaged. The degree of specularity, would also depend on the nature of the paper (glossy or non-glossy) and position/intensity of the light source in relation to the camera.

The initial pre-processing segment concerned with the selection and extraction/separation of specular features from a particular target image is based on a paper from literature connected with profiling of surveillance videos based on specular parameters [40] [2]:

Our contribution is not in the feature selection segment, but is from the second part which is linked to the manner in which the statistical model has been constructed. Generally, these specular features would have been extracted for real and print spoof samples to generate a 2-class SVM model or something equivalent. However, in our case the model-building is two layered wherein the first layer has an INTRINSIC-bias (natural face sample bias).

- Layer-1: Once the specular features from the NATURAL CLASS ALONE are isolated using the decomposition algorithm from [40], an EIGEN DECOMPOSITION is done, after the computation of the co-variance matrix based on these specular features. The top most significant eigenvectors are retained (which carry the specular details linked to the natural face). Owing

to the natural ruggedness associated with the natural face and the depth topography of the facial surface, the specular parameters are expected to be suppressed and non-coherent (rather diffused). Once the significant eigenvectors are determined, they are retained to form the main part of the NATURAL FACE SPECULAR profile.

- Layer-2: Samples from both classes: natural face samples and print-face samples are first processed to extract their specular features based on the two papers 1 and 2 from literature. Once the specular feature vectors are obtained, they are projected onto the EIGEN-VECTOR set from the natural space class to check their "natural-specular-alignment". This results in two sets of lower dimensional projected features connected with natural-on-natural-projected and print-spoof-on-natural-projected. This projected two class arrangement with a natural space-bias is used to form a 2-class SVM model [25].

The rest of the chapter is organized as follows: The proposed architecture is discussed in detail in Section. 3.2. The database selection, training and testing modes along with the performance evaluation is presented in Section. 3.3.

3.2 Proposed pipeline for effective use of specular features by pivoting around the natural face class

The proposed architecture has the following layers:

- Extracting the specular feature vectors from a set of natural faces (including illumination variations and partial pose variations).
- Using these specular feature vectors to characterize the space of natural faces, through an eigen-decomposition procedure.
- Extracting specular feature vectors from training sets corresponding to natural and spoof classes.
- Projecting these specular vectors onto the eigenspace of natural faces to learn the associativity profiles of the two classes (natural and spoof).
- Formulating an inferencing policy based on these learnt conditional densities.
- Deploying test images from both the classes and checking the mis-classification rate.

3.2.1 Specular feature extraction

Let $U = \{1, 2, 3, ..., n\}$ correspond to n legitimate subjects whose natural facial space needs to be characterized. Each subject i has M pose and illumination variations which are scanned into column vectors and concatenated into an $N^2 \times M$ sized matrix $D_i, i \in \{1, 2, 3, ...n\}$. Each subject specific data matrix D_i is decomposed into diffused component $\overline{I}_i$ and a specular component $\overline{S}_i$. The diffused component is richer and contains more visual information than the specular component which to some extent is an abstraction of the geometry of the surface and its orientation with respect to the illuminating source. While the diffused component is largely contiguous, the sparse components are more patchy but highly correlated within that curved patch. Hence the diffused component is expected to exhibit significant spatial structure and visual meaning while the specular component is expected to be noisy, patchy and to carry a certain amount of depth information [40]. When several partial pose and illumination variations of the same subject i, are presented in the form of the data matrix D_i, the column space of D_i corresponds to the common diffused component which approximately the centroid over all the variations, while the row space of D_i carries information regarding the geometry of the surface based on pixel intensity correlation as a function of the relative positions of the pixels. The latter forms the specular component and is isolated using the accelerated proximal gradient descent algorithm described in Algo. 1.

Algorithm 1. *Accelerated proximal gradient descent algorithm (APG) [41]*

Require: $D_i \in \Re^{\{N^2 \times M\}}, \lambda$

$\overline{I}_0 \leftarrow 0; \overline{S}_0 \leftarrow 0; t_0 \leftarrow 1 \overline{\mu} \leftarrow \delta\mu_0$

while *not converged* ***do***

$Y_k^{\overline{I}} \leftarrow \overline{I}_k + \frac{t_{k-1}-1}{t_k}(\overline{I}_k - \overline{I}_{k-1})$

$Y_k^{\overline{S}} \leftarrow \overline{S}_k + \frac{t_{k-1}-1}{t_k}(\overline{S}_k - \overline{S}_{k-1})$

$G_k^{\overline{I}} \leftarrow Y_k^{\overline{I}} - \frac{1}{L_f=2}(Y_k^{\overline{I}} + Y_k^{\overline{S}} - D_i)$

$(U, \Sigma, V) \leftarrow svd(G_k^{\overline{I}}); \overline{I}_{k+1} = U\beta_{\mu_k * 0.5}[\Sigma]V^T$

$G_k^{\overline{S}} \leftarrow Y_k^{\overline{S}} - \frac{1}{2}(Y_k^{\overline{I}} + Y_k^{\overline{S}} - D_i)$

$\overline{S}_{k+1} = \beta_{\frac{\lambda\mu_k}{2}}[G_k^{\overline{S}}]$

$t_{k+1} \leftarrow \frac{1+\sqrt{4t_k^2+1}}{2}$

$\mu_{k+1} \leftarrow max(\mu_k, \overline{\mu}); k \leftarrow k+1$

end while

end

Parameters of $\lambda = \frac{1}{\sqrt{M}}$, where M is number of pose varying samples. Thresholding function $\mu_0 = 0.99 \|D\|_2$ convergence parameter $\delta = 10^{-5}$. Where k is number of face samples, I is the intensity values of image, t is gradient parameter and S is extracted specular componenet of image I.

3.2.2 Genuine space characterization

In the earlier section, given a subject specific data matrix D_i, the sparse component is isolated and scanned as a t-point real valued column vector $\overline{S}_i, i \in \{1, 2, ...n\}$ using Algo. 1. Let $\mathbf{S}_{GEN}$ be the covariance matrix associated with specular feature set corresponding to the natural images. This is computed as,

$$\mathbf{S} = \sum_{r=1}^{n} (\overline{S}_r - \overline{\mu})(\overline{S}_r - \overline{\mu})^T \tag{3.1}$$

Where $\overline{\mu}$ is defined as the centroid over all the specular features from the natural photo set.

$$\mu = \frac{1}{n} \sum_{r=1}^{n} \overline{S}_r \tag{3.2}$$

The eigen-decomposition of this specular covariance matrix from the natural set, $\mathbf{S}$ is performed as,

$$\mathbf{S} = UDU^H \tag{3.3}$$

where, the eigenvector matrix is given by, $\mathbf{U} = [\bar{u}_1, \bar{u}_2, ...\bar{u}_t]$ which is a $t \times t$ matrix with t being the length of the specular feature vector,

$$\bar{u}_i^H \bar{u}_j = \left\{ \begin{array}{l} 1 \text{ IF } i = j \\ 0 \text{ IF } i \neq j \end{array} \right\} \tag{3.4}$$

and D is a diagonal matrix comprising of t eigenvalues $\gamma_1, \gamma_2, ..., \gamma_t$. Let $t_{sig} \leq t$ be the number of significant eigenvalues detected by first sorting all the eigenvalues and taking the modulus of successive eigenvalues. Thus, the specular eigenspace is eventually characterized by t_{sig} eigenvectors $\bar{u}_{s1}, \bar{u}_{s2}, ..., \bar{u}_{s(t_{sig})}$ with $\bar{u}_{s(j)} \in columns(U)$. The truncated eigenvector matrix is U_{sig}. This important matrix represents the EIGENSPACE of the natural face set in the specular feature space.

3.2.3 Learning the conditional densities

The eigenspace is characterized by the reduced eigenset, U_{sig} which is $t \times t_{sig}$ matrix comprising of t_{sig} significant eigenvectors in the descending order of significance $1, 2, 3..., t_{sig}$. The training operation is performed by collecting samples from three different classes of images:

- Natural/Genuine facial photos (N_G images).
- Printed photographs representing one form of spoofing (N_{PP} images).
- Photographs of people wearing 3D-masks build using paper folds (N_M images).

The corresponding specular feature sets derived using the illumination/pose variations for every subject using Algo. 1, can be defined as,

$$\begin{aligned}\mathbf{TR}_G &= \left\{\overline{S}_{G_1}, \overline{S}_{G_2}, ..., \overline{S}_{G_{N_G}}\right\} \\ \mathbf{TR}_{PP} &= \left\{\overline{S}_{PP_1}, \overline{S}_{PP_2}, ..., \overline{S}_{PP_{N_{PP}}}\right\} \\ \mathbf{TR}_M &= \left\{\overline{S}_{M_1}, \overline{S}_{M_2}, ..., \overline{S}_{M_{N_M}}\right\}\end{aligned}$$

It is important to know what form of information the diffused and sparse/specular components provide for not just the genuine images but also for the spoof models including both the printed photographs [2] (Fig. 3.1(a-d)) and also the 3D masks [8] (Fig. 3.2(a-d)).

In case of the genuine versus printed photographs, there is a significant similarity in the diffused components derived from genuine samples and also printed photos 3.1(a,b) since these images are concerned NOT with the facial topography or curvature, but rather with the the intensity profile. On the other hand the specular component shows considerable complexity and diversity in the case of genuine faces as compared to spoofed ones (Fig. 3.1(c,d)). There is more depth information and complexity in the specular map of genuine images as compared to the spoofed ones.

The results are much closer for 3D masks owing to the presence of depth information in the spoofing operation. Again the diffused components are very similar, Fig. 3.2(a-b) However the richness of the genuine specular component is greater than that of the 3D-mask specular component, despite the depth since the mask has to be smooth enough to fit multiple subjects to facilitate a many to one type of impersonation (Fig. 3.2(c,d)). Masks are prepared in such a manner so that they have to fit all potential clients. To ensure that the face cut of individual-A does not emboss onto the impersonation-mask of individual-B, the mask of individual-B ends up being constructed, as a smoothened version of the actual face cut of individual-B. This allows several individuals of type-A to wear a singular mask derived from individual-B. These specular vector sets are projected onto the eigenspace corresponding to natural images to obtain:

$$\begin{aligned}\overline{P}_{G_i} &= U_{sig}^H \overline{S}_{G_i} \\ \overline{P}_{PP_i} &= U_{sig}^H \overline{S}_{PP_i} \\ \overline{P}_{M_i} &= U_{sig}^H \overline{S}_{M_i}\end{aligned}$$

Figure 3.1: (a) Lowrank/Diffused component of attack face especially printed photo attack faces; (b) Lowrank/Diffused component of real legitimate faces; (c) Sparse/Specular component of printed photo attack faces; (d) Sparse/Specular component of real genuine faces. Images are taken from Wen et al. [2] printed photo database.

Figure 3.2: (a) Lowrank/Diffused component of spoofed faces wearing $3D$ masks; (b) Lowrank/Diffused component of real legitimate faces; (c) Sparse/Specular component of $3D$ mask faces; (d) Sparse/Specular component of real genuine faces.

for $\forall i$ in the respective training sets. The energies of the respective projections are computed as,

$$E_{G_i} = \| \overline{P}_{G_i} \|_2$$
$$E_{PP_i} = \| \overline{P}_{PP_i} \|_2$$
$$E_{M_i} = \| \overline{P}_{M_i} \|_2$$

for $\forall i$. A Gaussian fit is done to the Histograms emerging from the energies of the projected components corresponding to Genuine photos, printed photos and 3D mask images. The Gaussian for Genuine in relation to the conditional distribution for printed photos is shown in Fig. 3.3(a) while the comparison with the 3D mask distribution is shown in Fig. 3.3(b). Since the specular feature in

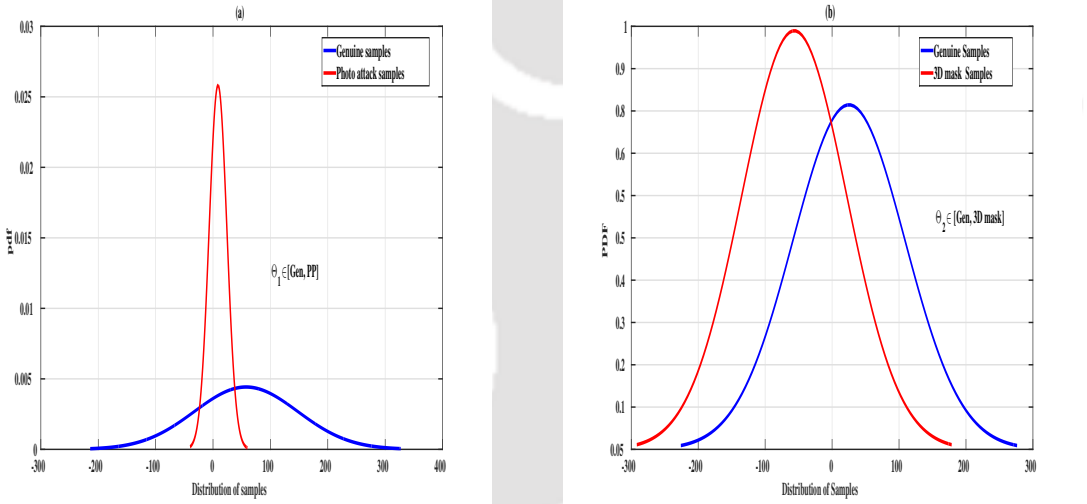

Figure 3.3: **(a)** Gaussian Distribution of Genuine and photo attack samples computed by extracting energy from eigenspace projected features. **(b)** Gaussian Distribution of Genuine and $3D$ Mask samples computed by extracting energy from eigenspace projected features.

essence is an abstraction of the facial surface topography and the reflectivity pattern, the conditional distribution connected with the projections of a genuine set onto the calibrated genuine set has a large variance (σ_G^2 large)indicating a complex association between the two topographies (the blue colored conditional density function in both Fig. 3.3(a) and Fig. 3.3(b)). The association between the genuine space and the printed photos is a simple one as the latter is a planar representation and this is described by a Gaussian having a relatively small mean and very small variance (mu_{PP} small and σ_{PP} small) as can be witnessed in Fig. 3.3(a) (red colored graph). The conditional distributions (Fig. 3.3(b)) get closer for the 3D mask case and $\sigma_M^2 < \sigma_G^2$ (simpler association as compared to genuine-genuine), since the masks essentially designed to fit multiple individuals leading to

a spatial smoothing effect, however the means μ_M and μ_G are close. Overall these conditional distributions clearly indicate that there is substantial variability across genuine and both spoof models, which can be used for constructing a suitable classifier.

3.3 Performance Evaluation

Two databases were deployed for testing the proposed architecture: i) Printed photo database [2] and ii) 3D mask database [8] based on paper craft masks. Two SVM models are constructed: (i) $SVM_{GENandPP}$: SVM for Genuine vs Printed photos and (ii) $SVM_{GENand3D}$: SVM for Genuine vs 3D Masks. The training and testing arrangement for $SVM_{GENandPP}$ was,

- Training: CLASS-GENUINE: 17 subjects (Genuine) and 150 variations per subject; CLASS-SPOOF: 17 subjects (printed photos) and 150 variations per subject;
- Testing: CLASS-GENUINE: 17 subjects (Genuine) and 150 variations per subject; CLASS-SPOOF: 17 subjects (printed photos) and 150 variations per subject;

The training and testing arrangement for $SVM_{GENand3D}$ was,

- Training: CLASS-GENUINE: 15 subjects (Genuine) and 100 variations per subject; CLASS-SPOOF: 15 subjects ($3D$ Masks) and 100 variations per subject;
- Testing: CLASS-GENUINE: 15 subjects (Genuine) and 100 variations per subject; CLASS-SPOOF: 15 subjects ($3D$ Masks) and 100 variations per subject;

The training and testing data was split into five folds K_1, K_2, K_3, K_4, K_5 (data rotation to ensure there is no bias in the formation of the SVM models). Within each fold, 80% of the data was used for training and 20% for testing. The accuracies and false alarm rates for SVM models $SVM_{GENandPP}$ and $SVM_{GENand3D}$ are shown in Table. 3.1 and Table. 3.2 respectively. For the printed photograph based SVM model, the optimal classification accuracy (corresponding to the Equal Error Rate - EER) was found to be 86.5% while for the 3D mask based SVM model the classification accuracy was 76.0%. The corresponding false alarm rates were: 11.21% and 13.53 respectively. The classification error rates as a function of the threshold are shown in Fig. 3.4 and the minimal error happens to be the point of minima in both the curves. 3D Masks do influence depth profile capturing, hence their spoof detection accuracy is on the lower side as compared to printed photos. Their projection variability however is

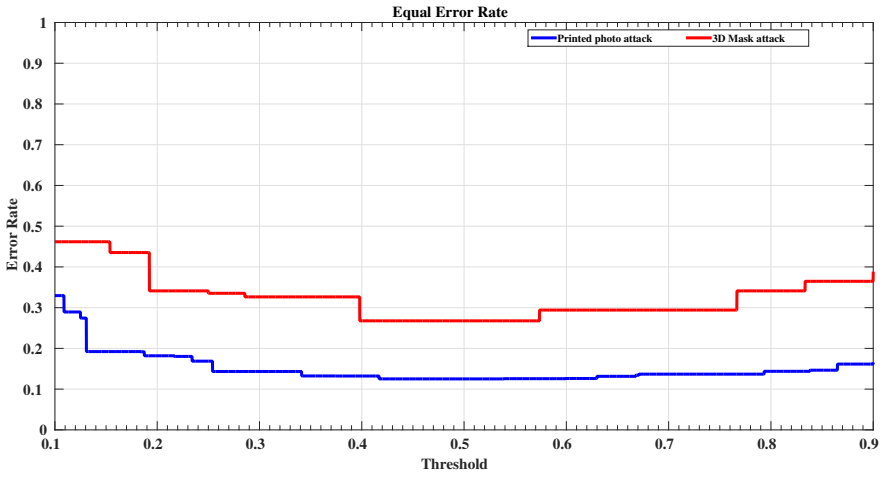

Figure 3.4: Expected performance curve (EPC) with RBF kernel of SVM classifier. Error as a function of the threshold for SVM model $SVM_{GENandPP}$ and SVM model $SVM_{GENand3D}$. The optimal threshold is the point of minima in the two curves.

small in comparison with genuine faces, which is reflected in a moderate detection rate as far as accuracy is concerned.

Table 3.1: Performance of $SVM_{GENandPP}$ SVM classifier evaluated over eigenspace projected features across each fold for printed photo attack face detection.

Measure	Accuracy @FAR	FAR
K_1	0.9523	0.035
K_2	0.9560	0.042
K_3	0.9584	0.0438
K_4	0.9658	0.036
K_5	0.9650	0.0330
Overall 50% of training and 50% of testing samples.	0.8651	0.1121

In both the Tables 3.1 and 3.2 old draft, the K-fold cross validation involved an 80% training and 20% testing. Hence, the accuracies were on the higher side. The last row involved a 50% training and 50% testing (hence an evident drop in accuracy in both tables). it had included an additional column to explain this clearly. The specular parameter is not our main contribution but the 2-layered biased model building process with respect to the natural face set is. Eventually the model is 2-sided with projected features. There have been approaches directed towards effective extraction of specular features.

- Since the specular feature is based on quantum of light reflected from the surface of the object based on the surface geometry and smoothness, glossy-planar-prints are likely to exhibit a higher quantum of specularity as compared to the natural face counter-parts.

Table 3.2: Performance of SVM $SVM_{GENand3D}$ classifier evaluated over eigenspace projected features across each fold for $3D$ mask face detection.

Measure	Accuracy @FAR	FAR
K_1	0.8972	0.070
K_2	0.8920	0.0583
K_3	0.8930	0.0507
K_4	0.8987	0.0537
K_5	0.8925	0.0470
Overall 50% of training and 50% of testing samples.	0.7603	0.1353

- Since the position of the local light source in relation to the object being illuminated may not always be fixed, there will be a natural variability for the print-versions. While the overall quantum associated with natural-face specularity is less, the variability is high owing to arbitrary surface topographies and roughness-profiles. It is a weak feature as it cannot be used as a UNIVERSAL SPOOF-indicator across different environments and spoofing modalities.

While it was acknowledge that in light of our other contributions, this one is incremental, this turned out to be partly exploratory. The 2-layered, EIGENSPACE characterization with natural-face bias was a novel twist in the model building process (an outcome of this work published in CVIP-2018 conference).

3.4 Conclusions

In this paper, we begin with a hypothesis that since the specular feature contains information pertaining to the topography and depth variation, in a face or a mask, the association between the specular sets from two genuine image sets, will exhibit significant complexity and diversity. On the other hand, the specular set from a genuine image set, is expected to have little similarity with that of a printed photo set and a moderate similarity with a 3D mask set (because of the introduction of depth information). This hypothesis has been ratified by testing the two spoof models (printed photo and 3D mask) onto two different SVM models. The classification accuracies for genuine vs printed photo and genuine vs 3D mask were 85.5% and 76.0% respectively.

4

PINHOLE camera model and analytical frame for sharpness analysis for depth profiling of natural images

Contents

Objective

Natural images particularly connected with portraits exhibit a diversity in the sharpness profile commensurate with varying depth associated with different parts of the face. This depth variation puts the object (the 3D face) slightly out of focus in different parts of the image. When a secondary image is created from this portrait, this blurring effect is amplified particularly if the plane of the printed photo is not aligned with the object plane of focus. Thus photos of photos exhibit greater homogeneity in the sharpness profiles and have overall lower sharpness values than their natural counterparts. This principle has been demonstrated through a lens model and has been applied towards face anti-spoofing. Proposed system was tested on the CASIA dataset and showed a recognition rate of 98.38% *corresponding to a false positive rate of* 10%*.*

4.1 Introduction

There are surveillance systems which deploy video cameras instead of still cameras to examine dynamism in the facial profile, such as natural variations in crease lines induced by a smile etc. to establish some form of liveliness in the individual. While natural video captures are hard to fake, in a network environment it is possible for someone to intervene and displace the actual video captured within a specific window with another pre-recorded one "internally". Hence in such a case, the problem of anti-spoofing translates to detection and flagging of pre-recorded videos as potential spoofed versions.

With videos there is a lot more information at the disposal of the verification unit to characterize genuineness of the video clip. It is rare for people to remain poker faced for long periods of time, hence every video snippet is expected to carry some natural variations of the face, such as frequent muscle twitches, the occasional frown, some form of anxiety to get in quickly and join a meeting or at the most a smile. When the facial muscles are activated, there are changes in the crease lines. By tracking these natural variations in the gradient profiles extracted from selective regions, it is possible tune and train a classifier to both models: (i) Models involving natural expressions and (ii) Models involving prosthetics. Aligned with this paradigm, noise in the gradient domain across frames was treated as a "visual rhythm" in Pinto et al. [42], wherein snippets could be classified as real or spoof based on the temporal trajectory of the visual rhythms.

In places where still cameras are deployed, the verification unit has to confront two forms of attacks:

- Impersonation using printed photographs [31].
- Impersonation using carefully designed prosthetic masks [33].

The use of prosthetic masks was explored in Nesli et al. [33]. It was observed that certain key point features undergo a transformation for both real and spoofed faces, particularly linked with gradient profiles. Local binary patterns (LBPs) are local directional gradient descriptors based on the histogram of directional gradients. The skew in the histogram quantifies the directional orientation of a specific feature descriptor. If only the significant descriptors are retained, the key feature points on the facial palette can be detected and compared. In Nesli et al. [33], the prosthetic masks (3D-masks) were created using paper craft which were used for both training and testing with LBP features extracted from normal faces and also prosthetics.

The same anisotropic features, based on directional skews associated with these local feature descriptors were quantified using a shearlet

transform in Feng et al. [43]. The approach worked for printed photographs [31]. Other variations of gradient profiles include the deployment of difference of Gaussian (DoG) based operators for feature extraction [31].

Garcia et al. [44] observed that taking a snapshot of a natural printed photograph was essentially a spatial re-sampling process. This spatial re-sampling introduced some regularized artifacts in the luminance profile. The regularity of these artifacts translated to a much more wide-band frequency representation with virtually periodic tones. By examining this high frequency information, it was possible to segregate normal images from printed photographs (or photos of photos). When there is a significant redundancy in the chosen feature vector, the demand for a larger number of training samples grows. Garcia et al. [44], claimed that their approach required fewer training samples as compared to the rest.

Amongst other model based approaches Gao et al. [37], conjectured that the specular component distribution from a printed photograph was expected to be more homogeneous as compared to that of a natural photograph. This was corroborated by the fact that when a 3-dimensional facial structure was illuminated, the specular component of the light reflected into the camera had a diverse distribution because of the natural curvature of the face. When this face is replaced with a printed photo, this diversity is curbed and the specular distribution becomes more uniform. This principle was used to detect spoofing through printed photos.

In an integrated approach wherein model based ideas were merged with texture profiles, padded with additional statistical features, Wen et al. [32] trained a multi-dimensional classifier to segregate

natural images from spoofed ones (which were now treated as distorted versions of the original images: model unknown).

The first time a snapshot is taken of an object a certain intensity profile of this object is captured in the form of an image. When this image is now presented to the camera and it re-captured, there is a innate degradation in quality. Galbally et. al. [23], attempted to attack this from the point of view of a quality assessment measure, wherein natural images were expected to have a higher quality as compared to printed photos (or photos of photos). The problem with their approach was that this quality assessment was semi-reference based and not completely blind, wherein the reference was derived from the same image (either original or degraded). This work therefore cannot be truly considered as a quality assessment measure, since the original natural reference is absent.

In KIM et al. [22], it was observed that planar print presentations lack depth, hence, when a real natural facial presentation is compared with a print-version, there will be a difference in the degree of sharpness and in particular the sharpness diversity seen in the images trapped in these two presentation modes. By ensuring a narrow depth of field during still image photography it was possible to generate differential sharpness profiles for natural images and print spoof versions.

The depth of field parameter in this still image experimentation was made small, so that the sharpness variation could be spotted in natural face images. They observed that if the camera was made to focus on two extremes of a human face (NOSE-TIP and EARS), the sharpness profile of a natural face image would exhibit a considerable change. For instance if the camera was focussed on the nose-tip, the region around the nose would appear clear but the regions near the sides of the head would appear blurred. The converse would take place of the camera was focussed at the ears. A Laplacian operator was used to trap this sharpness feature and profile it over entire face. The energy profile exhibited a parabolic structure for natural faces and was flat and uniform for planar prints. While this approach was discriminatory there were some issues with this:

1) Practicality: On field implementation would require multi-focal (or dual-focal) imaging for all test samples before performing the analysis. Either this is likely to involve manual intervention or a automated guiding algorithm is required, which detects the nose and ears reliably, irrespective of the presentation format. Furthermore, this expects the faces to be perfectly or partially registered in space so that, nose and ear detection becomes easier. Side-poses may not work.

2) Under diverse illumination conditions, pose variations and print spoofing variations, there is

expected to be a lot of subject and subject registration related content interference. Hence, even if the counter-spoofing is done in authentication mode, there will be a lot of noise.

Our proposed solution in [7], albeit based on the same physical phenomenon (i.e. based on the premise that natural facial presentations have depth while planar prints do not), has certain variations. Firstly the PINHOLE-camera model which depicts the infusion of the blur phenomenon when the object plane does not coincide with the plane of focus is distinct. The analysis connected with it clearly indicates that for a planar print version the overall blur is of a cumulative type since the print version inherits the blur diversity of the original image although in the suppressed form. Secondly there is no re-focussing done during the imaging process when a test sample is presented. This rules out issues related to perfect image registration. The sharpness is measured using a magnitude thresholded version of derivative of a Gaussian operator (which is termed as a GSM or gradient significance map [29]. No calibration is required here. But the GSM feature turns out to be weak since there is heavy subject related content interference, a problem resolved in Chapter-5 connected with random scans

Our chapter layout is as follows: A PIN-HOLE MODEL model has been presented to arrive at a sharpness profile based on the depth in a portrait image in Section. 4.2. The statistical contributions pertaining to this sharpness profiling are in Section. 4.3. Justification for applying this procedure towards face anti-spoofing is provided in Section. 4.4. Simulation results obtained from the CASIA dataset [31] and comparisons are presented in Section. 4.5.

4.2 PIN-HOLE MODEL for BLUR profiling and analysis

Depth profiling from single images still remains to be an open problem although some authors have attempted to accumulate and deploy depth-cues from various modes such as texture drifts [45] or geometric cues such as those based on vanishing points [46]. In this paper we have used a simple physical model to arrive at a sharpness profile (or a blur profile) as a function of the depth in the portrait image.

When the camera zooms into the face of an individual in any natural setting, the background goes out of focus and the face appears very clear. It is however to be noted that any lens system in a camera sets up the *object plane of focus* at a precise distance from the optical center. Objects or parts of them at distances nearer and further from this plane of focus, will appear slightly blurred in the image representation as compared to that part which resides in the *object plane of focus*. This part

is demonstrated using a lens model in Fig. 4.1, wherein a particular object OB is focussed onto the imaging grid (indicated as the screen), appearing as a sharp image IM. If the same object is moved away from the *object plane of focus* through a distance d (positioned at OB'), a crisp image will be formed in front of the screen at a lesser height (h_{Id} at position IM_d). On the screen, this object will appear as a blurred patch with the degree of blur being a function of the height from the optical axis. The point at the upper extreme will exhibit maximal blur while the point closest to the optical axis will have virtually no blur. The extent of blur (degree of uncertainty or fuzziness) with respect to

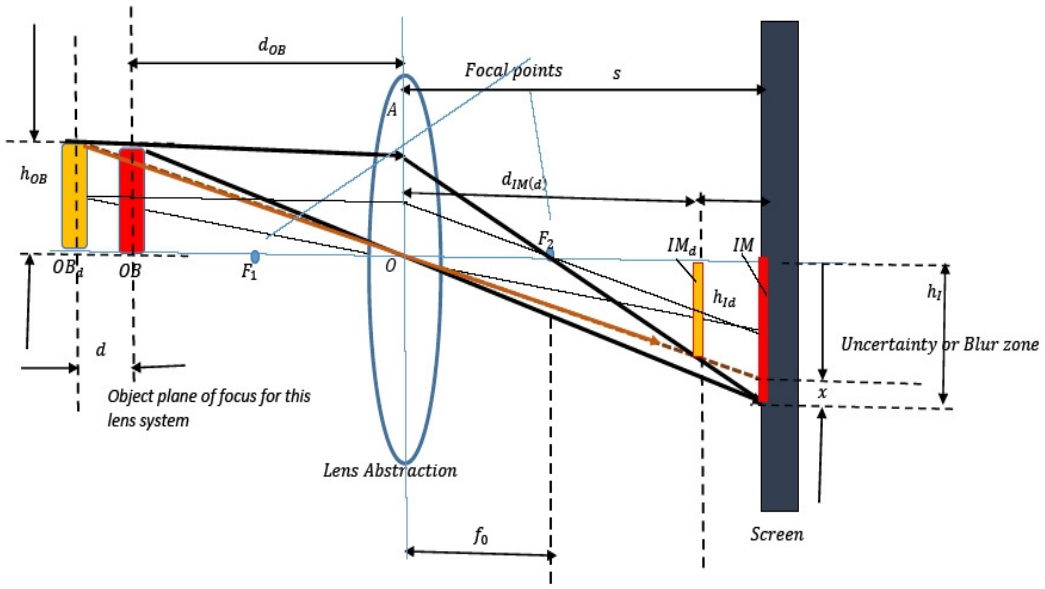

Figure 4.1: Fixed focal length system (focal length f_0) illustrating the effect of changes in depth on the image process. A change in depth is simulated as a shift in the location of the object OB to position OB_d (through a displacement d). The resultant effect is the formation of the new image at position IM_d and blur effect seen on the imaging screen as the collimation of beams at the screen is lost, with the divergence increasing with height $h_0 \leq h_{OB}$ [7]

a point on the object OB_d at a height ($h_0 \leq h_{OB}$) from the optical axis, the object positioned at a distance of ($d_{OB} + d$) from the optic center, can be derived as (for a lens system of equivalent focal length f_0),

$$x_0 = f(h_0, d) = \frac{d \times h_0 \times s}{d_{OB}\left[\frac{sf_0}{s-f_0} + d\right]} \tag{4.1}$$

Note that $x_0 = 0$ for all $h_0 \leq h_{OB}$ when $d = 0$ for this system. For all positive values of d, x_0 will increase with d for a fixed h_0 and will also increase with h_0 for a fixed d. This simple derivation is possible using the arrangement in Fig. 4.1. Thus for a natural image this change in depth through parameter d reflects as a blurring phenomenon.

4.2.1 Printed Photographs

When this natural image IM captured is printed out in the form of a photograph and presented to the same lens system for secondary image capturing, the following are the observations:

- The entire scene information is now contained in a two-dimensional planar surface placed at a specific distance from the optical center along the axis.
- The image intensity pattern distribution from the earlier image IM can be assumed to be virtually inherited through this printing process (viz. this printing process is assumed to be noise free).

Let the secondary image be IM_s. If the object plane of focus is not perfectly aligned with the printed photograph placed on the optical axis, there will be a secondary blurring effect. But since there is no depth variation, the uncertainty parameter x_s in contrast to x_0 will exhibit a more homogeneous blur at fixed distances from the optical axis. If δ is the mis-alignment error and assuming that the focal length is fixed at f_0,

$$x_s = f(h_0, \delta) = \frac{\delta \times h_0 \times s}{(d_{OB} + \delta)\left[\frac{sf_0}{s-f_0} + \delta\right]} \tag{4.2}$$

Owing to this cascade effect, the overall blur experienced corresponding to a point at a height h_0 above the optical axis can be expressed as a linear combination,

$$x_{tot} = x_0 + x_s + v \tag{4.3}$$

where, v is some noise term incurred from the printing of the first natural image. Thus the sharpness at a given point in space is expected to drop in the case of a printed photograph owing to this secondary blurring effect. Thus it conves to the following claim:

CLAIM CH 4.1: *The sharpness profile of a natural portrait photograph is expected to be higher than its corresponding secondary image. When there is perfect alignment the time of secondary capture, it will become very hard to segregate the primary and secondary photographs. However when the degree of misalignment increases because of the carelessness of the photographer, it becomes possible to segregate the primary and secondary photographs based on the extent of sharpness (or its antipodal property viz. blur).*

The proposed method, extracts a BINARY SIGNIFICANCE MAP using a GAUSSIAN GRADIENT and does a patch counting before generating secondary statistics. Consider 1D version of a

blurred/observed image:

$$I_{BLURRED}(x) = U(x - x_0) * G(x, \sigma) = I_{OBS}(x) \tag{4.4}$$

Assuming $G(x, \sigma) = e^{\frac{-x^2}{2\sigma^2}}$, If one takes a derivative of this,

$$I_{OBS}(x) = \delta(x - x_0) * G(x - x_0, \sigma) \tag{4.5}$$

if t is global threshold $t \in (0, 1)$.

$$BIN(x) = 1 \; if \; I_{OBS}(x) > t$$

$$BIN(x) = 0 \; otherwise$$

Area under the pulse is $Area(pulse) = 1 - 2Q[\sqrt{(2)}\sigma\sqrt{(ln(\frac{1}{t}))}]$. This is a monotonically increasing function of σ. Hence a good measure of local sharpness. $Area(\sigma \to \infty) = 1$, $Area(\sigma \to 0) = 0$.

4.3 Sharpness profiling

Here, it develops a simple statistical measure for quantifying the extent of sharpness around a given point in an image. At every point in an image, a region of interest (ROI) window (of size $w \times w$) is created with the reference pixel at the center. The gradients are computed along the X- and Y-gradients using the Sobel operator within the ROI window and then If the gray-scale image in question is represented by the function, $Y_n(x, y)$, where $x \in \{1, 2, 5, N_1\}$ and $y \in 1, 2, 5, N_2$, the results of the convolution with respect to the horizontal and vertical Sobel kernels are given by,

$$\begin{aligned} Y_h(x, y) &= Y_n(x, y) \star D_x(x, y) \\ Y_v(x, y) &= Y_n(x, y) \star D_y(x, y) \end{aligned}$$

where, the operator $\star$ indicates a two dimensional convolution of the image with a KERNEL. The gradient magnitude evaluated at a point (x_i, y_j), is given by the equation:

$$M_G(x_i, y_j) = \sqrt{[Y_h(x_i, y_j)]^2 + [Y_v(x_i, y_j)]^2} \tag{4.6}$$

This becomes a PRIMARY feature for establishing whether a given portrait has sufficient apparent depth. Associated with every point x_i, y_j is a gradient magnitude, $M_G(x_i, y_j)$.

If (x_{ref}, y_{ref}) represents the center of the ROI, the mean of the gradient magnitude profile within

this ROI is computed,

$$\mu_G = \frac{1}{w^2}\left(\sum_{(x_i,y_j)\in ROI} M_G(x_i, y_j)\right) \tag{4.7}$$

The ROI is first abstracted as a binary matrix B as an indication of whether the gradient distribution has a left-skew or a right skew. A left-skew would indicate low sharpness while a right skew would mean higher sharpness and less blur.

For $(x_i, y_j) \in ROI$,

$$\begin{aligned} \text{IF } M_G(x_i, y_j) > \mu_G \text{ THEN } B(x_i, y_j) &= 1 \\ \text{ELSE } B(x_i, y_j) &= 0 \end{aligned} \tag{4.8}$$

The fraction of pixels having gradient magnitudes larger than μ_G within the ROI is determined as,

$$S(x_{ref}, y_{ref}) = S(ROI) = \frac{1}{w^2}\sum_{(x_i,y_j)\in ROI} B(x_i, y_j) \tag{4.9}$$

4.4 Application to Anti-spoofing

Consider the example of two photographs (obtained using a google search) of (i) *Atal Vajpayee* and (ii) *An old woman* Fig.4.2. The picture of *Atal Vajpayee* is a photo of a photo (Fig. 4.2(a), viz. captured through a secondary effect). It is evident from its low contrast nature and the skew. The old woman's picture is a natural photograph (Fig. 4.2(d)). The sharpness profiles of both these pictures are shown in Fig. 4.2(b,e) and their exaggerated versions are in Fig. 4.2(c,f) respectively. It is very obvious from these two pictures that the sharpness profile in the case of the old woman, is much higher and exhibits a greater diversity as compared to the image of *Atal Vajpayee*. The latter exhibits a more homogeneous profile and has fewer sharper regions as compared to that of the old woman's picture.

A direct application of this proposed tool (or procedure for sharpness analysis) is anti-spoofing which was discussed in Section. 4.1, wherein in large organizations, people may attempt to get through unmanned surveillance/verification gateways by placing a planar photograph of some legitimate person over their face. Before the recognition engine can begin its matching process, it is imperative for the verification gateway to run a spoof-checking algorithm to establish genuineness or liveliness of the presented face. This anti-spoofing problem and solution has been tested using the CASIA dataset [31].

When applied to anti-spoofing, the system is presented with either an original and natural picture

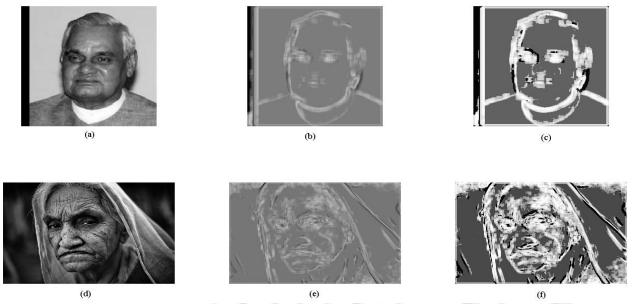

Figure 4.2: (a) Photo of photo of Atal Vajpayee (this secondary effect is easily seen because of the low contrast and the skew in the image); (b,c) Its sharpness profile and its exaggerated version respectively; (d) Natural photo of old woman; (e,f) Its sharpness profile and its exaggerated version respectively. The sharpness profile of the old woman shows a lot more depth and diversity as compared to that of the image of *Atal Vajpayee*.

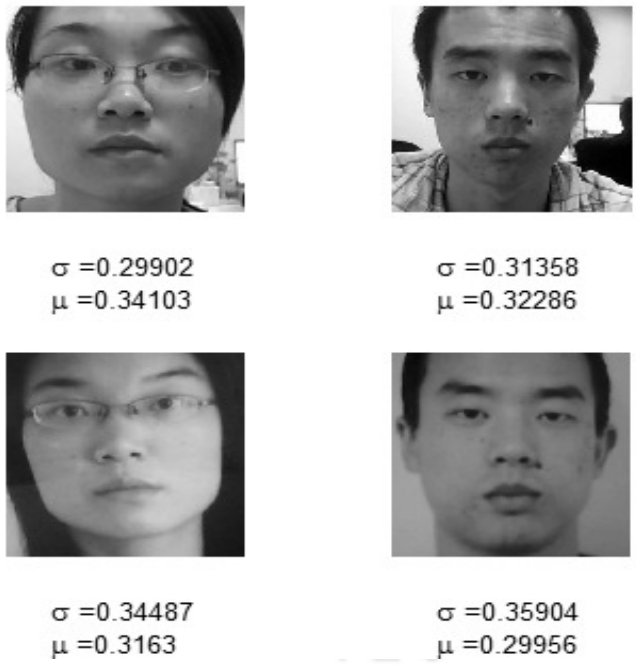

Figure 4.3: Mean and standard deviations of the sharpness parameter over the entire photograph. Note that the average sharpness of the original natural photo is greater than that of planar versions (or spoofed facial profiles). The CONDITIONAL MEANS and STANDARD DEVIATIONS have been computed based on the patch density statistic defined by Eqn. 4.9. This is done over the entire image. A HIGH MEAN indicates prominence of edges and relatively high contrast (or high contrast diversity). Hence as one may notice the means corresponding to the natural faces (top-row) are higher as compared to the means corresponding to print-faces (bottom-row), on a subject by subject basis. On an overall scale, the MEAN corresponding to the natural faces (across subjects) is much higher as compared to the MEAN for print-faces. The pattern linked to the standard deviation is hard to discern but values are conditionally discriminatory. $SD_{NAT} < SD_{print}$ Thus the patch scores when collected to form a registered feature vector serve as sufficiently discriminatory sharpness and contrast diversity feature for segregating the print-class from the natural one.

or with a planar photograph or a spoofed version of a face (analogous to a mask). As per the conjecture based on the physical model captured by Eqn. 4.3, the overall blur in the case of the planar photograph is expected to exceed that of a natural photo of the same object. When this is turned around, the sharpness of a natural photograph should be greater than that of a planar photo of the same registered object/face. Larger the blur, lower the sharpness and vice-versa.

To calibrate the feature extraction process, each facial image was split into smaller blocks of size N_B. The images were resized to 256×256 and the block size was set $N_B = 32$ ($1/8^{th}$ of the size of the image). This time the gradient thresholding is done with respect to the local gradient mean within the block $\mu_{G(r)}$.

For $(x_i, y_j) \in BLOCK_r; r = 1, 2, ...64$,

$$\begin{aligned} \text{IF } M_G(x_i, y_j) > \mu_{G(r)} \text{ THEN } B(x_i, y_j) &= 1 \\ \text{ELSE } B(x_i, y_j) &= 0 \end{aligned} \tag{4.10}$$

Now the patch statistic is computed for each block-r,

$$P_r = \frac{1}{(N_B)^2} \sum_{(x_i, y_j) \in BLOCK_r} B(x_i, y_j) \tag{4.11}$$

This is in-turn normalized as,

$$P_{Nr} = \frac{64 \times P_r}{\sum_{k=1}^{64} P_k} \tag{4.12}$$

since there are a total of 64 blocks. The mean and standard deviation of this gradient patch density statistic is computed. The mean value indicates the average sharpness over the entire image while the standard deviation indicates the sharpness diversity over the image. In sync with the BLUR equation (Eqn. 4.3), the average sharpness of the natural photograph is expected to be greater than that of the planar spoofed photo. This is witnessed in Fig. 4.3. When a natural facial photograph is registered with its planar spoofed version, comparisons can be made between statistics computed from different blocks. By registration, one refers to spatial registration of the elements of the face such as the eyes, nose, jaws etc. This registration process is severely impaired when the block size N_B either becomes too small or too large. When N_B is small, the correspondence is lost owing to significant micro-texture variations which includes expression changes, hair style changes and even pose variations. When N_B becomes too large, two corresponding blocks from the original and spoofed versions can never exhibit the same texture orientation and intensity mapping as the region of interest

has expanded. This subsequently makes syncing difficult. Hence, an intermediate value is necessary for N_B and experimentally this was found to be approximately 1/8 of the size of the facial image.

Table 4.1: Performance comparison with the state of the art.

Method	Database	TPR@ FPR =10%	TPR@ FPR =1%
LBP+SVM [33]	CASIA(150 Samples)	93.3	29.7
DoG+LBP+SVM [35]	CASIA (150 Samples)	66.7	53.3
IDA+SVM [32]	CASIA (150 Samples)	81.6	48.32
Sharpness proposed + SVM	CASIA(150-175 Samples)	**98.38**	**90.62**

4.5 Results and analysis

A trimmed version of the CASIA dataset [31] comprising of 400 natural photographs (including subject variations) and 400 spoofed images which are photos of natural photos of the same set of subjects has been used for training and testing. The natural subject variations include illumination changes, slight pose variations and expression changes. Furthermore subjects may exhibit cosmetic changes such as the presence or absence of glasses, hairstyle variations etc.

A fixed size sharpness profile matrix (or feature vector) is generated from each image (real or spoof) and is fed to an SVM classified for training. The optimal number of samples which gave best results (classification accuracy of 98.38%) on the top of the KNEE of the recognition rate curve was 175 out of 400 for real and 175 out of 400 for spoofed images. Fig. 4.4, shows the impact of changes in the size of the sharpness matrix on the overall recognition rate. Best results are obtained when original sharpness matrix derived for very point in the image (discussed in Section. 4.2) is then sub-sampled through a block abstraction procedure (or a fusion process), wherein all sharpness values within a specific window of size $m \times m$ are averaged with $m = 10$. The size of this window is a tradeoff between computational complexity and recognition accuracy. As this window becomes smaller and smaller (m decreases), more samples are fed to the classifier and so the accuracy is expected to increase. This trend is reflected in Fig. 4.4 for different window sizes. The ROC curve for a feature vector of length 676 (window size 10×10, all images are of size 252×252, with zero padding) is shown in Fig. 4.5. The beginning of the knee corresponds to 150 training samples (real and spoof each) and the top of the knee corresponds to 175 training samples (real and spoof each), less than half the size of the complete trimmed dataset, which is promising.

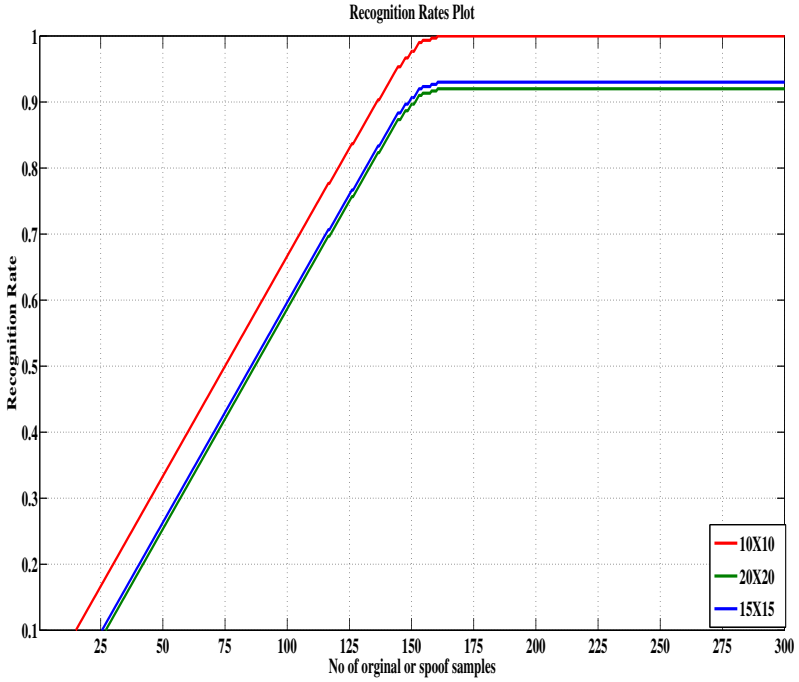

Figure 4.4: Recognition rates for different block sizes m. The end of the knee indicates the saturation point pointing to the optimal number of training samples. The optimal number of samples is 175 for real and spoof each (total of 350) and the window size resulting in the best recognition rate is $m = 10$.

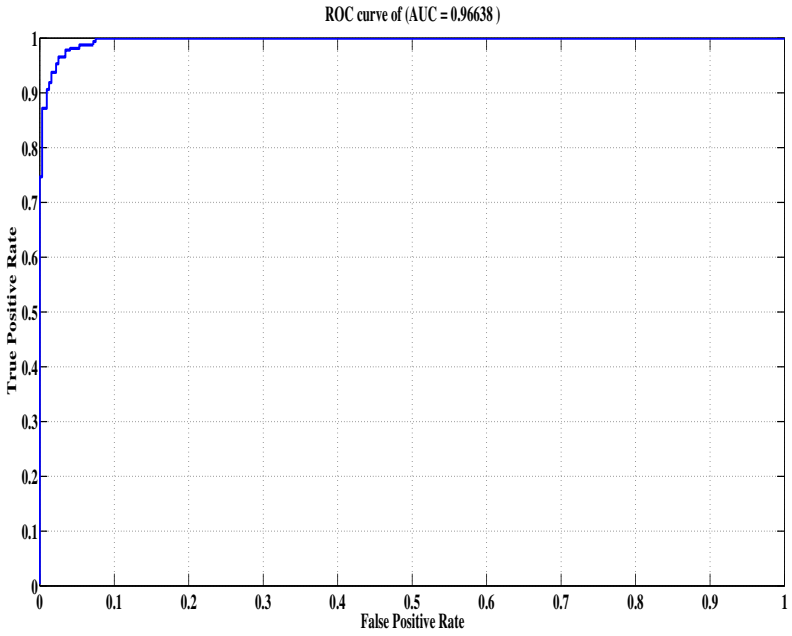

Figure 4.5: ROC curve of SVM classifier to segregate real and spoofed images. The size of the sharpness feature vector was 676.

Comparisons with the state of the art is shown in Table. 4.1. It is clear from the table that the proposed algorithm clearly outperforms most of the existing and computationally intensive algorithms. Most of the techniques deploy filters in the front (mostly of a gradient type such as LOCALIZED BINARY PATTERN etc.). However, the manner in which the secondary statistics are computed varies from paper to paper and with it the degree of complexity. Furthermore, some approaches mentioned in Table 4.1, such as that of Wen and Jain 2015, are multi-modal in nature. Since computational analysis was not the focal point of this thesis, it has not done a rigorous analysis on that front.

There are many blind estimation procedures for measuring clarity or sharpness and paper in [47] which uses polar maps with respect to edge orientation and magnitude is an example. Such a method may work on planar-print detection as these prints tend to exhibit a largely homogeneous blur. There are however some issues with the above method:

- Zhaoyang Liu, et.al [47] was designed to handle isometric Gaussian blur in natural images. This is true even with the proposed Gaussian gradient based significance map generation. The question however is degree of sensitivity as the planar-prints or planar-digital-images tend to exhibit subtle homogeneous blur. Can this be picked up by [47].

- On a more complete front it believes the simpler one (which is our proposed algorithm) may pick up more than what we think it captures.

4.6 Conclusions

This paper presents a novel procedure for arriving at a sharpness profile for performing spoof-checks on facial images (or portraits). Photos of photos tend to have lower sharpness diversity and overall mean sharpness as compared to natural photographs. This aspect has been quantified and used to train an SVM classifier. Proposed system was tested on the CASIA dataset and showed a recognition rate of 98.38% corresponding to a false positive rate of 10%. The proposed algorithm is computationally less intensive and better than most of the state of the art anti-spoofing systems.

5

Identity Independent Face Anti-Spoofing based on Random Scans

Contents

Objective

Existing architectures used in face anti-spoofing, tend to deploy registered spatial measurements to generate feature vectors for spoof-detection. This means that the ordering or sequence in which specific statistics are computed, cannot be changed, as one moves from one facial profile to another. While this arrangement works in a person specific setting, it becomes a major drawback when single sided training is done based on the natural face class alone. To mitigate subject identity linked content interference within the anti-spoofing frame, we propose a identity independent architecture based on random correlated scans of natural face images. The same natural face image can be scanned multiple times through independent correlated random walks before deriving simple differential features on the 1D scanned-vectors. This proposed frame tends to capture the pixel correlation statistics with minimal content interference and shows great promise, particularly when trained on natural face sets, using a one-class Support Vector Machine (SVM) and cross-validated on other databases. Performance measured interms of EER for detection of spoof face is found to be 3.8291% *with proposed random scan features, and* 2.02% *with auto population samples for inter database. We have deployed a 2-dimensional random walk for capturing lower order pixel correlation statistics from natural faces, with virtually no perceptual interference. The proposed identity independent frame has surpassed the state of the art with reference to a 3D mask dataset (image oriented, isolated frame setting), with an EER of 2.25without auto-population and an EER of 0.45*

5.1 Introduction

The problem with this unmanned recognition system is that a particular hidden person X may masquerade as another individual Y, either by wearing a prosthetic mask [15] or by deploying planar spoofing [9] by either presenting a printed photo of Y (or by replaying a an old video of Y's face). Facial recognition systems cannot tell the difference between a natural face, a prosthetic or a printed photo or for that matter even a caricature-sketch of the same individual. This is simply because the recognition system is not concerned with the format in which this facial visual information is presented, but is concerned with identification alone. Hence, a separate algorithmic layer which performs a check whether a spoofing operation exists, is included as the first line of defense before invoking the facial recognition system. Features/ measurements deployed in a typical facial recognition system may not be the same as those involved in anti-spoofing for the following reasons:

- Face recognition features rely heavily on perceptual similarity and to achieve this it is important to register the faces in the spatial domain. Once the faces positions of the key feature points/elements such as the eyes, nose and mouth are registered, either a suitable grid oriented hashing or an interest point matching algorithm can be devised to compare and match any two facial images.

- While face recognition features are designed to be discriminative across subjects, which stem naturally from the positional variations and changes in the structure of the patches or feature points being matched, they must remain robust to intra-class variations including illumination inconsistencies, natural ageing, presence or absence of embellishments such as beards, glasses and prosthetics, hair style changes etc. To ensure this robustness, the measurements must be localized and fixed across all subjects in space.

- The anti-spoofing system on the other hand seeks to answer the following questions: (i) Does the facial profile presented to the camera carry depth information? In the simplest possible terms, does it qualify as a face?; (ii) Through a rigorous semi supervised texture or colorimetric analysis, is it possible to establish whether the face is natural and there is no artificial impersonation involved?

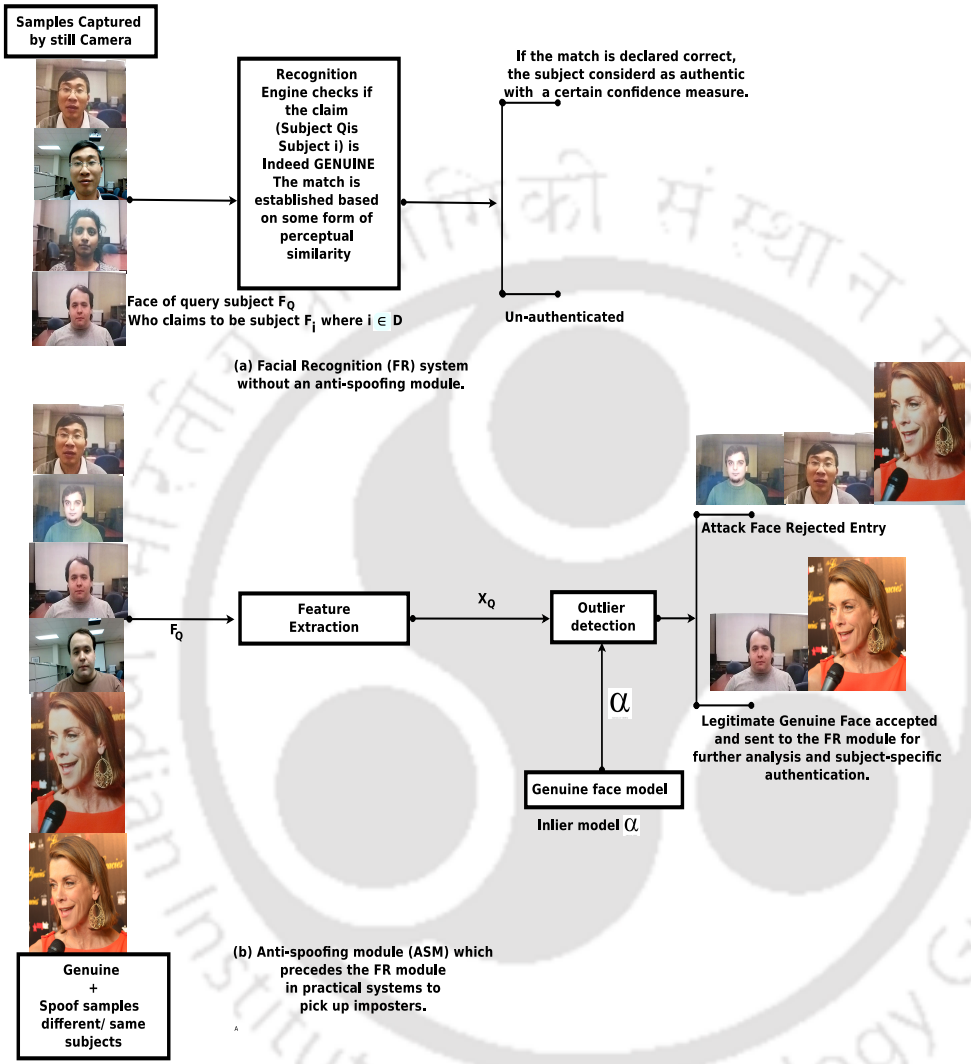

Figure 5.1: (a) Description of conventional face recognition (FR) system which does not include spoof detection module in the top block. (b) Anti-spoofing module (ASM) which examines the genuineness of the face presented and detects any form of artificial facial spoofing.

Fig. 5.1(a), shows the pipeline for a typical face recognition (FR) system, wherein, once the face is captured, the feature extraction and matching algorithm takes over to establish the identity of the individual purely based on perceptual grounds with respect to pre-stored facial images of subjects from a repository. However, when an impersonation attempt takes place either via planar spoofing wherein a printed photo of the target individual is presented to the camera or an attacker walks in with a prosthetic, the conventional FR system described in Fig. 5.1(a) ends up performing the same recognition on perceptual terms without looking for any form of cursory manipulation overriding the original natural face of the imposter. There is therefore a need for an algorithmic layer divorced from the actual FR-frame which searches for alterations in the natural structure of the face wherein person-X attempts to pass on as person-Y. This is the anti-spoofing frame which precedes the actual and standard FR frame shown in Fig. 5.1(b).

Fig. 5.1(b) shows a more generic and minimal anti-spoofing arrangement, wherein only one class of natural/genuine faces is available as a reference set. The natural reference set forms the inlier class, whose cluster boundary is decided based on the choice of feature set. The spoof model type we have assumed is that of a planar nature. Once the natural face space is characterized, outliers are picked up either through some form of a relativistic ranking procedure [3] or by checking conformity with a one-class natural space SVM-model [4]. In our proposed architecture, we shift the paradigm to a transformed space wherein the measurements are un-registered unlike the features extracted in Karthik et al. [3] and Arashloo et al. [4]. While we imbibe the one-sided training procedure for the natural face space from these two papers, we ensure that perceptual content related noise is suppressed, through a correlated but random scan of each facial profile. Our contributions are the following:

(i) Proposition, design and deployment of *Identity Independent Masking Paradigm* for effective anti-spoofing.

(ii) Design and development of a two dimensional random walk algorithm (inspired by Space Filling Curves (SPC) [48] originally devised to ensure compressibility of videos after encryption, since correlation statistics were conserved). This uniquely proposed interpretation of the 2D random walk is used to dissolve the identities of the subjects, while still capturing the pixel correlation properties.

(iii) In line with earlier work associated with one sided training from natural face sets and treatment of spoofed images as outliers [3], [4], we have extended this Identity independent masking algorithm

to include outlier detection by building a one-class SVM [4] from genuine facial features.

The rest of the chapter is organized as follows: In Section. 5.2, we scrutinize the literature in the context of planner spoofing. The motivation for our new paradigm related to identity independent anti-spoofing and positioning with respect to existing literature is discussed in Section. 5.3. The proposed identity independent anti-spoofing architecture whose basis is the two dimensional random walk is discussed in detail in Section. 5.4. The outlier detection frame is presented in Section. 5.7 and finally experimental results along with comparisons with the state of the art anti-spoofing algorithms in Section. 5.9.

Claim CH 5.1: Given an impersonator X and a target subject Y, the prosthetic is designed to mimic the surface contour of Y and has very little to do with the surface contour of X. This is to ensure identity masking from the point of the view of the attacker (X). The only way this can be achieved, is by ensuring that this physical facial re-mapping or (physical face-morphing), is of a many-to-one type. Thus a single prosthetic designed to impersonate Y, can fit multiple individuals of the X-type. In other words, one mask is designed to fit many. This makes the prosthetic, presented as a synthetic surface contour of X, an over-smoothed approximation of X's facial profile with some depth information. One therefore anticipates a reduction in facial image sharpness as far as image of the prosthetic of X is concerned. This sharpness variation can be captured by performing a gradient based analysis.

In this chapter, we focus on literature connected with prosthetic based facial spoofing. A 3D-mask dataset was developed using paper craft models in Erdogmus et al. [8], wherein the prosthetics were customized to target different subjects. Some examples of this are shown in Fig. 5.2. The natural faces of the subjects are shown in Fig. 5.2(b)(row-2), while their corresponding spoofed versions with prosthetics are shown in Fig. 5.2(a)(row-1). It is obvious that paper craft model has been cleverly designed to mimic the surface contours including the ocular and nasal profiles of each targeted subject. In Erdogmus et al. [8], the base feature used for recognition was the Local Binary Pattern (LBP) along with its variants. The 3D-Mask dataset was analyzed both as a sequence of static images, and also as a video sequence, by extending the LBP analysis to include both time and space differentials. A 2-class SVM was finally constructed by learning the prosthetic as well as the genuine face spaces, coupled with the decision boundary/surface. Spoof-detection was done by extracting the same features from a typical query test-face and checking its position with respect to the two reference clusters. In a video-based setting associated with the 3D mask database, more options exist, since it is possible to

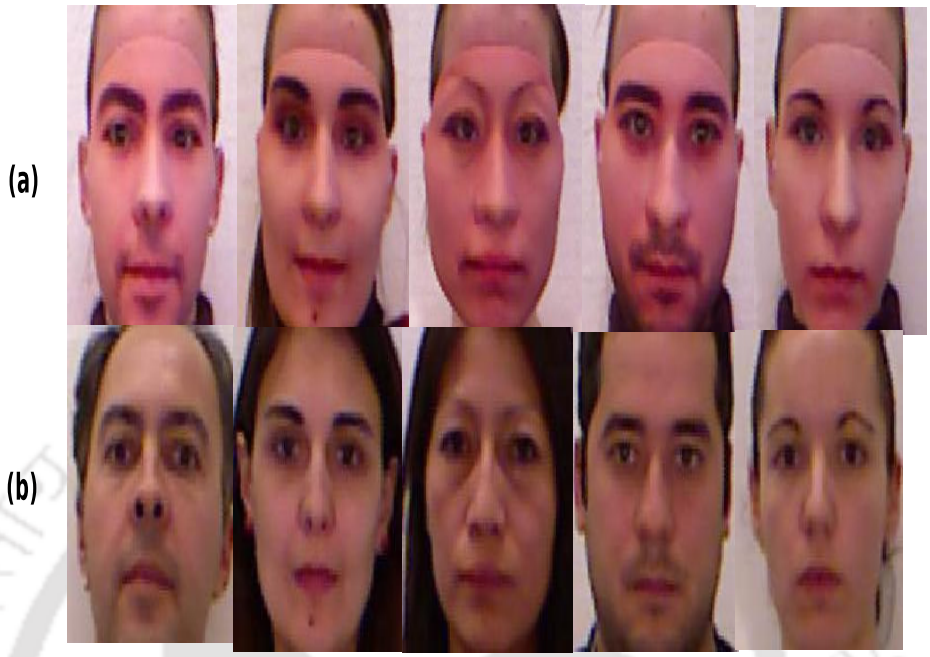

Figure 5.2: **(a)** Examples of 3D mask faces for different subjects, taken from 3D mask database [8] **(b)** Examples of their corresponding real genuine samples.

deploy space-time micro-feature analysis to search for liveliness in the facial profile, consistencies and naturalness in expression changes etc. Optical flow methods were used in Feng et al. [49] to detect differences in dynamism with respect to texture between an imposter and a genuine subject. A deep-learning network for attacking multi-biometric spoofing including facial spoofing was developed by Menotti et al. [50], but again the learning was two sided and assumed availability of samples related to the spoofing process. Wen et al. [2], proposed a mixed bag of features ranging from intensity and gradient all the way up to those which captured color and texture, with the objective of covering the complete gamut of statistics, which would help segregate the genuine face class from all forms of spoofing. However, once again, this arrangement demanded availability of spoof training samples, necessary for constructing a 2-class SVM. This existing frame had several issues:

- Very often the nature, texture and structure of the customized prosthetic may not be known. This implies that spoof-class training samples may not be available. Hence, it is important to shift and restrict the training process to the genuine face sample set, where the acquisition procedure, naturally captured facial profile coupled with the local statistics remains predictable.

- Since LBP features are highly localized in space and are registered, pose deviations and scale changes because of facial migrations with respect to the camera will lead to a contortion of

measurements. This will interfere with the counter-spoofing procedure. We term this form of interference as perceptual interference, which arise when the measurements are registered in space.

The first problem related to absence of a spoof model, can be addressed through an inlier space characterization procedure by learning the space spanned by genuine natural facial images from different subjects, for different poses and mild illumination variations. This inlier space characterization was done through a query feature ranking procedure, in relation to the genuine face feature set, to detect outliers in [3]. Genuine face space characterization coupled with anomaly detection in a much more general setting by constructing a one-class SVM was done in Arashloo et al. [4].

While this arrangement was designed to care of the first problem related to the absence of a proper spoofing model, they were applied to planar spoofing alone. In both these papers [3], [4], the measurements were registered in space either by gridding the image or by computing statistics in specific spatial zones, whose locations were largely static. They thus proved to be ineffective, when confronted with 3D-spoof models, wherein the prosthetics attempted to mimic the depth profile in the imposter's face.

Attacking this 3D-spoofing problem, with a single sided training procedure, involving only genuine face space characterization was the main challenge. This led to the proposed architecture which was placed on an identity independent setting. The rest of the chapter is organized as follows: In Section. 5.5 we propose a new paradigm based on identity independent feature auto-population through random scan patterns. Section. 5.6 validates the choice of randomly scanned feature and builds a one-class SVM to characterize the space of natural faces. Experimental results and comparison with the state of the art are in Section. 5.9. This chapter IMPLICITLY has three parts:

- Part-A: Formulation and Motivation for the Random scan solution towards face-counter spoofing (a technology transfer from an old 1987 paper (Matias and Shamir [48]) connected with Space Filling Curves (SPCs) directed to ensure compressibility of encrypted videos). This part has been tied to CH2 connected with the NATURAL SPACE CHARACTERIZATION and Outlier detection.

- Part-B: Validation of the differential statistics and feature vectors based on the random contiguous scans of images and this includes ONE-SIDED model building (on a multi-dimensional

front as in [4], i.e. 1-class SVM). But note that in [4], the features were registered, largely differential/image quality based and significantly weaker as compared to our random scan solution. The testing was done on both Planar-print and Planar-digital spoofing modalities. Since the NATURAL SPACE CHARACTERIZATION in CONJUNCTION with the RANDOM SCAN feature makes the model SPOOF MODALITY INDEPENDENT and also immune to changes in the camera acquisition environment, CROSS-VALIDATION is quite effective and superior to existing techniques including the closest compatriot of [4].

- Part-C: Here, we show that this method will also work against prosthetics based on an OVER-SMOOTHING assumption. While both natural faces and attackers wearing prosthetics will exhibit a depth variability, the depth-diversity profiles or face ruggedness profiles trapped using the differential random scan statistics are different in the two cases. This is precisely what we intent to exploit in this segment to pick off the prosthetics. This is the most interesting segment as our method has proven to be vastly superior to even CNN-based solutions developed on this front (with a two-sided training).

5.2 Literature review

The main challenge in such a framework is to devise a generic anti-spoofing system which works universally against all forms of spoofing ranging from printed photo attacks all the way to diverse forms of prosthetics. However, much of the solutions developed are heavily model driven and rely on both some prior knowledge regarding the spoofing model and its constraints, to segregate a spoofed face from a true one. We restrict our survey to planar spoofing, wherein the spoofed facial object happens to be either a printed photo of the target individual or a replayed video or an image from a tablet of that person. In line with this, the ideas prevailing in literature encompass the following paradigms:

- Quality assessment based facial image analysis: It has been found in literature that there is degradation in overall quality of the planar spoofed image in relation to the original image. This degradation can be quantified in either differential terms [36] or based on a reduction in the contrast scores evaluated over the entire image [3].

- Deploying physical constraints in printed photo frames: The very nature of the acquisition

process from a planar facial image or photo has certain intrinsic geometric limitations, which can be measured and quantified in relation to real and natural faces [51] [52] [19] [7].

- Statistical analysis by feature concatenation: A concatenation of features [2] from different modalities such as color, texture and intensity from both the true as well as the spoof-classes can be used to train classifiers. While such a concatenation attempts to combat planar as well as non-planar (or prosthetic) based spoofing, there is a loss in precision when compared with model specific approaches.

Quality assessment based approaches rely on the premise that the spoofed image under consideration is somewhat degraded in relation to the natural one. Either there is a compromise on sharpness or contrast or some other form of distortion which does not have a precise underlying model. The problem however with this frame is that there is no reference point against which the quality can be weighed. This problem was in part addressed by Galbally et al. [53], [36], wherein a reference image was derived from the test image through a blurring process. Blur differentials were computed and several quality measures were superimposed on this differential image. It was observed that the edge profile degradation for printed photos was much greater compared to natural faces. Subsequently, blur differential features from the two classes (spoof and natural) were used to train a classifier to segregate the classes based on the residual edge noise profile. While this arrangement can be extended to a more model independent setting, this two sided edge profile analysis and training demands a prior knowledge of certain types of spoofing and the availability of spoof-samples for building the classifier.

This problem was resolved in Karthik et al. [3], wherein the underlying premise was a degradation in contrast of a printed photo versus a natural image. Thus, it was anticipated that a natural facial image was expected to be contrast-rich carrying clear perceptual information in comparison with its counterpart spoofed version. The base feature involved the computation of contrast scores over the entire image and the mean of these contrast scores was used as a discriminatory statistic for detecting planar spoofings. A one sided training model was also developed in this work [3], wherein the natural facial images stored in the repository and their corresponding contrast statistics were pooled together to derive the conditional contrast distribution associated with natural face statistics. Relative to the inlier distribution a suitably calibrated threshold based on relative ranking, was set to check if the statistic computed from the test-query was a part of the tail of this distribution. It was found that query measurements which were a part of the tail were most likely to come from planar spoofings.

Similar work based on the inlier natural face space characterization using the idea of one class SVMs was done by Arashloo et al. [4] with Local Binary Patterns (LBPs) as a base feature, along with its variants. A more detailed review of this chapter is done in Section. 5.3.

The quality assessment model developed by Galbally et al. [54] was extended in Jourabloo et al. [55]. Jourabloo et al. suggested that the degradation model associated with a spoofed image can be treated as some form of noise, which can be characterized and learnt with examples. A convolution neural network(CNN), was trained to learn the differences between the natural and spoofed images in terms of quality, by mixing the images from natural faces and also pooling together the images from the spoof classes. It consisted of three convolution layers and pooling layers. Once the quality differentiation features were learnt using the CNN through supervised training, the classification boundary for this two-class problem was also learnt to pick up both 2D as well as 3D spoofings.

In model-specific spoofing, the spoofing frame is assumed to be known apriori, including its physical constraints. While the spoofing frame involving prosthetics is extremely diverse given the amount of variations in the texture and format associated with 3D mask structures, the frame associated with printed photos or pre-recorded videos is rather limited. Here, the differentiation between the natural and spoof sets is brought about because of physical constraints associated with the spoofing process (e.g. in planar spoofing depth information is absent).

In Garcia et al. [19], the printed photograph spoofing, was treated as spatial re-sampling problem leading to the introduction of a type of noise in the spatial domain termed as Moire patterns. Since this noise is wide-band, its power can be estimated by filtering the query images and then subtracting the original image from the filtered image.

In KIM et al. [22], it was observed that planar print presentations lack depth, hence, when a real natural facial presentation is compared with a print-version, there will be a difference in the degree of sharpness and in particular the sharpness diversity seen in the images trapped in these two presentation modes. By ensuring a narrow depth of field during still image photography it was possible to generate differential sharpness profiles for natural images and print spoof versions.

In Karthik et al. [7], the depth information trapped in the form of a natural and heterogeneous blur variation in natural facial photographs is tapped using a gradient based statistic. A printed photo (viz. a photo-print of an individual's face), is simply a glossy sheet presented to the camera which lacks depth information. Thus, the image of this photo-printed-face results in a homogeneous blur

when compared with blur-diversity seen in the case of a natural facial image.

Light field imaging is progressive in nature and has been known to provide sufficient information to characterize the depth and shape of objects [51]. This lack of depth information in planar spoofing has thus been exploited in Ji et al. [38] using light field cameras. Light field cameras, tend to accumulate information from different angles and formulate a depth map of the scene or the facial frame. This aspect has been used to detect planar spoofing in Ji et al. A similar application of light field imaging towards planar anti-spoofing was proposed by Alireza et al [52], wherein both color and texture features were used to learn a 2-class model. There are several issues, particularly associated with its complexity in hardware and practical deployment in smart-phones. This frame also assumed a constrained form of illumination so that the shapes of faces (including their depth maps) can be characterized in relation to this planar spoofed counterparts.

One way by which the anti-spoofing frame can be made virtually model independent, is by concatenating various types of features connected with contrast, colour, texture and intensity not only from natural photos but also from different spoofing models such as printed photos, videos, paper crafted prosthetics etc. Once the features are extracted from these two classes: one pure and the other in the mixed form by pooling together different forms of model specific spoofed images, a classifier is trained to detect a spoofing operation. Although this form of spoof-model mixing does not quite qualify as real model independence, some generalization is brought about through an intrinsic model interpolation gained via feature mixing. In line with this, Wen et al. [2] argued that spoofed faces tend to have a greater distortion as compared to a natural one. The only issue is that the nature of the distortion remains hidden. This distortion can be captured by computing and mixing a variety of features witnessed in spoofing models such as specular reflection, blurriness, chromatic moment and color diversity, but only with reference against a natural face class.

Similarly, Patel et al. [10], for a smart phone unlocking application, carried out several investigations to examine various types of spoofing, involving photo and video attacks. They extracted color moments and texture patterns associated with genuine and spoof samples for model training. It was spotted in Boulkenafet et al. [51], planar spoofed faces tend to have some form of a color aberration when compared with natural samples. In general, printing and display devices, have limited color gamut compared to whole bunch of visible colors, which is carried forward in this form of aberration or noise in the color space. Much of the earlier methods had been analyzed only in the luminance

channel or gray scale mode. Here, it was possible to detect fake faces simply by building color models for natural and planar-spoofed faces.

5.3 Motivation and problem formulation

Much of the literature covered in the previous section, has been driven towards a 2-class learning problem wherein prior information is available regarding both the genuine or natural faces as well as the underlying spoofing model. This makes it possible to synthesize experimentally spoofed variations of the faces of the same subjects held in the repository. The real differentiation comes from the manner in which the base features are derived: some based on physical models, some on quality constraints and some simply being an accumulation of all forms of coarse features from different modalities. Before the chapter addresses the core contribution or the frame under which the solution is constructed, we review three key papers critically and then dive into the specific formulation. There are three basic paradigms under which anti-spoofing can be performed:

(P1) Person specific setting (PS) [14]: Here, both natural as well as spoofed samples from the same subject class are either available or partially synthesized priori through some form of feature transference. The spoof-check is done at a subject-level when a test-query is presented in an authentication setting. However, a major caveat here is that the artificial synthesis of a spoofed set of images demands the availability of prior spoofed examples from a few subjects. The architecture cannot proceed unless firstly the spoofing model is known and secondly exemplar spoofed images of select subjects present in the database are available.

(P2) Outlier or Anomaly detection with respect to Natural faces (OADNF) [3], [4]: The inclination here is to learn only the inlier class which corresponds to the natural class of faces. Any deviation from this 'natural collective' can be treated as an anomaly and a form of spoofing. Of course some form of calibration is required for assimilating the illumination environment in which the natural faces are captured and along with the pose deviations of the subjects with respect to the camera. Once this is done, an outlier threshold or an outlier surface is generated in-house to pick up spoofings. Here, partial alignment of the spoofed images with respect to the natural images not necessarily at a subject level, but in terms of spatial locations of the features such as positions of eyes/nose/mouth etc., is required. When this pivoting or registration is guaranteed and the feature vectors are pooled together to form a cluster, the natural face statistics is captured. If

this pivoting fails or becomes more elastic due to extreme pose variations, the variability within this inlier space will cross certain bounds and the system will fail.

(P3) Proposed Identity Masking coupled with Outlier detection (PIMOD): This paradigm is essentially the previous paradigm P2 (OADNF) with a twist. The need for pivoting around key points in space becomes a major drawback for a generic anti-spoofing system which must tolerate a fair amount of pose variations. This problem associated with indirect registration can be avoided if the pixels are shuffled and the identities of the subjects dissolved on a perceptual scale using some form of a masking transformation. This perceptual masking transformation should be such that some statistical parameters or measurements from the original setting must be preserved even after the transformation. Note that the masking transformation must be randomized and most importantly key-independent [56]. The inlier class now comprises of masked and randomized natural faces along with the corresponding feature vectors carrying enough statistical transparency to qualify the natural-face class while weeding out outliers in the form of spoofing. Ideally this frame should work not just against planar spoofing but also to a lesser degree against prosthetic/3D-mask attacks (with or without slight pose variations).

5.3.1 Positioning

In the PS-paradigm of Yang et al. [14], the idea of model-specific spoof face synthesis based on exemplar images and prior model knowledge was introduced. This allowed the architecture to synthetically generate spoofed versions for subjects for which training images were not available. The solution assumed a heavily constrained incrementally linear frame wherein the features extracted from the natural face of one subject could be associated linearly with the features extracted from the natural face of another subject. A similar linear but structurally different form of mapping was assumed between the natural face and its corresponding spoofed version. The channel space was split into two components, a subtle registration component and a contrast/illumination alignment component. The former played a crucial role in synthesizing model-specific fake features. Another major drawback because of this linear dependence was the restriction of the spoofing frame to printed photo and video replay attacks, which are reproduced planar versions of natural faces. Since every person's facial depth map is expected to be distinct because of ancestral linkages, natural variations due to ageing and lifestyle, the two channel matrices based on the linear formulation are not adequate

enough to capture the associative map between any two distinct subjects. While this remains a 2-class SVM problem similar to much of the existing literature, the main difference stems from the fact that a 2-class classifier is being constructed for each and every subject. This de-linking helps provided the key features such as the eyes/nose and mouth, do not, migrate too much within the facial frame. However, extreme scale changes and pose variations may disturb the associative map and the registration of fake features with respect to the natural ones.

In the OADNF-paradigm [3], [4] the focus shifted from a 2-class problem to a single class problem comprising of natural facial images as inliers. It was understood that if the inlier class could be fully defined and completely represented, it would be very easy to pick up deviant images emerging as spoof-samples.

In Karthik et al. [3], the mean contrast score over several facial images patches was chosen as a discriminating statistic to separate planar spoofing from regular natural facial images. It was conjectured and observed that when a image is recaptured in a planar form, the intensity profile variation and subsequently contrast drops significantly. Hence, planar spoofed images were expected to register lower contrast scores as compared to natural faces. An outlier detection framework was built by learning the contrast distribution only for natural faces. Any query registering a contrast score near the tail of the distribution was declared a spoofed image. This was done through an implicit relative ranking procedure by checking the position of the query in relation to the natural images in the repository, through a sorting procedure and queries registering low ranks in terms of contrast scores (over the last 25%), were declared spoofed images. Cross-validation was done with MSU-MFSD database [3] and produced an Equal Error Rate (EER) of 21.56%. The main constraint or issue with this frame was the model specific assumption that planar printed photos exhibited a lower dynamic range as compared to natural photos and subsequent printing would lead to a further reduction in contrast. Natural illumination variations, pose changes can alter the intensity profile considerably. This would increase the inlier class diversity, leaving the contrast metric as a weak feature (highly sensitive) to separate natural faces from spoofed ones.

The same problem mentioned earlier was treated as an anomaly detection frame by Arashloo et al. [4], wherein a 1-class Support Vector Machine (SVM) was trained on top of a bag of features involving local binary patterns (LBPTOP), local phase quantization on three orthogonal planes (LPQ-TOP); and binarised statistical image features on three orthogonal planes (BSIF-TOP). While the

results were interesting for this one-sided training algorithm, there were certain limitations to this frame:

- The statistical measurements, computed in different parts of the image at different scales and organized to form the primary feature vector had a rigid structure. This means that the measurement modalities cannot be re-ordered for any image. An indirect implication of this is that the faces must be registered to a large extent in scale and pose. While to some extent pose and scale can be countered by block size adjustments, extreme pose variations cannot be tolerated.
- The affixture of the facial frame owing to a demand for approximate spatial registration of key facial features makes the process less elastic and more vulnerable to local and global geometric operations.
- The inlier class comprises of natural faces from different subjects carrying different facial features. Consequently with these registered measurements there is a natural variability across subjects owing to perceptual differences. This subject content dependent cross noise tends to interfere with the anti-spoofing frame. We term this form of cross-noise incurred due to feature registration as *perceptual interference*. This is the main reason why the EER's are expected to increase particularly when the training is one sided with respect to registered feature sets.

As in any application face counter-spoofing or otherwise, there are two fronts on basis of which counter-spoofing system optimization is done:

- Feature selection
- Model type and construction process (including classifier type)

These two parameters are in turn influenced by other factors such as:

- Are the features distortion phenomenon specific or is it a MIXED-bag?
- Are the features registered in space? (in all existing cases YES - except in our Random scan proposition).
- Are the features client/subject specific or subject independent?
- Are spoof samples used in the model building or training process? (Is the training one-sided or two-sided or partially two-sided?).

- Are there significant training samples available for building a robust model or a set/collection of client/subject specific sub-models?

This in turn indirectly decides whether a robust and effective CNN can be built on the data. CNNs cannot be easily built exclusively on NATURAL SPACES which makes our content agnostic natural spaced random scan approach superior to most (in a subject independent setting). In the grand setting, where does our approach fit in? This is what we meant by positioning in relation to prevailing implicit architectures connected with face-counter-spoofing.

