enhAnced Mobile BiomEtRics

DELIVERABLE D4.9

Academic Paper 4.4

Contract number:	675087
Project acronym:	AMBER
Project title:	Enhanced Mobile Biometrics
Project duration:	1 January 2017 - 31 December 2020
Coordinator:	Richard Guest, University of Kent, Canterbury, UK

Deliverable Number:	D4.9
Type:	Article
Dissemination level:	Confidential, only for members of consortium (including the Commision Services)
Expected submission date:	July 2020
Date submitted:	June 2020

Authors / contributors:	Anas Husseis, Judith Liu-Jimenez, Raul Sanchez-Reillo
Contributing partners:	University Carlos III de Madrid (UC3M)

This project has received funding from the European Unions Horizon 2020 research and innovation programme under the Marie Skodowska-Curie grant agreement No 675087

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.DOI

Fingerprint Presentation Attack Detection Utilizing Spatio-temporal Information

ANAS HUSSEIS, JUDITH LIU-JIMENEZ, RAUL SANCHEZ-REILLO, (SENIOR MEMBER, IEEE),
Electronic Technology Department, University Carlos III of Madrid, 28911 Leganés, Spain

Corresponding author: Anas Husseis (e-mail: ahusseis@ing.uc3m.es).

This work was supported by the European Union's Horizon 2020 for Research and Innovation Program under Grant 675087 (AMBER).

ABSTRACT This paper presents a novel algorithm for fingerprint dynamic presentation attack detection. We utilize five spatio-temporal feature extractors to efficiently eliminate and mitigate different presentation attack species. The feature extractors are selected such that the fingerprint ridge/valley pattern is consolidated with the temporal variations within the pattern in fingerprint videos. An SVM classification scheme, with a second degree polynomial kernel, is used in our presentation attack detection subsystem to classify bona fide and attack presentations, moreover the experiment protocol and evaluation are performed following the ISO/IEC 30107-3:2017 standard. Our proposed approach demonstrates efficient capability of detecting presentation attacks with significantly low BPCER where BPCER is 1.11% for an optical sensor and 3.89% for a thermal sensor at 5% APCER for both.

INDEX TERMS Fingerprint, Presentation Attack, Presentation Attack Detection, Anti-spoofing.

I. INTRODUCTION

FINGERPRINT recognition is one of the oldest and most prevalent biometric modalities. It has shown attractive features such as high accuracy and user convenience; accordingly, it is been applied in applications such as forensics, identity control, physical access control, and mobile devices. A recent study by Juniper anticipates having 4.5 billion mobile devices using fingerprint sensors by 2030 [1].

Unfortunately, the use of a biometric sub-system for authentication processes does not imply that the system is secured. The generic biometric scheme is vulnerable at different points starting from the sensor to the recognition score/decision [2]. Based on those vulnerabilities, biometric security is categorized in two main areas: (a) electronic security which concerns the digital process of the captured biometric trait (b) physical security which questions whether the biometric trait presentation is performed by a bona fide (i.e. genuine user) or by an attacker. This investigation is tended to focus on the second type and propose a potential software countermeasure.

Presentation Attack (PA), informally known as spoofing attack, is defined as a suspicious presentation that aims to manipulate the biometric decision using a Presentation Attack Instrument (PAI). The definition implicitly refers to two classes of attackers (a) concealer: aims to evade being recognized as him/herself (b) impersonator: seeks to claim an identity other than himself. In both cases, the attack might be performed with the bona fide cooperation, e.g. research studies, or without the bona fide consent, e.g. identity theft.

Despite the fact that fingerprint ridge/valley patterns are unique, fingerprints have other phenomena such as perspiration which causes the moisturized skin, consequently, fingerprints leave traces at touched surfaces. By using proper methods and tools, those traces can be captured and used to duplicate a PAI in order to impersonate one's identity. A group of forensic researchers has conducted an experiment demonstrating that fingerprint traces can be captured from problematic metal surfaces after over 26 days of deposition [3]. The experiment involves a sophisticated method along with advanced tools, which must be considered when calculating the attack potential, but it proves the possibility of capturing latent fingerprints when proper method and tools exist.

In order to overcome the issue of PA, researchers have been investigating Presentation Attack Detection (PAD) mechanisms that are capable of eliminating or mitigating PAs.

TABLE 1: Performance analysis for the SoA fingerprint PAD mechanisms (reported in [4] & [5])

PAD category	Reference	PAI species	Sensor/s	APCER (%)	BPCER (%)	APCER = BPCER
Distortion Analysis	Antonelli 2006 [6]	Gelatin, RTV silicon, white glue, and latex	Optical	-	-	11.24
	Zhang 2009 [7]	Silicon	Optical	-	-	4.5
	Jia 2007 [8]	Gelatin	Capacitive	-	-	4.87
Perspiration Analysis	Derakhshani 2003 [9]	Play-Doh, cadaver	Capacitive	-	-	11.11
	Parthasaradhi 2005 [10]	Play-Doh, cadaver	Capacitive	5-20	6.77-20	-
			Optical	4.6-14.3	0-26.9	
			Electro-Optical	0-19	6.9-38.5	
	Abhyankar 2009 [11]	Play-Doh, cadaver, and gummy	Capacitive	-	-	13.85
			Optical			
			Electro-Optical			
	Plesh 2019 [12]	Paper print, transparent film, wood glue, latex, Play-Doh, ecoflex, gelatin, dragonskin, ModelMagic, and SillyPutty	Optical	0.02	13.8-18.35	-
	Husseis 2020 [5]	Play-Doh, gelatin, white glue, spray rubber, nail hardener, nail polish, and latex.	Optical	5	19.5	13
			Thermal	5	18.1	9.5

PA and PAD on fingerprint recognition have been widely investigated in [4], [13]–[15]. In our previous work [6], we classified PAs considering the attacker's intention, the used materials for creating the PAI, and whether a PAI contains dynamic or static information. On the other hand, different taxonomies have been proposed to classify PAD mechanisms [15]: (a) hardware/software classification sorts the PAD mechanisms by implying the necessity of modifying the hardware design of the biometric sensor, (b) dynamic/static classification clarify whether the temporal biometric information is needed for a PAD mechanism, and (c) collateral-means/natural-phenomena classification investigate whether the PAD features are natural characteristics of the biometric trait or just collateral information.

A key observation on the literature of fingerprint PAD mechanisms is that most researches tend to study the static fingerprint's pattern, e.g. 2-D textures and fingerprint quality, rather than fingerprint dynamic features. This can be explained by the fact that collecting dynamic datasets requires extensive time, effort, and expertise which consequently had led to limited dynamic datasets. In addition, integrating dynamic PAD algorithms into the biometric system may require higher computational power and potentially adds more load to the overall system. Section II briefly demonstrates the literature studies about dynamic fingerprint PAD and conducts an accuracy performance comparison between those mechanisms.

In this paper, we propose a PAD Framework that exploits the dynamic texture of the fingerprint as the discriminative foundation. Five state-of-the-art dynamic texture descriptors are used for the PAD feature extractor and then evaluated after an SVM classification. The dynamic model was chosen because we have experimentally noticed that genuine fingerprint presentations demonstrate a unique development of the ridge/valley pattern due to natural phenomena such as elasticity and perspiration. Moreover, PAs have shown perceptual and statistical dynamic differences as shown in our previous work [5].

The experimental protocol and evaluation methodology have been conducted following the standard ISO/IEC 30107-3:2017 – "Information technology – Biometric presentation attack detection – Part 3: Testing and reporting" [16]. Our proposed PAD subsystem demonstrates the capability of detecting PAs while having a low proportion of misclassified bona fide presentation.

This paper is structured as follows. Section II presents a brief overview of the related work. In the third section, we describe the framework of the proposed PAD subsystem. The experiment is characterized in Section IV. Section V reports and discusses the experimental results. Finally, we draw our conclusions are in section VI.

II. RELATED WORK

In this section, we propose a two-level illustration for the State of the Art (SoA) investigations. We first focus on the SoA in dynamic fingerprint PAD mechanisms and report a performance analysis. The second level concerns the literature on dynamic texture applications in biometrics.

A. DYNAMIC FINGERPRINT PAD MECHANISMS

Existing dynamic PAD mechanisms can be categorized into two main classes: perspiration based and ridge distortion based mechanisms. Perspiration based mechanisms rely on the fact that genuine fingerprints naturally produce moisture from the pores, this moisture diffuses during the interaction with the sensor surface resulting in a darker image as time goes by. Ridge distortion mechanisms base on the claim that bona fide and attack presentations produce significantly different distortions under certain presentation circumstances such as pressure [6]. Table 1 conducts a performance analysis for literature researches on both categories and shows the used sensors and attack species.

B. DYNAMIC TEXTURE: APPLICATIONS IN BIOMETRICS

Dynamic textures are textures with motion [17]. Ideally, a dynamic texture descriptor consolidates 2-D textures in a scene with temporal variations, meaning that information of space and time are obtained simultaneously. There is a vast amount of literature on dynamic texture recognition with application to biometric recognition and analysis, this section highlights some related works in the domain.

In their seminal paper of 2007 [18], Zhao and Pietikäinen proposed a simple approach to extract dynamic textures using Volume Local Binary Patterns (VLBP) and Local Binary Patterns from Three Orthogonal Planes (LBP-TOP). The method had been proposed with application to facial expression recognition and reported over 95% accuracy. Moreover, a recent study on spontaneous facial micro-expression recognition suggested a deep learning model based on spatial and temporal streams and reported 63.53%-74.05% accuracy [19].

In 2018, an experiment had been carried out on the applications of the VLBP in face PAD [20]. The authors had evaluated their PAD mechanism considering printed and replay attacks (video attacks). The PAD mechanism had successfully eliminated all printed attacks with 100% accuracy and mitigated replay attacks with 97.38% accuracy.

Additionally, various dynamic descriptors were suggested to categorize human actions. Solmaz et al. [21] extended the GIST descriptor into GIST 3-D and evaluated the method on different datasets, the authors obtained 92% accuracy for classifying 6 action categories. Further, the authors in [22] suggested utilizing the binarized statistical image feature (BSIF) to extract the dynamic features from 3-D salient patches and reported 93.43% accuracy for classifying low-quality videos.

III. PROPOSED PRESENTATION ATTACK DETECTION SUBSYSTEM

The proposed PAD subsystem is designed in a fashion that leverages the dynamic information provided during the fingerprint presentation (Figure 1). Thus, the proposed feature extraction approach suggests exploiting the spatio-temporal features to achieve a robust description that characterizes the complete interaction between the fingerprint and the sensor's surface. Toward this end, we propose three modes to investigate fingerprint dynamics in frequency and time domains. Five feature extractors are therefore selected to achieve a description that discriminates genuine from attack presentations. By feeding the extracted features into a pre-trained classifier, the PAD subsystem finally decides whether the input video is a bona fide or attack presentation. The following subsections expound the processing modes, feature extractors, and classification method.

A. FEATURE EXTRACTION MODES

In order to investigate different aspects of fingerprint dynamics, three feature extraction modes are elaborated in this subsection. The first mode investigates dynamic fingerprint features in the frequency domain whereas a 3-D filter bank is utilized to extract spectral features in a diverse range of scales and orientations. As the video's frequency components effectively represent the static fingerprint pattern and the temporal variations, it is expected that the differences between natural skin and attack species produce frequency components in different planes. Hence, this mode captures the spatio-temporal information by filtering the video frequency spectrum in different orientations and center frequencies.

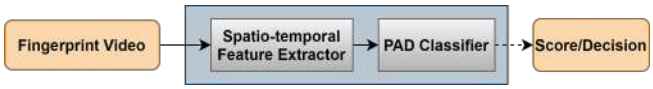


FIGURE 1: Dynamic PAD subsystem scheme.

The second mode samples the fingerprint video on space-time domain into small 3-D patches, extracts the spatio-temporal features from those samples, and provides the description as the frequency distribution of the extracted features. This mode has two main interesting features, primarily, it has the capacity to define local features in a stack of XY patches so that any anomalous formation in the fingerprint video is detected. Secondly, it provides the possibility of processing the 3-D patches in space-time and/or frequency domains.

The third mode resembles the second mode, a small brick is added after the sampling to decompose the 3-D patches into the Three Orthogonal Planes (TOP) XY, XT, and YT planes. Over the advantages of the second mode, the third mode had proved significantly reduced complexity for the adopted feature extractor while preserving a high accuracy [12].

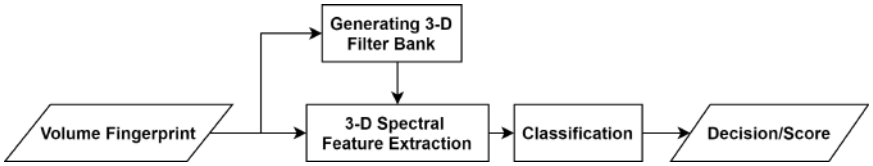
Figure 2 illustrates these modes and Figure 3 shows an example of a fingerprint video and its sampling into 3-D patches and TOPs.

B. FEATURE EXTRACTORS

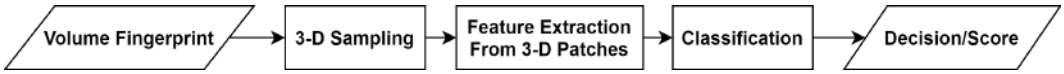
The feature extractors were selected in order to comply with the proposed modes, moreover, to analyze the features in spatio-temporal and spectral domains. Table 2 summarizes the proposed scenarios with the corresponding dynamic feature extractors and the following subsections reviews these algorithms.

1) GIST 3-D Descriptor

GIST 3-D is a global spatio-temporal descriptor that had been proposed for video classification problems. The method integrates the motion information and the scene structure in one feature vector without applying background subtraction or salient point detection at the input video, achieving performance better than SoA dynamic descriptors. In our experiment, the GIST 3-D works as follows: first, the frequency spectrum of the complete fingerprint video is achieved by applying 3-D Discrete Fourier Transform; as computed by Equation 1.



(a) Mode 1: spectral feature extraction using a 3-D filter bank



(b) Mode 2: spatio-temporal feature extraction from 3-D Patches

(c) Mode 3: spatio-temporal feature extraction from 3-D Patches decomposed into three orthogonal planes

FIGURE 2: Proposed PAD scheme in different modes.

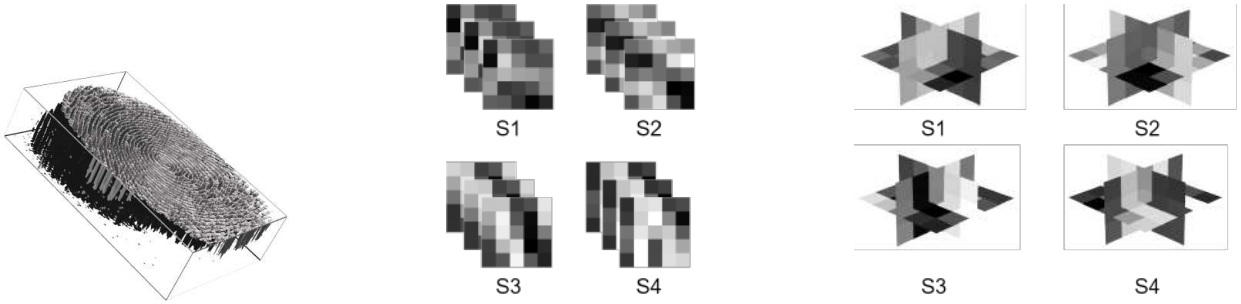


FIGURE 3: Illustration of 3-D sampling and decomposition. Fingerprint Video (a), 3-D patches sized (5x5x3) (b), and patches in b decomposed into xy, xt, and yt planes (c)

TABLE 2: Performance analysis for the SoA fingerprint PAD mechanisms (reported in [4] & [5])

FE algorithm	FE Mode	Domain of FE	Source of features	Reference
GIST 3-D	Mode 1	spatio-temporal frequency domain	Sub-volumes in the frequency domain	[21]
Volume Local Binary Patterns	Mode 2	spatio-temporal domain	3-D Patches	[18]
Local Binay Patterns from Three Orthogonal Planes	Mode 3	spatio-temporal domain	Patches of TOPs	[18]
Volume Local Phase Quantization	Mode 2	spatio-temporal frequency domain	3-D Patches	[23]
Local Phase Quantization from Three Orthogonal Planes	Mode 3	spatio-temporal frequency domain	Patches of TOPs	[23]

$$F(f_x, f_y, f_t) = \frac{1}{MNT} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} \sum_{t=0}^{T-1} f(x,y,t) \; e^{-j2\Pi(\frac{xf_x}{M} + \frac{yf_y}{N} + \frac{tf_t}{T})}, \quad (1)$$

Then, a bank of narrow band 3-D Gabor filters $G(fr, \theta, \phi)$ is generated and each 3-D filter $G_{\mathrm{i}}(f_{\mathrm{x}}, f_{\mathrm{y}}, f_{\mathrm{t}})$ is applied to the frequency spectrum as given by Equation 2. The filter bank is composed by 3-D filters with different orientations and scales, which allows capturing the components at various intervals of the video's frequency spectrum.

$$\Gamma_i(f_x, f_y, f_t) = F(f_x, f_y, f_t)[G_i(f_x, f_y, f_t)] \; , \quad (2)$$

After taking the inverse 3-D DFT as in Equation 3 for each filter in the bank, the output volume is quantized in fixed sub-volumes and the sum of each sub-volume is taken, thus, a feature vector is obtained to represent the video description.

$$H_i(x,y,t) = \sum_{f_x=0}^{M-1} \sum_{f_y=0}^{N-1} \sum_{f_t=0}^{T-1} \Gamma_i(f_x, f_y, f_t) \; e^{j2\Pi(\frac{xf_x}{M} + \frac{yf_y}{N} + \frac{tf_t}{T})}, \quad (3)$$

2) Volume Local Binary Patterns

The basic Local Binary Patterns method was extended to VLBP in order to describe the dynamic texture in a sequence of successive images [12]. The algorithm starts by sampling the gray level volume input into small 3D samples considering a certain number of local neighbors (P), time interval (L), and radius (R) in x-y plane, then every neighbor pixel in the 3D sample is given a binary value based on a comparison with the center pixel of the sample. Finally, each binary value is multiplied by a corresponding weight and all results are

summed to form the sample's $\text{VLBP}_{\text{L,P,R}}$ code; Equation 4. The distribution of the codes is used to compose the dynamic texture feature vector.

$$VLBP_{\text{L,P,R}} = \sum_{p=0}^{3P+1} s(g_p - g_c)\, 2^p \,, \tag{4}$$

where g_p and g_c correspond to the gray values of the central pixel and neighbours in the 3-D sample.

The authors in [18] also proposed two additional modes for the method: (1) rotation-invariant VLBP mode ($VLBP^{ri}_{L,P,R}$) which is based on the assumption that volume data rotates only around t-axis, (2) uniform VLBP mode ($VLBP^{u2}_{L,P,R}$), where the VLBP histogram consists of uniform patterns (i.e. patterns contain at most 2 bitwise transitions between 0 and1) and sums up all non-uniform patterns in 1 bin.

3) Volume Local Phase Quantizer

The VLPQ method [23] is an extension to the local phase quantization which was originally proposed as an image descriptor [24]. VLPQ essentially encodes local Fourier transform's phase information at low-frequency points. The method consists of three steps: (1) local Fourier transform is applied, using Short Term Fourier Transform (STFT), over $M \times M \times N$ neighborhood N_x centered at each pixel position x using 1-D convolutions for each dimension, (2) the dimensionality of the achieved data is reduced using Principal Component Analysis (PCA), and (3) a scalar quantization is applied to produce an integer value. The histogram of the binary codewords is computed to form the $\text{VLPQ}_{\text{M,N}}$ feature vector.

4) Local Binary Patterns from Three Orthogonal Planes

Although VLBP method is interesting, it suffers from two major issues. First, initializing the algorithm with a large number of neighbors P results in a very large number of patterns in the VLBP feature vector, limiting the method's applicability. Second, choosing a time radius L larger than 1 excludes the frames with a time variance less than L.

To address these issues, $\text{VLBP-TOP}_{\text{L,P,R}}$ method had been proposed in [18] to concatenate the local binary patterns on the three orthogonal planes: XY-LBP, XT-LBP, and YT-LBP. With this approach, spatial patterns are obtained from XY plane and space-time transitions information is attained from XT and YT planes. As a result, the number of patterns on the feature vector is significantly reduced from 2^{3P+2} to 3×2^P which allows considering a large number of neighbors with reduced computational cost, moreover, including neighbor pixels from frames with a time variance less than L, when L is larger than 1.

5) Local Phase Quantizer from Three Orthogonal Planes

$\text{LPQ-TOP}_{\text{Rx,Ry,Rz}}$ is implemented by calculating LPQ histograms from three orthogonal planes similar to LBP-TOP. The histograms are normalized and concatenated to form the LPQ-TOP descriptor [23].

C. PAD CLASSIFICATION

Through our experiment, we have tested different classification algorithms, specifically: Classification Trees, Discriminant Analysis, Naive Bayes, Nearest Neighbors, SVM Classification, and Classification Ensembles. SVM classification has been chosen due to its highest accuracy, while the other classification methods are not considered in this paper. Moreover, we have examined the impact of changing the SVM kernel whereas a second polynomial kernel demonstrated the best accuracy. A binary classification scheme has been utilized to evaluate the PAD subsystem performance and to assess the influence of specific PAI species on system security and usability.

IV. EXPERIMENT

To evaluate the performance of the proposed PAD subsystem, we use the dynamic dataset presented in [5]. In the initial stage of the experiment, a volume segmentation is applied to the database. This sets the input fingerprint videos to the feature extraction step. At this point, we utilize the scheme in figure 2 to extract the features and train the SVM model. As soon as these steps have been carried out, the testing process is performed, and the PAD subsystem accuracy is assessed.

A. DATABASE DESCRIPTION

The database had been collected to capture genuine and cooperative-attack presentations as videos using optical and thermal sensors. The database comprises 66 genuine fingerprints (thumb, index middle) taken from both hands of 11 independent subjects, and attacks using 7 PAI species. A definite characterization of the protocol applied to produce this database is introduced in [5].

Table 3 summarizes the 3564 bona fide and attack presentations in the database with the corresponding presentation type.

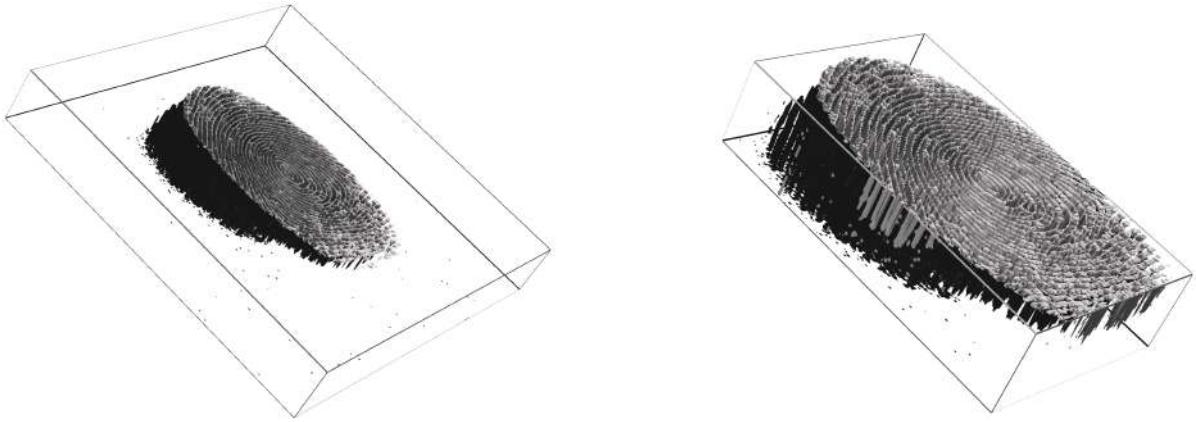
TABLE 3: Fingerprint presentation types in the database (the same protocol is applied for the 2 sensors)

Presentation type	Visit/PAI species	Number of presentations per sensor
Bona fide	visit 1	198
	visit 2	198
Attack	Play-Doh	198
	Gelatin	198
	White glue	198
	Spray rubber	198
	Nail hardener	198
	Nail polish	198
	Latex	198
Total	_	**1782**

The Common Criteria (CC) defines the attack potential as a function of expertise, resources, and motivation of the attacker. Reporting those aspects in biometric databases is therefore indispensable to the coherence of the PAD evaluation. We thus report that all attacks were carried out by one

TABLE 4: Comparison between the characteristics of the sensors and presentations in the database

Sensing technology	Resolution	Image size	Size of the segmented images	Scan time	number of frames per presentation
Optical	500 ppi	900 x 900 pixels	depends on the touched surface for each presentation	0.05 second/image	Varies with respect to the user's presentation time with average of 25 frames/presentation
Thermal	385 ppi	180 x 256 pixels	90 x 128 pixels	0.7 second/image	7



(a) (b)

FIGURE 4: Demonstration of a volume segmentation for a presentation consists of 29 successive frames. Before segmentation sized 375x400 (a), and after segmentation 234x145 (b). The figures do not reflect the real scale of the fingerprint.

attacker, he has an advanced knowledge in biometric systems and had proven practical experience in attacking fingerprint sensors embedded in smartphones. Furthermore, the attacker obtained all required materials from local shops and online stores for a very low cost. Accordingly, the attacker has prepared each PAI species with a particular recipe and determined that a PAI can be used multiple times for all species except the Play-Doh instrument where each attack requires a new PAI.

B. VOLUME SEGMENTATION

The dataset was collected using optical and thermal sensors where each sensor acquires images with different characteristics. Taking into account the sensors' features and database characteristics in Table 4, the following subsections illustrate the adopted segmentation techniques.

1) Thermal subset

The thermal sensor's SDK provides a capturing mode that acquires only the central region of the sensor sized 90×128 pixels. Thus, the acquired sequence is already segmented as a stack of 7 frames sized 90×128.

2) Optical subset

Since our study analyzes the formation of fingerprints, we have implemented a simple volume segmentation tool that creates the boundaries of the entire Interaction between a fingerprint and the sensor and crop the 3-D volume; an example is shown in Figure 5. Then, we have applied the segmentation to the entire subset of the optical sensor before feature extraction.

C. EXPERIMENTAL PROTOCOL

Each sensor subset is evaluated independently due to the differences in the sensors' technology, image size, resolution, noise, and capturing rate which produce different video characteristics. For a robust accuracy estimation, we have set a holdout validation scheme where the database is divided into training (55%) and testing (45%) sets. The database division into training/testing is randomized by independent subjects, meaning that presentations of each independent subject is either used for training or testing.

Since the work is focused on the PAD subsystem, we report the error rates following the recommendations of ISO/IEC 30107-3:2017 standard on PAD testing and reporting. The PAD subsystem evaluation determines the system's capability of detecting attacks taking into account the measurement of false detections.

The following metrics are used in the results to evaluate PAD mechanisms:

- Attack Presentation Classification Error Rate (APCER) presents the proportion of attack presentations incorrectly classified as bona fide presentations. Besides, $APCER_{PAIS}$ is outlined to denote the misclassified attack proportion for a given PAI species;
- Bona Fide Presentation Classification Error Rate (BPCER) presents the proportion of bona fide presentations incorrectly classified as attack presentations;
- Tradeoff Equal Error Rate (TEER) is when APCER and BPCER are equal. We introduce TEER, which is not defined in the standards, to compare with SoA mechanisms that were reported only in terms of TEER, and moreover to prevent the confusion with the conventional

EER.

The use of TEER to compare different PAD mechanisms is not recommended because it shows the systems BPCER at different APCER points. It is preferable to evaluate the PAD mechanism in terms of BPCER at fixed APCER, for instance, reporting a PAD mechanism's BPCER when APCER is 5% is standarized as BPCER20. Furthermore, showing the DET curves [25] provides a precise description of the relationship between APCER and BPCER at different thresholds, allowing better comparison between different mechanisms.

V. RESULTS AND DISCUSSION

In this section, we assess the accuracy of the proposed PAD scheme and analyze the influence of selecting the feature extractor on the PAD subsystem efficiency.

A. IMPACT OF PAD SUBSYSTEM MODE AND FEATURE EXTRACTION METHOD

The first set of analyses examined the impact of (i) the size of 3-D samples used in the processing mode, and (ii) selecting rotation invariant or uniform features, on the feature extractor performance. Figure 5 and Figure 6 show DET curves for VLBP, LBP-TOP, VLPQ, and LPQ-TOP with the corresponding sampling parameters. The figures confirm that 3-D spectral features (i.e. VLPQ and LPQ-TOP) performs better at smaller sampling size, and the accuracy degrades considerably when comparing the smallest and largest sampling size. An exception is noticed for the LPQ-TOP when executed on the optical sensor. On the other hand, 3-D spatio-temporal features (i.e. VLBP and LBP-TOP) have not revealed a general correlation between sampling size and accuracy. However, it is evident that rotation invariant and uniform features do not necessarily improve the accuracy in most of the cases but nonetheless no significant degradation has taken place after considering those features.

Table 5 and Table 6 detail the results categorized by the feature extraction method. We have selected multiple thresholds: (i) TEER, (ii) APCER= 5%, and (iii) APCER= 2.5% to evaluate the methods at different security levels. The table reveals the total number of the misclassified bona fide/attack presentations at each threshold. It is worthwhile noting that testing data, which corresponds to 5 independent subjects, consists of 630 attack and 180 bona fide presentations.

We then carry out a performance comparison between the five dynamic feature extraction methods (Figure 7) by selecting the methods' best parameters from tables 5 and 6. Note that those parameters had been chosen empirically, thus they might not be optimal for the suggested feature extractors in the context of our experiment.

The most striking result to emerge from Figure 7 is the achievement of significantly low BPCER20, where the system security remains high (low APCER) with low bona fide rejects (low BPCER), that is to say, these results offer powerful evidence for the fact that a genuine fingerprint provides sufficiently discriminative dynamic information that distinguishes it from attacks.

B. IMPACT OF SENSING TECHNOLOGY

We next investigate the robustness of the proposed PAD subsystem when different fingerprint sensing technologies are used, explicitly, we compare the PAD accuracy for the thermal and optical sensors (Figure 7) in terms of BPCER20. We observe from Table 7 that the accuracy of the PAD subsystem for the optical sensor has an advantage over the thermal sensor. The distinction appears to be well substantiated by the higher frame rate, image size, and resolution in the optical sensor which allows to precisely capture the fingerprint/PAI formation; i.e. spatio-temporal information. Moreover, each presentation in the thermal sensor is captured over roughly 5 seconds while in the optical sensor, a presentation can be captured in 0.5 second including 10 successive frames.

C. IMPACT OF ATTACK SPECIES

This section expounds the results in section A seeking to point out the attack potential for each PAI species. The classification results are shown considering the SVM classification decision in Tables 8 and 9.

As expected, the tables prove that different attack species have different attack potential considering a target sensor/PAD method. The PAD subsystem has been capable of eliminating some of the attack species and mitigate the rest of the species. Even though the overall performance for the optical sensor has been proven to be higher than the thermal sensor, a comparison between Table 8 and Table 9 demonstrates that the thermal sensor is notably vulnerable to white glue attacks but resistant to the rest of attack species. On the other hand, the optical sensor shows either low or 0% APCER for all attack species.

D. ACCURACY COMPARISON WITH SOA MECHANISMS

To conduct a comparison between different PAD mechanisms, we emphasize the importance of considering the differences between experimental protocols, used databases, and evaluation methodologies. These factors refer to a certain attack potential to specific database/technology and evaluated using defined metrics.

In the previous sections, these factors and the obtained results have been characterized to a considerable extent in order to allow the reader to compare our proposed PAD mechanism with SoA mechanisms in Table 1. We note that our results for the optical technology demonstrate significant improvement to the SoA methods.

E. TIME PERFORMANCE

Finally, we assess the computational cost of our PAD subsystem. The computation lies in the two main bricks: the feature extractor, and classifier. The evaluation is conducted using the MATLAB source codes provided by the authors of the dynamic descriptors and the Statistics and Machine Learning Toolbox – MATLAB [26]. The used machine is a Dell XPS/15/9560 at 2.80 GHz CPU, 16 GB RAM, and Windows 10 Pro 64-bit operating system. The codes had not been optimized for our use case and executed to verify

TABLE 5: PAD subsystem performance for the optical sensor

Discriptor	at TEER			at APCER = 5%			at APCER =2.5%		
	TEER	Successful attacks	Rejected bona fide	BPCER20	Successful attacks	Rejected bona fide	BPCER	Successful attacks	Rejected bona fide
$VLPQ_{3,3}$	**5.56%**	35	10	**5.56%**	31	10	**8.89%**	16	16
$VLPQ_{5,5}$	6.11%	39	11	6.67%		12	11.11%		20
$VLPQ_{7,7}$	9.21%	58	17	15.00%		27	29.44%		53
$VLPQ_{9,9}$	9.44%	60	17	12.22%		22	25.00%		45
$LPQ-TOP_{3,3,3}$	5.08%	32	9	5.56%	31	10	9.44%	16	17
$LPQ-TOP_{5,5,5}$	**3.89%**	25	7	**3.89%**		7	**7.22%**		13
$LPQ-TOP_{7,7,7}$	6.11%	39	11	6.67%		12	7.78%		14
$LPQ-TOP_{9,9,9}$	5.56%	35	10	6.11%		11	11.11%		20
			0						
GIST 3-D	**5.56%**	35	10	**6.67%**	31	12	**9.44%**	16	17
$VLBP_{1,4,1}$	3.65%	23	7	1.67%	31	3	7.22%	16	13
$VLBP_{1,4,3}$	4.76%	30	9	4.44%		8	11.67%		21
$VLBP_{2,4,1}$	**2.78%**	18	5	**1.67%**		3	**5.00%**		9
$VLBP_{2,4,3}$	3.65%	23	7	2.22%		4	5.56%		10
$VLBP^{ri}_{1,4,1}$	5.00%	32	9	5.00%	31	9	8.89%	16	16
$VLBP^{ri}_{1,4,3}$	6.67%	42	12	8.33%		15	13.33%		24
$VLBP^{ri}_{2,4,1}$	3.89%	25	7	3.33%		6	5.00%		9
$VLBP^{ri}_{2,4,3}$	4.92%	31	9	4.44%		8	8.89%		16
$LBP-TOP_{1,8,1}$	4.44%	28	8	3.89%	31	7	5.00%	16	9
$LBP-TOP_{1,8,3}$	3.97%	25	7	2.22%		4	5.56%		10
$LBP-TOP_{2,8,1}$	3.65%	23	7	2.78%		5	3.89%		7
$LBP-TOP_{2,8,3}$	3.89%	25	7	2.78%		5	3.89%		7
$LBP-TOP^{u2}_{1,8,1}$	4.76%	30	9	4.44%	31	8	7.22%	16	13
$LBP-TOP^{u2}_{1,8,3}$	**2.22%**	14	4	**1.11%**		2	**2.22%**		4
$LBP-TOP^{u2}_{2,8,1}$	5.56%	35	10	5.56%		10	7.22%		13
$LBP-TOP^{u2}_{2,8,3}$	3.49%	22	6	2.22%		4	6.11%		11

TABLE 6: PAD subsystem performance for the thermal sensor

Discriptor	at TEER			at APCER = 5%			at APCER =2.5%		
	TEER	Successful attacks	Rejected bona fide	BPCER20	Successful attacks	Rejected bona fide	BPCER	Successful attacks	Rejected bona fide
$VLPQ_{3,3}$	**8.10%**	51	15	**13.89%**	31	25	**27.78%**	16	50
$VLPQ_{5,5}$	13.02%	82	23	31.67%		57	46.11%		83
$VLPQ_{7,7}$	17.94%	113	32	46.67%		84	65.00%		117
$LPQ-TOP_{3,3,3}$	**4.92%**	31	9	**3.89%**	31	7	**14.44%**	16	26
$LPQ-TOP_{5,5,5}$	7.30%	46	13	9.44%		17	17.78%		32
$LPQ-TOP_{7,7,7}$	6.67%	42	12	8.89%		16	22.78%		41
GIST 3-D	**12.22%**	77	22	**28.89%**	31	52	**46.67%**	16	84
$VLBP_{1,4,1}$	**12.86%**	81	23	**30.00%**	31	54	**51.67%**	16	93
$VLBP_{1,4,3}$	16.19%	102	29	27.22%		49	48.33%		87
$VLBP_{2,4,1}$	16.03%	101	29	37.22%		67	61.11%		110
$VLBP_{2,4,3}$	19.44%	123	35	43.89%		79	57.78%		104
$VLBP^{ri}_{1,4,1}$	12.78%	81	23	23.89%	31	43	41.11%	16	74
$VLBP^{ri}_{1,4,3}$	12.70%	80	23	37.22%		67	72.22%		130
$VLBP^{ri}_{2,4,1}$	16.19%	102	29	33.33%		60	53.89%		97
$VLBP^{ri}_{2,4,3}$	13.33%	84	24	35.56%		64	56.67%		102
$LBP-TOP_{1,8,1}$	7.78%	49	14	10.56%	31	19	23.33%	16	42
$LBP-TOP_{1,8,3}$	8.33%	53	15	16.67%		30	24.44%		44
$LBP-TOP_{2,8,1}$	7.46%	47	13	10.00%		18	20.00%		36
$LBP-TOP_{2,8,3}$	**7.22%**	46	13	**15.00%**		27	**36.11%**		65
$LBP-TOP^{u2}_{1,8,1}$	7.78%	49	14	14.44%	31	26	28.33%	16	51
$LBP-TOP^{u2}_{1,8,3}$	8.33%	53	15	13.33%		24	28.33%		51
$LBP-TOP^{u2}_{2,8,1}$	8.33%	53	15	12.78%		23	29.44%		53
$LBP-TOP^{u2}_{2,8,3}$	8.25%	52	15	15.00%		27	37.22%		67

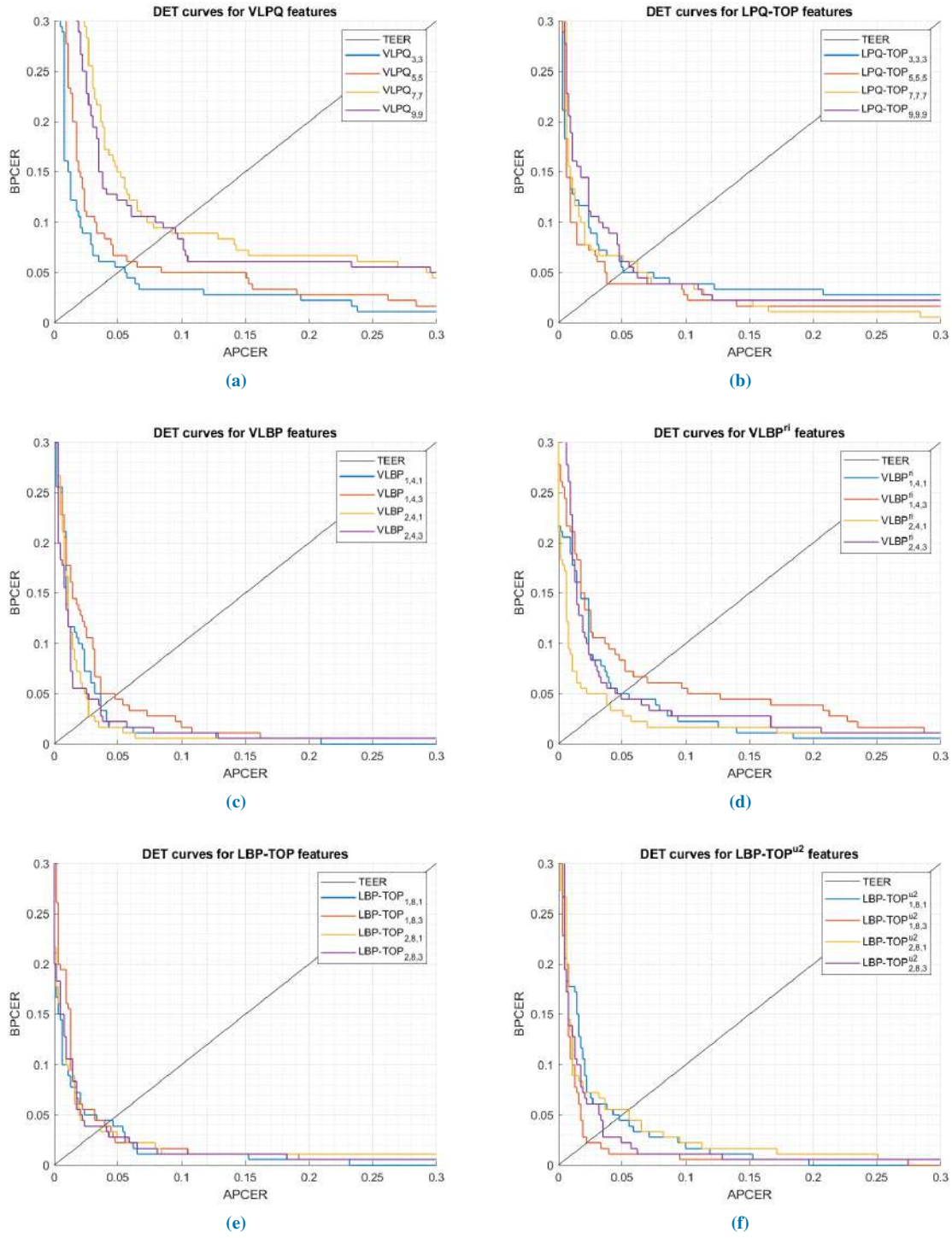


FIGURE 5: DET curves comparison of the proposed feature extraction algorithms using different parameters (optical sensor).

TABLE 7: BPCER20 comparison between the optical and thermal sensors

Sensor\FE	VLPQ		LPQ-TOP		GIST 3-D	VLBP		LBP-TOP	
Optical	$VLPQ_{3,3}$	5.56%	$LPQ-TOP_{5,5,5}$	3.89%	6.67%	$VLBP_{2,4,1}$	1.67%	$LBP-TOP^{u2}_{1,8,3}$	2.22%
Thermal	$VLPQ_{3,3}$	13.89%	$LPQ-TOP_{3,3,3}$	3.89%	28.89%	$VLBP_{1,4,1}$	30.00%	$LBP-TOP_{2,8,3}$	15.00%
Difference	8.33%		**0.00%**		22.22%	28.33%		12.78%	

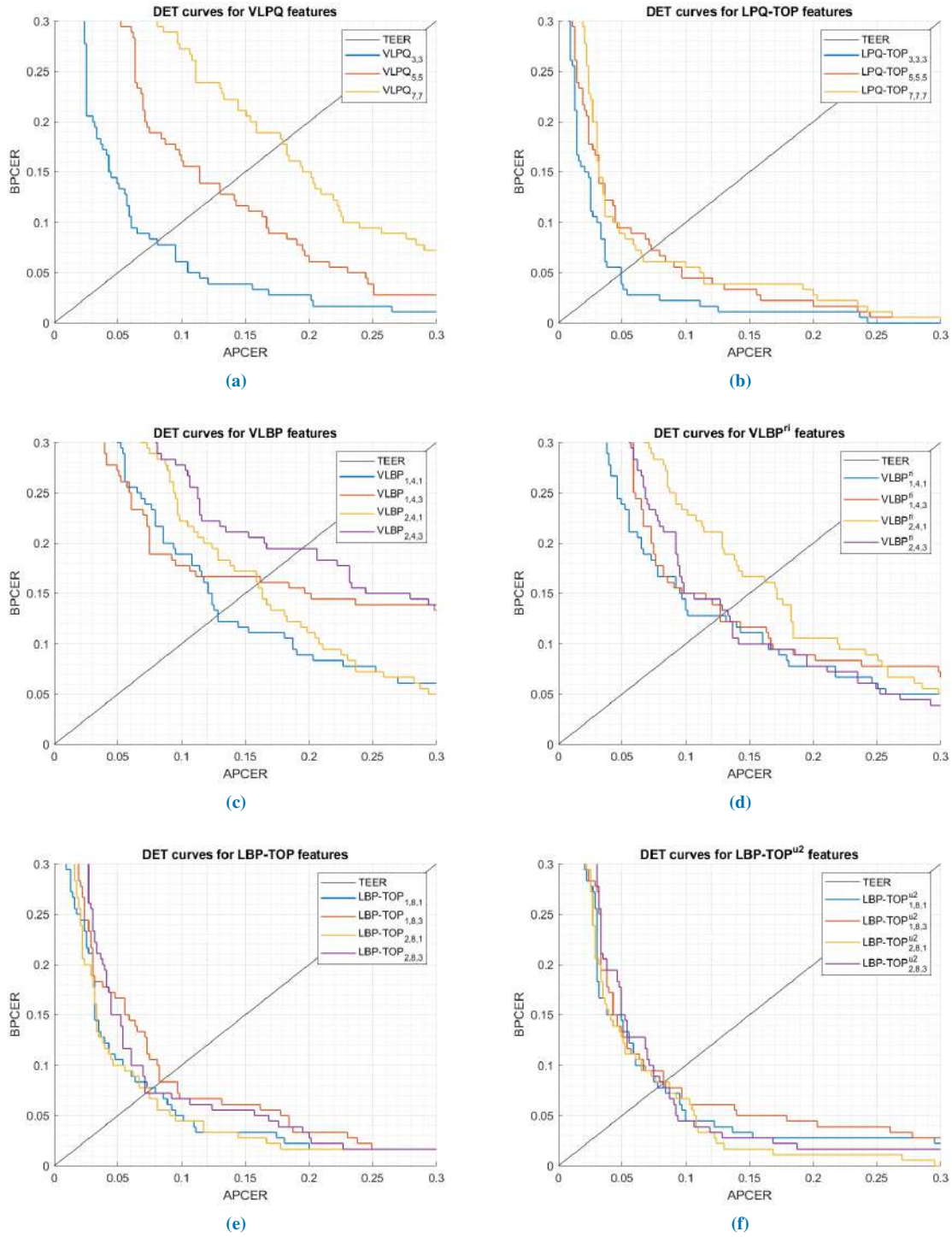
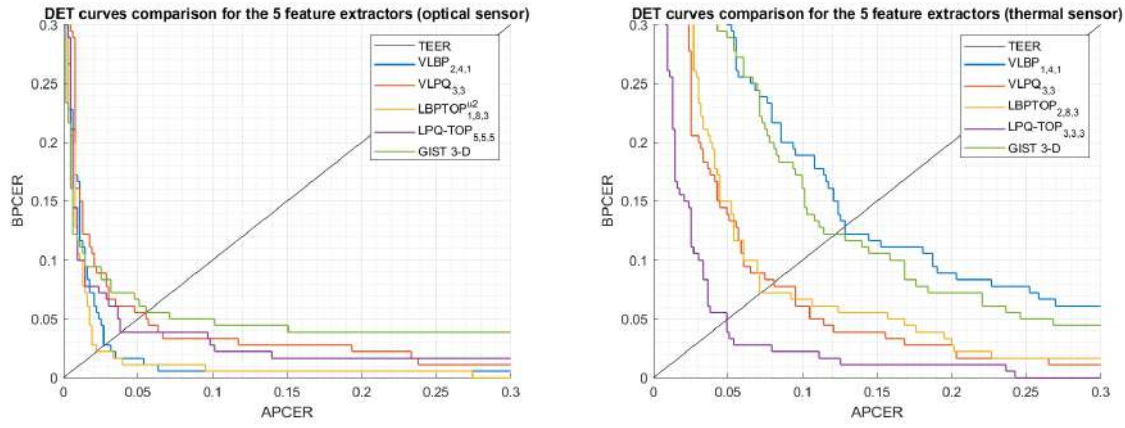


FIGURE 6: DET curves comparison of the proposed feature extraction algorithms using different parameters (thermal sensor).

the PAD mechanism efficiency rather than the computational complexity, nevertheless, the analyses in this section give an insight into our experimental work.

We separately evaluate the feature extraction time for optical and thermal sensors in Table 10 and Figure 7, and classification time in Table 11.

From Table 10 and Figure 8, we observe that computation time in 3-D frequency domain is significantly lower than that in 3-D spatial domain. Moreover, Figure 6 shows the influence of the presentation length, i.e. the number of frames per presentation, on computation time.



(a) (b)

FIGURE 7: DET curves comparison of the proposed PAD subsystem using five feature extractors.

TABLE 8: Attacks strength considering different PAI species (optical sensors).

Feature Extractor	SVM error rates		$APCER_{PAI}$						
	APCER	BPCER	PlayDoh	White glue	Spray rubber	Polish nail	Nails hardener	Gelatin	Latex
$VLBP_{2,4,1}$	1.75%	7.78%	**0.00%**	1.11%	1.11%	**0.00%**	8.89%	**0.00%**	1.11%
$LBP-TOP1,8,3$	1.59%	6.67%	3.33%	1.11%	**0.00%**	1.11%	1.11%	2.22%	2.22%
$VLPQ_{3,3}$	3.33%	6.67%	5.56%	**0.00%**	3.33%	1.11%	8.89%	4.44%	**0.00%**
$LPQ-TOP_{5,5,5}$	2.38%	11.67%	3.33%	3.33%	**0.00%**	**0.00%**	3.33%	4.44%	2.22%
$GIST\ 3-D$	1.43%	10.56%	4.44%	1.11%	2.22%	1.11%	**0.00%**	1.11%	**0.00%**

TABLE 9: Attacks strength considering different PAI species (thermal sensors).

Feature Extractor	SVM error rates		$APCER_{PAI}$						
	APCER	BPCER	PlayDoh	White glue	Spray rubber	Polish nail	Nails hardener	Gelatin	Latex
$VLBP_{1,4,1}$	1.59%	56.11%	**0.00%**	10.00%	1.11%	**0.00%**	**0.00%**	**0.00%**	**0.00%**
$LBP-TOP2,8,3$	4.44%	16.67%	1.11%	21.11%	6.67%	2.22%	**0.00%**	**0.00%**	**0.00%**
$VLPQ_{3,3}$	3.33%	18.33%	2.22%	15.56%	1.11%	1.11%	**0.00%**	3.33%	**0.00%**
$LPQ-TOP_{3,3,3}$	2.70%	11.11%	**0.00%**	8.89%	4.44%	2.22%	**0.00%**	3.33%	**0.00%**
$GIST\ 3-D$	4.76%	29.44%	8.89%	24.44%	**0.00%**	**0.00%**	**0.00%**	**0.00%**	**0.00%**

TABLE 10: number of extracted features for each method with the corresponding extraction time (thermal sensor)

	Feature extraction method				
	GIST 3-D	VLBP	LBP-TOP	VLPQ	LPQ-TOP
bins of FE histogram	34816	16384	768	1024	768
FE time (in seconds)	0.995	0.406	0.590	0.124	0.267

TABLE 11: Prediction time for a single presentation (in seconds).

Features	GIST 3-D	VLBP	LBPTOP	VLPQ	LPQTOP
Classification time	22.548	4.700	0.067	0.100	0.073

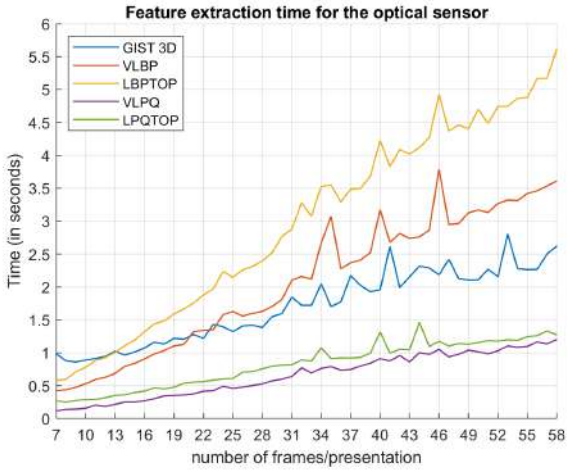


FIGURE 8: Average FE time for the optical sensor.

VI. SUMMARY AND CONCLUSIONS

In this paper, we present a novel fingerprint PAD approach in the dynamic scenario. We propose three modes to investigate the spatio-temporal and spectral features in fingerprint videos. We utilize five dynamic feature extractors to leverage the fingerprint features in space and time, then a binary SVM is used for classifying bona fide and attack presentations. The PAD mechanism is assessed using a database that was collected using optical and thermal sensors and consists of 792 bona fide presentations taken from 66 genuine fingerprints and 2772 attack presentations performed by an attacker using 7 PAI species.

The significance of the proposed approach is that it integrates the effect of all natural fingerprint phenomena from the acquired video using dynamic descriptors, for instance, the intensity variation caused by the perspiration and pressure, and the ridge/valley pattern's formation caused by the 3-D form and elasticity of genuine fingerprints. Moreover, the approach has the capacity to detect anomalous patterns caused by the various PAI species, concequently, enhance the PAD subsystem's accuracy.

The local spatio-temporal features were extracted using VLBP and LBP-TOP. On the other hand, spectral features were explored locally using VLPQ and LPQ-TOP, and globally using GIST 3-D. These feature extractors are evaluated for a thermal and an optical sensors showing an advantage for the latter due to its acquisition characteristics.

The experiment points out the importance of studying each sensing technology apart through comparing (i) the accuracy of the different feature extractors, and (ii) the potential of the attack species on the two sensors. The best accuracy is obtained by LBP-TOP for the optical sensor with 1.11 BPCER20, and by LPQ-TOP for the thermal sensor with 3.89 BPCER20.

These results would seem to suggest that our approach has an excellent capability of eliminating/mitigating PAs in different sensing technologies. Further, a comparison with SoA mechanisms shows that our method provides competitive error rates. However, given the small number of participants in the database, caution must be taken.

Our results are promising and should be validated by a larger database with additional attack species and sensing technologies. We recommend that further research should concentrate on fingerprint specific dynamic features such as the variation of fingerprint quality during the presentation.

ACKNOWLEDGMENT

We thank the authors of GIST 3-D, VLBP, and VLPQ for making the source codes available online. We also thank Ramon Blanco-Gonzalo for the valuable discussions and suggestions.

REFERENCES

[1] "Future Smartphone Payments to Rely on Software Security." [Online]. Available: https://www.juniperresearch.com/press/press-releases/future-smartphone-payments-rely-software-security
[2] N. K. Ratha, J. H. Connell, and R. M. Bolle, "Enhancing security and privacy in biometrics-based authentication systems," IBM Systems Journal, vol. 40, no. 3, pp. 614–634, 2001.
[3] T. D. Thandauthapani, A. J. Reeve, A. S. Long, I. J. Turner, and J. S. Sharp, "Exposing latent fingermarks on problematic metal surfaces using time of flight secondary ion mass spectroscopy," Science & Justice, vol. 58, no. 6, pp. 405–414, nov 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1355030618301515
[4] E. Marasco and A. Ross, "A Survey on Antispoofing Schemes for Fingerprint Recognition Systems," ACM Computing Surveys, vol. 47, no. 2, pp. 1–36, nov 2014.
[5] A. Husseis, J. Liu-Jimenez, I. Goicoechea-Telleria, and R. Sanchez-Reillo, "Dynamic Fingerprint Statistics: Application in Presentation Attack Detection," IEEE Access, vol. 8, pp. 95 594–95 604, 2020.
[6] A. Antonelli, R. Cappelli, D. Maio, and D. Maltoni, "Fake Finger Detection by Skin Distortion Analysis," IEEE Transactions on Information Forensics and Security, vol. 1, no. 3, pp. 360–373, sep 2006.
[7] Y. Zhang, J. Tian, X. Chen, X. Yang, and P. Shi, "Fake Finger Detection Based on Thin-Plate Spline Distortion Model," in Advances in Biometrics. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 742–749.
[8] J. Jia, L. Cai, K. Zhang, and D. Chen, "A New Approach to Fake Finger Detection Based on Skin Elasticity Analysis," in Advances in Biometrics. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 309–318.
[9] R. Derakhshani, S. A. Schuckers, L. A. Hornak, and L. O'Gorman, "Determination of vitality from a non-invasive biomedical measurement for use in fingerprint scanners," Pattern Recognition, vol. 36, no. 2, pp. 383–396, feb 2003.
[10] S. Parthasaradhi, R. Derakhshani, L. Hornak, and S. Schuckers, "Time-Series Detection of Perspiration as a Liveness Test in Fingerprint Devices," IEEE Transactions on Systems, Man and Cybernetics, Part C (Applications and Reviews), vol. 35, no. 3, pp. 335–343, aug 2005.
[11] A. Abhyankar and S. Schuckers, "Integrating a wavelet based perspiration liveness check with fingerprint recognition," Pattern Recognition, vol. 42, no. 3, pp. 452–464, mar 2009.
[12] R. Plesh, K. Bahmani, G. Jang, D. Yambay, K. Brownlee, T. Swyka, P. Johnson, A. Ross, and S. Schuckers, "Fingerprint Presentation Attack Detection utilizing Time-Series, Color Fingerprint Captures," in 2019 International Conference on Biometrics, ICB 2019, jun 2019.
[13] C. Busch and C. Sousedik, "Presentation attack detection methods for fingerprint recognition systems: a survey," IET Biometrics, vol. 3, no. 4, pp. 219–233, dec 2014.
[14] A. Husseis, J. Liu-Jimenez, I. Goicoechea-Telleria, and R. Sanchez-Reillo, "A survey in presentation attack and presentation attack detection," in Proceedings - International Carnahan Conference on Security Technology, oct 2019.
[15] M. S. Nixon, Handbook of Biometric Anti-Spoofing. Cham, Switzerland: Springer, 2019, no. April.
[16] "ISO/IEC 30107-3:2017 - Information technology – Biometric presentation attack detection – Part 3: Testing and reporting."
[17] M. Szummer and R. W. Picard, "Temporal texture modeling," in IEEE International Conference on Image Processing, vol. 3. IEEE, 1996, pp. 823–826.
[18] G. Zhao and M. Pietikäinen, "Dynamic texture recognition using local binary patterns with an application to facial expressions," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 29, no. 6, pp. 915–928, jun 2007.
[19] B. Song, K. Li, Y. Zong, J. Zhu, W. Zheng, J. Shi, and L. Zhao, "Recognizing spontaneous micro-expression using a three-stream convolutional neural network," IEEE Access, vol. 7, pp. 184 537–184 551, 2019.
[20] X. Zhao, Y. Lin, and J. Heikkila, "Dynamic Texture Recognition Using Volume Local Binary Count Patterns with an Application to 2D Face Spoofing Detection," IEEE Transactions on Multimedia, vol. 20, no. 3, pp. 552–566, mar 2018.
[21] B. Solmaz, S. M. Assari, and M. Shah, "Classifying web videos using a global video descriptor," Machine Vision and Applications, vol. 24, no. 7, pp. 1473–1485, sep 2013.
[22] S. Rahman and J. See, "Spatio-temporal mid-level feature bank for action recognition in low quality video," in ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings, vol. 2016-May. Institute of Electrical and Electronics Engineers Inc., may 2016, pp. 1846–1850.
[23] J. Päivärinta, E. Rahtu, and J. Heikkilä, "Volume local phase quantization for blur-insensitive dynamic texture classification," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 6688 LNCS. Springer, Berlin, Heidelberg, 2011, pp. 360–369.
[24] V. Ojansivu and J. Heikkilä, "Blur Insensitive Texture Classification Using Local Phase Quantization." Springer, Berlin, Heidelberg, 2008, pp. 236–243. [Online]. Available: http://link.springer.com/10.1007/978-3-540-69905-7_27
[25] A. Martin, A. Martin, G. Doddington, T. Kamm, M. Ordowski, and M. Przybocki, "The DET curve in assessment of detection task performance," in Proceedings of the European Conference on Speech Communication and Technology, 1997, pp. 1895—-1898. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.117.4489
[26] "Statistics and Machine Learning Toolbox - MATLAB." [Online]. Available: https://nl.mathworks.com/products/statistics.html

ANAS HUSSEIS received the B.S. degree in communication engineering from Yarmouk University, Irbid, Jordan, in 2012, and the M.S. degree in multimedia networking from the Telecom-ParisTech, Paris-Saclay University, Paris, France, in 2017. He is currently pursuing the Ph.D. degree in electrical engineering, electronics and automation with the University Carlos III of Madrid (UC3M), Madrid, Spain. From 2012 to 2016, he was a Communication Engineer with STC. He was a Visiting Researcher with the Warsaw University of Technology, Warsaw, Poland. He was also a Researcher with Next Biometrics Research and Development, Prague, Czech Republic, in 2018, in the framework of the European Project AMBER. He is a Marie Sklodowska-Curie Research Fellow with AMBER project. His research interests include biometric recognition with a focus on presentation attack detection, biometric systems evaluation, and computer vision. He is a member of the European Association for Biometrics (EAB).

JUDITH LIU-JIMENEZ received the degree in telecommunication engineering from the Polytechnic University of Madrid, in 2004, and the Ph.D. degree in electronics from the University Carlos III of Madrid (UC3M), in 2010. Since 2004, she has been with the UC3M. She has participated in several national and European funded projects, besides working on ID management, evaluation, and anti-spoofing mechanisms. Her focus of work is on biometrics and hardware/software codesign, specifically for iris biometrics.

RAUL SANCHEZ-REILLO received the Ph.D. degree. He is currently a Full Professor with the University Carlos III of Madrid. He is also the Head of the University Group for Identification Technologies (GUTI), where he is involved in project development and management concerning a broad spectrum of applications, ranging from social security services to financial payment methods. He has participated in several European projects, such as eEpoch and BioSec, by virtue of being the WP Leader. He is an expert in security and biometrics. He served as a member of the SC17, SC27, and SC37 Standardization Committees. He is also the Spanish Chair of SC17 and the Secretariat of SC37.