Face Anti-spoofing via Motion Magnification and Multifeature Videolet Aggregation

Samarth Bharadwaj, *Student Member, IEEE,* Tejas I. Dhamecha, *Student Member, IEEE,* Mayank Vatsa, *Member, IEEE,* and Richa Singh, *Senior Member, IEEE*

***Abstract*—For robust face biometrics, a reliable anti-spoofing approach has become an essential pre-requisite against attacks. While spoofing attacks are possible with any biometric modality, face spoofing attacks are relatively easy which makes facial biometrics especially vulnerable. This paper presents a new framework for face spoofing detection in videos using motion magnification and multifeature evidence aggregation in a windowed fashion. Micro- and macro- facial expressions commonly exhibited by subjects are first magnified using Eulerian motion magnification. Next, two feature extraction algorithms, a configuration of local binary pattern and motion estimation using histogram of oriented optical flow, are used to encode texture and motion (liveness) properties respectively. Multifeature windowed videolet aggregation of these two orthogonal features, coupled with support vector machine classification provides robustness to different attacks. The proposed approach is evaluated and compared with existing algorithms on publicly available Print Attack, Replay Attack, and CASIA-FASD databases. The proposed algorithm yields state-of-the-art performance and robust generalizability with low computational complexity.**

***Index Terms*—Face recognition, anti-spoofing, obfuscation, motion magnification**

I. Introduction

Biometrics based authentication is now being utilized in several applications including national identification schemes such as India's UID project. As the popularity of biometric systems grow, there is an increased threat of malicious attacks to circumvent the system. Corresponding to various stages of a biometric system's pipeline, Ratha *et al.* [1] have identified different points of vulnerability such as sensor attacks, overriding feature extraction, tampering feature representation, corrupting matcher, tampering stored template, and overriding decision. With such attacks, it is possible to circumvent a biometric system, gain unauthorized access, and impersonate another individual. Among different vulnerabilities of a biometric system design, the weakest link is the capture phase [2]. For example, surgically altered fingerprints [3] or face [4], fake fingerprints (using silicone, geltin, latex, and wood glue) [5], fake iris texture [6], cosmetic contact lenses [7], [8], disguise [9] or various spoofing approaches enable attackers to gain unauthorized access. Some such examples are shown in Fig. 1. To prevent these attacks, a biometric system must be fortified with special mechanisms that ensure the integrity of the system. In this research, we focus on spoofing techniques pertaining to 2D face biometrics.

IIIT-Delhi, India, e-mail: {samarthb, tejasd, mayank, rsingh}@iiitd.ac.in

[1]Image obtained from http://www.wikihow.com/

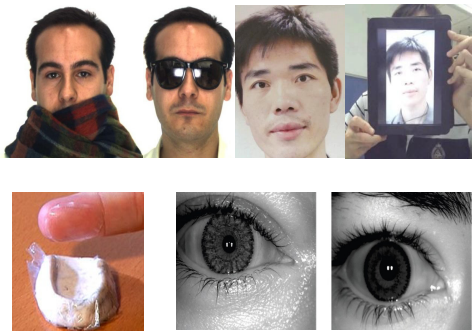


Fig. 1. An illustration of some known vulnerabilities for face (masks), fingerprint (gummy prints[1]) and iris (color/textured contact lens) biometrics at capture phase.

(a) (b) (c)

Fig. 2. An illustration of spoofing attacks on 2D face recognition systems, (a) real access, (b) print attack, and (c) replay attack.

Face spoofing is a simple and an efficient method to circumvent face recognition systems that are often unattended. The 'low-tech' nature of face spoofing techniques makes face biometric systems vulnerable. As illustrated in Fig. 2, literature on 2D spoofing detection generally discusses two types of spoofing attacks, namely *print* and *replay*. In print attack, printed photographs of a subject are used to spoof 2D face recognition systems, whereas a replay attack, a video of a person is presented to evade liveness detection. A replay attack video could also be of a digital photograph or a video replayed on a screen. To mitigate the spoofing attempts, anti-spoofing techniques are developed that can have several advantages; they can (i) help to increase the cost of obfuscating a biometric system, (ii) allow face recognition and biometrics in general to become truly operator independent, and (iii) facilitate non-repudiation as the user is unable to deny his/her physical presence [10].

A. Literature Review

Depending on the type of features used for information extraction and representation, face anti-spoofing techniques in

literature can be classified into liveness and texture analysis based approaches. Liveness techniques primarily encode signs of vitality, such as eye blinking and mouth movements. Such approaches are similar to anti-spoofing approaches in fingerprint that use body temperature and blood flow detection. Texture based approaches rely on the observation that face frames of a real person exhibit some unique spatiotemporal properties when compared to spoofed frames. Recent techniques also present combination approaches that combine evidence from multiple sources for robust anti-spoofing. The literature review is divided into three parts: datasets available for research, different types of approaches proposed to detect spoofing, and competitions organized by researchers to assess the state-of-art for spoofing.

Datasets: Anjos *et al.* [11] introduced the Print Attack dataset consisting of real and spoof videos of subjects with print attacks. A more challenging Replay Attack dataset consisting of both print and video replay spoofing captured using a display device kept in front of the camera, was presented in [12]. The CASIA-FASD spoofing dataset [13] consists of more challenging variants of both print attack and replay attack at different resolutions. The introduction of these three publicly available datasets (samples frames illustrated in Fig. 8) are the basis of experimental evaluation in recent literature.

Liveness based approaches: Early approaches to face anti-spoofing relied on liveness estimation usually by modeling eye-blinking. Pan *et al.* [14] modeled eye blinking by capturing blink behavior in a conditional random field framework. The different stages of blinking are learned from an annotated dataset of 20 subjects. Kollreider *et al.* [15] presented a combination approach to liveness detection by combining eye blinking and mouth movement with 3D properties of the face. Facial motion such as eye blinking and head movement were used to determine liveness by some participants of the IJCB facial spoofing competition [16]. To recognize subtle facial features in long videos, such as in case of person interrogation, Shreve *et al.* [17] presented a *temporal stain* metric computed from optical flow patterns obtained from facial regions. The problem of spoofing is particularly 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 [18], despite having a blinking based liveness detection. Recently, Anjos *et al.* [19] presented a motion correlation approach using optical flow. Wang *et al.* [20] re-constructed a 3D model from a single camera for liveness detection.

Texture based approaches: Texture analysis of face video has also been shown to provide evidence of spoofing. Määttä *et al.* [21] showed that concatenation of three Local Binary Patterns (LBP) descriptors of different configurations is more efficient than local phase quantization as well as Gabor wavelet based descriptor for print attack spoofing detection on the NUAA dataset [22] of 15 subjects. Määttä *et al.* [23] also proposed a score level fusion approach using LBP, histogram of oriented gradients, and Gabor wavelets computed from the local blocks of a face image. For each descriptor, the histogram computed from all the blocks were concatenated, thus resulting in three feature vectors. Kernel approximations of the three feature vectors were computed and a linear Support Vector Machine (SVM) was used for classification. Further, the match scores of all three SVMs were fused to provide the final result. The authors reported 0% Half Total Error Rate (HTER) on the Print Attack dataset. It was observed that Support Vector Machine outperformed both Linear Discriminant Analysis (LDA) and χ^2 distance based classification. The performance of various approaches were evaluated on the Print Attack dataset [11] and it was observed that texture based approaches resulted in 0% HTER. Power spectrum and LBP features were used in a fusion approach by Gahyun *et al.* [24] on a print attack dataset collected using a camera of an automated teller machine.

Pereira *et al.* [25] explored the utility of LBP from three orthogonal planes (termed as LBP-TOP) [26] for spoofing detection on the Replay Attack dataset. LBP-TOP explicitly utilizes the temporal information by computing LBP histograms in XT and YT planes along with spatial information in XY plane. In their experiments, multi-resolution LBP-TOP with SVM classifier achieved the best HTER of 7.6% on the Replay Attack dataset; however, it is computationally expensive. Several entries to the 2^{nd} ICB counter measure to 2D facial spoofing competition [27] presented variants of LBP texture analysis based approaches. The CASIA-FASD spoofing dataset [13] consists of more challenging variants of print attack as well as replay attack at different resolutions. A baseline of 17% equal error rate (EER) is reported using difference of Gaussian approach. However, the approach selects a random subset of frames from a video and does not utilize the temporal information of the video. The performance of LBP-TOP on the CASIA-FASD spoofing dataset yielded an EER of 21.59% [28]. Yang *et al.* [29] used component analysis for liveness detection using Fisher criterion analysis for pooling evidence from informative regions of the face. Recently, Pereira *et al.* [30] further analyzed the performance of LBP-TOP approach on the Replay Attack dataset and CASIA-FASD spoofing dataset. The analysis shows that the approach encodes temporal information that aids in spoofing detection.