5.3.2 Need for an Identity Independent Frame

The anti-spoofing problem is a typical frame wherein the nature of impersonation remains unknown in practice. By treating this problem as a form of planar image or printed photo spoofing, the problem becomes analytically tractable mainly because of physical constraints. On one side, there are a set of natural facial images which contain depth information embedded in the form of self-shadows and subtle heterogeneous blur variations [7]. On the other side, there are these planar spoofing arrangements which are heavily constrained geometrically.

Claim CH 5.2: *We claim that most anti-spoofing systems work best in an identity independent setting, wherein the measurements or features extracted are taken in such a way that perceptual relevance is given the least importance. However, the residual correlation or some other statistics which may be derived from this dissolved identity, should carry necessary information regarding the environment or channel in which the information has been captured to perform anti-spoofing. This identity dissolution in our case is performed using a constrained shuffle of pixels in the spatial domain using a 2-dimensional random walk. This 2-D random walk has been inspired by Space Filling Curves [48] which was originally devised for retaining the compressibility of video signals after encryption.*

5.4 Proposed Identity independent anti-spoofing architecture

In a registered anti-spoofing frame, the feature vectors derived from different facial images are *registered*. This implies that the local and global measurements over a particular facial image, must be ordered in the same fashion for all the other images. Usually this form of registration is required in facial recognition system, so that the key elements in any face, such as the eyes nose and mouth can be properly matched across subjects.

Consider samples of facial images F_i and F_j shown in Fig 5.3(a,c) respectively. $S_{i,k}$ and $S_{j,k}$ where $k \in [1, 2, ...]$ are the local statistics computed from these two images. The corresponding feature vectors are $\bar{X}_i = [S_{i_1}, S_{i_2....S_{i_n}}]^T$ and $\bar{X}_j = [S_{i_1}, S_{i_2....S_{i_n}}]^T$ respectively shown in Fig 5.3(b,d). Note that these two feature vectors are "registered" in space. For a given feature $\bar{X}_i$, the ordered measurements S_{i_k} contain two pieces of information:

- Perceptual information which allows the viewer to connect with the subject.
- Crucial background information connected with the manner in which the face has been captured using specific still camera. This information is abstract and not explicit. It has to be ferreted out carefully.

In a typical anti-spoofing problem, it is important to ensure that the statistics S_{i_k} are chosen and computed in such a way that the perceptual information (of the first type) is suppressed while the face capturing procedure linked information (of the second type) is amplified. For example, selecting S_{i_k} as a Sharpness features [7] as metric is a better choice as opposed to Contrast score [3]. However, withe this paradigm, there is always a danger that the feature vector carries some residual perceptual information. Implications are multi-fold:

(i) The intra class variability increases (natural face set alone or spoof set alone across subjects)

(ii) On a subject specific setting/ arrangement, the between class similarity (genuine vs spoof) increases.

Fig. 5.4(a) and Fig. 5.4(b) show exemplar conditional distributions for two statistics X_1 (Sharpness from Gaussian gradient profiles [7]) and X_2 (Contrast [3]), where, in the case of the former, the perceptual interference is less. Notice in the case of state X_2 shown in Fig 5.4(b), the conditionals acquire larger variations,leading to greater overlap and this means more perceptual interference. This provides us with the incentive for operating in an identity independent frame wherein perceptual features are completely dissolved and focus shifts to "scene linked statistics" and acquisition modality. To bring out the context, we intended to demonstrate both in our Pattern-Analysis-Springer-2020 [16] and also in the thesis, the importance of feature dependence on the content in the image.

- Some features which are intensity related such as CONTRAST, SELF-SHADOW diversity etc. are heavily local illumination environment dependent in relation to face-topographies and pose-types.

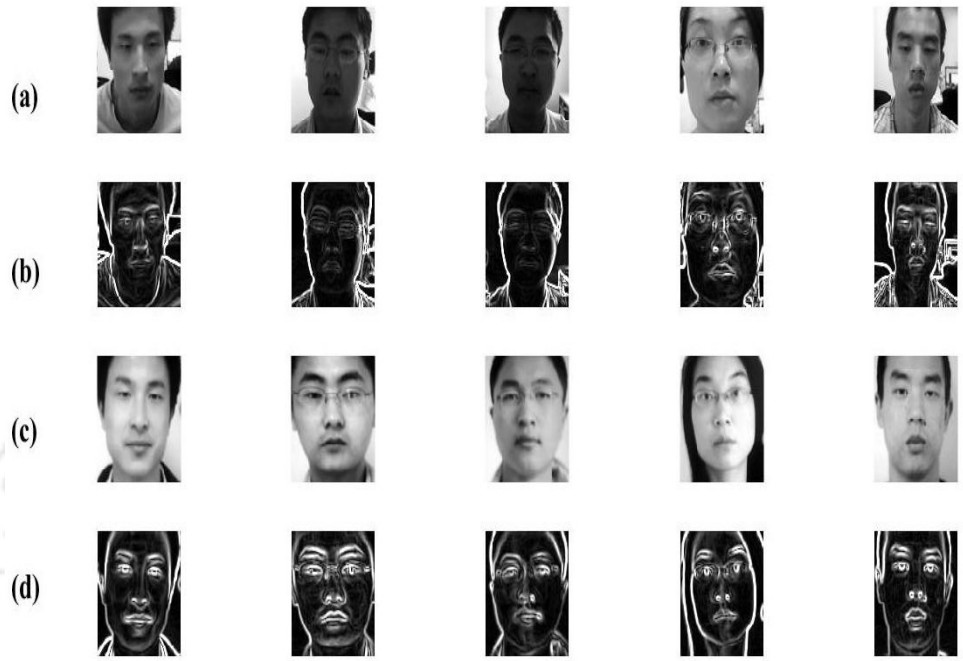

Figure 5.3: **(a)** Visualization of samples of genuine samples of CASIA face dataset **(b)** Gradient threshold based sharpness features extracted [7]. **(c)** Samples of photo attack faces. **(d)** Same sharpness features extracted [7].

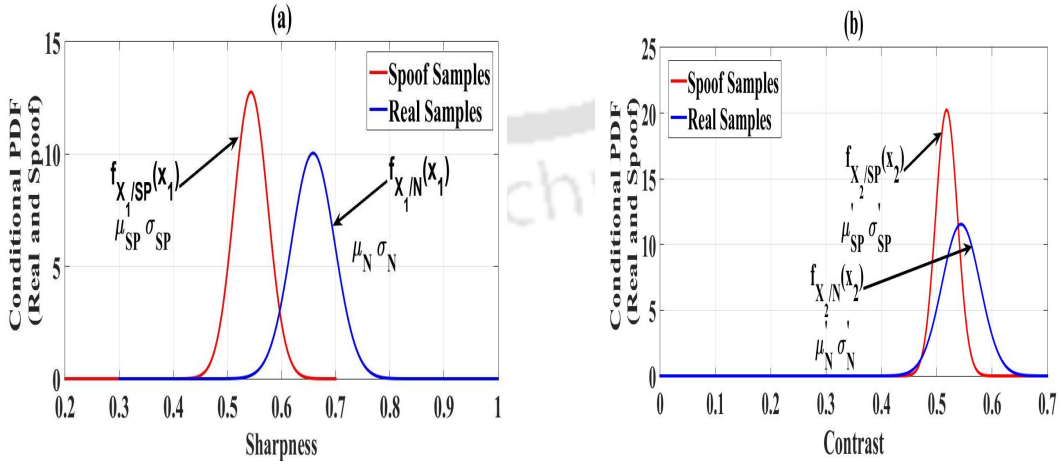

Figure 5.4: **(a)** The overlap between the conditional Gaussians is less when a sharpness metric [7] is used as a statistic. **(b)** The overlap between the conditional Gaussians increases considerably when contrast [3] is used as a discriminating statistic. Both the interpolated conditional histograms were computed for the CASIA dataset [9] (wherein the spoof-set comprised of printed photos).

- The moment we move towards differential features, the dependence on local illumination profiles and light source orientations drops substantially. Thus, class separation natural versus spoof improves considerably even in diverse datasets such as CASIA.

5.4.1 Random Scan Based Identity Dissolution

Consider two faces F_i and F_j, which are scanned in a certain fashion to produce a 1-D sequence. In the first case, we may assume the scan pattern to be fixed and to be of a "raster-type". Subsequently, the base feature vectors $\bar{X}_i$ and $\bar{X}_j$ become registered intensity profiles (Fig. 5.5). This arrangement would invite a significant perceptual interference. On the other hand, if pixels are scanned randomly through a spatial shuffle, both perceptual and structural information will be completely lost shown in Fig. 5.6. A balance between the two forms is a randomized, yet correlated scan. In the past, such correlated scans have been deployed to make encrypted videos compressible (post encryption). One type of scan pattern is a space filling curve(SPC) [48]. In our work we present our own interpretation of the SPC, in the form of a constrained two-dimensional random walk. The scan patterns for any two facial images F_i and F_j are expected to be distinct (irrespective of perceptual similarity or dissimilarity) to ensure identity dissolution. However, the first order, second order and third order pixel correlation statistics are conserved on a majority scale.

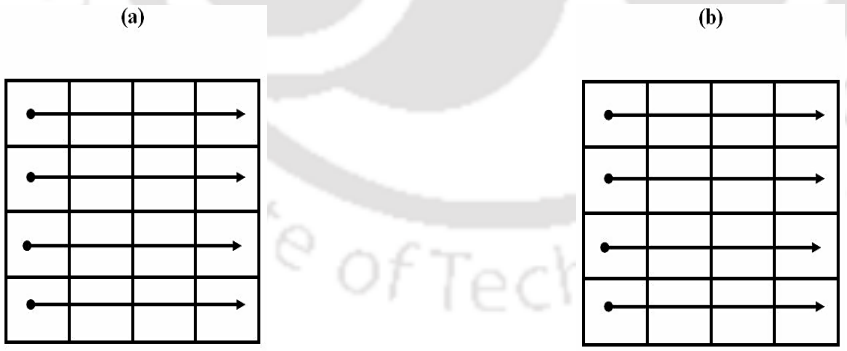

Figure 5.5: **(a-b)** Raster scan pattern for the two face images F_i, F_j.

Fig. 5.7, Fig. 5.8 shown are two typical sets of correlated scan patterns based on the random walk. While perceptual identity is lost in both unregistered feature vectors $\bar{X}_i$ and $\bar{X}_j$, format of the data captured is preserved. Gradient and sharpness features can now be computed on the top of this randomly scanned intensity feature. Let CScan$[F_i, \bar{k}_i]$ be a customized correlated scan with a key sequence $\bar{k}_i$. The key sequence carries information pertaining to the direction/ trajectory of the random walk. In case, there is an abrupt termination of the walk, the key sequence logs information related

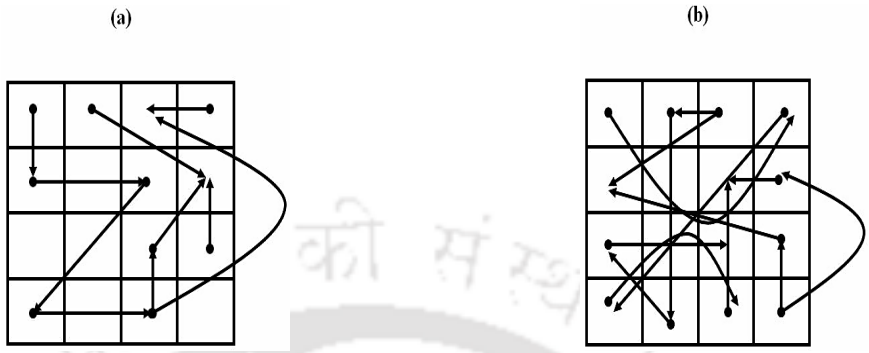

Figure 5.6: **(a-b)** Shuffle scan pattern for two images F_i and F_j

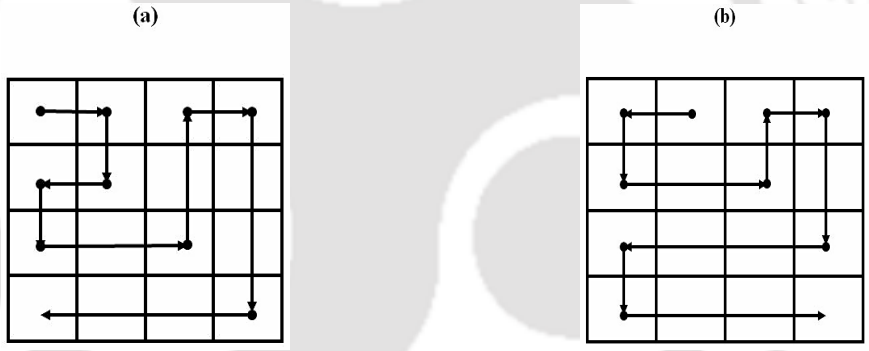

Figure 5.7: **(a-b)** Correlated scan for two images F_i and F_j respectively for instance 1

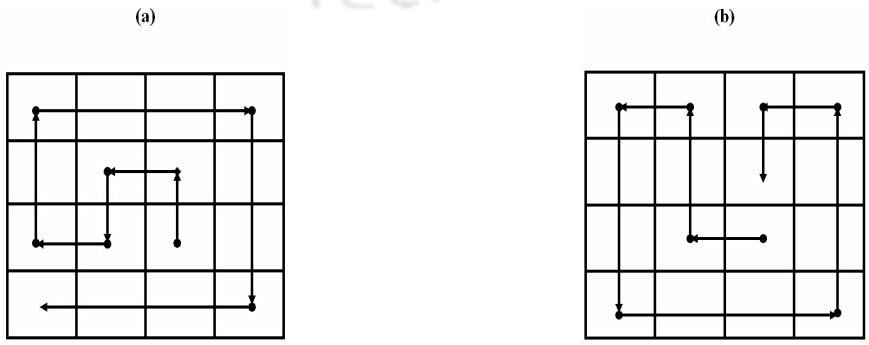

Figure 5.8: **(a-b)** Correlated scan for two images F_i and F_j respectively for instance 2

to the pixel jump. In a nutshell, the key sequence is a sequence of location pointers forming a linked list. Let $\bar{X}_i = CScan[F_i, \bar{k}_i]$ be the base unregistered intensity feature vector. To reduce scanning complexity, F_i is a down-sampled version of the parent facial image and $\bar{X}_i = [x_{i_1}, x_{i_2}, ...x_{i_n}]^T$. It is to be noted that not all the correlated scans are contiguous in terms of random walk. There will be several abrupt pointer terminations leading to random hops across the image grid.

5.4.2 Proposed 2D Random Walk Algorithm

The random scanning algorithm starts at the center of an odd size grid, and does a random walk based on the availability of free cells, with a depth of two hops. The algorithm not only checks which are the free neighborhoods with respect to the current pointer position, it also examines the status of the neighborhood of the immediate neighborhood. Nearest neighbours of PTR $\equiv (x_p, y_p)$ are cells ①, ② and ③ respectively. shown in Fig. 5.9.

$$① \equiv (x_p, y_{p-1})$$

$$② \equiv (x_{p-1}, y_p)$$

$$③ \equiv (x_p, y_{p+1})$$

We also have, $NN(PTR) = \{①, ②, ③\}$, Where$NN(.)$, denotes nearest neighbours.

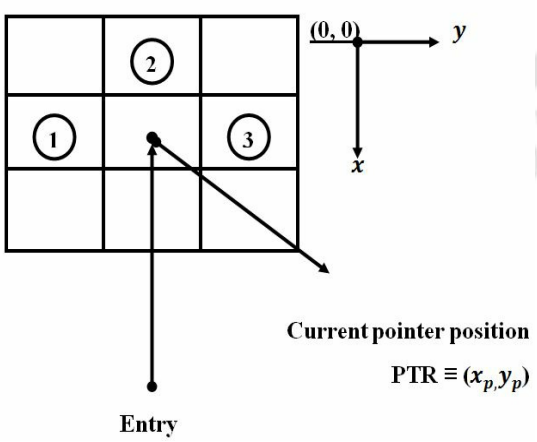

Figure 5.9: Nearest neighborhood of PTR $\equiv (x_p, y_p)$.

Define the scores of the nearest neighbors as:

$SNN(PTR) = scores\ of\ nearest\ neighbours\ of\ PTR$

$SNN(PTR) = \{S_1, S_2, S_3\}$

where $S_i \in \{0, 1, 2, 3\}$. The exit path is chosen by comparing the scores S_1, S_2, S_3.

$$e_s^* \in SET_e^* = \underset{i \in \{1,2,3\}}{ArgMax}\ S_i \tag{5.1}$$

If there is more than one solution, $e_s^* = Random\ selection\{SET_e^*\}$ the exit cell is randomly picked from the result of the previous stage SET_e^*. Note that $S_i = 3$, if cell i had 3 free neighbours; $S_i = 2$, if there are two free and so on... In case, there are no free neighbours $S_i = 0$. The whole process begins with an $N \times N$ grid, where N is odd. The PTR starts at $x_p = \frac{N+1}{2}$, $y_p \frac{N+1}{2}$. Each time a PTR moves to new cell that cell is flagged is 1. Initially, all cells inside the grid $N \times N$ except the center $(\frac{N+1}{2}, \frac{N+1}{2})$ are set to zero. Let TRACE (x, y) be the status of the random walk.

Figure 5.10: **(a-b)** Labeling in clock wise direction

INITIALIZATION

$$\begin{aligned}
&TRACE[x, y] = 0;\ x \in \{1, 2, ...N\};\ y \in \{1, 2, ...N\} \\
&(x, y) \notin [\frac{N+1}{2}, \frac{N+1}{2}] \\
&TRACE[\frac{N+1}{2}, \frac{N+1}{2}] = 1 \\
&x_p = \frac{N+1}{2};\ y_p = \frac{N+1}{2};\ [PTR_{START\ POS}]
\end{aligned} \tag{5.2}$$

REPEAT: TILL all grid points are covered.

Step-1: Determine $NN(x_p, y_p) = \{P_1, P_2, P_3\}$ labeling in clockwise direction shown in Fig 5.10(a-b).

Step-2: Determine the degree of freedom associated with P_1, P_2 and P_3, in other words compute the scores $S(P_1)$, $S(P_2)$ and $S(P_3)$, where $S(P_i)$ = Number of free cells adjacent to P_i.

Step-3: Compute:

$$SET_e^* = ARG\ \underset{i}{MAX}\ S[P_i]$$

Where, SET_e^* corresponds to the locations of cells having the highest scores or are maximally free. Note that $1 \leq |SET_e^*| \leq 3$ and at least one $S(P_i) > 0$.

Step-4: $PTR_{NEXT} = i^* \in SET_e^*$, provided,

$$IF\ Card[SET_e^*] = 1$$
$$i^* = SET_e^*$$
$$IF\ Card[SET_e^*] \in \{2, 3\}$$
$$i^* = random\ selection[SET_e^*]$$

In case, none of the neighbors are free,

$$IF\ SET_e^* = \phi[\text{ when }\ S(P_1) = S(P_2) = S(P_3) = 0]$$
$$PTR_{NEXT} = RANDOM\ HOP$$

where, Card(.) represents the cardinality or the number of elements in the set.

Step-5: Update TRACE matrix

$$TRACE[PTR_{NEXT}] = 1$$
$$\equiv TRACE[x_{P_{NEXT}}, y_{P_{NEXT}}] = 1$$

Step-6: RETURN to STEP-1

The contiguous random scan is a simpler version of the Space Filling Curve which deployed in 1987 to ensure compressibility of encrypted videos. The main idea in our interpretation is to prolong the walk as much as possible without biting into the tail of the same snake. Ideally, we wanted the pointer to expand and fan out as much as possible without curling into its own tail. There is also the issue of running into boundary walls as the image grid was finite in size.

Eventually, however, when this was deployed, we went with a uniform and completely UNBIASED crawling procedure, wherein the POINTER of the snake (or snakes including multiple hops) knows only two types of cells: TRESSPASSED ONES and UNUSED ONES.

5.4.3 Algorithm: UNBIASED CONTIGUOUS RANDOM WALK with some HOPS

- Resize image to SQUARE and ODD number of pixels $N \times N$.
- Initialize pointer at the centre of the image grid: Centre is at $(\frac{N+1}{2}, \frac{N+1}{2})$
- nitialize all $cells(x, N+1) = 1$ and $cells(N+1, y) = 1$ (MARKED BOUNDARIES, EAST and SOUTH assuming a standard image co-ordinate system located at the TOP-LEFT corner)
- All other cells are labelled 0.
- For PTR(x,y) in the range $[1, N]$, $[1, N]$ starting from the centre,
  - Define $NN(PTR) = N, S, E, W$ when PTR is at the centre.
  - NN(PTR) = Free cells among (N,S,E,W) when PTR is somewhere else.
- Given current position, PTR can move freely with equal likelihood to any one of the free nearest neighbours (N, S, E, W).
- f no FREE cells and IMAGE not fully scanned, do a RANDOM HOP to another FREE CELL and resume the walk.
- Repeat till complete image is scanned.

Fig. 5.11 illustrates the proposed random walk algorithm on a small scale (grid size 21×21). Even at this scale one can witness bends, several intermediate terminations and random hops. Since there are likely to be abrupt intensity transitions, when a specific trail encounters a termination point and starts off somewhere else on the grid, the scanned feature must be median filtered to suppress some of the singularities.

$$\bar{X_{MF_i}} = Medfilt[\bar{X}_i; w] \tag{5.3}$$

where w is window size of median filter and $w \in [3, 5]$. Feature vector $\bar{X_{MF_i}} = [x_{m,i_1}, x_{m,i_2}, ...x_{m,i_n}]^T$ where $x_{m,i_k} = median[x_{m,i_{k-1}}, x_{m,i_k}, x_{m,i_{k+1}}]$ for $w = 3$ and $k \in 2, .., n-1$. Once median filtered, intensity differentials can be computed to capture the sharpness diversity.

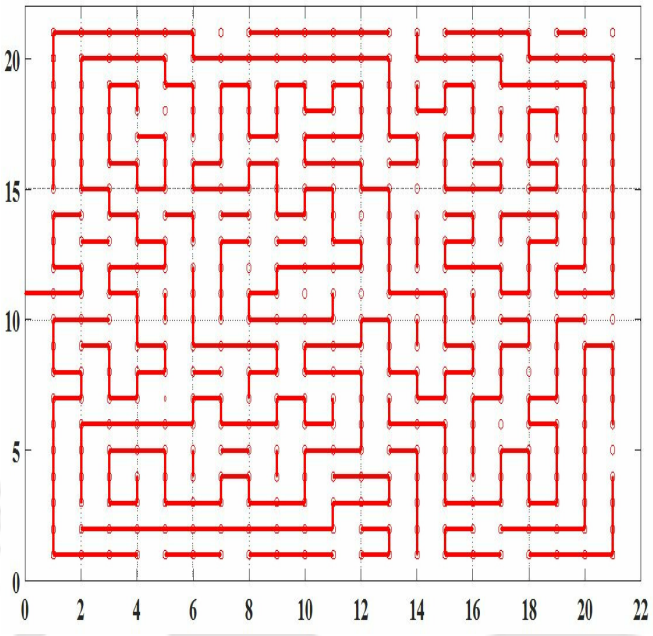

Figure 5.11: Random scan for patch size 21 × 21.

$$\bar{X_{D_i}} = Differentials[\bar{X_{MF_i}}, ORD] \tag{5.4}$$

where, the difference order is represented by $ORD \in 0, 1, 2, 3$ and $\bar{X_{D_i}} = [d_{i_1}, d_{i_2}, ...d_{i_n}]^T$. If $ORD = 0$, $d_{i_k} = x_{m,i_k}$; If $ORD = 1$, $d_{i_k} = x_{m,i_k} - x_{m,i_{k-1}}$; If $ORD = 2$, $d_{i_k} = x_{m,i_k} - x_{m,i_{k-1}} - x_{m,i_{k-2}}$. Thus, $\bar{X_{D_i}}$ becomes the final feature vector for calibration and tuning.

5.4.4 Application to facial dissolution

Each facial image F_i is first converted to gray scale and then resized to a grid of $N \times N$ (with N=101, odd number), keeping the computation complexity of the random scan in mind. A typical random scan implementation for $N = 21$ is shown in Fig 5.11. Observe in Fig. 5.11 that apart from long trails there are several short ones, because of abrupt transitions. Long trails tend to capture the statistics in the image better as compared to the shorter ones.

Fig. 5.14(a) shows the imapct of the random correlated scan on the perceptual quality of the image. If the 1D vector $\bar{X_{MF_i}} = Med\ filt[\bar{X_i}, w]$ is re-organized in the form of an image, the image will appear encrypted and distorted as shown in Fig. 5.14(b) for real image and Fig. 5.14(d) for spoof images. There is complete perceptual dissolution of the content, which indicates that the proposed approach is truly identity independent. The t-SNE plots of genuine faces from the database, genuine

test and spoof test faces for different discrete derivative orders are shown in Fig. 5.15. It is clear that maximum cluster separability and minimal intra class variance is observed for the first order derivative of the scanned vector $\bar{X_{MF_i}}$.

5.4.5 Feature Validation

To select and validate the right choice of feature vector for a given combination of parameters, we may express the final feature vector, $\bar{X_{D_i}}$ as a composition of several intermediate functionals,

$$\bar{X_{D_i}} = f_{SC}[\tilde{F}_i, \bar{K}_i, w, ORD] \tag{5.5}$$

Where $f_{SC}[.]$ is a composite function, including the correlated scan, median filtering and derivative process, with $w \in 3, 5$ and derivative order $ORD \in 0, 1, 2, 3$. The case $w = 1$ corresponds to an identity operation with respect to median filtering. For each combination of w and ORD and arbitrary $\bar{K}_i$ a slightly different condition distribution is produced for natural and spoof images. Consider a specific anti-spoofing database MSU-MFSD [2] described in Table. 5.1. In this Table. 5.1, variations within a particular subject class are in illumination direction change, scaling mild rotations and expression change. Let N_S be number of spoof samples and N_N the number of natural face samples. Construct the following parameters related to the feature vector or statistic $\bar{X_{D_i}}$:

$$\bar{X_{D_i}}|_{SP} \sim G[\bar{\mu_{SP}}, \Sigma_{SP}]$$

$$\bar{X_{D_i}}|_{N} \sim G[\bar{\mu_{N}}, \Sigma_{N}]$$

Where, μ is a n dimensional mean vector with $\bar{\mu_N}$ and $\bar{\mu_{SP}}$ representing the mean vectors for the genuine and spoof classes respectively. Σ represents an $n \times n$ covariance matrix, with Σ_N, Σ_{SP} being the covariance matrices for the genuine/natural and spoof classes. The Natural Space MODEL which is

Table 5.1: Description of a specific database namely MSU-MFSD [2]

MSU-MFSD	No of Subjects	No of poses/ Subject
Spoofed faces (Printed photographs)	35	50
Natural Faces	35	50

constructed using the RANDOM SCAN differential features based on the NATURAL FACE IMAGES, is expected to exhibit the following properties:

- It is expected to trap the depth profile in natural presentations reliably, across subjects.

- Being a differential statistic, the base random scan feature is expected to be illumination environment independent.
- With several random scans of the same face (i.e. auto-population degree high), the precision with which the ensemble differential feature set traps the depth and depth diversity profile increases considerably.
- This content agnostic scan feature is also expected to be immune to scale changes and subtle pose variations.

Both CASIA and MSU-MFSD were used for CALIBRATION and feature validation to check whether at the random scan feature level, the basic parameters remained stable. Tables 5.3 and 5.2 from the old-thesis draft indicated that registered features exhibit lower Mahalanobis-type distances (i.e. lower cluster separability) in comparison with random scan features. Feature validation was required to answer the following questions:

- Does the contiguous random scan really work in theory over registered features?
- How does one select the main parameter linked to the random scan (i.e. the differential order)? Should this be order 1 or 2 or 3 or higher?

The first question has been answered by the Mahalanobis-type cluster-distance scores generated for both raster scan features (registered features, Table 5.3) and random scan features (Table 5.2). Scores for the random scan features are much higher (cluster separability is promising).

The second part linked to the selection of the differential order had to be done to locate the differential order corresponding to the peak in Table 5.4 linked to the random scan features.

It was found that differential order '1' results in the highest Mahalanobis-type score across datasets (CASIA, MSU-MFSD and MSU-USSA). Go lower then one confronts illumination variability and interference. Go higher beyond '1' and one faces considerable subject-content variability. Thus, balance was found around an order of '1'. The overall testing procedure was split into three segments:

- CALIBRATION (both CASIA and MSU-MFSD was used)
- ONE-SIDED Training based on NATURAL SAMPLES alone and TESTING within the DATASET across both natural and spoof sets (again INTRINSIC model building and testing was done with respect to CASIA-CASIA and MSU-MSU)

- CROSS-VALIDATION (Natural face model with CASIA and TESTING on MSU-MFSD and vice-versa).

The last part linked to CROSS-VALIDATION was essential to check cross-porting of a model learnt in some other environment.

5.4.6 K L Divergence

In order, to measure the divergence associated with natural and spoof features (somewhat equivalent to the K-L divergence), we compute the distance

$D_1 = (\bar{\mu_{SP}} - \bar{\mu_N})^T [\Sigma_T]^{-1} (\bar{\mu_{SP}} - \bar{\mu_N})$, where, $\Sigma_T = \frac{\Sigma_{SP} + \Sigma_N}{2}$.

with Σ_T representing the total covariance matrix, which is a cumulation of the genuine and spoof covariance matrices. For the set of random scan features extracted using the relation given in Eqn. 5.5, the Mahalanobis distance D_1 is computed. Table. 5.3 gives the D_1 distance measure for different (w, ORD) out of which $w = 3$ and $ORD = 1$ yields the highest score. This implies that for the random scan algorithm, $w = 3, ORD = 1$ constitute the optimal parameter set. Along the same lines, when we extract fixed scan features for same set of constraints ($w = 3, ORD = 1$), the corresponding distances are tabulated in Table 5.2. When Table. 5.3 and Table. 5.2 are compared, the features from the random scan yield a higher score. This confirms the identity independent setting and we settle with the random scan features based on the parameter settings, $w = 3, ORD = 1$.

5.5 Proposed paradigm and architecture

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

Claim CH 5.3: We claim that most anti-spoofing systems work best in an identity independent setting, wherein the measurements or features extracted are taken in such a way that perceptual relevance is given the least importance. However, the residual correlation or some other statistics, which may be derived from this dissolved identity, carry necessary information regarding the environment or channel in which the information has been captured to perform anti-spoofing. This identity dissolution, in our case, is performed using a constrained shuffle of pixels in the spatial domain using a

2-dimensional random walk. This 2-D random walk has been inspired by Space Filling Curves [27], which was originally devised for retaining the compressibility of video signals after encryption.

Claim CH 5.4: By auto-populating the each facial image profile with these scans, it is possible to construct several variations of the same profile, which essentially carry the same pixel-correlation information, with minimal content interference. Thus a single facial profile is transformed into an ensemble of scans produced using independent 2-D random walk patterns. This ensemble carries significant information regarding the sharpness profile of the facial image and will have sufficient information to segregate 3D mask profiles from natural facial images.

5.5.1 Random scan algorithm

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

Fig. 5.12 shows two typical variations in the scan patterns, executed over the same image F_i. Eventually when the scan is completed, a 1-D vector $\bar{X}_i$ is generated as a function of the randomizer/key KEY_i:

$$\bar{X}_i = CSCAN(F_i, KEY_i) \tag{5.6}$$

where, $CSCAN(.,.)$, represents the randomized correlated scan algorithm based on a 2-dimensional random walk directed by a key sequence, KEY_i. While perceptual identity is lost in the unregistered

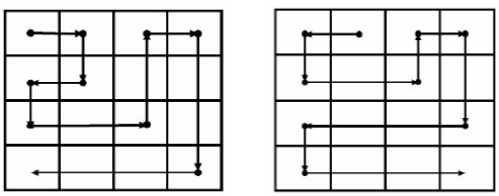

Figure 5.12: Random but correlated scans for same facial image F_i (two different walks executed on the same facial image).

feature vector $\bar{X}_i$ structure and format of the data captured is conserved. Gradient and sharpness

features can now be computed on the top of this randomly scanned intensity feature. The key sequence carries information pertaining to the direction/trajectory of the random walk. In case, there is an abrupt termination of the walk, the key sequence also stores information related to the pixel jump. In a nutshell, the key sequence is a sequence of location pointers forming a linked list. To reduce scanning complexity, F_i is a down sampled version of the parent facial image. Let $\bar{X}_i = [x_{i,1}, x_{i,2}, ..., x_{i,n}]$. It is to be noted that not all the correlated scans are contiguous in terms of random walk.

When the pointer in the 2-D random walk either walks into a corner or bumps into its own tail, it may encounter an abrupt termination. At this point, the pointer must hop to a new free cell within the same grid and resume the random walk. This process continues till all the pixels within the image grid are traversed. If $N \times N$ is the size of the image, the length of the scanned vector $\bar{X}_i$ is $n = N^2$ and the path length is $n - 1$ units. The walk is rectangular in nature and diagonal transitions are not allowed. Because of this, the primary scanned vector $\bar{X}_i$ must be median filtered using a 1×3 ($w = 3$) window, to iron out singularities. The scanned vector becomes smoother with a larger window size w at the expense of a loss of detail and an un-necessary alteration of natural pixel correlation statistics. It is important that the median filter does interfere with the accuracy of the natural image statistics. Hence, the optimal choice for w is three. The pre-processed statistic is given by,

$$\bar{X}_{MED,i} = MEDIAN[\bar{X}_i, w] \tag{5.7}$$

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

5.5.2 Final differential statistic

Based on earlier conjectures and observations, it is clear that the prosthetic arrangement is likely to have a smoother surface contour as compared to the natural face (partly owing to CLAIM CH 5.1 in Section. 4.1). This is based on the one-mask fits many assumption, the mask designed to dissolve the identity of the imposter (X), while emulating the identity of the target (Y), who is being impersonated. Hence, a simple differential feature which captures the first or second order pixel derivative, will be sufficient to discriminate between a natural face as compared to one which has a prosthetic. The natural face is expected to have a greater roughness (culminating in a greater and more heterogeneous sharpness profile) as compared to that of the prosthetic. Let $\bar{D}_{X,i}$ be the differential statistic computed on the median filtered 1D sequence. If $\bar{D}_{X,i} = [d_{i,1}, d_{i,2}, ..., d_{i,n}]^T$ and

$\bar{X}_{MED,i} = [x_{MED,i,1}, x_{MED,i,2}, ..., x_{MED,i,n}]$,

$$d_{i,r} = x_{MED,i,r} - x_{MED,i,(r-1)} \tag{5.8}$$

for $r \in \{1, 2, ..., n\}$ and with initial conditions, $x_{MED,i,(0)} = 0$. The vector, $\bar{D}_{X,i}$ is the final feature vector, extracted from the natural face image class alone, is fed to a one-class SVM [4] for characterizing the inlier space [3] (or the natural face space).

Table 5.2: Performance of fixed raster scan features with different specifications measured in terms of $log(D_1)$. Measurements are registered in space.

Method	MED Filter 1×1			MED Filter 3×3			MED Filter 5×5		
	CASIA	MSU-MFSD	MSU-USSA	CASIA	MSU-MFSD	MSU-USSA	CASIA	MSU-MFSD	MSU-USSA
n=0	29.2748	27.5097	26.4152	25.6972	27.8297	26.1409	27.4459	27.3714	24.1216
n=1	28.6931	23.7547	21.9418	29.3536	26.7972	29.1928	29.2719	26.9836	25.1412
n=2	26.9602	25.2325	22.8168	27.4580	27.7318	26.9710	20.1537	26.4615	30.7126
n=3	27.6434	26.6099	26.8128	29.0733	28.5846	26.7128	20.0624	21.0689	28.4146

Table 5.3: Performance of proposed random scan features with different specifications measured in terms of $log(D_1)$. This is an identity independent scan set.

Method	MED Filter 1×1			MED Filter 3×3			MED Filter 5×5		
	CASIA	MSU-MFSD	MSU-USSA	CASIA	MSU-MFSD	MSU-USSA	CASIA	MSU-MFSD	MSU-USSA
n=0	34.7132	35.3006	34.4264	34.807	36.3006	32.2472	35.9244	36.2600	34.0196
n=1	41.26	41.7222	40.8136	**45.0062**	**42.1048**	**42.6544**	41.6608	41.3704	40.3202
n=2	36.7148	34.0996	36.9064	37.9962	36.4998	35.9216	39.0524	37.9252	34.9546
n=3	37.511	36.0144	36.2368	38.5182	36.8546	39.9462	39.1704	36.0468	34.665

Criterion for identifying the derivative order and the median filter size to suppress spikes in the scanned data, is based on the largest registered Mahalanobis distance for three different databases.

5.6 Feature validation and training the one-class SVM

Feature validation is done by splitting the 3DMAD dataset [8] (composition given in Table. 5.4), into natural faces and prosthetic based images. The base feature used for this comparison is the norm of the final differential vector, $\bar{D}_{X,i}$, which is given by,

$$E_i = ||\bar{D}_{X,i}||_2 = \sqrt{d_{i,1}^2 + d_{i,2}^2 + ... + d_{i,n}^2} \tag{5.9}$$

The conditional distributions, $f_{E/NATURAL}(e)$ and $f_{E/SPOOF}(e)$ are computed on the same scale in Fig. 5.13, for the 3D-MASK dataset (these are essentially conditional histograms which have been interpolated to impart smoothness to the functions). In Fig. 5.13, the conditional distribution shown in blue corresponds to the genuine face space energy profile, while that shown in red corresponds to the

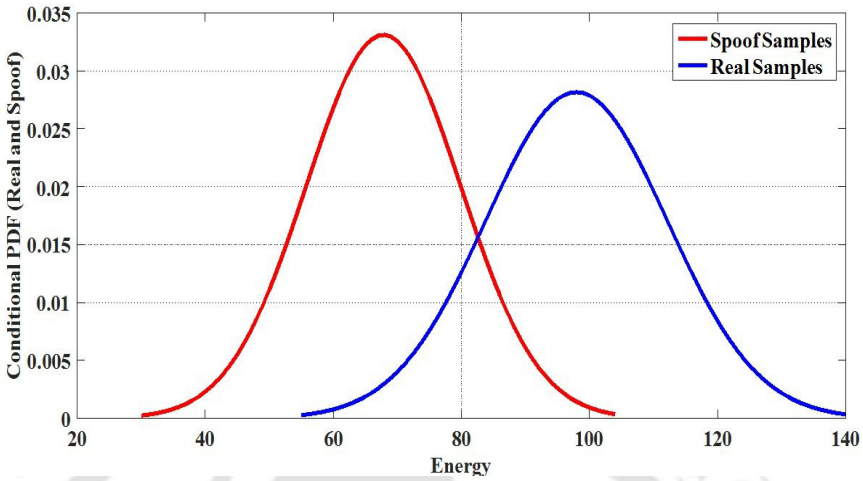

Figure 5.13: Conditional distributions of the differential energy feature for natural and spoof samples, computed from the 3DMAD dataset.

energy profile generated from the prosthetic samples. As expected, the differential statistics produced from the natural face space have a larger mean and larger variance (because of the increased roughness and intensity diversity), while that of the prosthetic shows a smaller mean and variance (owing to over-smoothing stemming from the one-mask fits all claim). While the conditional distributions

Table 5.4: Description and composition of 3D mask database [8]

3D Mask [8]	No of Subjects	No of poses/ Subject
Faces with 3D Masks	17	50
Natural Faces	17	50

demonstrate the feature separability and ability of the random scans to conserve the lower order correlation statistics present in the image, the impact of the of the random scan in obscuring the identity of the individual subjects is demonstrated in Fig. 5.16. Notice that the scanned versions presented for simplicity as a 2-D shuffled version in Fig. 5.16(b,d), have no resemblance to their corresponding un-scanned counterparts (Fig. 5.16(a,c)). Thus, the processing and feature extraction is done in truly an identity independent setting.

The set of natural faces from the 3DMAD database is split into a one-class training set for characterizing the inlier space and a test set which comprises of both natural faces as well as spoofed faces using the prosthetic. Given final differential base feature vectors $\bar{D}_{X,i}$, $i \in \{1, 2, ..., N_{GEN,T}\}$, where, $N_{GEN,T}$ is the number of training images from the genuine and natural face space. A one-class SVM [8], is constructed by building a hyper-sphere around the genuine multi-dimensional base differ-

ential features vector set corresponding to genuine face images, with an α-trim outlier fraction set to 10%.

5.7 Outlier detection frame

We could use the same transformed feature sets derived from natural faces through random scans, to train a one class SVM [4] [57]. The motive for training a one class SVM is the following:

It is anticipated that this feature set based on random scans contains minimal perceptual interference, while conserving natural scene linked information captured by the correlation profile. This correlation profile is tapped and amplified in the form of the final gradient feature vector $\bar{X_{D_i}}$. This scene-linked noise profile which is captured by the first, second and third order pixel correlation statistics can be used to segregate natural faces from planar spoofing.

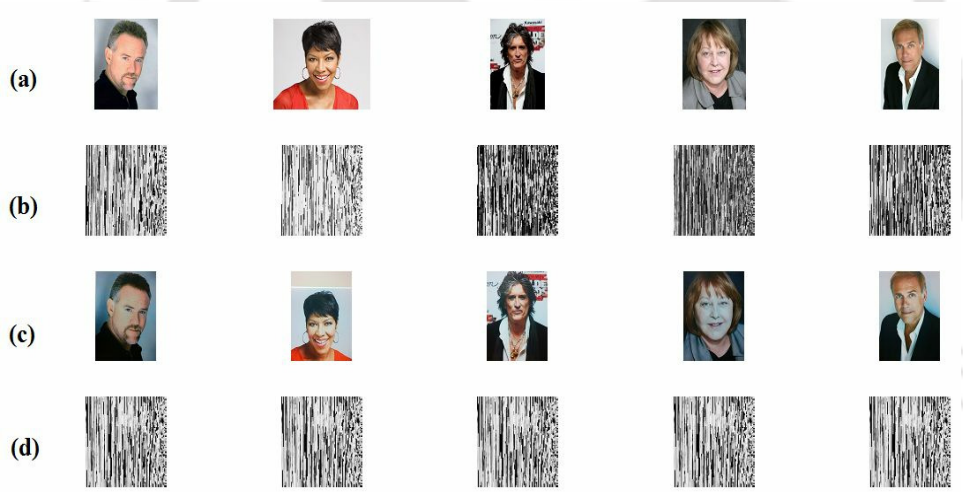

Figure 5.14: **(a)** Samples of real genuine faces recorded in live environment from the database analyzed in Patel at al. [10]. **(b)** Random scan features for corresponding real genuine faces. **(c)** Samples of images of printed photos. **(d)** Random scans extracted from these images taken from printed photos.

Once the base feature is crystallized, it is matter of characterizing the natural face space by pooling together the random scanned and filtered feature vectors $\bar{X_{D_i}}|_N$, where the sub-script 'N', refers to the natural face set. These feature vectors from the natural space form a constellation in $n-$dimensional space. The one-class SVM construction, builds a hyper-sphere over this constellation in such a way that the natural feature space is enveloped by a surface. This closed-surface is centered about the centroid computed over all the feature vectors used in training. The volume of the hyper-sphere and the shape/skew of the surface can be controlled by fitting the right form of kernel function

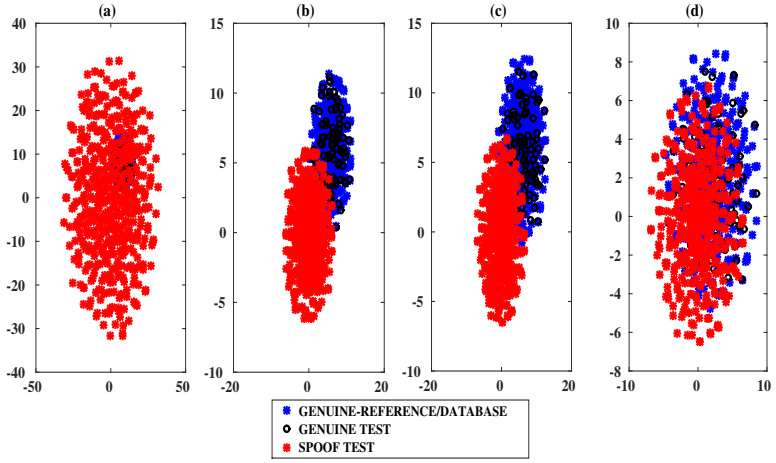

Figure 5.15: **(a)**TSNE plots of MEDIAN 3×3 with order of derivative $ORD = 0$ **(b)**TSNE plots of MEDIAN 3×3 with order of derivative $ORD = 1$ **(c)**TSNE plots of MEDIAN 3×3 with order of derivative $ORD = 2$ **(d)**TSNE plots of MEDIAN 3×3 with order of derivative $ORD = 3$. TSNE analysis computed for MSU-MFSD dataset [2]

which projects this lower dimensional feature data ($DIM_{orig} = n$) onto a higher dimensional space ($DIM_{target} = r); r > n$, such that the cluster points are uniformly distributed about the center in this higher dimensional space. The one-class SVM model is heavily customized according to the data-type and the type of statistic (or final feature vector). Here, we have chosen the RBF kernel for mapping the base feature onto a higher dimensional space. Three standard databases such as CASIA [9], MSU-

Table 5.5: Composition of databases used in our experimental analysis

Data bases	Genuine samples		Spoof samples	
	# subjects	Variations/ subjects	# subjects	Variations/ subjects
CASIA [9]	15	30	15	30
MSU-MFSD [2]	35	50	35	50
MSU-USSA [10]	1040	1	1040	1

MFSD [2] and MSU-USSA [10] were used for training and testing (Table. 5.5). A specific dataset was split into three sets (i) $D_{ref_{GEN}}$, set of genuine faces which were used for building the one-class SVM model; (ii) $D_{GEN-test}$; set of genuine faces which were used for testing and (iii) $D_{SPOOF-test}$; set of spoof faces (printed photographs or replayed images) used for testing. The fraction of genuine samples which used for inlier training was labeled as $x[\%]$, the remaining $(100 - x)[\%]$ was used for

testing. The entire spoofed set labeled by the fraction $y[\%]$ was used for testing. No part of the spoof set $D_{SPOOF-test}$ was used for constructing or even shaping the one-class SVM.

Given training data, the generation of the one-class SVM is thus a simple enough exercise, but requires one specific tuning parameter which is the trim-factor $\alpha[\%]$. Given a limited training set, it is always tricky to decide the radius of the hyper-sphere. If the radius is chosen such that almost all the points from the natural training set are engulfed by the hyper-surface, then α can be considered small (or virtually zero). Note that with this arrangement, points away from the centroid are less likely to be replicated as compared to those close to the center. Every training process in practice will involve select data samples which tend to exhibit extreme deviation from the normal structure. These can be termed as anomalous data points. These anomalous data points constitute the tail of this higher dimensional distribution and tend to simulate the incursion of a spoofing operation. The bigger question is what α-fraction may be deemed acceptable to characterize this tail. Too small a value of α leaves no room for understanding this tail while too large a value of α tends to bite into the inlier class leaving out normal samples. Thus an increase in α will invite false negatives while a decrease in α will increase the number of false positives. Between the two extremes is a compromise which can be witnessed in Table. 5.6 where the optimal $\alpha_{opt} = 10\%$, when this outlier fraction percentage was varied from $\alpha = 5\%$ to $\alpha = 20\%$. The Equal error rate (EER) for $x = 50\%$-natural space training, without involving any spoof samples was found to be $EER = 2.68\%$. This was for the MSU-MFSD [2] database.

5.7.1 Results with Auto-population

Since, we have the luxury of generating several randomized scans for the same image grid, it is possible to produce different scanned-versions of the same image. The every natural image can be scanned multiple times (N_S times) to produce several feature vectors. Each scanned vector presents a slightly different perspective regarding the original natural image. This, therefore amplifies the information pertaining to the natural face space. Beyond a certain number of scans per image, there is only a replication of information and there is no longer any new information. This type of auto-population is possible in our proposed identity independent setting and becomes very useful when the volume of training data is small. With the outlier trim factor crystallized and set as $\alpha_{opt} = 10\%$, we auto-populate the natural face set by generating N_S scans per image in Table. 5.7. The column with $N_S = 1$ in Table. 5.7 corresponds to the *NO auto-population* setting. The numbers for $N_S = 1$

Table. 5.7 (column-1) are therefore similar to the numbers in column-2 of earlier Table. 5.6. As

Table 5.6: Error rates for different α and database sizes with the MSU-MFSD database. The spoofing operation considered here is the printed photo attack.

Inlier ($x(\%)$)/ Outlier samples($y(\%)$)	EER@ $\alpha = 5\%$	EER@ $\alpha = 10\%$	EER@ $\alpha = 15\%$	EER@ $\alpha = 20\%$
$x = 10\%, (100 - x) = 90\%, y = 100\%$	11.9650	10.8039	11.7228	14.4198
$x = 20\%, (100 - x) = 80\%, y = 100\%$	10.3062	10.3557	11.5871	15.0497
$x = 30\%, (100 - x) = 70\%, y = 100\%$	7.1181	5.7552	13.2707	11.8132
$x = 40\%, (100 - x) = 60\%, y = 100\%$	7.0252	4.0280	9.6154	10.7266
$x = 50\%, (100 - x) = 50\%, y = 100\%$	5.4461	**2.6818**	8.9464	10.3142
$x = 60\%, (100 - x) = 40\%, y = 100\%$	4.3578	2.9455	6.9940	7.9579
$x = 70\%, (100 - x) = 30\%, y = 100\%$	3.3587	1.5297	4.2121	5.1661

For the row corresponding to the 50% referential test case, the lowest EER is registered when the inlier hypersphere radius for the 1-class SVM is calibrated in such a way that 10% of the natural face samples fall outside the hypersphere. This is a judicious trade-off between generalizability of the natural space versus suppression of the fraction of false positives with respect to the other unknown spoof class.

the number of scans N_S are increased for a given database-training fraction, the EER numbers drop and eventually saturate beyond a certain point. For $x = 50\%$, in Table. 5.7, the saturation point is $N_S = 20$, beyond which the results do not change significantly. The EER for $x = 50\%$ and $N_S = 20$ is 1.36%.

5.8 Comparison with the state of the art

We compare our proposed random scan based outlier detection algorithm with two papers: (i) The person specific (PS) anti-spoofing algorithm of Yang et al. [14] and (ii) Anomaly detection based anti-spoofing with single sided training of genuine faces, through a one-class SVM (Arashloo et al. [4]).

It is to be noted that in the person specific architecture of Yang et al. [14], both genuine/natural and spoof samples are required in the training phase. It is therefore a two-sided training procedure, likely to be more robust than single sided training architectures with respect to natural faces alone [3] [4]. One of the reasons why this is included in the comparison, is because it is an effective architecture when there are minimal pose variations. When the faces are registered in space, on a subject specific note, it becomes possible to learn the transformations from the natural face to the planar spoofed version of the same subject. This process when generalized, can be used to auto-populate the spoof set and grow the number of spoof samples. Comparison is done within the CASIA [9] database itself

Table 5.7: Performance evaluated with optimal parameter $\alpha_{opt} = 10\%$ with auto-populated samples (number of scans per image N_S varied) from the MSU-MFSD database. The spoofing operation considered here is the printed photo attack.

Inlier ($x(\%)$)/ Outlier samples($y(\%)$)	EER@ $\alpha = 10\%$				
	$N_S = 1$	$N_S = 5$	$N_S = 10$	$N_S = 20$	$N_S = 30$
$x = 10\%, (100 - x) = 90\%, y = 100\%$	10.8039	9.3076	9.3362	8.6838	7.303
$x = 20\%, (100 - x) = 80\%, y = 100\%$	10.3557	7.9208	7.7202	6.6667	5.1360
$x = 30\%, (100 - x) = 70\%, y = 100\%$	5.7552	4.4650	4.2437	3.9633	3.2021
$x = 40\%, (100 - x) = 60\%, y = 100\%$	4.0280	3.0973	2.2936	2.1690	1.3538
$x = 50\%, (100 - x) = 50\%, y = 100\%$	2.6818	2.8053	2.5330	**1.3644**	1.2743
$x = 60\%, (100 - x) = 40\%, y = 100\%$	2.9488	1.4807	1.2312	1.1440	0.9813
$x = 70\%, (100 - x) = 30\%, y = 100\%$	1.5297	0.9810	0.2722	0.2389	0.2731

Flagged EER represents the stagnation in the error rate beyond a certain number of random scans per image sample. Only the row corresponding to the referential 50% training and testing in the natural face space has been considered. Further increase in auto-population will not bring down the error rate much. This therefore gives a crude indication of the approximate number of scans per image required to tap the diversity in the face space for this specific database under the 50% testing scenario.

through a 50/50 split training/testing. The TEST(S) in Table. 5.8 considers the implementation on the CASIA dataset without auto-population and the use of projective transforms for producing the spoof samples. Best results are obtained for TEST(S) with the Histogram of Gradients (HOG) feature as $EER = 0.82\%$ [14]. The result is the same for both printed photo as well as digital image spoofing. However, this architecture has a fundamental flaw, wherein cross-validation is not possible since the subject profiles are expected to change in the new dataset. The model built using one dataset cannot be used for detecting spoofing in another dataset. For the TEST(T) which involves learning and then auto-populating a fraction of the spoof samples using projective transforms linked with natural faces, the error rate was found to be $EER = 2.26\%$ (printed photo and MSLBP) [14]. The proposed random scan algorithm however surpasses this with a lower EER score of 1.8920%.

On the other hand the work by Arashloo et al. [4] was tested on two databases CASIA [9] and MSU-MFSD [2], both with and without cross-validation. In the case of cross-validation the one-class SVM model devised for one natural face space (coming from a specific database), is tested as it is on another target database which has a completely different subject space. Since the subject profiles in the target set will not be the same as in the parent set, anti-spoofing is really tested on a person independent footing. The work by Arashloo et al. [4] does not deploy any transform or random scan

Table 5.8: Equal Error Rates (EERs) for both Intra as well as cross database testing: Comparison of the proposed random scan based algorithm with the state of the art, one-class [4] and two-class training methods[‡ [14]].

Features	Train	Test	Printed Photo	Digital Photo
IMQ+SVM [4]	CASIA	CASIA	23.07	28.00
		MSU-MFSD	25.20	38.05
	MSU-MFSD	MSU-MFSD	13.87	25.49
		CASIA	36.06	36.00
BSIF+SVM [4]	CASIA	CASIA	18.95	49.22
		MSU-MFSD	4.1	39.22
	MSU-MFSD	MSU-MFSD	11.69	4.91
		CASIA	35.19	12.56
LPQ+SVM [4]	CASIA	CASIA	14.71	45.67
		MSU-MFSD	3.51	40.42
	MSU-MFSD	MSU-MFSD	10.56	5.87
		CASIA	25.06	8.94
LBP+SVM [4]	CASIA	CASIA	15.11	48.68
		MSU-MFSD	4.77	33.60
	MSU-MFSD	MSU-MFSD	6.12	3.94
		CASIA	39.31	36.69
IMQ+SRC [4]	CASIA	CASIA	19.41	20.06
		MSU-MFSD	33.15	32.90
	MSU-MFSD	MSU-MFSD	13.44	8.10
		CASIA	31.44	30.06
BSIF+SRC [4]	CASIA	CASIA	13.65	35.90
		MSU-MFSD	5.15	23.64
	MSU-MFSD	MSU-MFSD	13.19	3.63
		CASIA	16.31	11.75
LPQ+SRC [4]	CASIA	CASIA	13.14	34.99
		MSU-MFSD	4.68	32.14
	MSU-MFSD	MSU-MFSD	26.06	6.50
		CASIA	36.69	15.75
LBP+SRC [4]	CASIA	CASIA	16.01	35.49
		MSU-MFSD	7.21	37.44
	MSU-MFSD	MSU-MFSD	6.75	3.37
		CASIA	43.63	40.69
MSLBP +SVM‡ [14]	CASIA (PS-iFAS)	Test(S)	5.60	5.60
		Test (T)	2.26	2.74
		Test	3.59	3.88
HOG +SVM‡ [14]	CASIA (PS-iFAS)	Test(S)	**0.82**	**0.82**
		Test (T)	5.045	3.08
		Test	3.35	2.175
RandomScanDerivative($N_S = 1$)+SVM	CASIA	CASIA	3.5122	4.1460
		MSU-MFSD	11.2327	14.2576
	MSU-MFSD	MSU-MFSD	2.6601	2.1266
		CASIA	3.0864	4.6398
RandomScanDerivative($N_S = 20$)+SVM	CASIA	CASIA	1.8920	2.1618
		MSU-MFSD	**3.3587**	**4.2121**
	MSU-MFSD	MSU-MFSD	**1.3436**	**1.1613**
		CASIA	**2.0127**	**2.7243**

~~Lowest registered EER rates for specific comparison segments (with or without cross-validation for different databases)~~ across both existing as well as proposed algorithms.

like ours. Instead, the anti-spoofing features are generated in a registered fashion over a fixed facial grid at different scales. For the CASIA-train, CASIA-test arrangement, best results for Arashloo et al. [4] were obtained for the feature/classifier LPQ + SVM of around $EER = 14.71\%$ (50% training from the natural face set and printed photo attack). However for digital image spoofing, the EER increased considerably to 45.67%. The reason for this alarming increase could be the following:

- Digital images of faces are usually back-lit and hence have a high contrast and minimal specular components. Hence, with a one-sided training procedure, with registered spatial measurements, it becomes extremely difficult to segregate the natural face set from the digital image set. This segregation however is possible either in a person specific setting [14] or with the proposed random scan arrangement witnessed in Table. 5.8.

For cross-validation with the MSU-MFSD database [2]: viz. training with CASIA and testing with MSU-MFSD (same SVM-model used there), the EER reported by Arashloo et al. was 3.51% for printed photo attack and 40.42% for digital image replay attack.

The highlight of the proposed algorithm is *identity independence, facilitated by the random scans, with a natural auto-population of genuine face samples*, eventually leading to the construction of a one-class SVM for characterizing the natural face space. The lowest EER scores for a particular dataset combination either intrinsic or cross-validated are highlighted in bold in Table. 5.8. The proposed algorithm registers the lowest scores for 6 out of 8 cases including the cross validations. The only two instances where the state of the art algorithms do better are for the intrinsic split CASIA-CASIA, wherein EER for (HOG feature and printed photo or digital image [14]) is 0.82%. The corresponding EERs for the proposed random scan algorithm ($N_S = 20$), are 1.8920 and 2.1618 respectively. The results are still promising considering the fact that this is purely based on single sided training. When compared with the state of the art one-sided training model of Arashloo et al. [4], the proposed random scan based outlier detection algorithm does better for both intra-database as well as cross-database checks. One thing to note is that while the one sided architecture of Arashloo et al. fails for digital image spoofing [for a majority of the test-cases] showing very high EERs, the proposed algorithm performs virtually uniformly for both forms of spoofing (printed photo as well as digital image spoofing). One of the reasons for this superior performance is because the spoofing-noise patterns are enhanced by the random correlated scans, while at the same time suppressing perceptual interference.

5.8.1 Importance of Cross-validations across datasets

In Yang et al. [14], every new subject added to the repository for a spoof-check, should have a distinct classifier model containing feature samples from normal faces and also exemplar samples (or synthetically projected spoof samples through domain adaptation). This model must be created each time a new subject is added. This becomes a computationally expensive affair. This problem does not exist in our frame for the simple reason it is identity independent. Hence, once trained on a sufficiently large number of natural facial samples, the model characterizes the natural face space on a holistic basis. The random scans offer the diversity and expand the natural feature space, while conserving the heterogenous blur characteristics [7], indirectly trapped to capture the depth map, through residual pixel correlations. **This not only permits cross-validations but also permits spoof checks for subjects not yet registered with the repository**. Thus, this frame once trained with sufficient number of subjects becomes both subject and subject type (to a certain extent, gender, age, cultural backdrop) independent. This is true to a large extent mainly because the pixel correlation statistics tend to capture the diversity in the blur profile stemming from a natural depth variation across the face at the time of imaging [7]. Planar photographs on the other hand are not expected to show the same diversity in blur profile and subsequently the spoof pixel correlation statistics will differ significantly from the natural face statistics.

Evidence of model transference and generalization (single shot training) is evident from the cross-validation results shown in Table 5.8. The CASIA set [9] comprised of subjects from East Asian origin (mostly Chinese), while the MSU-MFSD [2] dataset, had a heterogeneous subject mix with individuals from East Asia, Asia, Europe, Middle-East along with the local American population. Table. 5.8 clearly illustrates how a single trained model works on multiple datasets.

Row 11 of Table. 5.8, shows EER rates evaluated on MSU-MFSD [2] dataset when trained with the CASIA dataset [9] with our proposed random scan without auto-population ($N_{Scan} = 1$, diversification is minimal but the architecture induces identity independence). The converse with training on MSU-MFSD and testing on CASIA is also presented in presented. With auto-population, the EERs can be reduced further by generating $N_S = 20$ (Row-12, Table. 5.8) data-points (or feature vectors through independent scans) from the same image. This artificial increase in the number of data-points per image enhances the diversity of the natural face space and hence the results get better.

Case-1:Row-11(Table 5.8,PROPOSED RANDOM SCAN) $N_{Scan} = 1$ **wherein there is**

SUBJECT INDEPENDENCE, but no AUTO-POPULATION

The CASIA-CASIA scores for printed are 3.51%. However, the cross-validation results for printed photos (CASIA-MSU-MFSD) are poorer and show an increase in the EER as 11.23%. The generalization is better when the training is done on MSU-MFSD (which has a much more diversified subject class). The cross-validation results for printed photos [training with MSU-MFSD and testing on CASIA] are far better as compared to the previous case: The EER for this reverse cross-validation case is 3.08% (turns out to be better than the CASIA-CASIA testing, 3.51%).

The only way this lack of diversity in a particular parent database can be combated is through controlled randomization of data by executing independent random walks on the same image to produce $N_{Scan} = 20$ independent scanned versions which share similar correlation profiles but exhibit enough diversity to capture the variability in the face space.

Case-2: Row-12(Table5.8 Proposed random scan) $N_{Scan} = 20$, Auto-population with Subject-Independence

The cross-validation EER SCORES for CASIA training [the more culturally uniform dataset] and MSU-MFSD testing [the culturally diverse set], have now dropped to 3.35%, which is a great improvement.

5.9 Experimental results

The 3DMAD database [8], whose composition is presented in Table. 5.4, is split into three sets: (i) Genuine face set for training, x% of the total genuine face space; (ii) Genuine face set for testing, remaining $(100 - x)$% of the remaining genuine face space; (iii)Spoof samples ONLY for testing, from the paper-craft based prosthetic arrangement, $y = 100$% of the spoof set.

Selection of the trim factor in the one-class SVM α, is a careful tradeoff between extent of generalization of the natural face space versus weeding out spoof samples which are likely to be close in structure with respect to the natural space. With limited training samples, the need for generalization calls for an expansion of the hyper-sphere (or a reduction of α), while the urge to weed out almost all spoof samples as outliers, demands a compaction or a contraction of the hyper-sphere (or an increase in α). Either way there will be mis-classifications either in the form of false positives or in the form of false negatives. Somewhere in between there is compromise and this optimal trimming factor was found to be $\alpha = 10$%, as the outlier fraction. This is visible in Table. 5.10, wherein best results are

Table 5.9: State of the art comparison of difficulty levels associated with certain elements of the counter-spoofing architectural pipeline.

Method	MODEL TRAINING COMPLEXITY	MODEL ADAPTATION	TESTING COMPLEXITY
YANG-2015 [SUBJECT SPECIFIC 2-class SVM]	HIGH	REQUIRED and NARROW (cannot be generalized)	LOW
ARASHLOO-2017 [REGISTERED MEASUREMENTS 1- class SVM]	LOW	NOT REQUIRED but have limited scope Generalization because of registered measurements	LOW
Proposed [RANDOMIZED but Controlled Measurement Which preserve CORRELATION STATISTICS and 1-class SVM/Outlier detection]	HIGH	Not required at all because of SINGLE SHOT TRAINING (One dress if stiched carefully fits all)	LOW

obtained for a trim factor of $\alpha = 10\%$. For a specific inlier (genuine space) training fraction $x\%$ (viz. along a specific row in Table. 5.10), the Equal error rate (EER) decreases and then increases when α is varied from 5% to 20%, with the minima hovering around $\alpha = 10\%$. Note that in this table no auto-population is done using the random scans. The EER therefore is slightly on the higher side for $x = 50\%$, 50% natural face samples for training, wherein the minimum EER (corresponding to $\alpha = 10\%$) was found to be 2.25%.

5.9.1 Auto-population results and comparisons

It is natural to deploy the proposed random scan tool to derive statistically equivalent but identity independent representations of the same natural facial profile. Thus, every natural face training image is converted into an ensemble of scans which carry equivalent statistical information pertaining to the lower order pixel correlation profile. Since the walk is randomized each realization of the parent image is

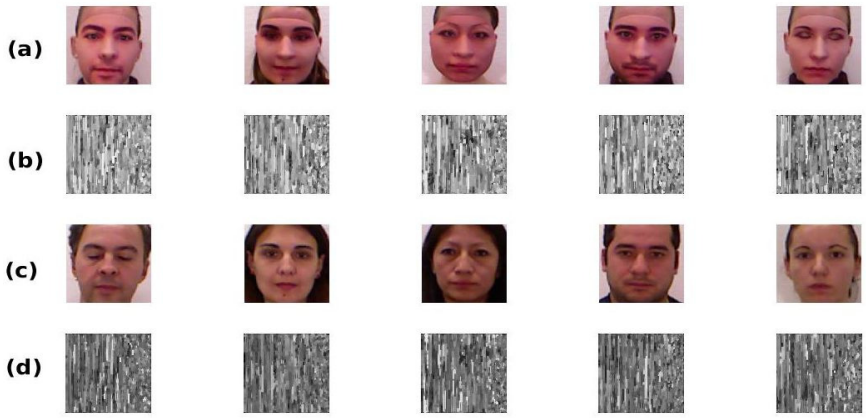

Figure 5.16: **(a)** Samples of 3D mask faces **(b)** Random walk features extracted for mask faces **(c)** Samples of real genuine face **(d)** corresponding random walk features.

Table 5.10: Error rates for different trim factors α and database splits with 3D Mask set. Note that best results are obtained for $\alpha = 10\%$.

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

distinct and provides a unique perspective. Results are therefore expected to improve considerably with this form of auto-population. The effective number of training samples is magnified by a significant amount, viz. by a scale factor N_{SCAN}. Impact of different ensemble sizes (or scale factor N_{SCAN}) is shown in Table. 5.11. Note that $N_{SCAN} = 1$, corresponds to results without auto-population and the EER numbers are expected to drop from left to right along a specific row. Saturation is expected beyond a certain point as the additional scans carry no new information for characterizing the inlier space. For $x = 50\%$, 50% face training, the lowest EER is obtained for $N_{SCAN} = 20$, highlighted in bold in Table. 5.11, with a percentage of $EER = 0.43\%$, which is way below the number obtained in the same row corresponding to $N_{SCAN} = 1$, which is, $EER = 2.25\%$.

A fair comparison is possible only when the state of the art algorithms are compared on an image analysis front (with or without implicit auto-population) but applied to the 3DMAD database. It is unfair to compare video processing algorithms which attempt to detect liveliness in faces by examining wrinkle and crease line dynamics to track consistency in emotional transitions of subjects. The only paper that fits this constraint is the original work by Erdogmus et al. [8]. Both the random scan

versions of the proposed algorithm with and without auto-population out-perform the state of the art. This validates the identity independent paradigm.

Table 5.11: Performance with optimal trim factor, $\alpha = 10\%$ and auto-population using the proposed random scan algorithm. EER results saturate beyond a certain point.

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

5.9.2 Computational Challenges

There are three processes (not necessarily all inclusive) to the anti-spoofing architecture represented in Table. 5.9.

- Model Training
- Model Adaptation when new subjects are included
- Testing

Model training is always done offline. This usually involves feature extraction and deriving measurements from the training set (which may or may not include exemplar spoof samples). However, when the number of training samples are limited, some architectures such as YANG-2015 and the proposed random scan one (shorter version in [26]applied to Prosthetics, 3D-MAD dataset), tend to deploy data-synthesis or auto-population to diversify the limited dataset in the repository. To re-create a natural feature or a spoof feature with similar statistics, yet with enough variability to qualify as another independent feature, one of the following needs to be done:

- Use regenerative neural networks [58], to replicate data, albeit with some variability to cover for pose, illumination, expression changes etc.
- Make use of the rigid geometric constraints associated with PLANAR SPOOFING models to derive planar-projections of natural faces by learning the subject-specific transformations [14].
- Recreate statistically equivalent yet independent variations of the same data (single feature gets mapped to multiple features), by deploying a form of controlled randomization which dissolves

perceptual information but conserves lower level information such as pixel correlation statistics. This is done in our present work and one of our recent publications [26].

Data-synthesis is a computationally intensive procedure, an optimization problem driven to meet two conflicting objectives: Conserving statistical similarity while ensuring diversification or natural variability in the synthesized feature. But, since this is usually done offline, this will not impact the speed associated with the counter-spoofing algorithm. However, in cases where the subjects do not possess exemplar spoof-samples for 2-sided training apriori [14], this auto-population must be done on the fly and on an on-demand basis and slows down the spoof-checking procedure. This is not a problem for the proposed random scan architecture as training is single-shot and one time. The same model can be tested on subjects not present originally in the training set and is also applicable to different image acquisition environments, provided the training set is sufficiently diverse.

The three architectures [14], [4] and the Proposed Random scan are compared and summarized on three parameters: Training, Adaptation and Testing in Table. 5.9. The higher initial single shot training cost for the proposed architecture comes with benefit of excellent generalization across a wide variety of datasets. The testing cost for all the three architectures is small as this is simply a binning problem.

Table 5.12: Comparison with the state of the art, which has used the 3DMAD dataset as a sequence of images. Training fraction, 50% from the natural face space.

Algorithm	Classifier	EER %
Ergodomus et al. [8]	SVM	4.92
Wang et al.. [59]	CNN	4.55
Proposed Random Scan (NO auto-population $N_{SCAN} = 1$)	SVM	**2.25**
Proposed Random Scan (with auto-population $N_{SCAN} = 20$)	SVM	**0.4310**

5.10 Conclusions and discussions

This work proposes an identity independent architecture for face anti-spoofing, wherein natural face images are scanned and auto-populated using a 2-dimensional random walk. This random walk minimize content interference connected with the subject's identity and focusses on the pixel correlation profiles. Since this process stays identity independent, the model developed for characterizing natural faces in one database can be tested on other databases with alternate subject profiles. The proposed algorithm outperforms the state of the art (Arashloo et al. [4]) for all the cross-validation cases.

In terms of error rates, the proposed random scan work matches the person specific analytical work of Yang et al. [14]. The work of Yang.et al. however has other limitations related to the practicality of implementation. In organizations where the subject profiling varies on a day-to-day basis and the repository undergoes continuous changes, it is extremely difficult to add/train and drop new person-specific models this frequently. One reason for this is because the spoof-model may either not be available at all or it may not be feasible to generate images of photo prints all the time for new subjects. When individuals leave, their models need to be decompiled and discarded for privacy reasons.

The proposed random scan based algorithm and architecture which operates in the subject independent domain and on real faces has the following advantages in a practical setting:

- Once a one-class model is trained on real faces across a number of subjects, with different pose variations, it can be used as a generic benchmark for picking up outliers, not just for same facial sub-space but also for subjects not necessarily a part of this setup. If a new subject is included in an organization, his/her face or facial features need not be included in the anti-spoofing module. Hence, for a sufficiently large natural face training set carrying sufficient variations in pose and illumination, re-training or model-adaptation need not be done to include more subjects. While the frame of Arashloo et al. [4] satisfies this criterion (since cross validation was done for their architecture), it fails on two counts: (i) Because of registered spatial measurements, their method incurs high error rates for planar digital image spoofing; (ii) Their architecture is unlikely to work on databases involving subjects with pose variations;

- The other major advantage of the proposed random scan is its impact on conserving the privacy of the individual's facial profile. Since the random walk is a form of constrained shuffle designed to preserve correlation statistics, privacy of the subjects is preserved. This can witnessed in the extreme perceptual secrecy visible in Fig. 5.14(b) and Fig. 5.14(d).

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

6

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

Contents

Objective

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 subjects and faces are registered, secondary life trail differential 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. A simple statistical model leading to an an elaborate tuning procedure, based on a reduced set of training images was developed to first identify the optimal parameter set corresponding to a specific dataset and then adapt the feature-vectors so that the error-rates were minimized. 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.

6.1 Introduction

There are many applications, particularly involving smart phones, where, prosthetic based spoofing is unlikely. This is mainly because the customized design 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 [60](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 [16] [26]. 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.

6.1.1 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 [61], [62], [7]; (ii) The reprinting process to synthesize a planar print brings about a progressive degradation in contrast [3], clarity, specularity [25], quality [4] or color-naturalness [21].

One type of statistical lens for detecting planar spoofing is a specularity check [24]. If the paper printing of the target's face is done on a glossy type of paper, this results in a dominant specular component [24] [25] 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 [24]. In Emmanuel et al. [40], 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. [25]. 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 [3] or sharpness (or its opposite blur) [7], [61], [62].

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 [3]. This reduces the dynamic range in the intensity profile considerably, eventually resulting in a more homogeneous contrast profile through-

out the image. This contrast homogeneity can be measured by fusing local contrast statistics, using a global variance measure [3]. 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 [61] [22] [7] in the captured image. In the work of Kim et al. [61], 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. [61] to detect planar spoofing.

In another blur variability detection procedure [22], 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 [63], 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 [63]), 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 [63].

- 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" [16].

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 [21]. 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 [64]. 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 [9] was found to be on the lower side.

6.1.2 Counter-spoofing based on Image Texture and Quality Analysis

It was proposed in Maatta et al. [65], 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) [65] [64] [14], 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 [66]. 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 [20]. 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 unstable (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.

6.1.3 Mixed bag techniques

Apart from model based approaches, in Wen et. al. [2], 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 [65], 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 [16] 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 [64] [14] tend to outshine the subject independent ones [3] [4].

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 [36, 53]. 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" [16].
- 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.

6.1.4 Subject mixing noise

Overall, in the approaches discussed so far, features connected with intensity, contrast [21], [3], blur/sharpness [62], [7], specularity [25] 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 [16]. 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 [64] [14].

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.

6.1.5 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 [16] [26].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 [16] 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 [16]. 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 [26], 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.

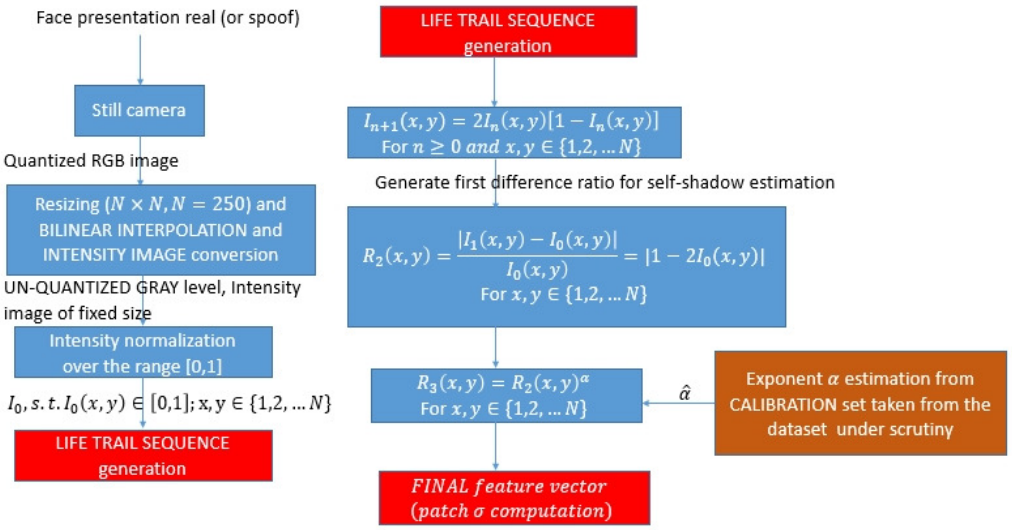

Figure 6.1: Block-diagram of feature extraction procedure

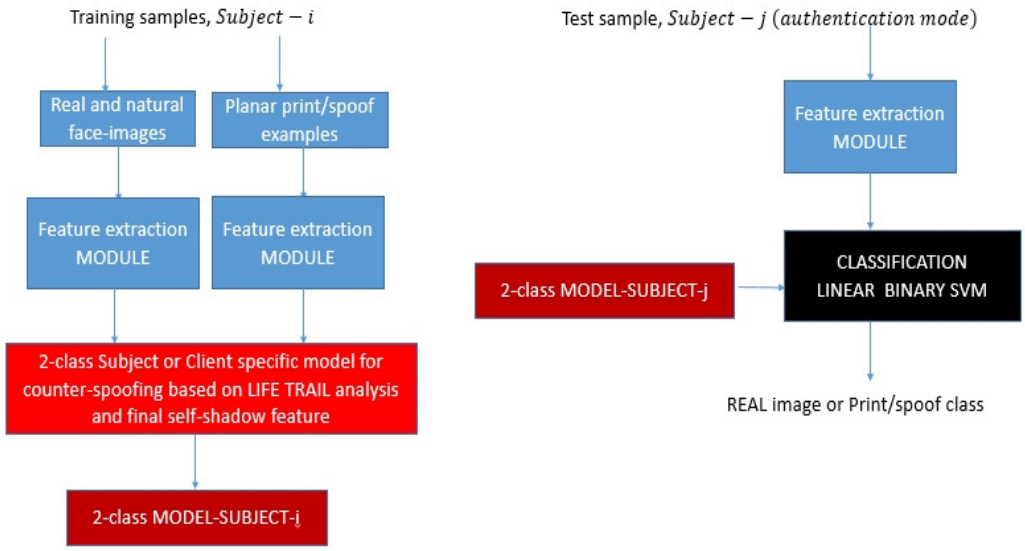

Figure 6.2: Block-diagram of TRAINING and CLASSIFICATION/DETECTION module.

6.1.6 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 [14] [64], 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 [67]. 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. 6.1) and (ii) The training, subject-specific model building and final testing procedure shown in Fig. 6.2.

The section-specific organization is as follows: The proposed self-shadow formulation, which is facilitated by a non-linear iterated function mapping [67], is discussed in detail in Section. 6.2 along

with an analytical model. The analytical model is directed towards the selection of the optimal exponential parameter α, when coupled with measurements taken across some exemplar images. This calibration protocol (phase-1) is discussed in Section. 6.3. Feature extraction and secondary statistics which sit on top of the enhanced self-shadow images are discussed with analytical justifications and preliminary classification results in Section. 6.4. Finally, a second level calibration phase followed by positioning and error rates of the proposed architecture in relation to the state of the art based on test-results, are presented in Section. 6.5.

6.2 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. 6.3 and Fig. 6.4. 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. 6.3.

The images captured with this arrangement for three different source locations are shown in Figs. 6.4(a,b,c). In Fig6.4(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 6.4(b), the source is positioned towards the left of the person and partly in front (west position), while in Fig. 6.4(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).

Claim CH 6.1: The first claim is that these self-shadows can be enhanced by first deploying an

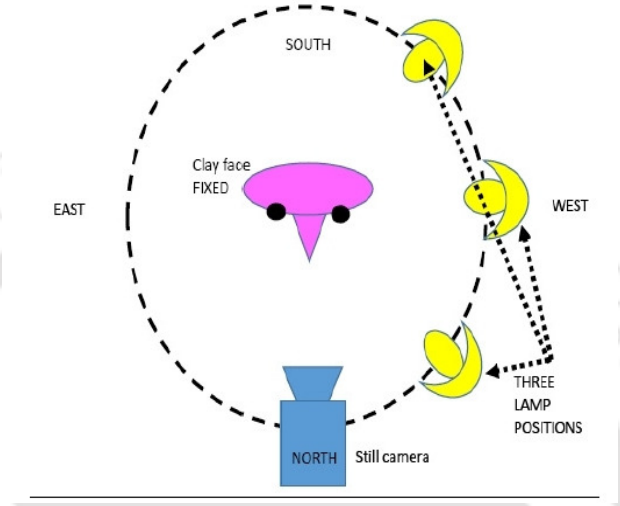

Figure 6.3: Experimental setup using a clay model and a fixed cell-phone camera for producing natural images with self-shadows.

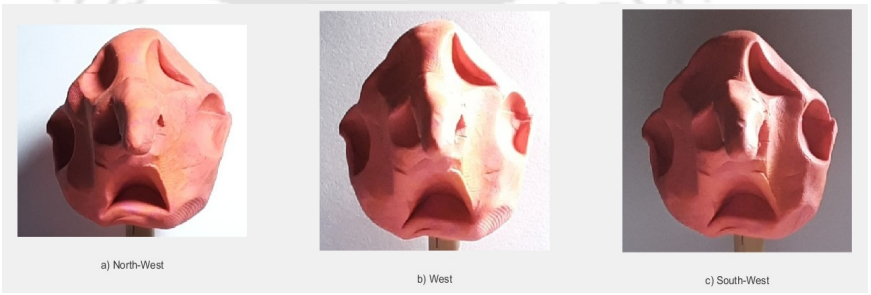

Figure 6.4: Images captured using the experimental setup (Fig. 6.3), for three different table lamp positions (north-west, west and south-west)

.

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 CH 6.2: 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 [67] 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).

6.2.1 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 [67],

$$I_{n+1}(x, y) = 2I_n(x, y)(1 - I_n(x, y)) \tag{6.1}$$

with the initial value, $I_0(x, y) \in (0, 1)$ and $I_n(x, y)$ is the value at the $n^{th}, n > 0$ 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 images (Fig. 6.7). 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).

6.2.2 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.

It was surmised in [3], 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) [68]. 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,

$$I_{NZ}(k) = I_0(x, y); \text{ provided } I_0(x, y) > \epsilon_1 \tag{6.2}$$

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

$$\begin{aligned} \mu_G &= \frac{1}{M}\sum_{k=1}^{M} I_{NZ}(k) \\ \sigma_G &= \sqrt{\frac{1}{M}\sum_{k=1}^{M}(I_{NZ}(k) - \mu_G)^2} \end{aligned} \tag{6.3}$$

The final contrast score can be computed as [3], with a slight modification to account for images with very dark foregrounds:

$$CON = \frac{\sigma_G}{\mu_G} \tag{6.4}$$

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. 6.5. 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 [3], 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,

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

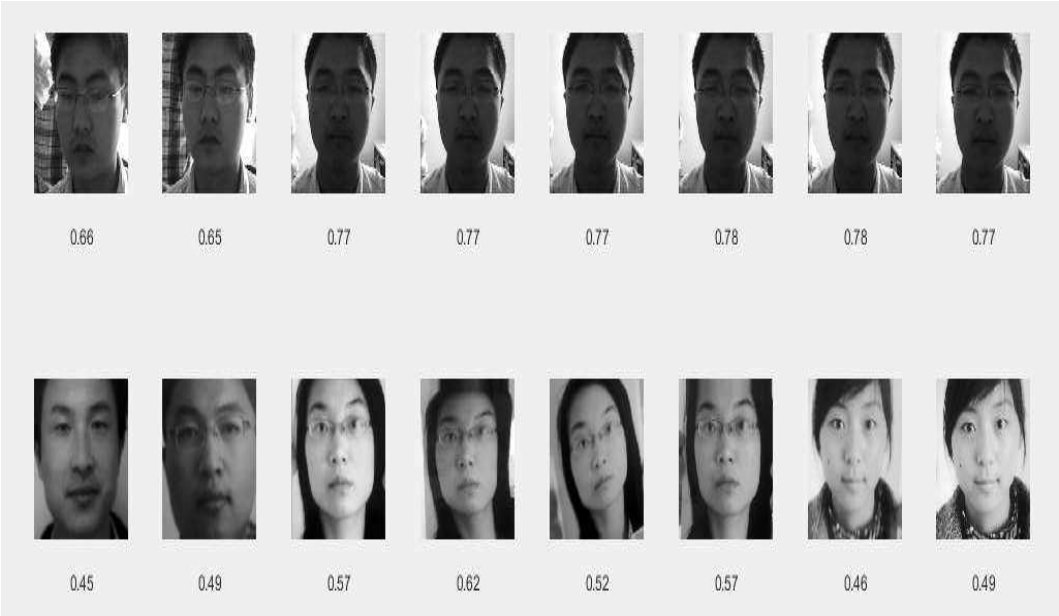

Figure 6.5: Real and planar prints with contrast scores as per Eqn. 6.4).

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. 6.6. 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.

6.2.3 Fixed point analysis based on a simple statistical model

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. To capture this mathematically in a crude way and to use this basic formulation to direct many important aspects connected with the model building process, we consider the following frame:

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,

$$Y_0 = X_0^{\gamma} \tag{6.5}$$

Figure 6.6: 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 $\gamma = 1$ (no transformation) and for $1.5, 3, 5$

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 [67] 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,

$$X_n = 2X_{n-1}(1 - X_{n-1}); n \geq 1 \tag{6.6}$$

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

$$f_{X_n}(x) = \left(\frac{1}{2}\right)^{n-1} \left[\frac{1}{\sqrt{1 - 2x}}\right]^n \tag{6.7}$$

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,

$$X_n \rightarrow 0.5 \text{ with prob. '1' for large } n \tag{6.8}$$

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:

$$\frac{1}{2} - Y_n = 2\,(Y_0 - 0.5)^{2n}\,; n \geq 1 \tag{6.9}$$

Let,

$$Z = (Y_0 - 0.5)^2 \tag{6.10}$$

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

$$Q_n = Z^n; n >> 1 \tag{6.11}$$

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

$$f_{Q_n(z)} \longrightarrow \delta(z) \text{ for } n >> 1 \tag{6.12}$$

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

$$Prob(Z^n \longrightarrow 0) \longrightarrow 1; n >> 1 \tag{6.13}$$

This implies that based on Equations. 6.9 to 6.11,

$$Prob(0.5 - Y_n \longrightarrow 0) \longrightarrow 1; n >> 1 \tag{6.14}$$

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

$$Y_n \longrightarrow 0.5 \text{ with probability '1' for } n >> 1 \tag{6.15}$$

6.2.4 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 (modelled by random variable X and print abstractions (modelled 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,

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

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

Furthermore, it can be shown that,

$$\begin{aligned} G_n - G_{n-1} &= 2[X_{n-2} - X_{n-1}] \\ &= 2\left[\frac{X_{n-2} - X_{n-1}}{X_{n-2}}\right] X_{n-2} \quad (6.16) \\ &= -2G_{n-1}X_{n-2}; n \geq 2 \quad (6.17) \end{aligned}$$

Can show that,

$$G_n = G_{n-1}(1 - 2X_{n-2}) = G_{n-1}^2; n \geq 1 \quad (6.18)$$

It follows that,

$$G_n = (G_0)^{2n}; n \geq 1 \quad (6.19)$$

Similarly,

$$H_n = (H_0)^{2n} = (2Y_0 - 1)^{2n}; n \geq 1 \quad (6.20)$$

as $n \geq 1$. Where, G_0 and H_0 are initialized as

$$G_0 = \frac{X_0 - 0.5}{0.5} = (2X_0 - 1) \quad (6.21)$$

$$H_0 = (2Y_0 - 1) \quad (6.22)$$

It can be easily seen that the PDF of difference ratio, random variable, G_0 is $f_G(g) = \frac{1}{2}$ for $g \in [-1, 1]$ and that the PDF of its print version counterpart is H_0 is $f_H(h) = \frac{a}{2}$ for $h \in [-1, (-1 + \frac{1}{2a})]$. Now, let the expected value, $E[H_n] = \mu_H(n)$. Given $E[H_n] = \mu_H(n)$ and $Prob(H_n < 0) < 0$ for $n \geq 1$ it

can be shown that,

$$E[H_n] = \int_{u=0}^{u=1} \left(\frac{2u}{a} - 1\right)^{2n} du \tag{6.23}$$

$$= \frac{(\frac{2}{a} - 1)^{2n}(a - 0.5) + a/2}{2n + 1} \tag{6.24}$$

Taking the limit as $a \longrightarrow 1^+$,

$$E[H_n] \longrightarrow E[G_n] \tag{6.25}$$

$$\longrightarrow \frac{1}{2n + 1} \tag{6.26}$$

Define a ratio-statistic as,

$$R_{HG}(a) = \frac{E[H_n]}{E[G_n]} \tag{6.27}$$

It can shown that this ratio statistic, is monotonic increasing, by computing and simplifying the derivative,

$$\frac{dR_{HG}(a)}{da} = 0.5 + \left(\frac{2}{a} - 1\right)^{2n} \left(1 + \frac{2n}{a}\right) \tag{6.28}$$

which strictly positive. Also since,

$$R_{HG}(a \longrightarrow 1^+) = 1 \tag{6.29}$$

it follows that,

$$E[H_n] > E[G_n] \ \forall a > 1 \tag{6.30}$$

which implies that the convergence rate of a natural-version is much more slower as compared to that of a print-version.

6.2.5 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

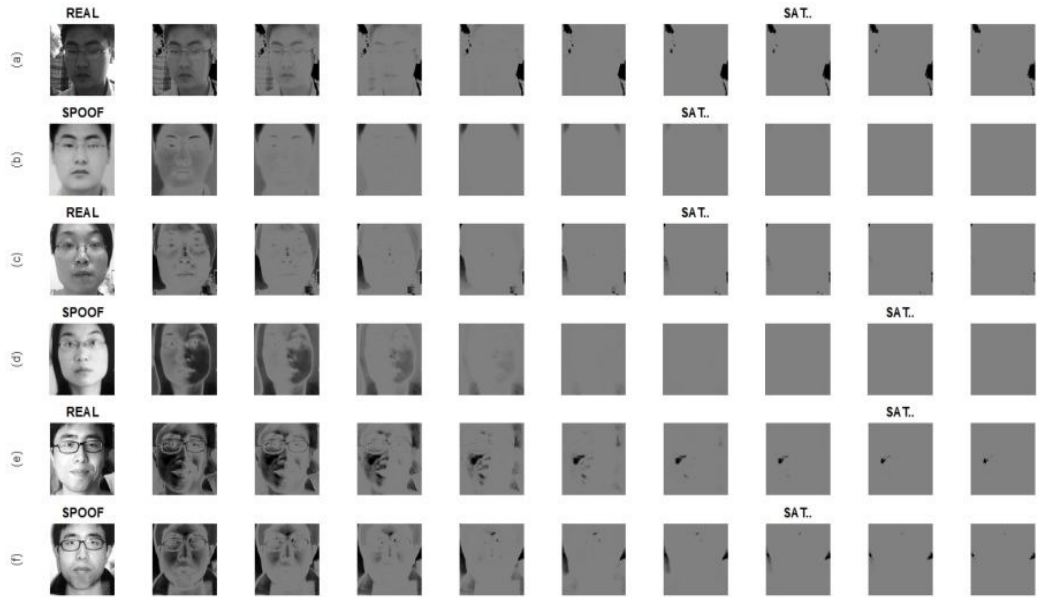

Figure 6.7: Contrast reductionist life trails for real and spoof image samples using the logistic map.

a saturation point if,

$$\frac{|I_n - I_{n-1}|}{I_n} < \epsilon \tag{6.31}$$

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. 6.31 individually. This swarm convergence trend has been tapped using a saturation curve based on a function $P(n)$ (Fig. 6.8), where n is the iteration number. Typical saturation curves for natural and spoof images are shown in Fig.6.8.

Fig.6.7 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

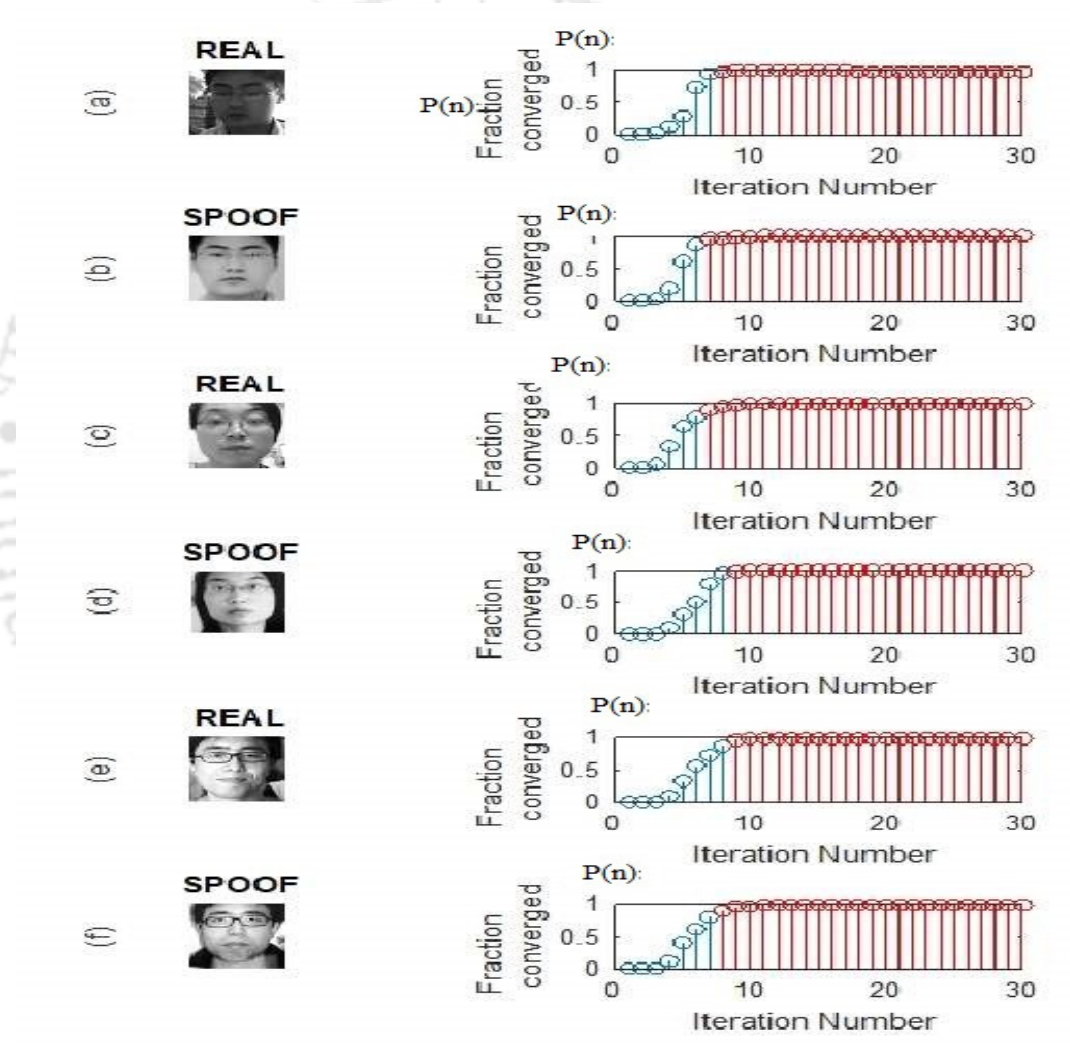

Figure 6.8: Saturation curves which bring out the trends linked to the rate at which the initial image samples (either natural or spoof) converge to a zero-contrast image in the life-trail.

print counter parts. The planar print versions tend to usually be of a lower quality, typically lower contrast [3] and limited color [64] 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.6.7.

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. 6.7(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 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.

6.2.6 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

Figure 6.9: TWIN-image where one version is taken under normal outdoor lighting and the other one under diffused sunlight.

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

$$I_1(x, y) = 2I_0(x, y)[1 - I_0(x, y)] \tag{6.32}$$

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

$$R_1(x, y) = |I_1(x, y) - I_0(x, y)| = |I_0(x, y) - 2\left[I_0(x, y)\right]^2| \tag{6.33}$$

$$R_2(x, y) = \left[\frac{|I_1(x, y) - I_0(x, y)|}{I_0(x, y)}\right] = |1 - 2I_0(x, y)| \tag{6.34}$$

$$R_3(x, y) = [R_2(x, y)]^{\alpha} \tag{6.35}$$

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 [11] in Fig. 6.9, 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,

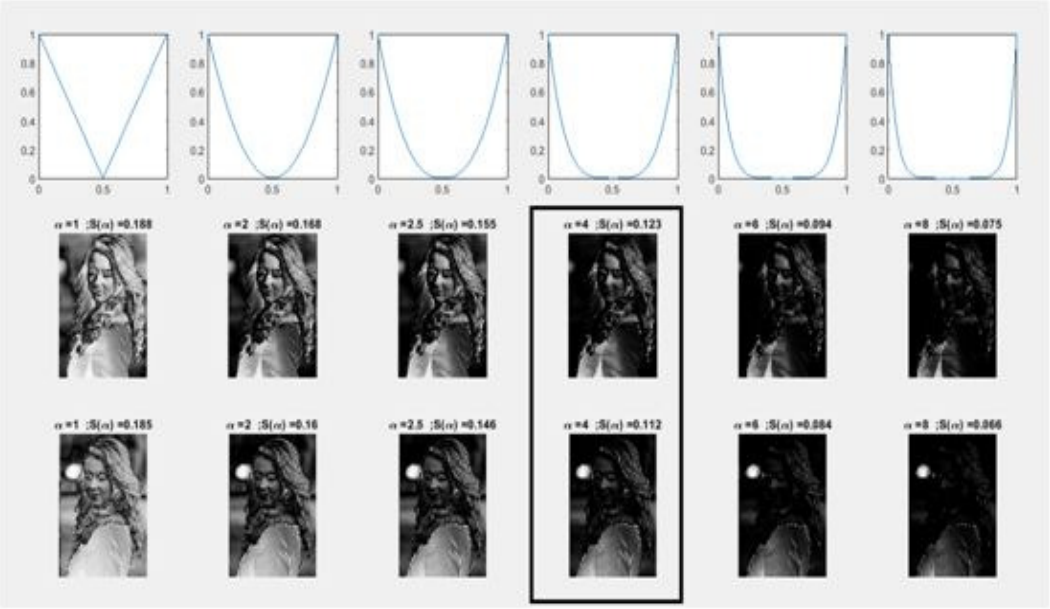

Figure 6.10: Impact of changes in the exponential parameter α on both the versions from the TWIN-image set [11]. As the exponent increases, the self-shadows become much more discernible for the version where the lighting is normal. Beyond a certain point the ratio images corresponding to both the normal version and the diffused version become dark.

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 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.

6.2.7 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 [69] as,

$$f_{G_n}(g) = \frac{1}{2n} g^{(\frac{1}{2n} - 1))} \tag{6.36}$$

where, $g \in [0, 1]$

. The continuous/differential entropy([70]) of G_n can be evaluated as,

$$H[G_n] = -E_{G_n}\left[ln\left(f_{G_n}(G)\right)\right] \tag{6.37}$$

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.

6.2.8 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 $H[R_{REAL}] = -E_R[lnf_{REAL}(R)]$. 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 CH 6.3: 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. [69],

$$f_{R_{REAL}}(r) = (\frac{1}{\alpha})r^{\frac{1}{\alpha}-1} \tag{6.38}$$

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

$$f_{R_{SPOOF}}(r) = (\frac{a^{\alpha}}{\alpha})r^{\frac{1}{\alpha}-1} \tag{6.39}$$

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

$$H[R_{REAL}] = ln[\alpha] + 1 - \alpha \tag{6.40}$$

$$H[R_{SPOOF}] = ln(\frac{\alpha}{a}) + (1-\alpha)(1 + ln(a)) \tag{6.41}$$

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

$$-\Delta H(\alpha) = H[R_{SPOOF}] - H[R_{REAL}] \tag{6.42}$$

$$= -\alpha \times ln(a) \tag{6.43}$$

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

$$E_{REAL}(\alpha) = ln(\alpha) + 1 - \alpha \tag{6.44}$$

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. 6.11. 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.

6.3 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. 6.11) is dependent on 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

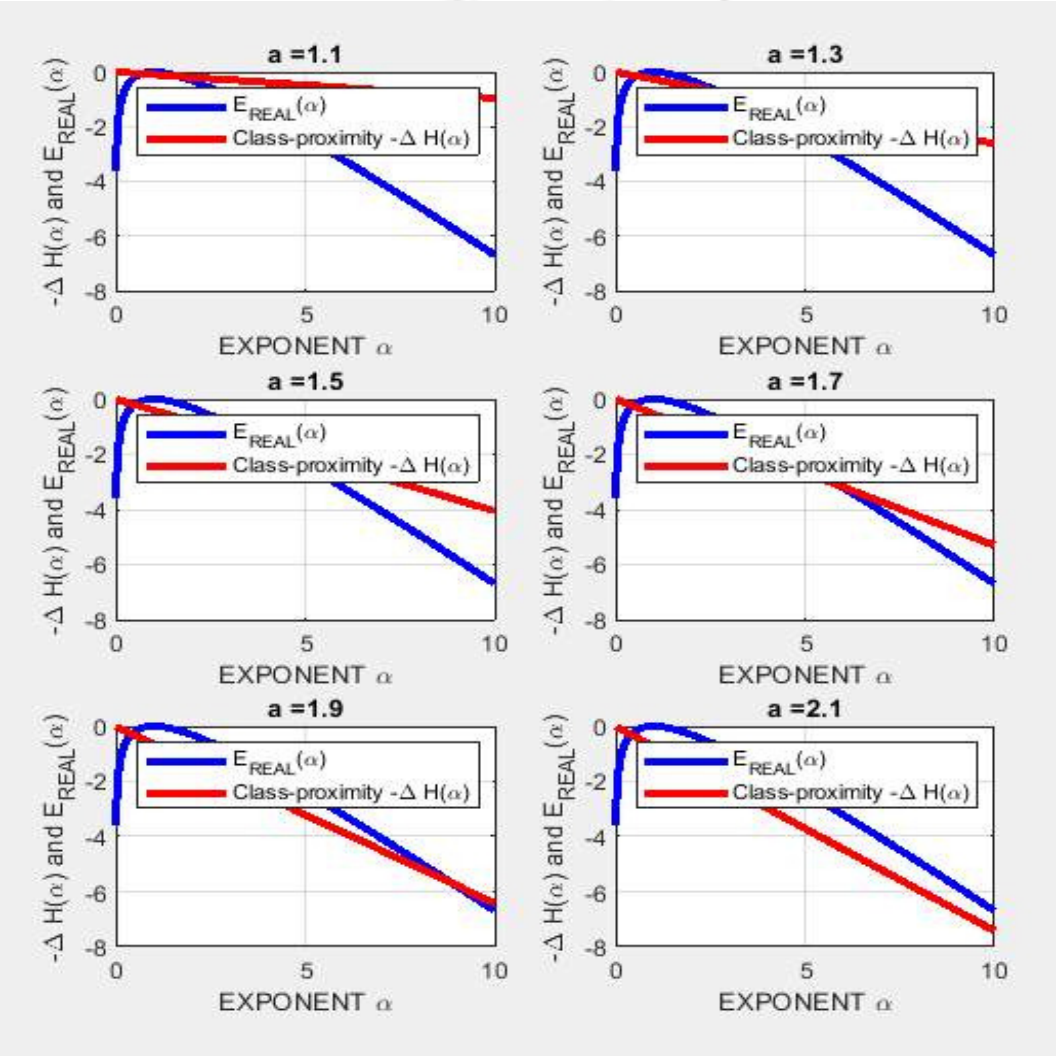

Figure 6.11: The selection of the operating point, as the point of intersection towards the right of the RED and BLUE curves, to maximize class separation (not the one on the left) is shown for different values of a

being scrutinized.

- For a particular image sample in the real-class, generate the global contrast score [3], (obtained from Equation. 6.4).

$$CR_i = \sigma_i / \mu_i \tag{6.45}$$

- The mean contrast score for natural faces is,

$$CON_{REAL(E)} = \frac{1}{N_{CALIBREAL}} \sum_{i \in SET_{CALIBREAL}} CR_i \tag{6.46}$$

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,

$$CON_{SPOOF(E)} = \frac{1}{N_{CALIBSPOOF}} \sum_{i \in SET_{CALIBSPOOF}} CS_i \tag{6.47}$$

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. 6.11, the mean contrast score of the real-calibration set is referenced against the spoof set taking a ratio of the two:

$$\hat{a}_F = \frac{CON_{REAL(E)}}{CON_{SPOOF(E)}} \tag{6.48}$$

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. 6.11 (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 $\alpha_{CASIA} = 2.7$.

6.4 Final feature extraction procedure and Client Specific Classification

Block-diagrams of the feature extraction procedure following by the classification and testing are shown in Chapters/6/figures. 6.1 and 6.2 respectively.

6.4.1 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:

$$E_\alpha(x, y) = R_3(x, y) = \left[\frac{|I_1(x, y) - I_0(x, y)|}{I_0(x, y)}\right]^{\hat{\alpha}} \tag{6.49}$$

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,

$$S = \begin{pmatrix} \sigma_{1,1} & \sigma_{1,2} & \dots & \sigma_{1,n} \\ \sigma_{2,1} & \sigma_{2,2} & \dots & \sigma_{2,n} \\ \dots & \dots & \dots & \dots \\ \sigma_{n,1} & \sigma_{n,2} & \dots & \sigma_{n,n} \end{pmatrix} \tag{6.50}$$

with,

$$\sigma_{i,j} = \sqrt{\frac{1}{W^2} \sum \sum_{(x,y) \in PATCH(i,j)} (R_3(x, y) - \mu_{i,j})^2} \tag{6.51}$$

where,

$$\mu_{i,j} = \frac{1}{W^2} \sum \sum_{(x,y) \in PATCH(i,j)} R_3(x, y) \tag{6.52}$$

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

6.4.2 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.

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

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

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 $n_{TYPICAL} = 30$.

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

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

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

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

Step-4: Computing the patch-wise intensity diversity statistic Let $\beta \in (0,1)$ be the frac-

tional 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

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

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

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,

$$STAT_{RAW}(I_0) = \frac{1}{N_{patches}} \sum_{\forall patches} \sigma_p \ \textbf{TYPE1} \tag{6.53}$$

$$LSTAT_{NORM}(I_0) = \frac{2}{N_{patches}} \sum_{\forall patches} [|ln(\sigma_p)|] \ \textbf{TYPE2} \tag{6.54}$$

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

6.4.3 2-class SVM Models for each client/subject

The original CASIA set [9] 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 [12] and CASIA-SURF [13].

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. 6.12 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. 6.12(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. 6.12(a), third and fourth images).
- In Subject-11, the problem was very different and existed in the spoofing segment (Fig. 6.12(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 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 [64] [14]. The remain 50% of the samples from both the natural and spoof segments were used for testing. The t-SNE maps [71] of the reduced CASIA set test set on a subject specific basis are shown in Fig 6.13 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

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

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

Figure 6.12: 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.

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 [71] (a stochastic map which presents a fairly realistic lower dimensional representation of higher dimensional data) in Fig. 6.13. In all the subject specific subplots of the test-data, Fig. 6.13(a-n), the cluster separation was found to be excellent, attesting and reinforcing CLAIMS CH 6.1 and CH 6.1.

6.5 Experimental results and comparisons

In this section, a description of three different datasets, CASIA [9], OULU [12] and CASIA-SURF [13] 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 phase-1 discussed in Section. 6.3. Based on these optimized parameters, subject specific model-building, testing and comparisons form the last few subsections.

6.5.1 Description of Databases

A summary of the datasets used for final round testing of the proposed life-trail algorithm is provided in Table. 6.1. The original CASIA face dataset [9] shown in Fig. 6.14 which was created from Chinese individuals showed significant variability on both the natural face front as well as the

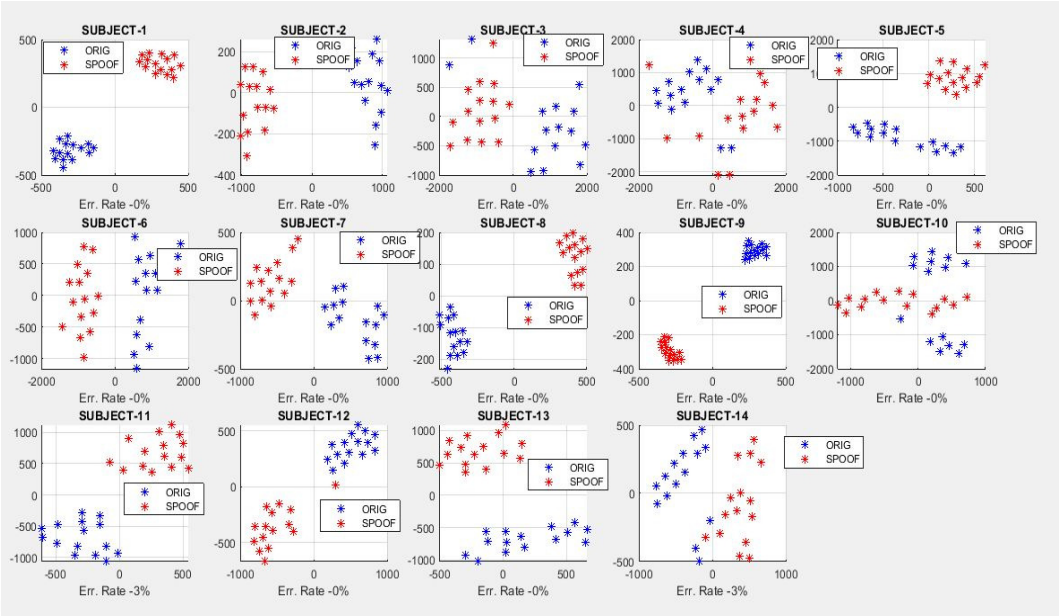

Figure 6.13: Cluster separation (subject-wise) in a 2-class setting, for the reduced CASIA-dataset comprising of 14-subjects (out of a total of 50) in which 50% of the variations per subject, were used for testing.

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 [12] shown

Table 6.1: Selective face anti-spoofing datasets and related parameters.

Datasets	Year	No of Subjects	Camera	Image representation	Spoof-type	Race
CASIA [9]	2012	50	VIS	RGB	Print photo, cut photo	Chinese
MSU-MFSD [2]	2015	35	Phone Laptop	RGB	printed photo	Asia, Middle east Hispanic, Europeans Latan Americans
Oulu NPU [12]	2018	20	VIS	RGB	Printed photo	Europeans and Middle east
CASIA SURF [13]	2018	1000	Real sense	RGB/Depth	Print and cut photos	Almost all the races

in Fig. 6.15, 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

Figure 6.14: Examples (both natural and spoof-print versions) from the original CASIA dataset [9].

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 [13] shown in Fig. 6.16, 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.

Figure 6.15: Examples (both natural and spoof-print versions) from the original OULU-NPU dataset [12]

Figure 6.16: Examples (both natural and spoof-print versions) from the original CASIA-SURF dataset [13].

6.5.2 Parameter Estimation

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 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 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. 6.4.2(Step-5).

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

$$L_p = |ln\left(\epsilon + \sigma_p\right)| \tag{6.55}$$

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

$$LSTAT = \frac{1}{N_{patches}} \sum_{p=1}^{N_{patches}} L_p \tag{6.56}$$

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:

$$\begin{aligned} \mu_{REAL} &= \frac{1}{r}\sum_{k=1}^{r} u_k \\ \mu_{SPOOF} &= \frac{1}{r}\sum_{k=1}^{r} v_k \\ \sigma^2_{REAL} &= \frac{1}{r}\sum_{k=1}^{r} (u_k - \mu_{REAL})^2 \\ \sigma^2_{SPOOF} &= \frac{1}{r}\sum_{k=1}^{r} (v_k - \mu_{SPOOF})^2 \end{aligned} \tag{6.57}$$

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 [72], 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:

$$SEPARATION_{KLD} = \left(\frac{\sigma^2_{REAL}}{\sigma^2_{SPOOF}} + \frac{\sigma^2_{SPOOF}}{\sigma^2_{REAL}}\right) + (\mu_{REAL} - \mu_{SPOOF})^2 \left(\frac{1}{\sigma^2_{REAL}} + \frac{1}{\sigma^2_{SPOOF}}\right) \tag{6.58}$$

Table 6.2: 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

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. 6.2. 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 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. 6.3.

Table 6.3: 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

6.5.3 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 [54] [9] [22] [2] [16]; (ii) The subject identity is known to the counter-spoofing system (more like an authentication environment) [64] [14];

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. 6.6)as compared to the subject-independent error scores (Table. 6.5). The best among them is the random walk/scan based algorithm [16] [26] 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 0.310% Table 6.6. 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 [12] was 1.192% and that for the CASIA-SURF [13] was found to be 2.246%. These numbers were comparable with the

Table 6.4: EEE for optimal values of α^* and $beta^*$ for

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

Convolutional Neural Network (CNN) based solutions shown in Table. 6.6. Notice that in the case of CASIA-SURF, the CNN based solutions available in [73], depth map information was augmented with RGB information to support the learning process. With pure RGB information these error-numbers will be higher.

Table 6.5: 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 [74]	GMMUBM	CASIA	CASIA	C_{gb}	21.69
	Two Class SVM	CASIA	CASIA	C_{gb}	15.42
LBP-TOP [74]	GMM UBM	CASIA	CASIA	C_{gb}	12.65
	Two Class SVM	CASIA	CASIA	C_{gb}	8.53
Motion [74]	GMMUBM	CASIA	CASIA	C_{gb}	12.52
	Two Class SVM	CASIA	CASIA	C_{gb}	11.53
IMQ [4]	One class SVM	CASIA	CASIA	-	23.07
BSIF [4]	One class SVM	CASIA	CASIA	-	36.06
LPQ [4]	One class SVM	CASIA	CASIA	-	35.19
LBP [4]	One clas SVM	CASIA	CASIA	-	25.06
Random Scan [16]	One class SVM	CASIA	CASIA	$N_S = 1$	3.5122
Random Scan [16]	One class SVM	CASIA	CASIA	$N_S = 20$	1.8920

This solution was designed to combat a very specific form of spoofing "Planar-Print-spoofing". The accuracies witnessed from the point of view of a client specific, 2-class model building procedure are comparable with some of the state of the art CNN based methods.

It will however not work against planar-digital-spoofing owing to the backlighting.

The LIFE-TRAILS approach uses a non-linear iterated functional mapping to generate contrast reductionistic life trail. In each successive iteration, the contrast is reduced and eventually the "IMAGE SWARM" gets watered down to a zero-contrast image.

It was observed that the normalized difference between the original and the image generated after the first iteration carried significant information regarding the self-shadows.

The more diverse the parent image in terms of intensity profile, the longer the LIFE-TRAIL.

Table 6.6: 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 same C_{sp} mode has been deployed.

Method	Classifier	Train Data	Test Data	Threshold	EER
LBP [74]	GMMUBM	CASIA	CASIA	C_{sp}	10.09
	Two Class SVM	CASIA	CASIA	C_{sp}	9.87
LBP-TOP [74]	GMM UBM	CASIA	CASIA	C_{sp}	6.36
	Two Class SVM	CASIA	CASIA	C_{sp}	3.95
Motion [74]	GMMUBM	CASIA	CASIA	C_{sp}	9.66
	Two Class SVM	CASIA	CASIA	C_{sp}	11.27
MSLBP [14]	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 [14]	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 [75]	Deep CNN	CASIA	CASIA	C_{sp},α, β, γ	1.85
Radiometric Distortion [21]	Two class SVM RBF	CASIA	CASIA	C_{sp}	0.00
CPqDN [76]	CNN	OULU-NPU	OULU-NPU	*ProtocolI* [77]	6.9
GRADIANT [76]	CNN	OULU-NPU	OULU-NPU	*ProtocolI* [77]	6.9
STASN [78]	CNN	OULU-NPU	OULU-NPU	*ProtocolI* [77]	1.9
FaceDs [55]	CNN	OULU-NPU	OULU-NPU	*ProtocolI* [77]	1.5
STPM [59]	CNN	OULU-NPU	OULU-NPU	C_{sp}, RGB+Depth	1.0
NHF [13]	CNN	CASIA-SURF	CASIA-SURF	C_{sp}, RGB+Depth	4.7
Single-scale SEF [13]	CNN	CASIA-SURF	CASIA-SURF	C_{sp}, RGB+Depth	2.4
Multi-scale SEF [13]	CNN	CASIA-SURF	CASIA-SURF	C_{sp}, RGB+Depth	0.8
PSMM-Net [73]	CNN	CASIA-SURF	CASIA-SURF	C_{sp}, RGB+Depth	0.4
PSMM-Net(CeFA) [73]	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

Print-spoof presentations tend to have shorter trails as compared to Natural face trails. This trend analysis was quite difficult and largely in-conclusive and thus we decided to focus on the first order, first-difference which carried significant self-shadow information. Natural faces tend to exhibit significant self-shadows as compared to spoof faces. But the generation of a self-shadow takes place due to an interplay between the following elements:

- Facial surface topography.
- Positioning in relation to light source which can be variable even for the same subject.
- Positioning of the camera in relation to the face-object and the light source.

Thus, because of this extreme variability, one must focus on the quantum and spread of self-shadow

information trapped in the image taken and extracted from the LIFE-TRAIL. This QUANTUM is higher for Natural images/faces and very less for print-spoofing [79]. Table 6.6. shows how our distortion model-specific approach (self-shadow based) can compete with the state of the art CNN based techniques. In some of the cases only when depth map information is included along with other parameters have these techniques achieved accuracies close to zero-percent.

Note that the availability of the depth map is LUXURY and it is impractical to assume all acquisition-environmental settings will have systems that will record depth details of all their clients.

If one leaves out the depth parameters in the CNNs, the numbers are comparable. Thus, this contribution of ours is a significant one not just on the result-front but also on the analytical front.

6.6 Summary and Conclusions

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. 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 α^* 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 α^* if the dynamic range parameter $\hat{(a)}_F = \hat{a}$ can be extracted via measurements.

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 α^* 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 β^*.

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. The algorithm may not be effective for digital planar image presentation cases based on tablets and laptops, as the back-lighting tends to enhance the self-shadow profiles present even in digitally spoofed-segments.

7

Conclusions and Future work

Contents

7.1 Summary

The main research problem posed in this thesis was to attack all forms of face spoofing modalities with one single monolithic solution. The only phenomenon that is common to both planar spoofing and prosthetic spoofing which can be attacked to produce a discriminatory feature is the BLUR which stems from the PINHOLE camera effect. Two main issues observed in the existing frame were:

(i) The spoofing segment or frame is largely un-predictable, whether digital planar, print-planar or prosthetic based.

(ii) The problem linked counter-spoofing is very different from plain face-authentication or identification, in the sense that what needs to be checked is NOT the identity but rather whether the format of the face-object or face-like object presented to the camera is indeed NATURAL.

This called for a fine-grained face image analysis, performed in a content and subject agnostic mode. Inspired by the space-filling curves which were designed for compressing encrypted videos, in this thesis, a contiguous random walk process was proposed to trap the first, second and third order pixel intensity correlation statistics in natural face alone along. Thus, the final solution was a subject and content agnostic random scan based natural face space representation or model. To implement this intrinsic natural face model in the outlier detection mode and in a higher dimensional space, a standard 1-class SVM was deployed from literature [4]. Accuracies obtained for this identity independent random scan based natural space model were very promising.

Apart from this a secondary, an analytically heavy contribution was the frame based on contrast reductionistic image life-trails. Here, it was observed that there exist logistic maps, wherein starting off from any initial point over a certain range of values, repeated iterations will result in a convergence to the same so called FIXED POINT. In the context of an image, an image in the normalized grey-scale mode, is a SWARM of pixel intensities having different initial positions over the range $(0, 1)$. Let the image intensity in the beginning at location x, y be and the logistic map be,

$$I_n(x, y) = 2I_{n-1}(x, y)[1 - I_{n-1}(x, y)] \tag{7.1}$$

Eventually after a few iterations this IMAGE SWARM collectively converges to a ZERO CONTRAST image wherein virtually all pixel intensities become 0.5. This sequence, $I_n(x, y) : n = 1, 2, ...$ has been termed as a LIFE TRAIL. It was observed from careful scrutiny that the first difference

obtained from this life-trail,

$$D_0(x, y) = I_1(x, y) - I_0(x, y) \tag{7.2}$$

carries crucial self-shadow information, which is prevalent in natural face images, however, suppressed in printed face images. A client specific [64] [14] 2-class model was built using the proposed self-shadow feature to detect print-spoofing alone. Accuracies reported for print spoof detection were comparable with the state of the art CNN-techniques.

7.2 FUTURE WORK and DIRECTIONS

The contiguous random walk has many benefits. Over a very small local zone in the image, it is possible to gather statistically equivalent sequences which exhibit the same first, second and third order pixel intensity correlation. This compact ENSEMBLE like profiling can be used on multiple fronts:

(i) To generate DEPTH MAPS from NATURAL SINGLE IMAGES very precisely via random scan auto-population in different parts of the image.

(ii) Quality assessment in a BLIND fashion is also possible to the extent that one many be able to use the auto-populated statistics to detect the nature of the distortion and profile it.

(iii) Fingerprint counter-spoofing via natural space characterization in a CONTENT AGNOSTIC fashion can be designed using RANDOM WALKS .

Apart from this, since these random scans are largely irreversible without the KEY or the LINKED LIST, they can be used to construct KEY-INDEPENDENT IMAGE DESCRIPTORS for IMAGE RETRIEVAL from LARGE DATABASES. Every image can be converted into a BLOB of vectors (or a cluster of points) [80] and this representative BLOB or cluster or data-driven model can be used in PRIVACY PRESERVING IMAGE RETRIEVAL.

Bibliography

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

[2] D. Wen, H. Han, and A. K. Jain, "Face spoof detection with image distortion analysis," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 4, pp. 746–761, 2015.

[3] K. Karthik and B. R. Katika, "Image quality assessment based outlier detection for face anti-spoofing," in *Communication Systems, Computing and IT Applications (CSCITA), 2017 2nd International Conference on*. IEEE, 2017, pp. 72–77.

[4] S. R. Arashloo, J. Kittler, and W. Christmas, "An anomaly detection approach to face spoofing detection: A new formulation and evaluation protocol," *IEEE Access*, vol. 5, pp. 13 868–13 882, 2017.

[5] S. Bhattacharjee and S. Marcel, "What you can't see can help you-extended-range imaging for 3d-mask presentation attack detection," in *2017 International Conference of the Biometrics Special Interest Group (BIOSIG)*. IEEE, 2017, pp. 1–7.

[6] C. Chen, W. Yuan, X. Lu, and L. Ma, "Spoof face detection via semi-supervised adversarial training," *arXiv preprint arXiv:2005.10999*, 2020.

[7] K. Karthik and B. R. Katika, "Face anti-spoofing based on sharpness profiles," in *Industrial and Information Systems (ICIIS), 2017 IEEE International Conference on*. IEEE, 2017, pp. 1–6.

[8] N. Erdogmus and S. Marcel, "Spoofing face recognition with 3d masks," *IEEE transactions on information forensics and security*, vol. 9, no. 7, pp. 1084–1097, 2014.

[9] Z. Zhang, J. Yan, S. Liu, Z. Lei, D. Yi, and S. Z. Li, "A face antispoofing database with diverse attacks," in *Biometrics (ICB), 2012 5th IAPR international conference on*. IEEE, 2012, pp. 26–31.

[10] K. Patel, H. Han, and A. K. Jain, "Secure face unlock: Spoof detection on smartphones," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 10, pp. 2268–2283, 2016.

[11] C. Bell, "Bright sunlight," 2015. [Online]. Available: https://in.pinterest.com/pin/290552613441811631/

[12] Z. Boulkenafet, J. Komulainen, L. Li, X. Feng, and A. Hadid, "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)*. IEEE, 2017, pp. 612–618.

[13] S. Zhang, X. Wang, A. Liu, C. Zhao, J. Wan, S. Escalera, H. Shi, Z. Wang, and S. Z. Li, "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*, 2019, pp. 919–928.

[14] J. Yang, Z. Lei, D. Yi, and S. Z. Li, "Person-specific face antispoofing with subject domain adaptation," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 4, pp. 797–809, 2015.

[15] N. Erdogmus and S. Marcel, "Spoofing in 2d face recognition with 3d masks and anti-spoofing with kinect," in *2013 IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS)*. USA: IEEE, 2013, pp. 1–6.

[16] B. R. Katika and K. Karthik, "Face anti-spoofing by identity masking using random walk patterns and outlier detection," *Pattern Analysis and Applications*, pp. 1–20, 2020.

[17] D. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, pp. 91–110, 2004.

[18] A. K. J. S. P. David M, Dario. M, *Handbook of Fingerprint Recognition*. Springer on biometrics New York, 2012.

[19] D. C. Garcia and R. L. de Queiroz, "Face-spoofing 2d-detection based on moiré-pattern analysis," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 4, pp. 778–786, 2015.

[20] T. Wang, J. Yang, Z. Lei, S. Liao, and S. Z. Li, "Face liveness detection using 3d structure recovered from a single camera," in *2013 international conference on biometrics (ICB)*. IEEE, 2013, pp. 1–6.

[21] T. Edmunds and A. Caplier, "Face spoofing detection based on colour distortions," *IET biometrics*, vol. 7, no. 1, pp. 27–38, 2017.

[22] S. Kim, Y. Ban, and S. Lee, "Face liveness detection using defocus," *Sensors*, vol. 15, no. 1, p. 1537–1563, Jan 2015. [Online]. Available: http://dx.doi.org/10.3390/s150101537

[23] J. Galbally and S. Marcel, "Face anti-spoofing based on general image quality assessment," in *2014 22nd International Conference on Pattern Recognition*, Aug 2014, pp. 1173–1178.

[24] X. Gao, T.-T. Ng, B. Qiu, and S.-F. Chang, "Single-view recaptured image detection based on physics-based features," in *2010 IEEE International Conference on Multimedia and Expo*, 2010, pp. 1469–1474.

[25] B. R. Katika and K. Karthik, "Face anti-spoofing based on specular feature projections," in *Proceedings of 3rd International Conference on Computer Vision and Image Processing*, B. B. Chaudhuri, M. Nakagawa, P. Khanna, and S. Kumar, Eds. Singapore: Springer Singapore, 2020, pp. 145–155.

[26] K. Karthik and 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)*. India: Springer, 2019.

[27] Y. Matias and A. Shamir, "A video scrambling technique based on space filling curves," in *A Conference on the Theory and Applications of Cryptographic Techniques on Advances in Cryptology*, ser. CRYPTO '87. London, UK, UK: Springer-Verlag, 1988, pp. 398–417. [Online]. Available: http://dl.acm.org/citation.cfm?id=646752.704721

[28] G. . Ekman, "Weber's law and related functions," p. 343–351. [Online]. Available: https://psycnet.apa.org/record/1960-04949-001

[29] K. Karthik, S. Chakraborty, and S. Banik, "Muzzle analysis for biometric identification of pigs," in *2017 Ninth International Conference on Advances in Pattern Recognition (ICAPR)*. IEEE, 2017, pp. 1–6.

[30] A. Anjos and S. Marcel, "Counter-measures to photo attacks in face recognition: A public database and a baseline," in *2011 International Joint Conference on Biometrics (IJCB)*, Oct 2011, pp. 1–7.

[31] Z. Zhang, J. Yan, S. Liu, Z. Lei, D. Yi, and S. Z. Li, "A face antispoofing database with diverse attacks," in *2012 5th IAPR International Conference on Biometrics (ICB)*, March 2012, pp. 26–31.

[32] D. Wen, H. Han, and A. Jain, " Face Spoof Detection with Image Distortion Analysis," *IEEE Trans. Information Forensic and Security*, vol. 10, no. 4, pp. 746–761, April 2015.

[33] N. Erdogmus and S. Marcel, "Spoofing in 2d face recognition with 3d masks and anti-spoofing with kinect," in *2013 IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, Sept 2013, pp. 1–6.

[34] J. Li, Y. Wang, T. Tan, and A. K. Jain, "Live face detection based on the analysis of fourier spectra," vol. 5404, 2004, pp. 296–303. [Online]. Available: http://dx.doi.org/10.1117/12.541955

[35] J. Maatta, A. Hadid, and M. Pietikainen, "Face spoofing detection from single images using micro-texture analysis," in *Biometrics (IJCB), 2011 International Joint Conference on*, Oct 2011, pp. 1–7.

[36] J. Galbally and S. Marcel, "Face anti-spoofing based on general image quality assessment," in *Pattern Recognition (ICPR), 2014 22nd International Conference on*. IEEE, 2014, pp. 1173–1178.

[37] X. Gao, T.-T. Ng, B. Qiu, and S.-F. Chang, "Single-view recaptured image detection based on physics-based features." in *ICME*. IEEE Computer Society, 2010, pp. 1469–1474.

[38] Z. Ji, H. Zhu, and Q. Wang, "Lfhog: a discriminative descriptor for live face detection from light field image," in *Image Processing (ICIP), 2016 IEEE International Conference on*. IEEE, 2016, pp. 1474–1478.

[39] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *Journal of the ACM (JACM)*, vol. 58, no. 3, p. 11, 2011.

[40] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *CoRR*, vol. abs/0912.3599, 2009. [Online]. Available: http://arxiv.org/abs/0912.3599

[41] H. Li and Z. Lin, "Accelerated proximal gradient methods for nonconvex programming," *Advances in neural information processing systems*, vol. 28, 2015.

[42] A. Pinto, W. R. Schwartz, H. Pedrini, and A. d. R. Rocha, "Using visual rhythms for detecting video-based facial spoof attacks," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 5, pp. 1025–1038, May 2015.

[43] Y. Li, L. M. Po, X. Xu, L. Feng, and F. Yuan, "Face liveness detection and recognition using shearlet based feature descriptors," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2016, pp. 874–877.

[44] D. C. Garcia and R. L. de Queiroz, "Face-spoofing 2d-detection based on moire pattern analysis," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 4, pp. 778–786, April 2015.

[45] A. Saxena, S. H. Chung, and A. Y. Ng, "3-d depth reconstruction from a single still image," *International Journal of Computer Vision*, vol. 76, no. 1, pp. 53–69, 2008. [Online]. Available: http://dx.doi.org/10.1007/s11263-007-0071-y

[46] M. Luck and U. Kirchmaier, "Geometrically based depth assignment using a single image," *Technical Report*, 2012.

[47] Z. Liu, H. Hong, Z. Gan, J. Wang, and Y. Chen, "An improved method for evaluating image sharpness based on edge information," *Applied Sciences*, vol. 12, no. 13, p. 6712, 2022.

[48] Y. Matias and A. Shamir, "A video scrambling technique based on space filling curves," in *Conference on the Theory and Application of Cryptographic Techniques*. Springer, 1987, pp. 398–417.

[49] L. Feng, L.-M. Po, Y. Li, and F. Yuan, "Face liveness detection using shearlet-based feature descriptors," *Journal of Electronic Imaging*, vol. 25, no. 4, pp. 043 014–043 014, 2016.

[50] D. Menotti, G. Chiachia, A. Pinto, W. R. Schwartz, H. Pedrini, A. X. Falcao, and A. Rocha, "Deep representations for iris, face, and fingerprint spoofing detection," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 4, pp. 864–879, 2015.

[51] Z. Boulkenafet, J. Komulainen, and A. Hadid, "Face spoofing detection using colour texture analysis," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 8, pp. 1818–1830, 2016.

[52] A. Sepas-Moghaddam, L. Malhadas, P. L. Correia, and F. Pereira, "Face spoofing detection using a light field imaging framework," *IET Biometrics*, vol. 7, no. 1, pp. 39–48, 2017.

[53] J. Galbally, S. Marcel, and J. Fierrez, "Image quality assessment for fake biometric detection: Application to iris, fingerprint, and face recognition," *Image Processing, IEEE Transactions on*, vol. 23, no. 2, pp. 710–724, 2014.

[54] ——, "Biometric antispoofing methods: A survey in face recognition," *IEEE Access*, vol. 2, pp. 1530–1552, 2014.

[55] A. Jourabloo, Y. Liu, and X. Liu, "Face de-spoofing: Anti-spoofing via noise modeling," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 290–306.

[56] K. Karthik and S. Kashyap, "Transparent hashing in the encrypted domain for privacy preserving image retrieval," *Signal, Image and Video Processing*, vol. 7, no. 4, pp. 647–664, 2013.

[57] C.-C. Chang and C.-J. Lin, "Libsvm: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 2, no. 3, p. 27, 2011.

[58] S. Minaee and A. Abdolrashidi, "Finger-gan: Generating realistic fingerprint images using connectivity imposed GAN," *CoRR*, vol. abs/1812.10482, 2018. [Online]. Available: http://arxiv.org/abs/1812.10482

[59] Z. Wang, Z. Yu, C. Zhao, X. Zhu, Y. Qin, Q. Zhou, F. Zhou, and Z. Lei, "Deep spatial gradient and temporal depth learning for face anti-spoofing," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 5042–5051.

[60] S. M. Nesli Erdogmus, "Spoofing in 2d face recognition with 3d masks and antispoofing with kinect," in *IEEE BATS, USA*, 2013.

[61] S. Kim, S. Yu, K. Kim, Y. Ban, and S. Lee, "Face liveness detection using variable focusing," in *Biometrics (ICB), 2013 International Conference on*. IEEE, 2013, pp. 1–6.

[62] S. Y. J. L. Y. Kim, Jaekenun, "Masked fake face detection using radiance measurments," in *Conference on Optical Society OF America*, vol. 26, no. 4, 2009, pp. 1054–1060.

[63] X. Zhang, X. Hu, M. Ma, C. Chen, and S. Peng, "Face spoofing detection based on 3d lighting environment analysis of image pair," in *2016 23rd International Conference on Pattern Recognition (ICPR)*, 2016, pp. 2995–3000.

[64] I. Chingovska, A. R. Dos Anjos, and S. Marcel, "Biometrics evaluation under spoofing attacks," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 12, pp. 2264–2276, 2014.

[65] J. Määttä, A. Hadid, and M. Pietikäinen, "Face spoofing detection from single images using texture and local shape analysis," *IET biometrics*, vol. 1, no. 1, pp. 3–10, 2012.

[66] J. M. Saragih, S. Lucey, and J. F. Cohn, "Deformable model fitting by regularized landmark mean-shift," *International Journal of Computer Vision*, vol. 91, no. 2, pp. 200–215, 2011.

[67] E. W. Weisstein, "Logistic map." [Online]. Available: https://mathworld.wolfram.com/LogisticMap.html

[68] Z. Zhang, D. Yi, Z. Lei, and S. Z. Li, "Face liveness detection by learning multispectral reflectance distributions," in *Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on*. IEEE, 2011, pp. 436–441.

[69] A. Papoulis, *Probability, random variables, and stochastic processes*. McGraw-Hill, 1991.

[70] T. M. Cover and J. A. Thomas, *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. USA: Wiley-Interscience, 2006.

[71] L. Van der Maaten and G. Hinton, "Visualizing data using t-sne." *Journal of machine learning research*, vol. 9, no. 11, 2008.

[72] "Kullback-leibler divergence." [Online]. Available: https://en.wikipedia.org/wiki/Kullback%E2%80%93Leibler_divergence

[73] A. Liu, Z. Tan, J. Wan, S. Escalera, G. Guo, and S. Z. Li, "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*, 2021, pp. 1179–1187.

[74] I. Chingovska and A. R. Dos Anjos, "On the use of client identity information for face antispoofing," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 4, pp. 787–796, 2015.

[75] Y. Sun, H. Xiong, and S. M. Yiu, "Understanding deep face anti-spoofing: from the perspective of data," *The Visual Computer*, vol. 37, pp. 1015–1028, 2021.

[76] Z. Boulkenafet, J. Komulainen, Z. Akhtar, A. Benlamoudi, D. Samai, S. E. Bekhouche, A. Ouafi, F. Dornaika, A. Taleb-Ahmed, L. Qin *et al.*, "A competition on generalized software-based face presentation attack detection in mobile scenarios," in *2017 IEEE International Joint Conference on Biometrics (IJCB)*. IEEE, 2017, pp. 688–696.

[77] 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. 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, and A. Hadid, "A competition on generalized software-based face presentation attack detection in mobile scenarios," in *2017 IEEE International Joint Conference on Biometrics (IJCB)*, 2017, pp. 688–696.

[78] X. Yang, W. Luo, L. Bao, Y. Gao, D. Gong, S. Zheng, Z. Li, and W. Liu, "Face anti-spoofing: Model matters, so does data," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 3507–3516.

[79] B. R. Katika and K. Karthik, "Image life trails based on contrast reduction models for face counter-spoofing," *EURASIP Journal on Information Security*, vol. 2023, no. 1, p. 1, 2023.

[80] K. Karthik, "Constructing intrinsic image signatures by random downsampling and differential mean-median information," *Special Issue on Information and Communication Technology International Journal of Tomography and Simulation (IJTS)*, vol. 22, no. 1, pp. 112–144, 2013.

List of Publications

Journal Publications

1. Balaji Rao Katika and Kannan Karthik "Face Anti-spoofing by Identity Masking usingRandom Walk Patterns and Outlier Detection" Pattern Analysis and Applications (PAA),Status: DOI https://doi.org/10.1007/s10044-020-00875-8

2. Katika, B.R., Karthik, K. Image life trails based on contrast reduction models for face counter-spoofing. EURASIP J. on Info. Security 2023, 1 (2023). https://doi.org/10.1186/s13635-022-00135-8

Conference Publications

1. Karthik, Kannan and Balaji Rao Katika. "Face anti-spoofing based on sharpness profiles.",IEEE International Conference on Industrial and Information Systems (ICIIS), (Srilanka),2017 , IEEE, 2017

2. Karthik, Kannan and Balaji Rao Katika. "Image quality assessment based outlier detec-tion for face anti-spoofing.", 2nd International Conference on Communication Systems,Computing and IT Applications (CSCITA), Mumbai(India), 2017 . IEEE, 2017.

3. Balaji Rao Katika and Kannan Karthik. "Face Anti-spoofing based on Specular FeatureProjections ", 3rd International Conference on Computer Vision anbd Image Processing(CVIP), (IIITDM Jabalpur), 2018 . Springer, 2018

4. Karthik, Kannan and Balaji Rao Katika. "Identity Independent Face Anti-Spoofing using2D Random Walk Patterns ", 8th International Conference on Pattern Recognition andMachine Intelligence, Tezpur(India), 2019.