Face Antispoofing Using Shearlets: An Empirical Study

Dustin Terence van der Haar, *Member, IEEE,*

***Abstract*—Face recognition - A promise made to the modern technologists as the ultimate access control or surveillance technology. However, similar to its fingerprint spoofing counterpart, current face antispoofing technology is still vulnerable to inexpensive spoofing attacks, which pose a threat to security. Although basic face spoofing attacks that use photographs and video are common in attack scenarios, they are still not addressed appropriately, thereby making security in these environments a difficult task to achieve. Newer face antispoofing attacks, such as 3D mask-based antispoofing have emerged and further complicated face antispoofing matters. Although methods have improved over the last decade, a robust solution that can accommodate changing environments is still out of reach. More so, these methods have not been assessed across multiple publicly available datasets and very little work has been done to perform a fair comparison across multiple face spoofing methods. Face spoofing attacks introduce an object into the scene, which presents curvilinear singularities that are not necessarily portrayed in the same way in different lighting conditions. We present a solution that addresses this problem by using a discrete shearlet transform as an alternative descriptor that can differentiate between a real and a fake face without user-cooperation across multiple environmental conditions. We have found the approach can successfully detect blurred edges, texture changes and other noise found in various face spoof attacks. In order to prove that discrete shearlet transforms are a valid descriptor and to perform a fair comparison of other methods, an empirical study is conducted with multiple experimental parameters and concrete results. Our benchmarks on the publicly available CASIA-FASD, MSU-MFSD, OULU-NPU, and HKBU-MARs datasets, show that our approach portrays good results and improves on the most popular methods found in the field on modest computer hardware, but requires further improvement to beat the current state of the art for basic face antispoofing efforts, such as the photo, cut and video attacks. However, where it succeeds is for detecting 3D mask-based antispoofing methods. Discrete Shearlet Transforms achieved very good accuracy on the HKBU-MARs 3D mask dataset and exhibited excellent precision, recall and f1-score, thereby showing it is an excellent descriptor for the task. The approach also achieves real-time face spoof discrimination with minimal resource overhead, which makes it a practical solution in real-time applications and a viable augmentation to current face recognition methods.**

***Index Terms*—Face Recognition, Face Antispoofing, Presentation Attack Detection**

I. Introduction

We are entering the age of automation. Machines can replace people with repetitive tasks and differentiate between people using face recognition. By giving machines the ability to identify or authenticate people automatically, it can prevent unauthorised users from accessing secure areas or provide a tailored user experience. Face recognition is already being used in public spaces for surveillance monitoring of citizens and to maintain watch-lists in airports, malls, border gates, and casinos [1]. Other similar applications include its use in mobile computing for face tracking and in affective computing to detect the emotion of users [2].

However, as technology and methods have progressed for face recognition, so have the attacks to subvert face recognition systems. Attacks, such as using photos, modifying them in some way and video attacks [3], have become common in various subversion scenarios. Technology has also assisted attackers through the advent of cheaper 3D printing and increased computational resources. These combined with higher quality screens, which have improved resolutions and colour ranges, allow them to facilitate better replay attacks, where a counterfeit biometric is presented to the face recognition system.

In order to keep up with these developments, more liveness and antispoofing research and systems have been introduced to combat these attacks [4], [5], [6], [7], [8], [9], [10], [11], [12]. These approaches either determine liveness, analyse movement or perform image quality assessment. Both sides of technology improvements have resulted in an ebb and flow of attacks and defences. However, a robust generic solution without significant computational resource overhead, which performs well under various conditions is yet to be realised. The work discussed here poses another defence in the war against face spoofing. It provides an alternative real-time presentation attack detection (PAD) method for describing the facial region of interest, which can tolerate blur, texture changes, and low-quality frames in order to achieve improved face antispoofing. A comparison is then made with it and other popular methods with multiple datasets to get a fair measure of its performance amongst its peers.

The article begins by unpacking the problem at hand and face antispoofing related work. The proposed approach is then discussed by outlining the method with an appropriate discussion. The experiment methodology followed by the researcher to validate the approach is then described for all four datasets, followed by the achieved results. The paper is then concluded, along with future work.

II. Related Work

The prevalence of face recognition systems in society has motivated criminals to find ways of subverting these systems to avoid identification. The ideal method for achieving this is for them to undergo facial plastic surgery that changes fundamental components of their face. Technology has also

D.T. van der is with the Academy of Computer Science and Software Engineering, University of Johannesburg, Johannesburg, Gauteng, 2006 South Africa e-mail: dvanderhaar@uj.ac.za.

Based on: *“Real-Time Face Antispoofing Using Shearlets”*, by D.T. van der Haar which appeared in
Communications in Computer and Information Science (volume 973). © 2019 Springer International Publishing

made it possible to achieve these attacks at a lower cost. Although plastic surgery is a pressing concern, there have been attacks that prove it is not necessary to go to that extent to subvert a face recognition system [13].

There are more simple, non-intrusive attacks that can be used to spoof face recognition systems. These attacks include (and examples can also be seen in Figures 1 and 2):

1) A *photo attack* where a printed image of a legitimate user is presented in front of the user's face.
2) The *warped photo attack* where the photo attack is extended to include movements and minor folding of the photo.
3) A *cut photo attack* where holes are cut out by the eyes are in the photo to fool blink-based liveness systems.
4) The *video replay attack* where a user places a video of a legitimate user in front of their head using a tablet or similar device [6], [3].
5) A *mask attack* where the attacker wears a 3D mask of a legitimate user [10].

Thankfully researchers have become cognizant of these attacks and research has been pursued to address them appropriately. There have also been face antispoofing competitions to promote more novel solutions within the research community, as seen in the International Joint Conference on Biometrics (IJCB) 2011, 2013 and 2017. Datasets have been created that emulate these attacks, such the NUAA Photograph Impostor Database (which only contains photographs) [14], REPLAY-ATTACK Database [6], CBSR database [3], CASIA-FASD [8], MSU Mobile Face Spoofing Database (MFSD) [15] and more recently OULU-NP [16] (used in one of the 2017 IJCB competitions) along with the HKBU-MARs mask-based face antispoofing dataset [17] to serve as a benchmark for any proposed methods.