Combination approaches: The use of a combination of experts to determine spoofing has been acknowledged in literature. Schwartz *et al.* [31] combined several low level texture features to form a high dimensional vector (feature length over a million) and classified it using partial least squares approach. An EER of 1.67% is reported on the Print Attack dataset. Komulainen *et al.* [32] suggested fusion of computationally inexpensive linear classifiers for robust anti-spoofing. Using motion correlation analysis and LBP, HTER of 5.1% is reported on the Replay Attack dataset. The top performing teams in the 2^{nd} ICB counter measure to 2D facial spoofing competition [27] combined motion and texture features. Both the approaches used variants of LBP and background to foreground motion estimation from an input video and obtained 0% HTER.

Competitions: To promote the research in face spoofing, researchers have organized several competitions in different

conferences. Chakka *et al.* [16] evaluated the performance of six different spoofing detection algorithms as part of the IJCB counter-measures to 2D facial spoofing competition. Chingovska *et al.* [27] presented the results from the 2^{nd} ICB counter measure to 2D facial spoofing competition. Both the competitions evaluated various liveness based, texture based, and combination approaches. However, texture based approaches have been found to provide better performance for the Print and Replay datasets.

B. Research Contributions

Wide availability of portable display devices with high resolution has brought spoofing attacks into the purview of face biometrics. To enable deployment of unattended face recognition systems in access control situations, it is imperative that they must be robust to spoofing attacks. It is our assertion that face recognition systems must be equipped with a preprocessing stage that evaluates an input video of the subject for possible spoofing. This paper presents a new framework for face spoofing detection. The proposed framework is based on the observation that different types of spoofing attacks have different effects on the face. For instance, photo attack will not have facial movements across the video whereas replay attack will have the motions but the texture may vary due to the replay attack. Therefore, the algorithm should be able to encode both the information to be resilient to multiple types of attacks. The key contributions of this paper can be summarized as follows:

- A motion magnification based preprocessing algorithm to enhance facial motion exhibited by a person.
- A multiscale configuration of LBP (referred as multi-LBP) is proposed that encodes the texture of videos and SVM is used for classifying into *spoof* and *non-spoof*.
- A novel spoofing detection algorithm is proposed based on motion estimation using optical flow encoded with a Histogram of Oriented Optical Flow (HOOF) [33].
- The proposed framework is a combination approach that utilizes both texture and motion estimation along with preprocessing using motion-magnification. HOOF based motion estimation approach provides evidence of liveness, whereas, multi-LBP based texture analysis of motion magnified videos distinguishes between texture of spoofed faces and real faces. The combination of these diverse approaches, aggregated in small number of video frames, termed as *videolets*, provides an effective anti-spoofing approach.

An evaluation on three spoofing databases, namely, the Print Attack, Replay Attack, and the CASIA-FASD databases, using the official protocols, show state-of-the-art performance along with lower computation time. A cross-database experiment is also performed to analyze the generalizability of the proposed algorithm [28].

II. Proposed Framework

Fig. 3 illustrates the steps involved in the proposed framework for spoofing detection. It consists of three main steps: (1) preprocessing using motion magnification to enhance the *liveness* nature of a face video, (2) feature extraction using multi-LBP and HOOF, and (3) evidence fusion aggregated over videolets (video frames of short durations). Details of the proposed framework are discussed in the subsections below.

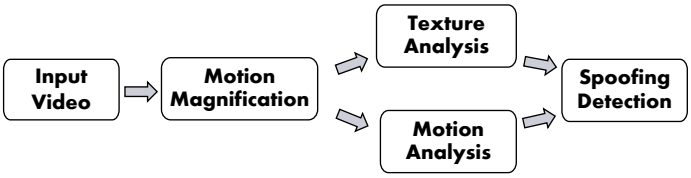


Fig. 3. The steps involved in the proposed framework for detecting face spoofing. Motion magnification of input video may accentuate facial expressions thereby aiding spoofing detection techniques. This research combines the evidence of both motion and texture to perform spoofing detection.

A. Motion Magnification

Faces exhibit micro-movements primarily near the lip and eye regions which are visible only on close inspection of the video. It is our assertion that enhancing these subtle facial movements such as blinking, saccadic and conjugate eye motion can provide evidence of liveness. Therefore, we propose a motion magnification based preprocessing algorithm that operates on a video to enhance these subtle motions exhibited by a face.

Motion magnification techniques are primarily of two types: Lagrangian and Eulerian approaches. Lagrangian approaches are based on explicitly tracking a pixel's trajectory over time. These are computationally expensive and difficult to compute around occlusion boundaries. On the other hand, an Eulerian approach to motion magnification directly amplifies temporal intensity changes at a given position without the need for explicit estimation [34]. Therefore, the proposed preprocessing algorithm utilizes Eulerian approach to motion magnification.

Eulerian motion magnification combines appropriate temporal and spatial filtering to localize and magnify the desired motion. Consider a video signal V, such that $V(x, y, t) = f(x + \delta_x(t), y + \delta_y(t))$, where $V(x, y, 0) = f(x, y)$ and $\delta_x(t), \delta_y(t)$ are the displacement functions in x and y directions respectively. The goal of video magnification, for an amplification factor α, can be expressed as

$$\hat{V}(x, y, t) = f\left(x + (1 + \alpha)\delta_x(t), y + (1 + \alpha)\delta_y(t)\right) \quad (1)$$

Under the first order Taylor series expansion about x and y directions, the video V can be represented as,

$$V(x, y, t) \approx f(x, y) + \delta_x(t)\frac{\partial f}{\partial x} + \delta_y(t)\frac{\partial f}{\partial y} \quad (2)$$

A temporal bandpass filter $B(x, y, t)$ is applied on the input video V such that all the components except $f(x, y)$ are filtered. This filter can be expressed as,

$$B(x, y, t) = \delta_x(t)\frac{\partial f}{\partial x} + \delta_y(t)\frac{\partial f}{\partial y} \quad (3)$$

Assuming that motion $\delta_x(t)$ and $\delta_y(t)$ are captured within the band, the motion magnified video $\hat{V}$ can be constructed as,

$$\hat{V}(x, y, t) = V(x, y, t) + \alpha B(x, y, t) \quad (4)$$

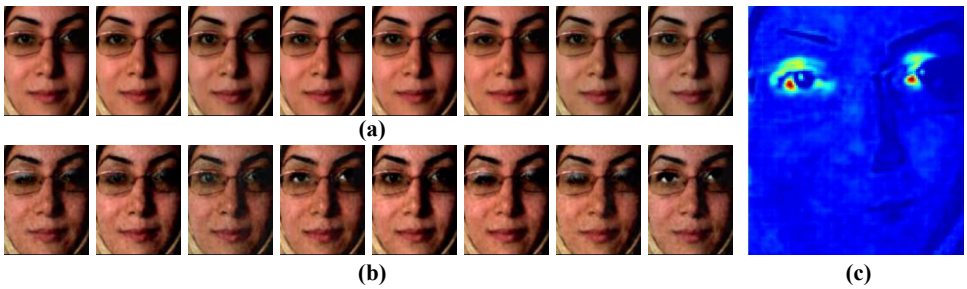


Fig. 4. Sample frames at regular intervals of a subject from the Replay Attack dataset. Frames in (a) are original while (b) contains corresponding frames from the corresponding motion magnified video. (c) illustrates the heat map of the mean of absolute differences of corresponding frames. The activity in eye regions represent the blinking motion of the eyes. More videos are available at http://research.iiitd.edu.in/groups/iab/facespoofing.html for better visualization.

Combining Eqs. 2, 3, and 4 we have,

$$\hat{V}(x,y,t) = f(x,y) + (1+\alpha)\left[\delta_x(t)\frac{\partial f}{\partial x} + \delta_y(t)\frac{\partial f}{\partial y}\right] \quad (5)$$

Assuming that Taylor series expansion holds for a magnified video, Eq. 5 is approximated to the desired form of Eq. 1. To avoid undesirable artifacts during magnification, the magnification factor α is suitably attenuated with respect to a spatial cut-off frequency. This reduces α for bands of higher frequencies and minimizes the artifacts in the resultant video. Laplacian decomposition is applied to the input video spatially and then magnification process is performed on each level with varying α. It must be noted that the effect of magnification is highly dependent on the filter and the magnification factor α used. In this research, an optimal value of α is chosen by visual inspection of processed videos from the training set.

Fig. 4 demonstrates the results of the proposed preprocessing algorithm for motion magnification in face videos. Fig. 4(a) shows the frames taken at equal intervals from an original video where the micro-movements are not visible but after applying the proposed preprocessing algorithm, the movement is clearly visible in the corresponding motion magnified frames shown in Fig. 4(b). The heat map pertaining to the mean absolute differences of all the corresponding frames of the sample video is shown in Fig. 4(c). From the heat map it is evident that the magnification approach enhances the blinking motion of eyes. Further, some regions in cheek areas show that the approach is also able to magnify micro-facial movements.

B. Feature Extraction

Motion magnified video of a subject can be utilized for spoofing detection using either texture or motion based features. In this research, we propose a texture and a motion based feature extraction algorithms followed by their combination to efficiently classify spoofed and non-spoofed videos.

1) Multi-LBP: In literature, LBP has been used to encode texture information in several applications. Depending on the application, LBP can be configured to provide a coarser or finer encoding. Existing spoofing detection algorithms have proposed feature level concatenation of global LBP features. However, with motion magnification, comparatively coarser features may be sufficient for spoofing detection. Based on this hypothesis, we 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. 5 illustrates the steps involved in the proposed multi-LBP feature extraction algorithm.

$\text{LBP}^{u2}_{P,R}$ represents uniform local binary pattern computed at P sampling points on a circle of radius R. A uniform local binary pattern at (x,y) with sampling points $(x_p,y_p), p = 0,1,2,\ldots,P-1$, is computed as,

$$LBP^{u2}_{P,R} = \begin{cases} \Sigma_{p=0}^{P-1} s(g_p - g_c)2^p & \text{if } U((x,y),P,R) \leq 2 \\ P(P-1)+3 & \text{otherwise} \end{cases} \quad (6)$$

Here g_p and g_c are the pixel intensities at positions (x_p,y_p) and (x,y) respectively, and functions s and U are defined as

$$s(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

$$U((x,y),P,R) = |s(g_P - g_c) - s(g_{P-1} - g_c)| + \Sigma_{p=1}^{P-1}|s(g_p - g_c) - s(g_{p-1} - g_c)| \quad (8)$$

U computes the number of bitwise transitions for circular bit patterns. It encodes the definition that a local binary pattern is called uniform if the binary pattern contains at most two bitwise transitions from 0 to 1 or vice versa. The resultant binary pattern at each pixel can take one of the distinct $(P(P-1)+3)$ values which translates to a histogram with the same number of bins. As illustrated in Fig. 5, the feature vector corresponding to the k^{th} frame, F_k, of the input video V is computed as

$$l^F_k = [LBP^{u2}_{8,1}(F_k)\ LBP^{u2}_{8,2}(F_k)\ LBP^{u2}_{16,1}(F_k)] \quad (9)$$

As opposed to Määttä *et al.* [21] 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). The algorithm is applied individually at every frame and the features of an input video consisting of n frames can be represented as:

$$lbp_V = [l^F_1\ l^F_2\ l^F_3 \ldots l^F_n] \quad (10)$$

For classification, SVM [35] with Radial Basis Function (RBF) kernel is used. Further, each frame is pre-processed to remove illumination based variations [36].

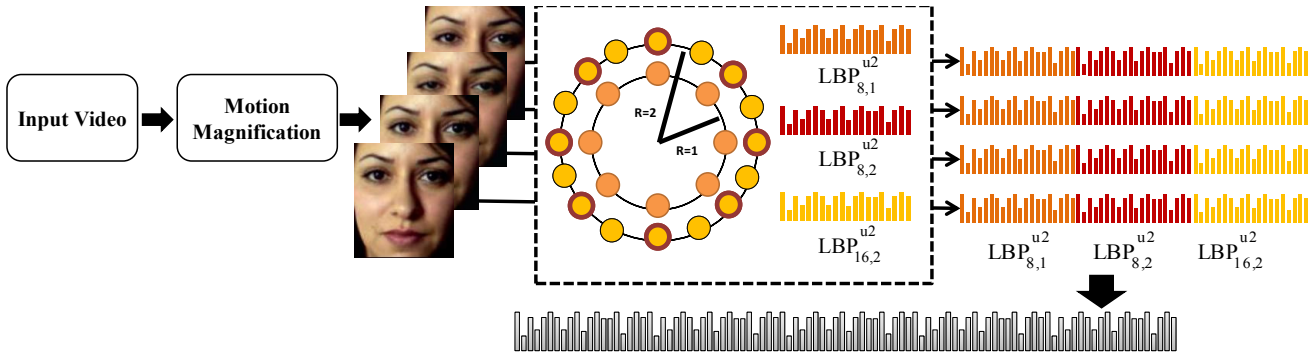


Fig. 5. Illustrating the proposed texture based spoofing detection approach with motion magnification.

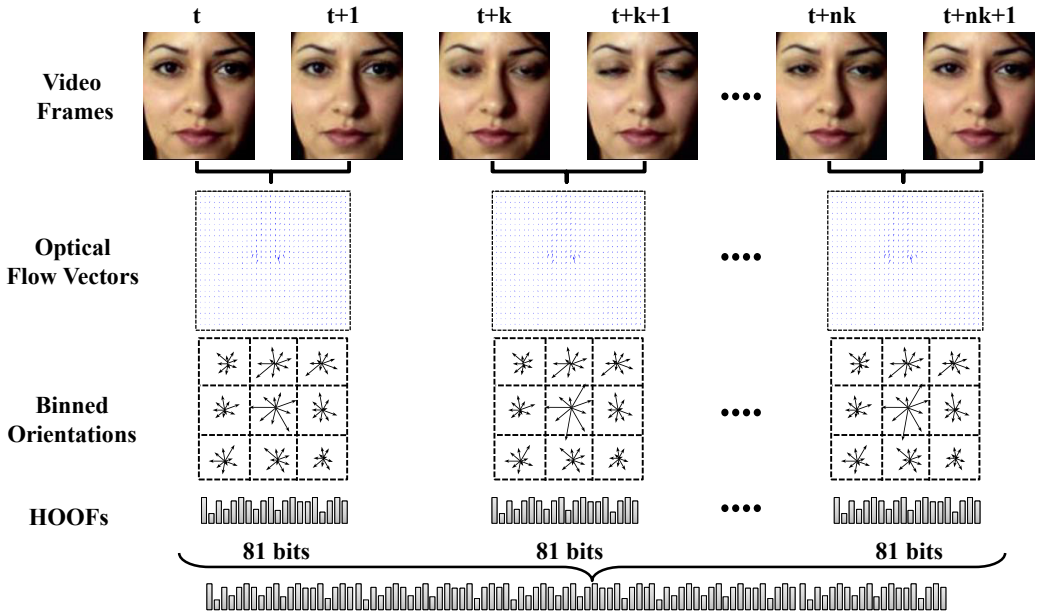


Fig. 6. An illustration of the proposed liveness based feature extraction approach. HOOF descriptors obtained between pairs of frames at a fixed interval are concatenated to create a single feature vector.

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 efficient features for spoof detection. Optical flow is a dense motion estimation technique that computes the motion of each pixel. Optical flow has already been used for motion estimation in several applications including identification of facial micro-expressions in videos [17].

$$\frac{\partial V}{\partial x}G_x + \frac{\partial V}{\partial y}G_y + \frac{\partial V}{\partial t} = 0 \tag{11}$$

$$\theta = \arctan\frac{G_y}{G_x}, \quad m = \sqrt{G_x^2 + G_y^2} \tag{12}$$

In an image, the flow in both horizontal (G_x) and vertical (G_y) directions are used to compute the orientation based flow vector by solving the optimization problem shown in Eq. 11. In this research, conjugate gradient approach [37] is used to solve the optimization problem due to its low computational complexity. Raw optical flow per pixel may be too spatially constrained and encode redundant background or unwanted motion. Therefore, as illustrated in Fig. 6, 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. 12, 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. Instead of using the histogram of optical flow magnitude, magnitude weighted orientation bins are utilized, and the vector thus obtained is termed as HOOF [33]. The final feature vector ($hoof_{V,k}$) 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 as shown in Eq. 13.

$$\begin{aligned} h^V_{p,q} &= HOOF(F_p, F_q) \\ hoof_{V,k} &= [h^V_{1,1+k}\ h^V_{2+k,2+2k} \cdots\ h^V_{n-k,n}] \end{aligned} \tag{13}$$

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.

C. Videolet Score Aggregation and Evidence Fusion

Thus far, this research presents two approaches to spoofing detection, namely texture analysis with multi-LBP and motion analysis using HOOF descriptor. HOOF based motion analysis encodes micro-movements of a face. It is proficient as a *liveness* approach and provides a *real-face* classification. On the other hand, LBP texture analysis encodes the texture of a facial region which manifests differently for a live face than a spoofed face. For instance, the texture of facial region is affected by the reflective properties of the printed surface, mask or display device used for spoofing. Since there are different types of attacks such as print and replay, using only texture based or only motion based feature may not always yield the best results. Therefore, we propose to combine both texture and motion features for improved performance. The fusion is performed by combining the prediction scores obtained by SVM classification of motion magnified multi-LBP ($P_{Mag-LBP}$) and HOOF ($P_{Mag-hoof}$) features.

In several existing 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 feature vector. As illustrated in Fig. 7, 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 divided into overlapping windows of size w, similar to [32], with a step size of half the window size. The frames corresponding to a single window are termed as *videolet*.

Motion magnified video is divided into $\eta = (\frac{2N}{w} - 1)$ videolets followed by extraction of multi-LBP and HOOF features from each of these videolets. These scores are combined using sum rule to generate the fused prediction scores (P) per videolet. The prediction scores are then combined using Eq. 14 to generate the final score (F).

$$F = \frac{1}{\eta}\sum_{j=1}^{\eta}(P_t^j) \tag{14}$$

A threshold T is applied on F for classification of video as either spoofed or real. The proposed videolet aggregation approach with LBP texture analysis combined with HOOF motion analysis on motion magnified videos is summarized in Algorithm 1.

Algorithm 1 Spoofing Detection in video V

input: A video V=$\{F_1, F_2, \ldots F_N\}$, trained models: SVM^{hoof} and SVM^{lbp}, videolet size w, interval k, and decision threshold T.
V^{mag} = Motion magnification of input video V
$F^{Mag-LBP} = lbp_{F^{mag}}$ (as in Eq. 10)
$F^{Mag-hoof} = hoof_{F^{mag},k}$ (as in Eq. 13)
$\eta = (\frac{2N}{w} - 1)$ (number of videolets)
$s = 1$
iterate: $i = 1$ **to** η **do**
$videolet_{Mag-LBP}$= $\{F_j^{Mag-LBP} | s \leq j \leq (s+w)\}$
$videolet_{Mag-hoof}$= $\{F_j^{Mag-hoof} | s \leq j \leq (s+w)\}$
$s = s + (\frac{w}{2})$
$P_{Mag-LBP}$= $SVM^{lbp}(videolet_{Mag-LBP})$
$P_{Mag-hoof}$= $SVM^{hoof}(videolet_{Mag-hoof})$
P_t^i= $P_{Mag-LBP}$ + $P_{Mag-hoof}$ (Evidence fusion)
end iterate.
F= $\frac{1}{\eta}\sum_{j=1}^{\eta}(P_t^j)$ (Videolet aggregation)
Output: report **if** $(F > T)$ "spoof" **else** "non-spoof"

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) Print Attack dataset [11], (2) Replay Attack dataset [12], and CASIA-FASD dataset [13]. Fig. 8 shows sample images from the three databases and the details of these datasets are summarized in Table I. The details of the structure of the databases and predefined (official) protocols are provided below.

TABLE I
Details of the databases used for evaluation.

	Print Attack [11]	**Replay Attack** [12]	**CASIA FASD** [13]
Size (real/attack)	200/200	200/1000	150/600
Length	14 sec	14 sec	1-19 sec
Resolution	640 × 480		Varied
Background	Controlled/ Adverse Illumination		Lab Setting
Attacks	Print	Print, Replay	Print, Wrap, Replay

- The Print Attack dataset [11] consists of 200 real access and 200 printed-photo attack attempt videos of 50 subjects. For spoofing detection, the dataset is split into training (120 videos), development (120 videos), and testing (160 videos) subgroups. The training and development subgroups contain 60 real access videos and 60 print attack videos each, whereas the testing subgroup contains 80 real access and 80 print attack videos. The videos are captured under both controlled and adverse lighting conditions.
- The Replay Attack dataset [12] consists of 1200 videos that include 200 real access videos, 200 print attack videos, 400 phone attack videos, and 400 tablet attack videos. The evaluation protocol splits the dataset into training (360 videos), development (360 videos), and testing (480 videos) subgroups. The training and development subgroups contain 60 real access videos and 300 attack videos each, whereas the testing subgroup contains 80 real access and 400 attack videos. Fig. 8 illustrates some sample frames from the Print and Replay datasets.
- The CASIA-FASD dataset [13] consists of 600 videos corresponding to 50 subjects, separated as 240 videos in training and 360 videos in testing. In addition to

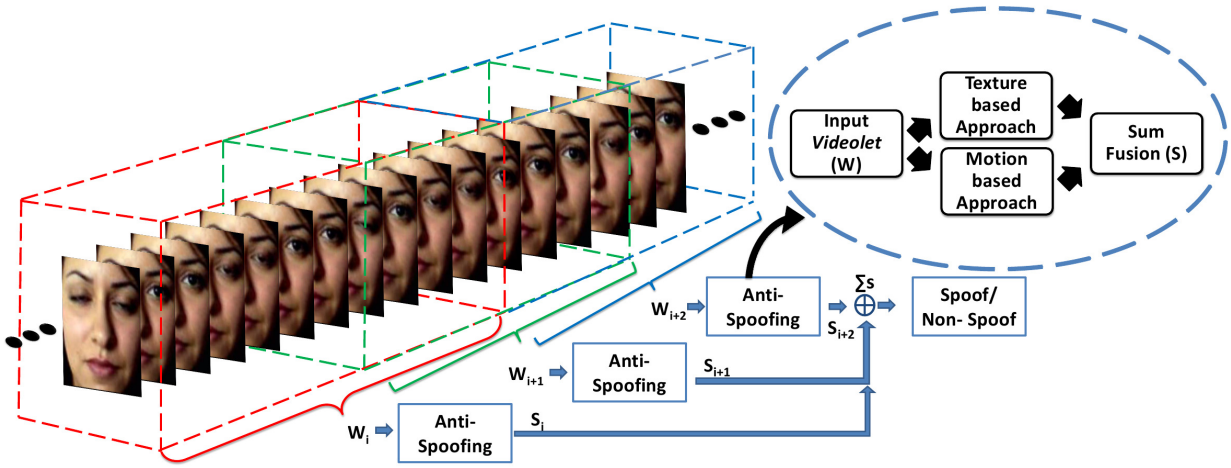


Fig. 7. An illustration of the windowed approach. The prediction scores obtained for each *videolet* are aggregated over the entire length of a video.

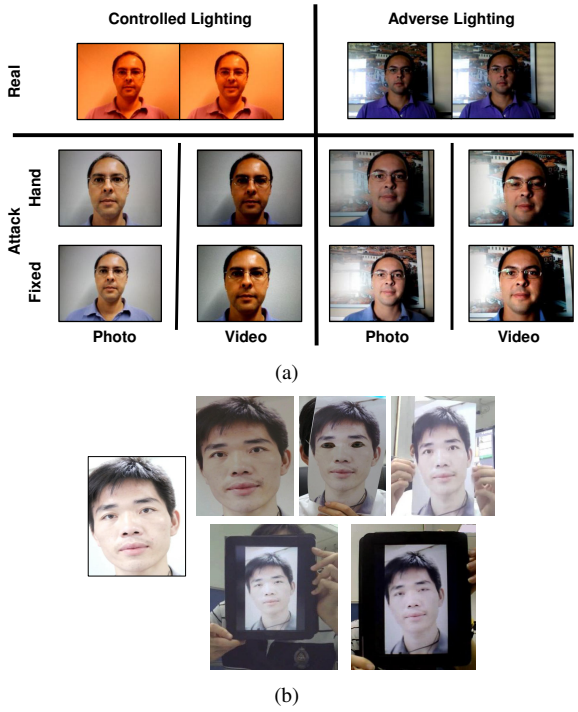


Fig. 8. Samples images from a) Print and Replay Attack datasets [11], [12] and b) CASIA-FASD dataset [13].

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. Sample frames from this database are shown in Fig. 8. The challenging nature of the dataset is furthered by variations in resolution, quality, and video length (ranging from 1 to 19 seconds).

To demonstrate the effectiveness of the proposed framework, three experiments are performed as described below.

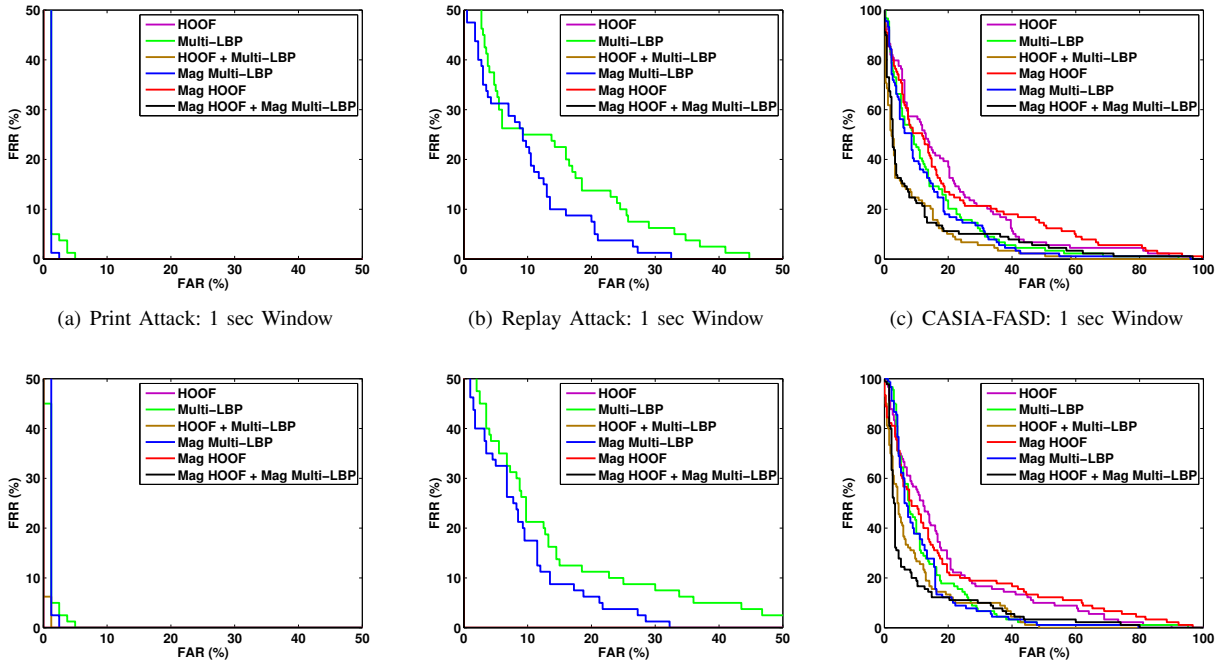
1) *Experiment 1*: The Print and Replay Attack databases are used with the pre-defined experimental protocols described earlier in the section. The same experiment is performed using both normal and motion magnified videos.
2) *Experiment 2*: The performance of the proposed approach is evaluated on the challenging CASIA-FASD dataset. Since no development set is provided, cross validation is used by randomly dividing the training data into five folds, as described in [30]. The average test HTER that is obtained from the five training models is reported. The EER results for the same dataset are reported as per the pre-defined protocol of the dataset [13].
3) *Experiment 3*: A study by Pereira *et al.* [28] on existing techniques shows low generalization across different datasets. In order to evaluate the *generalizability* of the proposed approach, two evaluations are performed:
(i) CASIA $\rightarrow$ REPLAY: The training set of the CASIA-FASD dataset is used as per Experiment 2. The proposed approach is evaluated on the development as well as test sets of REPLAY attack.
(ii) REPLAY $\rightarrow$ CASIA: The training and development sets of the REPLAY attack are used as per Experiment 1. The performance of the proposed approach is computed on the test sets of CASIA-FASD dataset.

All the input frames are first pre-processed by cropping the face region based on eye coordinates obtained from a commercial face recognition system to a fixed resolution (130×150). In order to correct for small inconsistencies in eye detection, global image registration [39] is applied with the first frame as reference. This process also minimizes the motion in videos that are not facial motion. The normalization process may also reduce the effect of hand motion in spoofing attacks (while holding up a printed image or a display device). For computing texture based features, all the frames are converted to gray scale.

The same pre-processed videos (eye-detection and frame registration) are used in all the experiments performed in this

TABLE II
EXPERIMENT 1: CLASSIFICATION PERFORMANCE OF VARIOUS APPROACHES IN TERMS OF HTER (%) AND EER (%) ON THE PRINT AND REPLAY ATTACK DATABASES.

Algorithm	Print Attack						Replay Attack					
	Normal Videos			Motion Magnified Videos			Normal Videos			Motion Magnified Videos		
Window Size = Video Length												
	HTER		EER	HTER		EER	HTER		EER	HTER		EER
	Dev	Test	Test	Dev	Test	Test	Dev	Test	Test	Dev	Test	Test
Multi-LBP [38]	5.00	4.37	5.00	3.33	3.12	3.75	15.00	20.62	20.25	6.67	5.62	5.00
HOOF [38]	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.63	0.00
Window Size = 1 second (25 frames)												
Multi-LBP	3.33	2.50	3.75	3.33	1.87	1.25	18.50	17.25	17.50	16.66	13.75	13.00
HOOF	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
HOOF + Multi-LBP	0.00	0.00	0.00	**0.00**	**0.00**	**0.00**	0.00	0.00	0.00	**0.00**	**0.25**	**0.00**
Window Size = 2 seconds (50 frames)												
Multi-LBP	5.00	2.50	2.50	3.33	1.87	2.50	18.33	19.00	14.50	16.66	12.75	12.00
HOOF	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
HOOF + Multi-LBP	1.66	1.87	1.25	0.00	0.625	0.00	1.66	1.25	0.00	0.33	0.00	0.00



(a) Print Attack: 1 sec Window (b) Replay Attack: 1 sec Window (c) CASIA-FASD: 1 sec Window

(d) Print Attack: 2 sec Window (e) Replay Attack: 2 sec Window (f) CASIA-FASD: 2 sec Window

Fig. 9. ROC plots of the proposed approach on the test sets of the Print, Replay, and CASIA-FASD datasets with different window sizes. The curves with 0% EER overlap with axes.

evaluation. To be consistent with existing literature, the results of spoofing detection are reported in terms of both Half Total Error Rate (%) and Equal Error Rate (%) on the test set. However, the HTER on the test set at a threshold computed on a development set represents a more realistic scenario.

The parameters of SVM are determined using a grid search where the objective of grid search is defined in terms of optimizing the equal error rate on the development set. The threshold used for computing HTER on the test set is also obtained from the development set. For motion magnification, optimal parameters are empirically determined to be $\alpha = 50, \lambda_c = 10$, and an ideal bandpass filter with band $100 - 120$ Hz.

IV. EXPERIMENTAL EVALUATION

Using the protocols described in the previous section, three experiments are performed and the results are described below. Later the results are also compared with those reported in literature on the same experimental protocol.

A. Experiment 1: Print and Replay Attack Datasets

Table II illustrates the results of Experiment 1 for HOOF, multi-LBP, and the proposed spoofing detection approach. The error plots on the test set of both Print and Replay Attack datasets are also illustrated in Fig. 9. Salient observations are presented below.

Performance of Multi-LBP and HOOF:

- Texture based multi-LBP and SVM classification yields 2.50% HTER on the Print Attack dataset and 17.25% on the Replay Attack dataset for a window size of one second.
- The proposed HOOF + SVM provides perfect classification performance on both the datasets (0% HTER on both Print Attack and Replay Attack).

Effect of Motion Magnification on Multi-LBP:

- We investigate the effect of motion magnification on the two feature extraction algorithms. The HTER of multi-LBP reduces to 1.87% (from 2.50%)and 13.75% (from 17.25%) for Print and Replay attacks when the videos are preprocessed using motion magnification. We also observe that motion magnification significantly improves the performance of each component of multi-LBP.
- The distribution of SVM scores of the multi-LBP descriptors shown in Fig. 10 indicates an improved separation between the real and spoof classes when using motion magnification. For further analysis, we compute the absolute difference of descriptors from a small video snippet as shown in Fig. 11. The histograms illustrate that the effect of motion magnification on LBP descriptor is more evident for real videos than for attack videos. Motion magnification enhances the subtle variations in texture of faces obtained from a real person and those obtained by spoofing.
- As shown in Eq. 5, given a video V, the motion magnified form $\hat{V}$ is a function of the given frame (f) and the residual from a bandpass filter (B). From Eq. 3, the components of the filtered signal are a function of $\delta_x(t)$ and $\delta_y(t)$ that describe the displacement of each pixel in x and y directions. It is our assertion that the displacement functions for attack videos are different from a real video. Hence, the motion magnification operation changes the frame texture differently, leading to more separated LBP texture patterns.
- Another important difference between a real face presented in front of a camera, compared to a replay attack video (played on a portable video screen) is the variation in sampling rate of capture. Specifically, real face can be viewed as a continuous signal available for sampling from a video camera. On the other hand, a replay attack video presented on a display screen (refresh rate of 30 frames/sec) is re-sampled by a video camera at 25 frames/sec. The disparity in this capture rate leads to a different texture that is further enhanced by motion magnification. Hence, the LBP descriptor obtained from a motion magnified video may be more discriminating for spoofing attacks.

Effect of Motion Magnification on HOOF:

- Since HOOF already yields 0% error rate, applying motion magnification does not change the accuracy. However, the separation of SVM scores of the two classes increases on preprocessing with motion magnification.
- The proposed HOOF approach encodes the motion variations at block level by estimating the displacement vectors, $\delta_x(t)$ and $\delta_y(t)$, via optical flow. When motion magnification is applied, the displacements are magnified and the separation between spoof and non-spoof classes increases.

Performance of the Proposed Approach:

- The proposed fusion approach (using motion magnified HOOF and multi-LBP with videolets) provides 0% EER with uncontrolled illumination and background on both the datasets. As mentioned earlier, motion estimation using HOOF encodes the liveness features of a face thereby acting as a good *live-face* classifier. On the other hand, multi-LBP encodes the texture of real faces differently from that of spoofed faces. Hence, sum rule fusion of these classifiers provides a robust and accurate anti-spoofing measure.
- As shown in Fig. 9(a), (b), (d) and (e), it is consistently observed that the window size of 1-2 seconds provides better classification performance on the Print and Replay attack videos, and also accommodates for varying video lengths. It is also observed that the performance of the proposed approach saturates after accumulation of 12 videolets in the case of Print Attack and 8 videolets for Replay Attack datasets.

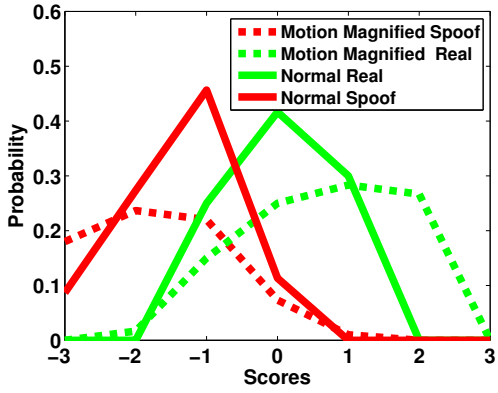


Fig. 10. Histograms of SVM scores for multi-LBP on Replay dataset. Motion magnification helps to reduce the overlap between the real and attack distributions.

- **Computational Efficiency**: The time is reported for MATLAB2012 implementation on a machine with Intel Quad Core CPU Q8300 at 2.5GHz and 4GB RAM. For a video with 375 frames (15 seconds in length), 14 videolets are created, each of 25 frames. The proposed approach involves registration (293.8s), motion magnification (28.4s), HOOF extraction (15.2s), and multi-LBP feature extraction (14.3s) requires a total of 351.7s to process the entire video serially. We believe that a parallel implementation can further reduce the processing time.

B. Experiment 2: CASIA-FASD Dataset

The results of Experiment 2 on the CASIA-FASD dataset are shown in Table III and Fig. 9(c) and (f) show the ROC plot. With motion magnification, the proposed multi-LBP approach provides 21.11% HTER with a window size of one

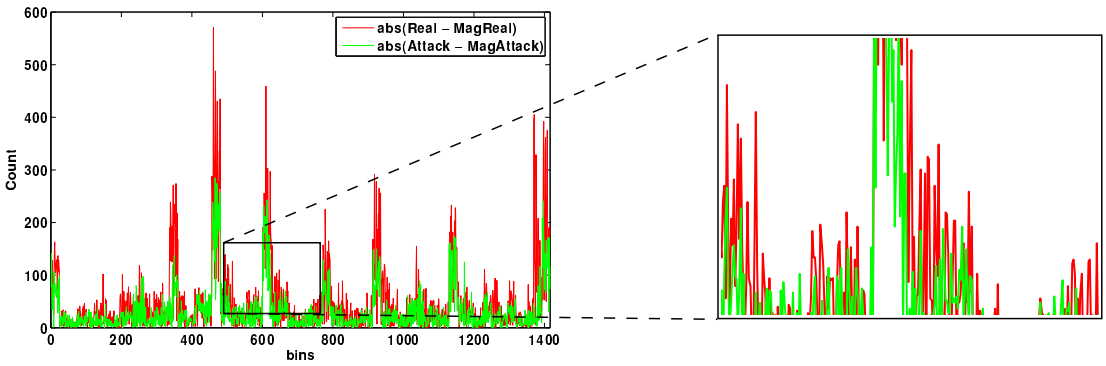


Fig. 11. An illustration on the effect of motion magnification on LBP descriptor of a single videolet from real and attack videos. The absolute difference between the histograms shows a larger change in the descriptor of the real videos compared to an attack video.

TABLE III
EXPERIMENT 2: CLASSIFICATION PERFORMANCE OF VARIOUS APPROACHES IN TERMS OF HTER (%) AND EER (%) ON THE CASIA-FASD DATASET.

Window Size	**Algorithm**	**CASIA-FASD**					
		Normal Videos			**Motion Magnified Videos**		
		HTER		EER	HTER		EER
		Dev	Test	Test	Dev	Test	Test
1 second (25 frames)	Multi-LBP	13.31	19.00	17.77	12.77	21.77	15.74
	HOOF	14.51	32.07	22.22	16.98	28.37	21.11
	HOOF + Multi-LBP	10.12	19.51	15.55	**10.13**	**17.59**	**14.44**
2 seconds (50 frames)	Multi-LBP	17.16	21.95	20.29	13.32	22.05	18.99
	HOOF	18.50	34.14	24.95	16.66	27.66	23.64
	HOOF + Multi-LBP	14.76	19.49	15.64	13.67	17.86	14.52

TABLE IV
EXPERIMENT 2: CLASSIFICATION PERFORMANCE IN TERMS OF EER (%) ON SUB-SETS OF THE CASIA-FASD DATASET [13].

Test Protocol	**Low quality**	**Normal quality**	**High quality**	**Warped photo attack**	**Cut photo attack**	**Video attack**	**Overall**
Multi-LBP	12.77	16.66	26.66	15.55	25.55	17.77	17.77
Mag-Multi-LBP	7.22	13.33	29.44	14.44	22.22	13.33	15.74
HOOF	16.66	30.00	26.11	15.55	17.77	38.88	21.11
Mag-HOOF	17.22	33.33	22.77	12.22	20.00	36.60	22.22
HOOF + Multi-LBP	9.44	20.55	16.66	10.00	16.66	24.44	15.55
Proposed	**6.11**	23.33	**13.88**	**10.00**	14.44	**20.00**	**14.44**
Zhang *et al.*(2012) [13]	13.00	**13.00**	26.00	16.00	**6.00**	24.00	17.00

second whereas the HOOF approach provides 28.80% HTER. On combining the two, the proposed fusion approach yields 17.59% HTER and 14.44% EER at a window size of 25 frames. The windowed approaches provide better performance since they ensure that the evidence obtained from the entire duration of the videos is considered. Comparatively larger error rates on the CASIA-FASD dataset may be attributed to the challenging nature of the dataset, specially, varied length videos, variations in capture devices, resolution, and types of spoofing attacks. Table IV shows the performance of the proposed approach on the six sub-sets of the CASIA-FASD test dataset. The subsets are divided based on the type of spoofing attack (print, cut, and wrapped) and quality of the video samples (low, normal, and high). SVM is trained from the training set as before. The results showcase the advantage of fusing HOOF and multi-LBP. In each of the six experiments, the performance of Mag(HOOF + multi-LBP) is superior to that of HOOF + multi-LBP, which suggests that motion magnification improves the performance of the proposed approach in all cases. On this database, as observed previously, motion magnification amplifies the difference in texture from real and spoofed video frames.

C. Experiment 3: Cross Datasets

Experiment 3 evaluates the *generalizability* of the proposed approach with an inter-database experiment and the results are shown in Table V. In the first evaluation, the trained models from Experiment 2 of the CASIA-FASD dataset are used to evaluate the performance on the development and test sets of Replay attack dataset. Since the CASIA dataset consists of instances from three different types of attacks, the proposed fusion of HOOF and multi-LBP yields an EER of 41.50% and 41.12% on the print, cut and video attacks respectively. Further, a marginal improvement in performance (both in terms of HTER and EER) is observed when motion magnification is applied before feature extraction. In the second evaluation, trained models obtained from Experiment 1 of the Replay attack dataset are used to evaluate the performance on the CASIA-FASD dataset. The performance of the proposed

TABLE V
EXPERIMENT 3: CLASSIFICATION PERFORMANCE OF THE PROPOSED FUSION ALGORITHM IN TERMS OF HTER AND EER (%). THE ALGORITHMS ARE TRAINED USING THE CASIA-FASD DATASET AND TESTED ON THE REPLAY ATTACK DATASET, AND VICE VERSA. FOR EACH EXPERIMENT, TOP TWO RESULTS ARE HIGHLIGHTED.

Algorithm	**Test: Replay Attack (Train: CASIA)**				**Test: CASIA-FASD (Train: Replay)**		
	Dev		**Test**		**Dev**	**Test**	
	HTER	EER	HTER	EER	HTER	HTER	EER
Multi-LBP	53.93	46.66	52.47	48.75	53.58	55.55	52.22
Mag-Multi-LBP	51.16	60.00	50.97	61.00	53.07	53.14	55.55
HOOF	**48.30**	43.33	**49.62**	42.50	**48.29**	**49.81**	**43.51**
Mag-HOOF	**35.76**	**32.00**	**37.45**	**37.50**	51.36	52.59	**43.51**
HOOF+Multi-LBP	51.10	41.66	51.10	41.25	48.43	51.66	47.77
Proposed	50.00	**41.50**	50.20	**41.12**	**43.86**	**50.37**	46.66

TABLE VI
COMPARISON WITH SOME EXISTING ALGORITHMS IN TERMS OF HTER (%) AND EER (%).

	Algorithm	**Print Attack**		**Replay Attack**		**CASIA-FASD**	
		HTER	EER	HTER	EER	HTER	EER
Anjos *et al.*(2011) [11]	Motion correlation	8.98	–	–	–	–	–
Schwartz *et al.*(2011) [31]	Partial least squares	–	1.67	–	–	–	–
Zhang *et al.*(2012) [13]	DoG baseline	–	–	–	–	–	17.00
Pereira *et al.*(2013) [28]	Motion correlation	–	–	11.79	11.66	30.33	26.65
	$\text{LBP}^{u2}_{8,1}$	–	–	15.45	14.41	23.19	24.63
	$\text{LBP-TOP}^{u2}_{8,8,8,1,1,1}$	–	–	8.51	8.17	23.75	21.59
Yang *et al.*(2013) [29]	Component approach	–	1.20	–	–	–	11.80
Pereira *et al.*(2014) [30]	Motion correlation	–	–	11.79	–	–	–
	$\text{LBP}^{u2}_{8,1}$	–	–	15.16	–	–	16.00
	$\text{LBP-TOP}^{u2}_{8,8,8,1,1,1}$	–	–	8.51	–	–	–
	$\text{LBP-TOP}^{u2}_{8,8,8,1,1,1}$ with average of feature	–	–	–	–	–	10.00
	$\text{LBP-TOP}_{8,8,8,1,1,[1-2]}$	–	–	7.60	–	–	–
Proposed	Mag-HOOF + Mag-Multi-LBP	**0.00**	**0.00**	**0.25**	**0.00**	**17.59**	**14.44**

approach on the development set is 43.86% HTER, whereas, the performance on the test set is 50.37% HTER and 46.66% EER. Note that in this case, EER is reported only on the test set of CASIA-FASD dataset since there is no development set for this dataset under the pre-defined experimental protocol [13]. The comparatively larger error rates in the second evaluation may be attributed to the lack of training samples in Replay attack dataset to tackle to various types of spoofing and varying quality of videos in the CASIA-FASD dataset. Also, the results indicate that the motion-based approach (mag-HOOF) is more resilient to cross dataset settings compared to texture-based approach (multi-LBP). Compared to the published results of Pereira *et al.* [28], these results are marginally better.

D. Comparison with Existing Approaches

As stated previously, the performance of the proposed algorithm has been evaluated using the pre-defined (official) experimental protocols provided with the databases. Therefore, the error rates of the proposed algorithm can be directly compared with that of the existing results results. Table VI presents the comparison with some existing approaches on the three datasets. On the Print and Replay attack datasets, the proposed algorithm outperforms existing algorithms with near perfect performance. On the CASIA-FASD dataset, the proposed algorithm provides comparable results in terms of EER and outperforms existing approaches in terms of HTER.

V. CONCLUSION AND FUTURE DIRECTIONS

For secure face recognition systems, it is important that face anti-spoofing techniques are robust and computationally efficient to improve the practicality of face biometrics. This research presents a novel framework for facial spoofing detection using motion (liveness) and texture (anti-spoofing) features. Using motion magnification, an input video of a subject is enhanced to exaggerate subtle macro- and micro-facial expressions usually presented by a real person. Our experiments indicate that motion magnification improves the performance of LBP texture features, including that of the proposed computationally efficient configuration of LBP features (multi-LBP). Further, a motion descriptor is computed using HOOF to encode liveness features. Finally, combining evidence from both texture and motion analysis ensures efficient solution that is robust to a diverse range of spoofing attacks. The proposed approach achieves state-of-the-art accuracy on different publicly available databases and requires reasonable computational time. Since the majority of the computational time required for registration step, time complexity can be further reduced by parallelizing the registration process. However, cross dataset performance improvement and new attack methods such as face masks [40] are important future research directions.

VI. ACKNOWLEDGEMENT

A shorter version of this manuscript [38] is published in the Proceedings of IEEE Conference on Computer Vision

and Pattern Recognition Workshops, 2013. This research is supported in part by Department of Electronics and Information Technology, Government of India. The authors thank the reviewers for their feedback and comments.

REFERENCES

[1] 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.

[2] ——, "An analysis of minutiae matching strength," in *Proccedings of Audio- and Video-Based Biometric Person Authentication*, vol. 2091, 2001, pp. 223–228.

[3] S. Yoon, J. Feng, and A. K. Jain, "Altered fingerprints: Analysis and detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 3, pp. 451–464, 2012.

[4] R. Singh, M. Vatsa, H. Bhatt, S. Bharadwaj, A.Noore, and S. Nooreyezdan, "Plastic surgery: A new dimension to face recognition," *IEEE Transaction on Information Forensics and Security*, vol. 5, no. 3, pp. 441–448, 2010.

[5] 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, 2006.

[6] S. Venugopalan and M. Savvides, "How to generate spoofed irises from an iris code template," *IEEE Transactions on Information Forensics and Security*, vol. 6, no. 2, pp. 385–395, 2011.

[7] S. E. Baker, A. Hentz, K. W. Bowyer, and P. J. Flynn, "Degradation of iris recognition performance due to non-cosmetic prescription contact lenses," *Computer Vision and Image Understanding*, vol. 114, no. 9, pp. 1030–1044, 2010.

[8] D. Yadav, N. Kohli, R. Singh, and M. Vatsa, "Revisiting iris recognition with color cosmetic contact lenses," in *Proceedings of IAPR International Conference on Biometrics*, 2013, pp. 1–7.

[9] T. I. Dhamecha, A. Nigam, R. Singh, and M. Vatsa, "Disguise detection and face recognition in visible and thermal spectrums," in *Proceedings of IAPR International Conference on Biometrics*, 2013, pp. 1–7.

[10] C. Roberts, "Biometric attack vectors and defences," *Computers & Security*, vol. 26, no. 1, pp. 14–25, 2007.

[11] A. Anjos and S. Marcel, "Counter-measures to photo attacks in face recognition: a public database and a baseline," in *Proceedings of IEEE/IAPR International Joint Conference on Biometrics*, 2011, pp. 1–7.

[12] I. Chingovska, A. Anjos, and S. Marcel, "On the effectiveness of local binary patterns in face anti-spoofing," in *Proceedings of International Conference of the Biometrics Special Interest Group*, 2012, pp. 1–7.

[13] Z. Zhang, J. Yan, S. Liu, Z. Lei, D. Yi, and S. Z. Li, "A face antispoofing database with diverse attacks," in *Proceedings of IAPR International Conference on Biometrics*, 2012, pp. 26–31.

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

[15] K. Kollreider, H. Fronthaler, and J. Bigun, "Verifying liveness by multiple experts in face biometrics," in *Proceedings of Computer Vision and Pattern Recognition Workshops*, 2008, pp. 1–6.

[16] M. Chakka, A. Anjos, S. Marcel, R. Tronci, D. Muntoni, G. Fadda, M. Pili, N. Sirena, G. Murgia, M. Ristori, F. Roli, J. Yan, D. Yi, Z. Lei, Z. Zhang, S. Li, W. Schwartz, A. Rocha, H. Pedrini, J. Lorenzo-Navarro, M. Castrillon-Santana, J. Määttä, A. Hadid, and M. Pietikainen, "Competition on counter measures to 2-D facial spoofing attacks," in *Proceedings of IEEE/IAPR International Joint Conference on Biometrics*, 2011, pp. 1–6.

[17] M. Shreve, S. Godavarthy, D. Goldgof, and S. Sarkar, "Macro- and micro-expression spotting in long videos using spatio-temporal strain," in *Proceedings of IEEE International Conference on Automatic Face Gesture Recognition and Workshops*, 2011, pp. 51–56.

[18] R. D. Findling and R. Mayrhofer, "Towards face unlock: on the difficulty of reliably detecting faces on mobile phones," in *Proccedings of International Conference on Advances in Mobile Computing & Multimedia*, 2012, pp. 275–280.

[19] A. Anjos, M. M. Chakka, and S. Marcel, "Motion-based counter-measures to photo attacks in face recognition," *Institution of Engineering and Technology Journal on Biometrics*, pp. 1–7, 2013.

[20] T. Wang, J. Yang, L. Zhen, L. Shengcai, and S. Z. Li, "Face liveness detection using 3D structure recovered from a single camera," in *Proceedings of IAPR International Conference on Biometrics*, 2013, pp. 1–7.

[21] J. Määttä, A. Hadid, and M. Pietikainen, "Face spoofing detection from single images using micro-texture analysis," in *Proceedings of IEEE/IAPR International Joint Conference on Biometrics*, 2011, pp. 1–7.

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

[23] J. Määttä, A. Hadid, and M. Pietikainen, "Face spoofing detection from single images using texture and local shape analysis," *IET Biometrics*, vol. 1, no. 1, pp. 3–10, 2012.

[24] K. Gahyun, S. Eum, J. K. Suhr, D. I. Kim, K. R. Park, and J. Kim, "Face liveness detection based on texture and frequency analyses," in *Proceedings of IAPR International Conference on Biometrics*, 2012, pp. 67–72.

[25] T. F. Pereira, A. Anjos, J. M. De Martino, and S. Marcel, "LBP-TOP based countermeasure against facial spoofing attacks," in *Proceedings of Asian Conference on Computer Vision Workshops*, 2012, pp. 121–132.

[26] G. Zhao and M. Pietikainen, "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, 2007.

[27] I. Chingovska, J. Yang, Z. Lei, D. Yi, S. Z. Li, O. Kähm, C. Glaser, N. Damer, A. Kuijper, A. Nouak, J. Komulainen, T. Pereira, S. Gupta, S. Khandelwal, S. Bansal, A. Rai, T. Krishna, D. Goyal, M.-A. Waris, H. Zhang, I. Ahmad, S. Kiranyaz, M. Gabbouj, R. Tronci, M. Pili, N. Sirena, F. Roli, J. Galbally, J. Fierrez, A. Pinto, H. Pedrini, W. S. Schwartz, A. Rocha, A. Anjos, and S. Marcel, "The 2nd competition on counter measures to 2D face spoofing attacks," in *Proceedings of IAPR International Conference on Biometrics*, 2013, pp. 1–7.

[28] T. F. Pereira, A. Anjos, J. M. De Martino, and S. Marcel, "Can face anti-spoofing countermeasures work in a real world scenario?" in *Proceedings of IAPR International Conference on Biometrics*, 2013, pp. 1–7.

[29] J. Yang, L. Zhen, L. Shengcai, and S. Z. Li, "Face liveness detection with component dependent descriptor," in *Proceedings of IAPR International Conference on Biometrics*, 2013, pp. 1–7.

[30] 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, pp. 2–10, 2014.

[31] W. R. Schwartz, A. Rocha, and H. Pedrini, "Face spoofing detection through partial least squares and low-level descriptors," in *Proceedings of IEEE/IAPR International Joint Conference on Biometrics*, 2011, pp. 1–8.

[32] J. Komulainen, A. Anjos, S. Marcel, A. Hadid, and M. Pietikäinen, "Complementary countermeasures for detecting scenic face spoofing attacks," in *Proceedings of IAPR International Conference on Biometrics*, 2013, pp. 1–7.

[33] 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 *Proceedings of International Conference on Computer Vision and Pattern Recognition*, 2009, pp. 1932–1939.

[34] H.-Y. Wu, M. Rubinstein, E. Shih, J. Guttag, F. Durand, and W. T. Freeman, "Eulerian video magnification for revealing subtle changes in the world," *ACM Transactions on Graphics*, vol. 31, no. 4, pp. 65:1–65:8, 2012.

[35] C. Cortes and V. Vapnik, "Support-vector networks," *Springer Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[36] X. Tan and B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," *IEEE Transactions on Image Processing*, vol. 19, no. 6, pp. 1635–1650, 2010.

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

[38] S. Bharadwaj, T. I. Dhamecha, M. Vatsa, and R. Singh, "Computationally efficient face spoofing detection with motion magnification," in *Proceedings of Computer Vision and Pattern Recognition Workshops*, 2013, pp. 1–7.

[39] S. Periaswamy and H. Farid, "Elastic registration in the presence of intensity variations," *IEEE Transactions on Medical Imaging*, vol. 22, no. 7, pp. 865–874, 2003.

[40] S. Marcel and N. Erdogmus, "Spoofing in 2D face recognition with 3D masks and anti-spoofing with kinect," in *Proceedings of IEEE Biometrics Theory, