2016 23rd International Conference on Pattern Recognition (ICPR)
Cancún Center, Cancún, México, December 4-8, 2016

Face Anti-Spoofing with Multifeature Videolet Aggregation

Talha Ahmad Siddiqui[1,2], Samarth Bharadwaj[2], Tejas I. Dhamecha[2], Akshay Agarwal[1], Mayank Vatsa[1,*], Richa Singh[1], and Nalini Ratha[2]
[1] *IIIT-Delhi, Delhi, India*, [2] *IBM Research Labs*

***Abstract*—Biometric systems can be attacked in several ways and the most common being spoofing the input sensor. Therefore, anti-spoofing is one of the most essential prerequisite against attacks on biometric systems. For face recognition it is even more vulnerable as the image capture is non-contact based. Several anti-spoofing methods have been proposed in the literature for both contact and non-contact based biometric modalities often using video to study the temporal characteristics of a real vs. spoofed biometric signal. This paper presents a novel multi-feature evidence aggregation method for face spoofing detection. The proposed method fuses evidence from features encoding of both texture and motion (liveness) properties in the face and also the surrounding scene regions. The feature extraction algorithms are based on a configuration of local binary pattern and motion estimation using histogram of oriented optical flow. Furthermore, the multi-feature windowed videolet aggregation of these orthogonal features coupled with support vector machine-based classification provides robustness to different attacks. We demonstrate the efficacy of the proposed approach by evaluating on three standard public databases: CASIA-FASD, 3DMAD and MSU-MFSD with equal error rate of 3.14%, 0%, and 0%, respectively.**

I. Introduction

Biometric systems have different points of vulnerability such as sensor attacks, overriding feature extraction, tampering feature representation, corrupting matcher, tampering stored template, and overriding decision [18]. With such attacks, it is possible to circumvent, gain unauthorized access, and impersonate another individual. Among the different points of a biometric system design, capture phase vulnerabilities lead to the possibility of spoofing attacks. To mitigate the spoofing attempts, anti-spoofing techniques are developed that can be advantageous to (i) help increase the cost of obfuscating a biometric system, (ii) allow biometrics to become truly operator independent, and (iii) facilitate non-repudiation as the user is unable to deny his/her physical presence.

Face spoofing is a simple yet effective method to circumvent unattended face recognition systems. The 'low-tech' nature of these techniques makes face biometric systems particularly vulnerable to spoofing attacks. The problem of spoofing is also compounded with mobile devices enabled with face recognition. For instance, the *Face Unlock* feature, that uses face recognition to unlock a phone, is vulnerable to spoofing attacks [9], despite having a blinking based liveness detection. Additionally, sophisticated high quality 3D masks of persons have also become cheaper to obtain [7].

To prevent spoofing attacks, a face biometric system must be fortified with special mechanisms that ensure the integrity of the system. In this research, we present a single approach for efficiently detecting a plethora of possible 2D & 3D face spoofing techniques such as *print*, *replay*, *wrap*, and *mask* attacks that have been shown to be effective at breaching face biometric systems.

A. Literature Review

The face anti-spoofing problem is extensively studied in literature, particularly with the introduction of Print Attack dataset [1], Replay Attack dataset [5], CASIA-FASD spoofing dataset [21], 3DMAD database [7], and MSU mobile face spoofing database [20]. Each of these datasets offer different types of spoofing attacks with varying degrees of sophistication and quality, and have been a valuable asset to the scientific community in driving this area of research forward. Depending on the type of features used for information extraction and representation, face anti-spoofing techniques in literature can be classified into image texture analysis and temporal evidence based approaches.

Texture analysis approaches rely on the observation that video frames (individually) exhibit some unique image properties that help distinguish when compared to spoofed frames. Early approaches showed Local Binary Patterns (LBP) descriptors of different configurations are effective for print attack detection [1], [14]. Orthogonal to the LBP texture descriptors based approaches, quality assessment metrics such as specular reflection, blurring and color density are also explored for anti-spoofing [10], [20]. Another approach to replay spoofing utilizes detection of Moirè patterns, which are typically manifested in video recapturing, to detect replay video attacks [17].

In contrast, temporal evidence techniques encode spatial and temporal evidence across videos for cues such as signs of vitality [16], for spoofing detection. Dynamic texture features such as LBP-TOP [22] are studied in this regard. Top performing teams in the 2^{nd} ICB counter measure to 2D facial spoofing competition [6] combined motion and texture features to obtain interesting results. Bharadwaj *et al.*[2] present a combination approach leveraging texture and motion descriptors across the entire video frame. The approach also leverages Eulerian motion magnification to further enhance the subtle motions in the video. The approach is effective on PRINT and REPLAY attack databases, however, an equal error rate of 14.4% is reported on the CASIA database [3]. Tirunagari *et al.*[19] combine dynamic mode decomposition with LBP

978-1-5090-4847-2/16/$31.00 ©2016 IEEE

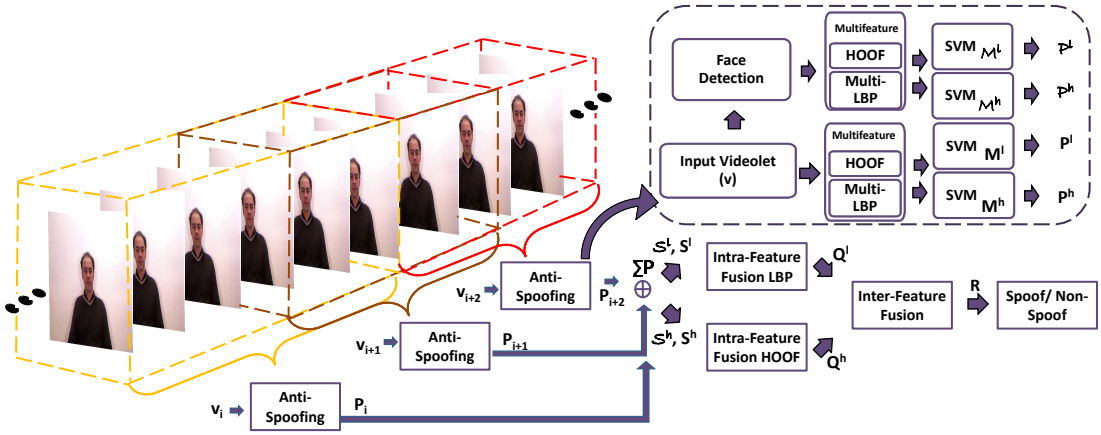


Fig. 1. An illustration of the windowed approach. The prediction scores obtained for each *videolet* are aggregated over the entire length of a video.

for spoofing detection. Recently, deep learning neural network architectures are also explored to encode liveness and texture for spoofing detection [8], [15].

B. Research Contributions

To enable deployment of unattended face recognition systems in access control applications, it is imperative that they are robust to spoofing attacks. This paper presents a unified approach to face spoofing detection based on the observation that different types of spoofing attacks have varying effects on the video frame that can be leveraged as evidence of spoofing. The key contributions of this paper can be summarized as follows:

- A multifeature encoding approach is proposed consisting of a multiscale configuration of LBP (referred as multi-LBP), that encodes the texture of videos temporally, and a motion estimation based encoding approach using optical flow, termed as Histogram of Oriented Optical Flow (HOOF) [4]. Support Vector Machines (SVM) are used for classification of an encoded video into *spoof* and *non-spoof*.
- In the proposed algorithmic pipeline, multifeature encoding is performed simultaneously over both full frame and detected face region. The additional information provides evidence of spoofing at the *scene* level of the video frames.
- Further, the evidence of spoofing obtained from the face and scene analysis is aggregated in a windowed approach for a small number of video frames, termed as *videolets*. Collection of evidence in videolet fashion enables short motion analysis which provides an effective anti-spoofing detection across various types of spoofing attacks and video duration.
- An evaluation on three publicly available spoofing databases, namely, the CASIA-FASD, MSU-MFSD, and, 3DMAD databases, using the official protocols, show state-of-the-art performance along with lower computation time.

II. Proposed Framework

In this research, we propose a temporal evidence based anti-spoofing algorithm that consists of a multifeature extraction method, followed by a combination approach to efficiently classify spoofed and non-spoofed videos. Further, we show that leveraging temporal evidence simultaneously from the face region and the *scene* of the video further enhances performance. Fig. 1 illustrates the overview of the algorithmic pipeline. Details of the proposed framework are discussed in the subsections below.

A. Feature Extraction

The proposed multifeature extraction consists of (1) feature extraction using multi-LBP and HOOF, (2) multifeature extraction on both face regions and full frames of the video, and (3) feature aggregation over videolets (video frames of short durations) followed by intra and inter feature evidence fusion.

1) Multi-LBP: In literature, LBP has been used to encode texture information in several applications including spoofing detection. LBP can be configured to provide a coarser or finer encoding depending on the intended application. Existing spoofing detection algorithms have proposed feature level concatenation of LBP features. We hypothesize that comparatively coarser LBP features may be sufficient for spoofing detection and propose to encode texture information at multiple scales via feature concatenation of three LBP configurations: $\text{LBP}^{u2}_{8,1}$, $\text{LBP}^{u2}_{8,2}$, and $\text{LBP}^{u2}_{16,2}$, collectively termed multi-LBP. Fig. 2 illustrates the steps involved in the proposed multi-LBP feature extraction algorithm. As opposed to Määttä *et al.*[13] that computes overlapping local histograms of $\text{LBP}^{u2}_{8,1}$, resulting in a feature vector of size 833; multi-LBP computes global histograms at three scales, thereby resulting in a descriptor of size 361 (i.e. 59+59+243).

2) Histogram of Oriented Optical Flows: Micro-movements in the consecutive frames of a face video are unique characteristic of liveness and challenging to imitate in spoofing. Therefore, encoding such variations in consecutive frames can provide effective features for spoof

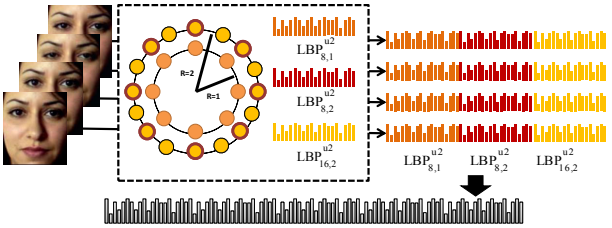


Fig. 2. Illustrating the proposed texture based spoofing detection approach.

detection. Optical flow is a dense motion estimation technique that computes the motion of each pixel.

$$\frac{\partial V}{\partial x}G_x + \frac{\partial V}{\partial y}G_y + \frac{\partial V}{\partial t} = 0 \tag{1}$$

$$\theta = \arctan\frac{G_y}{G_x}, \quad m = \sqrt{G_x^2 + G_y^2} \tag{2}$$

where (G_x) and (G_y) represent flow (gradient) in horizontal and vertical directions, respectively. The orientation based optical flow vector is computed by solving the optimization problem 1 using conjugate gradient method [12]. Raw optical flow per pixel may be too spatially constrained and encode redundant background or unwanted motion. Therefore, as illustrated in Fig. 3, the flow vectors are computed and pooled over local block regions weighted by the corresponding magnitude. Specifically, optical flow is computed between the frames at a fixed interval (k). From Eq. 2, the histogram of optical flow orientation angle (θ) weighted by the magnitude (m) is computed over local blocks and concatenated to form a single vector. Histogram of the magnitude weighted orientation bins are utilized, and the vector thus obtained is termed as HOOF [4]. The final feature vector $(hoof_k(V))$ for a video V with n frames $F_{1,\ldots,n}$ and a sampling interval of k is obtained by concatenating the HOOF vector for all the sampled frames (Eq. 3).

$$\begin{aligned} h^V_{p,q} &= HOOF(F_p, F_q) \\ hoof_k(V) &= [h^V_{1,1+k}\ h^V_{2+k,2+2k} \cdots\ h^V_{n-k,n}] \end{aligned} \tag{3}$$

In this research, $k = 2$ sampling interval is chosen empirically which results in a feature vector of size 81 per frame pair. Low interval ensures that small differences in motion between consecutive frames are also encoded. The feature vector is classified using SVM with RBF kernel.

B. Frame Level Feature Encoding

Thus far, this research presents a multifeature extraction method, namely, multi-LBP and HOOF descriptors, collectively termed *multifeature*. HOOF based motion analysis encodes motion of a face and is proficient as a *liveness* approach. On the other hand, multi-LBP encodes the temporal evidence in texture which manifests differently for a real video than a spoofed video. In addition to computing the multifeature vectors for the face region in the frame, the proposed approach also computes the multifeature vectors for the entire frame region, to encode the *scene* evidence that is available.

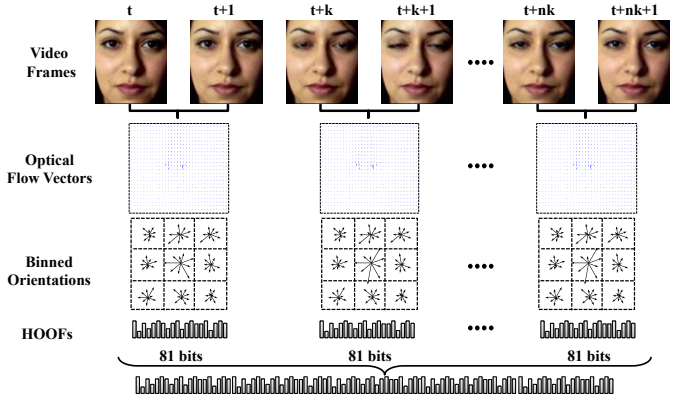


Fig. 3. An illustration of the proposed liveness based feature extraction approach. HOOF descriptors obtained between pairs of frames at a fixed interval are concatenated to create a single feature vector.

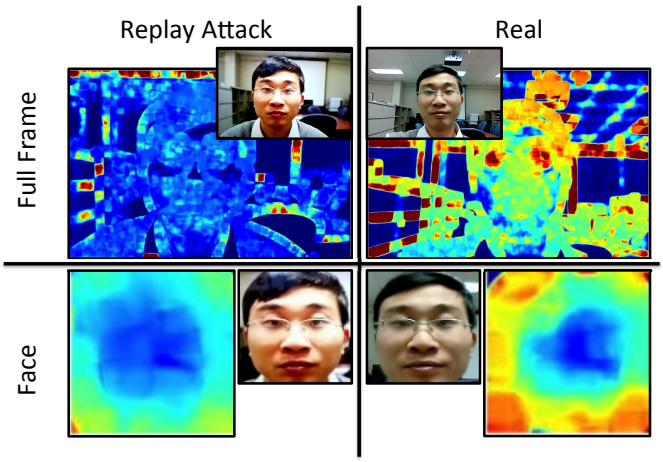


Fig. 4. An illustration on the information encoded by HOOF from a *videolet* of real and replay attack videos of MSU-MFSD database. The accumulation of absolute difference between the optical flow maps of consecutive frames shows a larger amount of activity potentially captured by the descriptor from the real video compared to an attack video. Inset, the corresponding mean frame also shows texture artifacts such as reflections and low contrast.

The process of video spoofing injects minute artifacts into a spoofed video that are manifested throughout the video frame. Fig. 4 is drawn to illustrate such artifacts being manifested in the encoded features. The heat maps show accumulated absolute difference between optical flow maps of a single videolet from a video in the MSU-MFSD dataset (inset corresponding mean frame for the same videolet). By removing the bounding box for face region, the multifeature vectors are able to encode a sufficient amount of such artifacts to favourably affect the spoof vs non-spoof decision boundaries. These artifacts include Moirè patterns [17], fading effects, blurring, video encoding artifacts and aliasing effects (from varying sampling rate). Additionally, artifacts in some datasets can include the spoofing medium, such as paper, tablet screen bezel, etc,

that are also encoded in the full frame. The observation also explains the performance of anti-spoofing techniques in literature that leverage image properties via quality assessment metrics of frames.

Algorithm 1 Spoofing Detection in a video

input: A video V=$\{F_1, F_2, \dots F_N\}$, corresponding video consisting of cropped face regions $\mathcal{V}$=$\{\mathcal{F}_1, \mathcal{F}_2, \dots \mathcal{F}_N\}$ trained SVM models: $\mathcal{M}^h$, M^h, $\mathcal{M}^l$ and M^l. (M and $\mathcal{M}$ corresponds to models for frames and faces, respectively.), videolet size w, interval k, and decision threshold T.

procedure:

▷ Compute LBP features

$\mathcal{X}^l = lbp(\mathcal{V})$
$X^l = lbp(V)$

▷ Compute HOOF features

$\mathcal{X}^h = hoof_k(\mathcal{V})$ (as in Eq. 3)
$X^h = hoof_k(V)$ (as in Eq. 3)
$\eta = (\frac{2N}{w} - 1)$ ▷ number of videolets
$s = 1$
iterate: $i = 1$ **to** η **do**

▷ Obtain videolet features

ν^l= $\{\mathcal{F}_j^l | s \leq j \leq (s+w)\}$
v^l= $\{F_j^l | s \leq j \leq (s+w)\}$
ν^h= $\{\mathcal{F}_j^h | s \leq j \leq (s+w)\}$
v^h= $\{F_j^h | s \leq j \leq (s+w)\}$
$s = s + (\frac{w}{2})$

▷ Score Computation

$\mathcal{P}^l$= $\mathcal{M}^l(\nu^l)$, $\mathcal{P}^h$= $\mathcal{M}^h(\nu^h)$
P^l= $M^l(v^l)$, P^h= $M^h(v^h)$
end iterate.
$\mathcal{S}^l$= $\frac{1}{\eta}\sum_{j=1}^{\eta}(\mathcal{P}_j^l)$, $\mathcal{S}^h$= $\frac{1}{\eta}\sum_{j=1}^{\eta}(\mathcal{P}_j^h)$
S^l= $\frac{1}{\eta}\sum_{j=1}^{\eta}(P_j^l)$, S^h= $\frac{1}{\eta}\sum_{j=1}^{\eta}(P_h^j)$

Q^l = $w_1^l\mathcal{S}^l$ + $w_2^l S^l$ ▷ Weighted Intra-Feature Fusion
Q^h = $w_1^h\mathcal{S}^h$ + $w_2^h S^h$ ▷ Weighted Intra-Feature Fusion

R = $w_a Q^l$ + $w_b Q^h$ ▷ Weighted Inter-Feature Fusion
Output: report **if** $(R > T)$ "spoof" **else** "non-spoof"

C. Videolet Score Aggregation and Evidence Fusion

As discussed earlier, a unified spoofing approach must be robust to various types of face spoofing attacks such as print, replay, wrap, and 3D mask. Therefore, we propose to combine both feature extraction algorithms across full frame and face region for improved performance. Fusion is performed by combining the video level aggregation of the predicted scores obtained from SVM classification of both multi-LBP (lbp) and HOOF ($hoof$) features separately.

In several existing temporal evidence based approaches, feature extraction is performed on the entire length of the available video. The extracted features are then concatenated to create a single descriptor of fixed length. However, with videos of varying length, only the minimum number of frames can be considered to maintain fixed length of the feature vector. As illustrated in Fig. 1, we propose a windowed approach to effectively utilize all the information present in a video (V), without constraining the size of the input video. In this approach, both HOOF and multi-LBP features are computed on a video divided into overlapping windows of size w, similar to [11], with a step size of half the window size. The frames corresponding to a single window are termed as *videolet*.

Each video is divided into $\eta = (\frac{2N}{w} - 1)$ videolets followed by extraction of multi-LBP and HOOF features from each of these videolets for both full frame (ν^l, ν^h) and face region (v^l, v^h) separately. These scores are separately averaged across all videolets ($\mathcal{S}_l, \mathcal{S}_h, S_l, S_h$) and further combined at two level, i) Intra-Feature Fusion and ii) Inter-Feature Fusion.

- **Intra-Feature Fusion**: A weighted combination of full frame scores and face region scores is computed for HOOF (Q^h) and Multi-LBP (Q^l) separately.
- **Inter-Feature Fusion**: The evidence scores obtained separately for HOOF and LBP features for a given video are further combined with weighted combination to obtain the prediction score (R).

A threshold T is applied on prediction score R for classification of video as either spoofed or real. The proposed videolet aggregation approach combines evidence from multi-LBP texture analysis with HOOF motion analysis over short w sized intervals, is summarized in Algorithm 1.

III. Dataset and Protocol

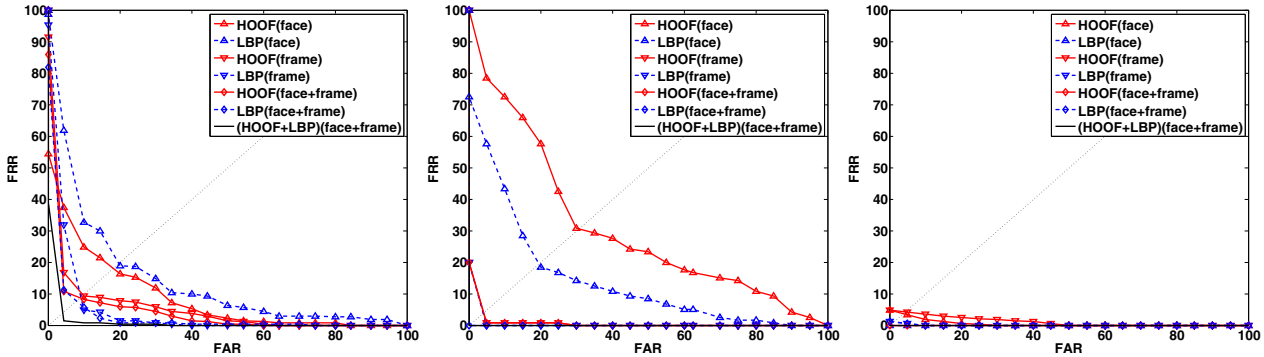
A spoofing detection technique must be robust to different types of attacks. Therefore, the experiments are performed on three publicly available databases, namely (1) CASIA-FASD dataset [21], (2) MSU mobile face spoofing database [20], and (3) 3D-MAD [7]. An overview of the structure of the databases and pre-defined (official) protocols are provided below.

- The CASIA-FASD dataset (CASIA) [21] consists of 600 videos corresponding to 50 subjects, separated as 240 videos in training and 360 videos in testing. In addition to print and replay attacks using photos and replayed videos from tablets, wrapped photos are used to simulate the cylindrical nature of the face. Further, print attack photos are manually cut around the eyes to deter eye-blinking based techniques. The challenging nature of the dataset is furthered by variations in resolution, quality, and video length (ranging from 1 to 19 seconds). The equal error rates (EER) are reported as per the pre-defined protocol of the dataset [21].
- MSU Mobile Face Spoofing Database (MSU) [20] consists of 280 videos corresponding to 35 subjects, captured with two devices, a built-in webcam and an Android phone camera. The database consists of spoofing by print attack using a HD color printer and replay attack using video captured with a high resolution SLR camera and an iPhone 5S back-facing camera. The wide variation in the possible resolution of the videos poses a practical challenge of the database. In this research, we use the standard (predefined) experimental protocol of the paper.
- 3D Mask Attack Database (3DMAD) [7] is a spoofing database recorded for 17 subjects using Microsoft Kinect sensors. It consists of three sessions, each containing 5 videos of 17 subjects (in all 255 videos). Two of these sessions consist of real videos of the subject and one session contains spoofing attack performed using 3D masks of the subject applied to some other user. For the

TABLE I
EER(%) FOR DIFFERENT TECHNIQUES ON MULTIPLE DATABASES.

Approach	Datasets											
	CASIA				MSU				3DMAD			
	HOOF		LBP		HOOF		LBP		HOOF		LBP	
	Face	Frame	Face	Frame	Face	Frame	Face	Frame	Face	Frame	Face	Frame
Individual Features	16.80	9.81	19.26	7.78	30.41	2.50	20.00	2.50	2.35	2.35	0.00	0.00
Face and Frame Fusion	7.96		6.67		2.50		0.00		0.00		0.00	
Feature Fusion	**3.14**				**0.00**				**0.00***			
Feng *et al.*[8]	5.83				–				0.00			
Wen *et al.*[20]	12.9**				8.58				–			
Patel *et al.*[17]	0.00^{+}				–				–			
Menotti *et al.*[15]	–				–				0.00			

* Averaged over 17 cross validation folds. ** Results reported on high resolution subset of the dataset. $^{+}$ reported 0% HTER.



(a) CASIA: 3.14% EER (b) MSU: 0.00% EER (c) 3DMAD: 0.00% EER

Fig. 5. ROC plots of the proposed approach on the test sets of CASIA, MSU, and 3DMAD datasets (averaged over 17 folds) at various stages of the pipeline. The curves with 0% EER overlap with axes.

purpose of our research we have used only the colour 2D component of the database.

The EER results on 3DMAD database are presented in a leave-one-out-cross-validation (LOOCV) fashion. The experiment is performed on 17 folds, one for each subject. For each fold, all videos of the respective subject from all three sessions in the test set are utilized and the mean EER across all 17 folds is reported.

To obtain face region, frames are cropped using publicly available eye coordinates and the features are extracted from histogram normalized image sequences. For computing texture based features, all the frames are converted to gray scale. To be consistent with existing literature, the results of spoofing detection are reported in terms of the respective protocols and EER (%) on the test set[1]. The parameters of SVM are determined using a grid search where the objective is to minimize equal error rate on the training set. In this research, w is set to 24 frames or 1 second of video, determined experimentally to be optimal.

IV. Experimental Analysis

Using the protocols described in the previous section, three experiments are performed. Table I and Fig. 5 illustrate the results for HOOF, multi-LBP, and the proposed spoofing detection approach.

[1]HTER is another metric used in spoofing literature but not common for these three datasets.

- **Performance of the Proposed Approach**: The proposed fusion approach (using HOOF and multi-LBP with face and scene aggregated over videolets) provides 0% EER with uncontrolled illumination and background on both MSU and 3DMAD datasets. The approach also provides high performance with 3.14% EER on CASIA dataset. A weighted sum rule fusion of these classifiers has resulted in a robust and accurate anti-spoofing measure. The observation is consistent with the ROC plots shown in Fig. 5(a), (b), (c).
- **Face Region vs. Full Frame:** It is observed that on MSU and CASIA datasets, utilizing the full frames yield better performance than only the face regions. As discussed earlier, this affect can be explained due to *scene* characteristics, which are more pronounced in full frames than face regions. On 3DMAD dataset, both, full frames and face regions, yield same performance. 3DMAD involves the mask attack videos only; it does not involve replay attack. Such mask attack can be considered as advanced print attack. Thus, there are no spoofing medium (such as iPad screen) specific characteristics to be encoded. Therefore, we observe that in the 3DMAD *scene* characteristics are very similar for real and attack videos.
- On MSU dataset, HOOF obtains tremendous improvement in EER (from 30.41 to 2.50%) when utilizing full frames as compared to only face regions. For CASIA, the improvement is relatively smaller (from 16.80 to 9.81%). MSU dataset contains a higher fraction of replay

attack videos compared to CASIA. Thus, it is observed that HOOF is better suited for replay attack than print attack; probably, due to an explicit encoding of temporal information.

- With LBP features, the full frame yields better performance than the face regions in CASIA (7.78% compared to 19.26% EER), and MSU (2.50% compared to 20.00% EER) datasets, whereas in 3DMAD a perfect classification is obtained with both face regions and full frames. This observation provides evidence that the LBP features are efficiently encoding the *scene* characteristics to differentiate between spoof and real videos.
- **Effectiveness of Fusion:** The proposed approach benefits from the score level fusion by combining evidence from various types of feature encoding. In order to quantitatively analyze the effect of the fusion, Spearman Rank correlation is computed between scores at both stages of fusion.

  The scores utilized in Intra Feature Fusion yield following correlation values for both the features: HOOF (CASIA: 0.25, MSU: 0.26, 3DMAD: 0.70) and Multi-LBP (CASIA: 0.41, MSU: 0.46, 3DMAD: 0.66). Similarly, at the Inter Feature Fusion stage, the correlation of 0.51, 0.62, and 0.66 is observed for CASIA, MSU, and 3DMAD datasets, respectively. Overall, a low correlation is observed for CASIA and MSU datasets, providing the quantitative indication of effectiveness of fusion.
- This research presents a unified approach to face spoofing detection that can provide robust results on multiple datasets comprising of various types of 2D and 3D spoofing methods. As shown in Table I, this research is the first instance of a single approach showcasing best results on all three datasets.
- **Computational Efficiency**: The OpenCV implementation runs on a machine with Intel Quad Core CPU Q8300 at 2.5GHz and 4GB RAM. For a video with 375 frames (i.e., 15 seconds in length), 14 videolets are created, each of 24 frames. The proposed approach involves HOOF extraction for full frames (15.8s), and for face region (0.54s). Similarly, for multi-LBP extraction of full frames (20.0s), and for face region (0.65s) requires a total of 38.6s to process a single videolet serially. The computational time varies slightly on different databases (difference under 0.1s). We believe that a parallel implementation can further reduce the processing time.

V. Conclusion

It is imperative that face recognition systems be equipped with a pre-processing stage that evaluates an input video for possible spoofing. This research presents a face spoofing detection that leverages temporal evidence aggregation over face region and scene of a video. The proposed approach achieves state-of-the-art accuracies on different publicly available databases. Development of new anti-spoofing approaches is a continuous process, to stay ahead of malicious intent towards robust universal anti-spoofing techniques. We are currently exploring the effectiveness of the proposed approach in cross dataset experiments, presented in recent literature [20].

VI. Acknowledgement

The authors acknowledge the Open Science Collaboration initiative between IBM India Research Labs and IIIT-Delhi, India for facilitating this research.

References

[1] A. Anjos and S. Marcel, "Counter-measures to photo attacks in face recognition: a public database and a baseline," in *IEEE/IAPR IJCB*, 2011, pp. 1–7.

[2] S. Bharadwaj, T. I. Dhamecha, M. Vatsa, and R. Singh, "Computationally efficient face spoofing detection with motion magnification," in *IEEE CVPRW*, 2013, pp. 1–7.

[3] ——, "Face anti-spoofing via motion magnification and multifeature videolet aggregation," in *IIITD-TR-2014-002*, 2014, pp. 1–14.

[4] R. Chaudhry, A. Ravichandran, G. Hager, and R. Vidal, "Histograms of oriented optical flow and Binet-Cauchy kernels on nonlinear dynamical systems for the recognition of human actions," in *IEEE CVPR*, 2009, pp. 1932–1939.

[5] I. Chingovska, A. Anjos, and S. Marcel, "On the effectiveness of local binary patterns in face anti-spoofing," in *BIOSIG*, 2012, pp. 1–7.

[6] I. Chingovska *et al.*, "The 2nd competition on counter measures to 2D face spoofing attacks," in *IAPR ICB*, 2013, pp. 1–7.

[7] N. Erdogmus and S. Marcel, "Spoofing in 2D face recognition with 3D masks and anti-spoofing with kinect," in *IEEE BTAS*, 2013.

[8] 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," *JVCIR*, vol. 38, pp. 451 – 460, 2016.

[9] R. D. Findling and R. Mayrhofer, "Towards face unlock: on the difficulty of reliably detecting faces on mobile phones," in *MoMM*, 2012, pp. 275–280.

[10] J. Galbally and S. Marcel, "Face anti-spoofing based on general image quality assessment," in *IAPR ICPR*, 2014, pp. 1173–1178.

[11] J. Komulainen, A. Anjos, S. Marcel, A. Hadid, and M. Pietikäinen, "Complementary countermeasures for detecting scenic face spoofing attacks," in *IAPR ICB*, 2013, pp. 1–7.

[12] C. Liu, "Beyond pixels: Exploring new representations and applications for motion analysis," Ph.D. dissertation, Massachusetts Institute of Technology, 2009.

[13] J. Määttä, A. Hadid, and M. Pietikainen, "Face spoofing detection from single images using micro-texture analysis," in *IEEE/IAPR IJCB*, 2011, pp. 1–7.

[14] ——, "Face spoofing detection from single images using texture and local shape analysis," *IET Biometrics*, vol. 1, no. 1, pp. 3–10, 2012.

[15] D. Menotti, G. Chiachia, A. Pinto, W. R. Schwartz, H. Pedrini, A. X. Falco, and A. Rocha, "Deep representations for iris, face, and fingerprint spoofing detection," *IEEE TIFS*, vol. 10, no. 4, pp. 864–879, April 2015.

[16] G. Pan, L. Sun, Z. Wu, and S. Lao, "Eyeblink-based anti-spoofing in face recognition from a generic webcamera," in *IEEE ICCV*, 2007, pp. 1–8.

[17] K. Patel, H. Han, A. K. Jain, and G. Ott, "Live face video vs. spoof face video: Use of moire; patterns to detect replay video attacks," in *IAPR ICB*, May 2015, pp. 98–105.

[18] 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.

[19] S. Tirunagari, N. Poh, D. Windridge, A. Iorliam, N. Suki, and A. T. S. Ho, "Detection of face spoofing using visual dynamics," *IEEE TIFS*, vol. 10, no. 4, pp. 762–777, April 2015.

[20] D. Wen, H. Han, and A. Jain, "Face spoof detection with image distortion analysis," *IEEE TIFS*, vol. 10, no. 4, pp. 746–761, 2015.

[21] Z. Zhang, J. Yan, S. Liu, Z. Lei, D. Yi, and S. Z. Li, "A face antispoofing database with diverse attacks," in *IAPR ICB*, 2012, pp. 26–31.

[22] G. Zhao and M. Pietikainen, "Dynamic texture recognition using local binary patterns with an application to facial expressions," *IEEE TPAMI*, vol. 29, no. 6, pp. 915–928, 2007.