In the last decade, there has been significant progress in face antispoofing methods. Earlier methods used Gabor wavelets [14] or eye blinking in order to determine whether there was potential face spoofing occurring [4]. The approach had basic liveness detection with minimal computational overhead. However, blink detection still fell victim to the cut photo attack and poor accuracy. One of the breakthroughs in the field was the use of micro-texture analysis with local binary patterns (LBP) [6]. It encapsulated the change of texture in the scene attributed to the foreign object, was also computationally fast and did not require cooperation from the user to blink. However, fine detail that can be used for discriminating between real and fake faces is lost when using LBP-based histograms around larger areas of the face. As shown by [5], there is value in dividing it into overlapping sub-regions (such as 3 by 3 pixels) and calculating sub-region-based LBP histograms. However, it comes at the cost of more sparse feature space, additional resources and is affected by blur present in the region of interest.

Another improvement uses LBP from Three Orthogonal Planes (LBP-TOP) to achieve face antispoofing [7]. The improvement increases accuracy, by leveraging the power of dynamic textures, analysing motion and texture, along with a score level fusion-based framework. However, the approach showed incremental improvement and it struggled with lower quality frames found in the CASIA-FASD data set.

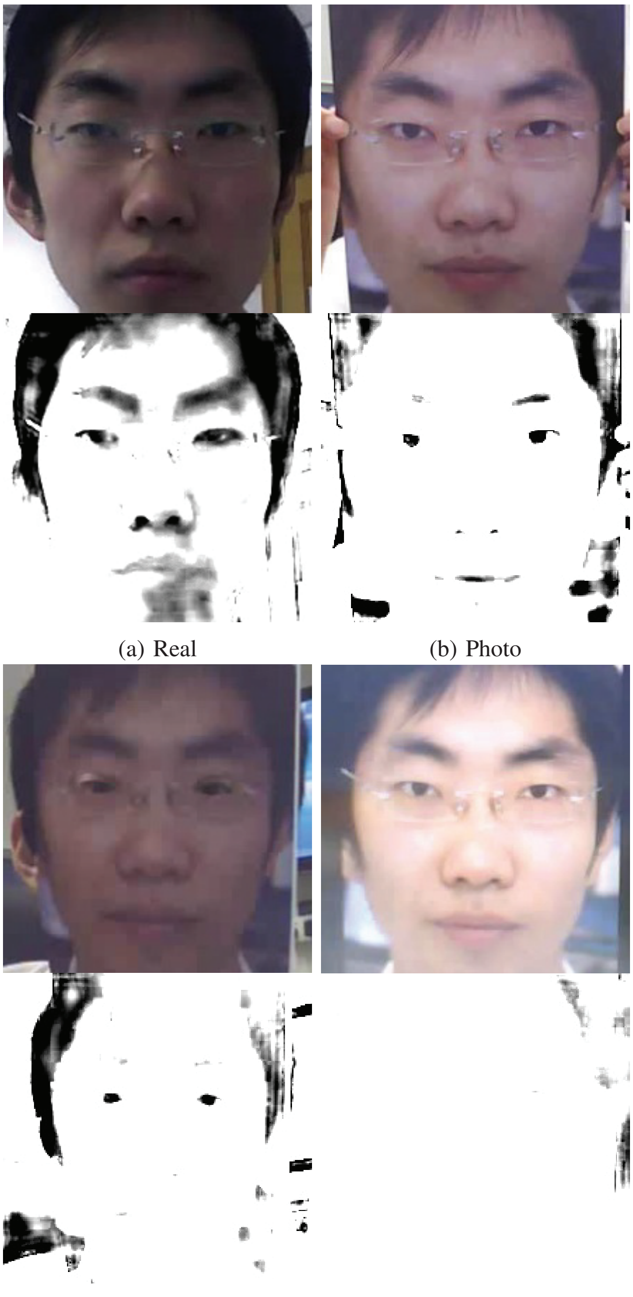

(a) Real
(b) Photo
(c) Cut
(d) Video

Fig. 1: Examples of various attack scenarios and their mean shearlet representation for the region of interest derived from the CASIA-FASD dataset.

The latter part of the decade was then spent on mostly building on existing methods, performing more image quality assessments and combining multiple methods. In [8], a score fusion framework was proposed and provided additional insights into dataset quality. However, only methods that are statistically independent can be used in the score fusion framework, so any potential methods would need to be statistically vetted before it can be included. The work in [9] proposes a context-based antispoofing method. It showed

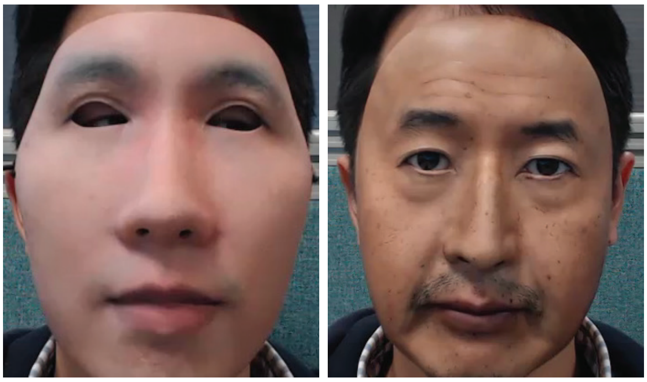

(a) ThatsMyFace mask (b) Real-F mask

Fig. 2: Examples for the type of mask attacks taken from the HKBU-MARs dataset.

that scene cues detected with Histogram of Oriented Gradients (HOG) descriptors and upper body analysis could be used for face antispoofing. The results show an improvement in other methods, especially in video replay attacks. However, it is constrained by particular attack scenarios and is limited to close-up environments.

In [18], multiple feature vectors are created from twelve different components that form what they call their holistic face (or H-Face). The H-Face-based approach performs well, but there are many measures to calculate, thereby limiting its potential in real-time applications. The approach by [19] extends traditional LBP approaches by performing motion magnification using optical flow vectors first before deriving LBP sub-sampled histograms. However, it only provides an incremental improvement and is subject to similar issues experienced by other LBP-based descriptors. Another approach by Galbally and Marcel [10] uses general image quality assessment to achieve face antispoofing, by using a combination of different image quality measures, such as measures used to determine pixel difference, correlation and edges. These features, such as signal to noise ratio (SNR), average distance (AD) and total edge difference (TED) present a low degree of complexity and can potentially be used for real-time applications. However, the approach exhibits only a marginal improvement over other methods.

Antispoofing based on colour texture analysis is achieved in [11] by analysing LBPs in various colour spaces, such as HSV, RGB and YC_bC_r. The approach shows promising results but suffers in varying lighting and environmental conditions. Patel et al. shows in [20] that Moiré pattern aliasing present in spoof face videos can be used for face antispoofing. The approach worked well on video but lacks results for photo-based spoofing methods. Agarwal also proposes in [21], the use of Haralick features to achieve face antispoofing with good accuracy, but it comes at the cost of very large feature space and computational overhead, which results from the subsequent dimensional reduction required.

More recently in [12], a fusion of various approaches is used to achieve face antispoofing. Face and scene optical flow-based motion features coupled with an image quality feature are used to train an artificial neural network for spoof discrimination. It results in very good accuracy, but at the cost of significant computational and memory resources. The use of convolutional neural networks (CNN) has also been shown to exhibit excellent accuracy for face antispoofing, which make it the current state of the art [22]. Although it achieves very good results, it too consumes a great deal of memory and is slower than methods with a lower level of complexity.

Methods in face antispoofing have improved, but it is still clear that there is still no robust solution that performs consistently under various changes (such as a change in camera, resolution or lighting) without incurring significant computational and memory overhead. Little attention is paid to the resource usage of methods, its potential use in real-time applications and its performance under varying environmental conditions. The approach discussed in the next section remains cognizant of these requirements (especially the real-time application aspect) and attempts to address them appropriately.

III. Face Antispoofing using Shearlets

In a spoof scenario, there are certain elements present in a video frame present, which will help with face-based spoof discrimination. Attacks introduce a rigid object within the scene to mimic a legitimate object. Much of the work done so far attempts to differentiate between these rigid objects from a non-rigid face by deriving a feature space that can classify for anomalies. However, many of them rely on colour space-based methods and are greatly affected by lighting changes. However, there is value in analysing curvature and significant edges found in the scene that are more robust to lighting changes. Our approach uses a discrete shearlet transform-based (DST) descriptor to achieve this.

A. DST for Face Antispoofing

Shearlets were introduced to overcome the traditional wavelet limitation of describing directionality. They are a natural extension to wavelets that can efficiently represent anisotropic features, such as edges in images, and serve as a good sparse approximation of multidimensional data. These properties make it an excellent candidate for face antispoofing because the rigid objects introduced in the scene exhibit anomalies in the form of curvilinear singularities in a compact and computationally efficient form. Instead of analysing the high dimensional space found in the original sample, a succinct representation made up of a wavelet-like function that can portray directionality. A discrete shearlet transform (DST) is defined as [23]:

$$SH_\psi f(j,k,m) = \langle f, \psi_{j,k,m} \rangle, j > 0, k \in \mathbb{Z}, m \in \mathbb{Z}^2 \qquad (1)$$

Where SH_ψ maps the function f to the shearlet coefficients, j is associated with the scale index, k the orientation index and m the positional index. Shearlets have been used for edge detection [24] and non-reference image quality assessment [25], but it has the potential for face antispoofing. The DST can highlight blurred edges, texture changes and other noise found in the frame, which can be attributed to the rigid object used in a typical face spoofing attack.

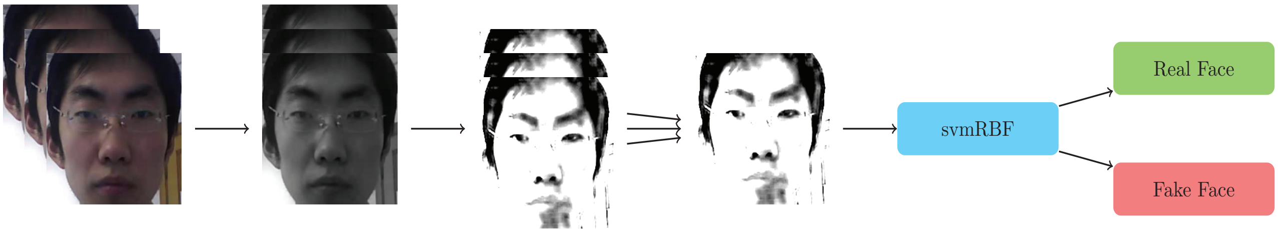
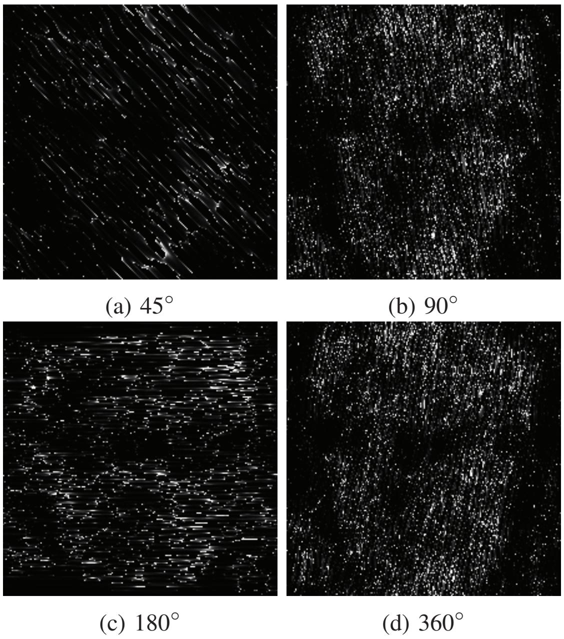


Fig. 3: The proposed approach that uses a discrete shearlet transform to achieve face antispoofing.

Other work exists that shows initial attempts to achieve face antispoofing using shearlets, but they lack evidence that validates the proposed use case. Li et al. in [26]and [27] follow a more integrated approach by using shearlet spectrum for both live detection and face recognition. They take the shearlet transform, feed it into stacked autoencoders and classify it using a support vector machine. They test the approach on the CASIA database and derive Detection-Error Trade-off (DET) curves to show their results. However, they do not test it with other attack vectors or provide enough detail to reproduce the method and test it with higher quality datasets and derive other performance metrics, so that it can be compared against other methods.

When applied to a video frame captured in the scene that contains a face, it allows us to perform face spoof discrimination. As seen in Figure 1, the DST results can be seen for each attack scenario. In 1a the sharp outline of significant curvature found on the real face can be seen, along with subtle details of the face. Whereas in 1b and 1c key details of the face are missing, which can be attributed to the lack of detail portrayed in the photograph and cut attack. Lastly, it can also be seen that there is a significant amount of detail lost in the video attack portrayed in 1d. These examples show that there is clear value in using DST for face antispoofing.


(a) 45° (b) 90°

(c) 180° (d) 360°

Fig. 4: Partial shearlet orientation coefficients.

B. The Proposed Method

In order to use DST within the context of face antispoofing (as seen in the process found in Figure 3) there are specific steps that need to be made to perform face antispoofing that maximises performance and make it tolerant of environmental changes. Each captured video frame undergoes face region of interest (ROI) segmentation. The Viola-Jones face detector [28] is used to capture the face, but instead of using the method for every subsequent frame, a basic colour histogram check is performed and if it passes the check, the coordinates of the previous ROI is used. In order to compensate for motion artefacts that cause ROI drift, minor translation smoothing is applied by analysing prior ROI coordinates found in previous frames.

Once the face ROI is segmented, the DST is used to derive the features. It begins by going through a calibration phase, where the shearlet spectrum is derived for the ROI. The precomputed shearlet spectrum during the calibration phase is then used in subsequent frames for deriving the shearlet coefficients in a significantly faster manner. Frames that fail the previous colour test trigger another shearlet spectrum derivation event. This optimisation allows the approach to achieve face antispoofing in real-time. Further speed optimisations can be achieved on much higher resolutions (such as 4k video) by using partial shearlet coefficients at a specific direction only, as seen in Figure 4. Directionality present in the DST can also be leveraged further if details of the scene are known, such as subject pose and camera position. However, for this investigation, the partial orientation coefficients are not explored, but a combination of all the orientations are explored.

In the last step before classification, the feature space is converted into a more succinct representation that can be compared. Unlike LBP-based systems, which lose specificity across large areas, the DST-based feature space is not changed into a histogram. Instead, a DST-based feature mean (as seen in Figure 3) is calculated across the orientations for a defined window of video frames (the results use a 100 frame window) and fed into the classifier for assessment.

Once the DST-based features have been calculated, we investigate the classification accuracy using support vector machines (SVM). We first used a Linear SVM classifier and then similar to other methods [5], [7], [11] we use a Radial Basis Function (RBF)-based SVM for determining whether the DST features depict a real or fake face. The use of the SVM classifier instead of a neural network-based classifier allowed us to achieve the most memory and computational gains. The SVM classifier is first trained using a set of real and fake faces with ground truth labels according to each respective dataset's assessment protocols.

IV. Experiment Analysis

In order to validate the approach an experiment is used, which is cognizant of environmental changes and the real-time constraint. In order to ensure more robust results, unlike many other approaches which portray results for a favourable data set, the proposed approach was validated against three datasets. The experiment data and setup is discussed in the subsections that follow.

A. Experimental Data

In order to provide more robust, repeatable results publicly available datasets were used for the experiment. The CASIA-FASD, MSU-MFSD, OULU-NPU and HKBU datasets were all selected for the experiment and compared because they exhibited the most common attacks under varying environmental conditions and portrayed a good range of quality for samples. Each of the datasets contains at least three types of videos used for face antispoofing assessment. At least one video with the subject's real face, which would constitute the true videos, along with videos that facilitate a print attack and video attack (with CASIA-FASD also containing the cut attack) to represent the spoofed video. By using various datasets, it also allows us to assess results under varying lighting conditions with a diverse amount of subjects.

The CASIA-FASD dataset contains 20 subjects for training and 30 subjects for testing. The base videos are captured at a resolution of 640 by 480 pixels, and their higher resolution videos are captured at a resolution of 720 by 1280. Both the standard and higher resolution videos were captured using a web camera at a frame rate of 25 frames per second under varying lighting conditions for at least 6 seconds. The dataset contains photo, warp, cut and video attacks.

The MSU-MFSD dataset contains 35 subjects (in the publicly available version) captured using two types of cameras (a built-in Macbook Air camera at 640 by 480 pixels and a front facing Google Nexus 5 camera at 720 by 480 pixels) to capture video at 30 frames per second. The dataset includes photo attacks and video replay attacks using an iPad and an iPhone.

The OULU-NPU dataset contains 20 subjects for training and 15 for development. The videos are captured at a resolution of 1080 by 1920 at 30 frames per second using the front camera of multiple devices (Samsung Galaxy S6, HTC Desire EYE, MEIZU X5, Asus Zenfone, Sony XPERIA C5 ultra Dual and OPPO N3) and under varying lighting conditions. The dataset contains photo attacks using two types of printers and video attacks using two types of displays. For this study, evaluation protocol one was used when evaluating how face antispoofing methods impact performance under illumination and background scene changes.

The HKBU-MARs dataset contains eight subjects captured with a web camera with a resolution of 1280 by 720 and a frame rate of 30 frames per second. The dataset primarily focuses on mask attacks and uses two types of masks, namely ThatsMyFace and REAL-F, for face spoofing (as seen in Figure 2).

B. Experimental Setup

The videos of the subjects for each data set corresponding to the training and test sets were used to facilitate the benchmark. All the captured ROI images are normalised to 256 by 256 pixels before the DST-based mean or derivative for each window of samples is derived. The results of our approach are then compared against a local binary pattern (LBP) approach under the same conditions (using both a linear and RBF-based SVM). Time traces and memory usage-based tests are performed between frames and an average is calculated on a modest computer with an i7 920 CPU with 2.67 GHz and 4 GB of RAM all using un-optimised python code. Each benchmark is done independently of each other and given the same resources to provide objective results.

Since using the original feature vector with all the shearlets and their respective orientations would incur additional memory and computational overhead, a DST derivative would be more practical for face antispoofing deployment. In order to determine which DST derivative to use for describing the feature space, an additional experiment was performed with different variations of processing pipelines using the MSU-MFSD data set. The experiment included deriving histograms, the mean and principal components (using PCA) from the DST representation. The results for this experiment can be seen in Figure 5.

From the experiment results, we can see some DST derivatives did not perform well in terms of accuracy. The worst performer was using principal component analysis (PCA) on the raw shearlets to achieve dimensional reduction. Showing that the dimensional reduction comes at the cost of feature specificity. Alternatively, in a similar vein to using LBP-based histograms, the other DST derivative we assessed computes histograms for the original shearlets. It performed better than PCA and resulted in a compact representation, which may be beneficial in certain contexts, but was not as consistent as the dstMean in our study. The best performer was the DST mean-based derivative, which achieved an AUC of 93%. Another finding was that radial basis function or gaussian kernel-based support vector machine (SVM) classifiers performed better in terms of accuracy at the cost of longer training and testing times.

C. Experiment Results

We evaluated the performance of our approach using three different datasets and compared its results with LBP. The

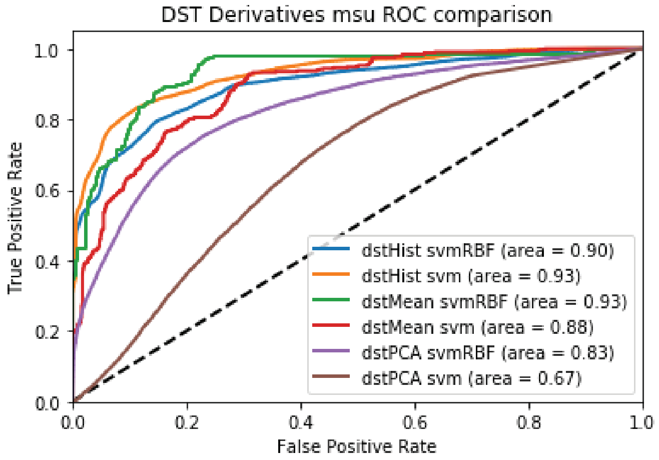


Fig. 5: The receiver operating characteristic (ROC) curves and Area Under Curves (AUC) for different DST derivatives on the MSU data set.

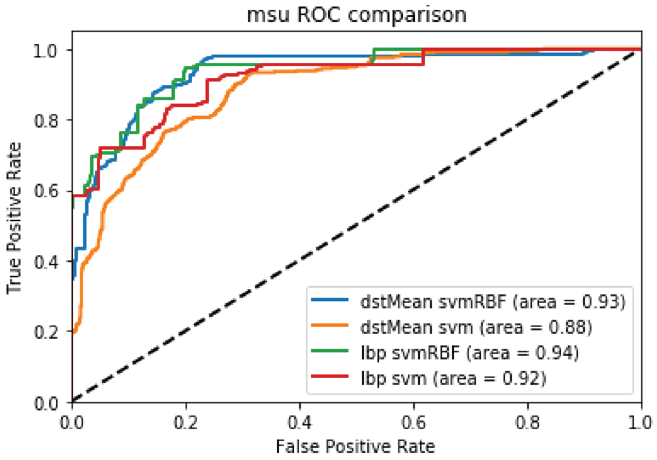


Fig. 7: The receiver operating characteristic (ROC) curves and Area Under Curves (AUC) for our various approaches for face antispoofing on the MSU-MFSD data set.

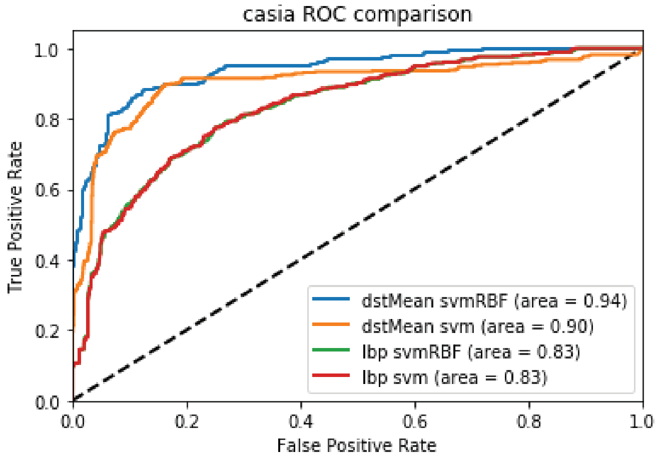


Fig. 6: The receiver operating characteristic (ROC) curves and Area Under Curves (AUC) for our various approaches for face antispoofing on the CASIA-FASD data set.

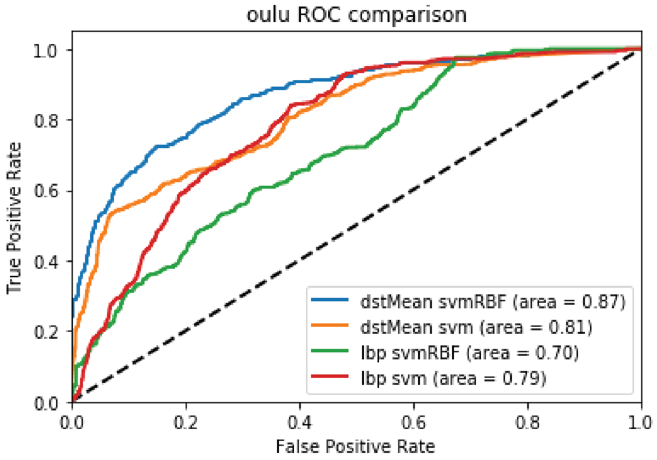


Fig. 8: The receiver operating characteristic (ROC) curves and Area Under Curves (AUC) for our various approaches for face antispoofing on the OULU-NPU data set.

performance of the approach is determined in terms of the receiver operating characteristic (ROC) curve to determine its accuracy for each respective data set (CASIA, MSU, OULU and HKBU). The equal error rate (EER) is then calculated for each data set and compared with current approaches (as seen in table I) to provide perspective on the results. Additional metrics, namely time taken, memory usage, accuracy, precision, recall and f1-score, are then derived for the best approach to form the evidence for our empirical study.

As seen in Figures 6, 7, 8 and 9, our DST approach performs consistently across the low, medium and even high quality ranges, along with different types of attacks. Overall the feature descriptor fared well against the various datasets showing robustness in the face of environmental and attack changes.

In the CASIA dataset, the results (Figure 6) showed a margin of improvement over LBP-based face antispoofing was between 7 and 11%. Thereby showing consistent improvement in the lower quality image band. The choice of classifier did not impact the LBP-based descriptor, but the RBF SVM showed incremental improvement for the dstMean descriptor.

For the OULU dataset results (Figure 8) the significant environmental changes can be seen to impact performance. The LBP variants could not breach an AUC of 80%, whereas the dstMean variants fared relatively well. Interestingly enough, as seen in Figure 7 and 8 the choice of linear and RBF-based SVM's exhibit different results for LBP and DST by a margin of 5-6%, thereby validating it as a better choice for face antispoofing methods than LBP-based methods. However, as investigated by the author, face antispoofing efforts at the higher resolution scale still have room for improvement even when using DST's.

In the MSU dataset results (Figure 7), we can see closer competition, due to attack materials quality and change in

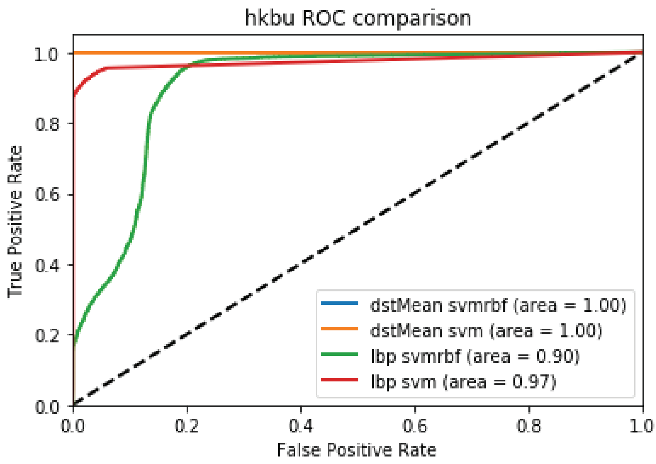


Fig. 9: The receiver operating characteristic (ROC) curves and Area Under Curves (AUC) for our various approaches for face antispoofing on the HKBU data set.

TABLE I: A table comparing the EER results of DST-based face antispoofing with other approaches on different datasets, where the approach results in italics are derived from this study.

Approach	**CASIA**	**MFSD**	**OULU**	**HKBU**
IQA+LDA [10]	32.4%	-	-	-
LBP+SVM(RBF)	24.4%	14.3%	37.75%	7%
LBP+SVM	24.1%	17%	29.5%	5%
Gabor Filters+LTVfused [14]	17.0%	-	-	-
dstMEAN+SVM(Linear)	15.7%	20.3%	20.9%	0%
LBP-HOOF+SVM [19]	14.4%	-	-	-
LBP-TOP+SVM(RBF) [29]	10.0%	-	-	-
dstMEAN+SVM(RBF)	12.1%	14%	22%	0%
IDA+SVM [15]	-	8.58%	-	-
CNN [30]	7.34 %	-	-	-
Haralick+PCA+ SVM[21]	6.7%	2.9%	-	-
Multi-cue integration+ NN [12]	5.83%	-	-	-

TABLE II: A table comparing the average time trace results (excluding training time) for LBP (in ascending quality order) and DST on various face antispoofing datasets applied on the experiment computer.

Approach	**Resolution**	**Frame rate**	**LBP**	**DST**
CASIA	640 by 480	25fps	0.03ms	9ms
MSU-MFSD	720 by 480	14-29fps	0.03ms	9ms
OULU-NPU	1080 by 1920	30 fps	0.04ms	9ms

TABLE III: A table comparing the average memory usage results (excluding training time) for LBP (in ascending quality order) and DST on various face antispoofing datasets applied on the experiment computer.

Approach	**LBP**	**DST**
CASIA	144MB	290MB
MSU-MFSD	76MB	310MB
OULU-NPU	189MB	301MB

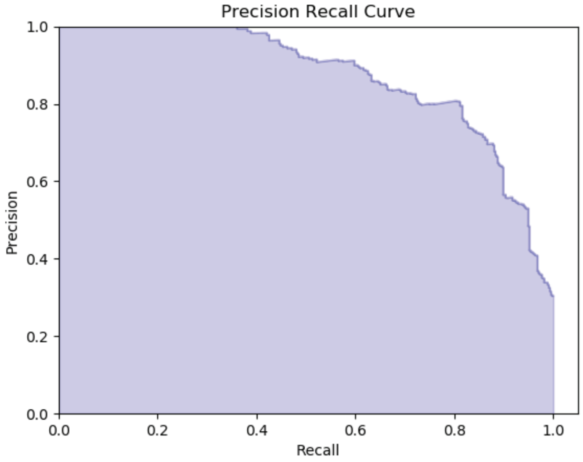


Fig. 10: The precision recall curve for the CASIA-FASD data set.

cameras. As confirmed by other work [11], lighting changes and camera quality impact LBP-based descriptor performance, but its dstMean-based counterpart performed incrementally better. It also shows there is value in further exploring the role camera artefacts play in face antispoofing systems.

As shown by the HKBU dataset results (9) the dstMean-based descriptor performed very well especially when dealing with face mask-based attacks. The LBP-based descriptor also performed well against the ThatsMyFace-based masks (as previously shown in Figure 2a), but struggled with the Real-F-based masks (Figure 2b). Since the amount of Real-F masks is quite low in the HKBU, there is value in exploring the dstMean-based descriptor on a more extensive Real-F face mask dataset.

The EER in table I show that it outperforms many methods, but still needs further improvement to beat very recent methods. As higher quality video becomes more prevalent, it would be interesting to see how new methods tolerant high definition video, lighting differences and different camera artefacts. The performance can be attributed to DST's blur tolerance and much of the error can be attributed to significant over-exposed frames. These over-exposed frames show there is a limit to how tolerant methods can be with regards to lighting.

In tables II and III, we can also see that the timing and memory results on a modest computer are reasonable and the approach allows for real-time detection. As expected, when the quality of the video increases, so does the time it takes to derive the results using LBP. As seen in table II our DST approach achieves relatively stable timings irrespective of the image quality difference, thereby allowing it to scale better at higher resolutions. When looking at the memory usage, LBP does use more memory for the OULU data set, but not for the MSU-MFSD data set. Interestingly, as seen in table III when using LBP, the memory usage was not as consistent as the DST approach's memory usage. Upon further inspection, the MSU-MFSD dataset's drop in memory usage during processing can be attributed to a lack of textures derived when using their specific mobile device cameras, thereby affecting accuracy and specificity. The resource usage shows that the DST approach

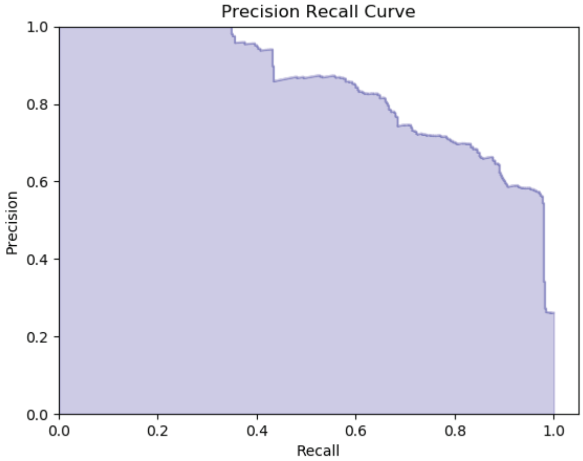


Fig. 11: The precision recall curve for the MSU-MFSD data set.

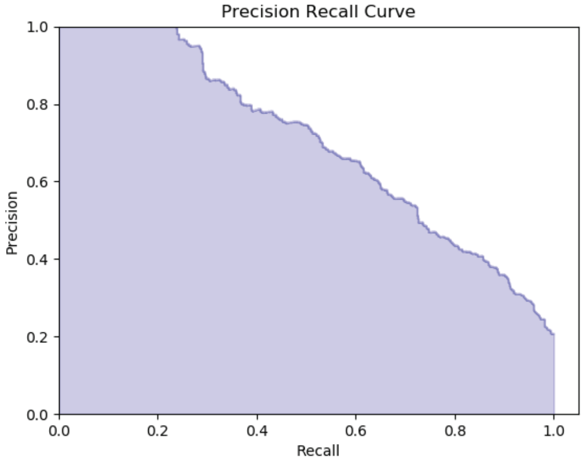


Fig. 12: The precision recall curve for the OULU-NPU data set.

works well considering the 4GB memory and processing power constraint.

Additionally, as seen in table IV, along with Figures 10, 11, 12 and 13 the dstMean with a SVM(RBF) fares well across the accuracy, precision, recall and F1 score metrics for the different datasets. In the context of face antispoofing systems, we argue that detecting face spoofing is commonly an imbalanced classification problem and metrics such as EER (HTER) do not provide us with a complete measure of its performance. The accuracy metric does provide a full picture of the bias the approach exhibits towards positive and negative cases. It will also provide a basis on which to focus optimisation efforts on a fundamental level instead of resorting to heuristic-based tweaks.

By analysing table IV and the precision-recall curves for all the datasets, we can draw insights into how dstMean-based descriptors are used for face antispoofing. From the f1 score, we can see that the datasets contain little class imbalance issues and that the descriptor is fairly stable when comparing precision to recall. From the results, we can see that although dstMean-based descriptors perform well overall and it is tolerant to lighting changes, but as seen in the precision-recall curve of the OULU dataset in Figure 12, the varying backgrounds contribute towards error.

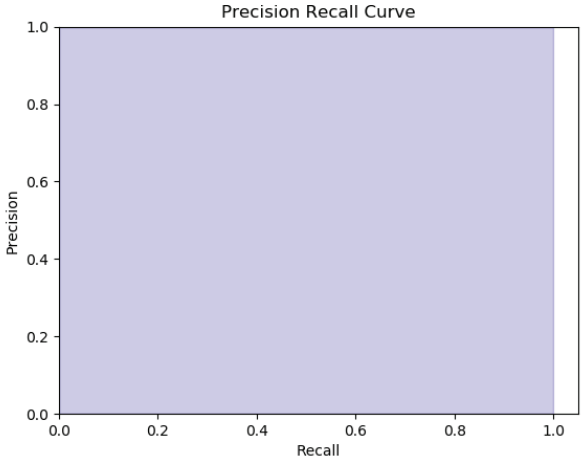


Fig. 13: The precision-recall curve for the HKBU data set.

TABLE IV: A table comparing the average meanDST accuracy, precision, recall and f1 score on the datasets.

Dataset	Accuracy	Precision	Recall	F1 Score
CASIA	90%	89%	90%	89%
MSU-MFSD	84%	85%	84%	82%
OULU-NPU	86%	86%	86%	86%
HKBU	100%	100%	100%	100%
Average	**90%**	**90%**	**90%**	**89.3%**

Alternatively, other quality measures can be included in order to improve accuracy. However, it comes at the cost of additional computational resources and more time required to achieve face antispoofing. Using the mean DST for a window of frames allowed us to break the real-time barrier at a minor cost to accuracy. We have also seen that it is also possible to improve accuracy by deriving shearlets at multiple scales (i.e. increasing the j index in equation 1). However, the scale reaches a saturation point that is directly proportional to the resolution of the video frame. Any increases to the scale after that point has diminishing returns on accuracy. Currently, the scale for the DST which maximises accuracy and still maintains real-time (based on the hardware constraints) is at:

$$DST_s = \frac{log_2(max(width, height))}{2} \tag{2}$$

The various datasets used to test the approach allowed us to gauge better performance across varying image qualities, population diversity, and lighting conditions. The CASIA low-quality data set with an average resolution of 640 by 480 covers the low-quality band. The MSU-MFSD datasets with 720 by 480 respectively for the medium quality range with different sensors. The OULU-NPU dataset with 1080 by 1920 resolution to cover the high definition range with even more environmental changes. Thereby showing consistent DST's performance even in varying difficult conditions.

Overall, the DST-SVM(RBF)-based implemented system detected most of the spoofing videos under varying environmental conditions and video qualities with a minimal resource footprint. Deriving DST features on average consumed around 311MB of RAM on average took 9ms to complete. This provides proof that our proposed approach can potentially be used in a low resource environment and still achieve real-time detection. Whereas other methods would require a networked solution to offload computation to a server or require additional computation or memory resources.

V. Conclusion

The prevalence of face recognition systems has made face antispoofing research an important concern. As face recognition systems remain vulnerable to face spoofing, a practical solution is in dire need. Part of making face antispoofing methods feasible in the real world is making sure that the results can be achieved in real-time. By achieving faster results, critical security environments can become more proactive in mitigating potential spoof attempts and prevent any potential attackers from entering a secure area or masquerading as another user. These results also need to be validated on varying image qualities to gain better insights into its robustness, as well as its applicability within different contexts.

By taking into account how curvilinear singularities can be derived in shearlets, we found it to be a good solution for achieving face spoof discrimination. By precomputing the shearlet spectrum during a calibration phase, we achieved real-time performance for the full assessment process at a little cost to complexity and resources. The discrete shearlet transform-based mean features are fed into a linear or RBF SVM, which are very computationally efficient. The overall results from the three datasets are good in varying environmental conditions and also show that it can compete with existing methods, thereby showing that our approach is viable in practical real-time face antispoofing application.

Further preliminary results also show that our approach has great potential in higher quality video, such as 4k video, without a significant computational footprint. We believe that further improvements to displays with higher resolutions and high dynamic range (HDR), along with cheaper 3D face masks are going to be the next challenge ahead for face antispoofing methods and we need to be ready for them.

References

[1] K. W. Bowyer, "Face recognition technology: security versus privacy," *IEEE Technology and society magazine*, vol. 23, no. 1, pp. 9–19, 2004.

[2] M. S. Bartlett, G. Littlewort, I. Fasel, and J. R. Movellan, "Real time face detection and facial expression recognition: Development and applications to human computer interaction." in *2003 Conference on Computer Vision and Pattern Recognition Workshop*, vol. 5, June 2003, pp. 53–53.

[3] 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.

[4] G. Pan, L. Sun, Z. Wu, and S. Lao, "Eyeblink-based anti-spoofing in face recognition from a generic webcamera," in *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*. IEEE, 2007, pp. 1–8.

[5] J. Määttä, A. Hadid, and M. Pietikäinen, "Face spoofing detection from single images using micro-texture analysis," in *Biometrics (IJCB), 2011 international joint conference on*. IEEE, 2011, pp. 1–7.

[6] 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.

[7] T. de Freitas Pereira, A. Anjos, J. M. De Martino, and S. Marcel, "Lbp- top based countermeasure against face spoofing attacks," in *Asian Conference on Computer Vision*. Springer, 2012, pp. 121–132.

[8] ——, "Can face anti-spoofing countermeasures work in a real world scenario?" in *Biometrics (ICB), 2013 International Conference on*. IEEE, 2013, pp. 1–8.

[9] J. Komulainen, A. Hadid, and M. Pietikainen, "Context based face anti-spoofing," in *Biometrics: Theory, Applications and Systems (BTAS), 2013 IEEE Sixth International Conference on*. IEEE, 2013, pp. 1–8.

[10] J. Galbally and S. Marcel, "Face anti-spoofing based on general image quality assessment," in *22nd International Conference on Pattern Recognition (ICPR)*. IEEE, 2014, pp. 1173–1178.

[11] Z. Boulkenafet, J. Komulainen, and A. Hadid, "Face anti-spoofing based on color texture analysis," in *Image Processing (ICIP), 2015 IEEE International Conference on*. IEEE, 2015, pp. 2636–2640.

[12] L. Feng, L.-M. Po, Y. Li, X. Xu, F. Yuan, T. C.-H. Cheung, and K.-W. Cheung, "Integration of image quality and motion cues for face anti-spoofing: A neural network approach," *Journal of Visual Communication and Image Representation*, vol. 38, pp. 451–460, 2016.

[13] X. Liu, S. Shan, and X. Chen, "Face recognition after plastic surgery: a comprehensive study," in *Asian Conference on Computer Vision*. Springer, 2012, pp. 565–576.

[14] X. Tan, Y. Li, J. Liu, and L. Jiang, "Face liveness detection from a single image with sparse low rank bilinear discriminative model," *Computer Vision–ECCV 2010*, pp. 504–517, 2010.

[15] 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.

[16] Z. Boulkenafet, J. Komulainen, L. Li, X. Feng, and A. Hadid, "Oulu-npu: A mobile face presentation attack database with real-world variations," in *Automatic Face & Gesture Recognition (FG 2017), 2017 12th IEEE International Conference on*. IEEE, 2017, pp. 612–618.

[17] S. Liu, P. C. Yuen, S. Zhang, and G. Zhao, "3d mask face anti-spoofing with remote photoplethysmography," in *Computer Vision – ECCV 2016*, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds. Cham: Springer International Publishing, 2016, pp. 85–100.

[18] J. Yang, Z. Lei, S. Liao, and S. Z. Li, "Face liveness detection with component dependent descriptor," in *2013 International Conference on Biometrics (ICB)*, June 2013, pp. 1–6.

[19] S. Bharadwaj, T. I. Dhamecha, M. Vatsa, and R. Singh, "Face anti-spoofing via motion magnification and multifeature videolet aggregation," Tech. Rep., 2014.

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

[21] A. Agarwal, R. Singh, and M. Vatsa, "Face anti-spoofing using haralick features," in *Biometrics Theory, Applications and Systems (BTAS), 2016 IEEE 8th International Conference on*. IEEE, 2016, pp. 1–6.

[22] O. Lucena, A. Junior, V. Moia, R. Souza, E. Valle, and R. Lotufo, *Transfer Learning Using Convolutional Neural Networks for Face Anti-spoofing*. Cham: Springer International Publishing, 2017, pp. 27–34.

[23] W.-Q. Lim, "The discrete shearlet transform: A new directional transform and compactly supported shearlet frames," *IEEE Transactions on Image Processing*, vol. 19, no. 5, pp. 1166–1180, 2010.

[24] G. Kutyniok and P. Petersen, "Classification of edges using compactly supported shearlets," *Applied and Computational Harmonic Analysis*, vol. 42, no. 2, pp. 245–293, 2017.

[25] Y. Li, L.-M. Po, X. Xu, and L. Feng, "No-reference image quality assessment using statistical characterization in the shearlet domain," *Signal Processing: Image Communication*, vol. 29, no. 7, pp. 748–759, 2014.

[26] 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)*. IEEE, 2016, pp. 874–877.

[27] L.-M. Po, Y. Li, F. Yuan, and L. Feng, "Face liveness detection using shearlet-based feature descriptors," *Journal of Electronic Imaging*, vol. 25, no. 4, p. 043014, 2016.

[28] P. Viola and M. J. Jones, "Robust real-time face detection," *International journal of computer vision*, vol. 57, no. 2, pp. 137–154, 2004.

[29] T. de Freitas Pereira, J. Komulainen, A. Anjos, J. M. De Martino, A. Hadid, M. Pietikäinen, and S. Marcel, “Face liveness detection using dynamic texture,” *EURASIP Journal on Image and Video Processing*, vol. 2014, no. 1, p. 2, 2014.

[30] Z. Xu, S. Li, and W. Deng, “Learning temporal features using lstm-cnn architecture for face anti-spoofing,” in *2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR)*, Nov 2015, pp. 141–145.

Dustin van der Haar received his Masters in Information Technology on face recognition and monitoring video telecommuincation systems in 2011 and PhD degree in Computer Science on authentication using human physiological signals in 2014 from the University of Johannesburg. He is currently an Associate Professor in the Academy of Computer Science and Software Engineering at the University of Johannesburg. His research interests include biometrics, computer vision and machine learning and have collaborated with multiple universities.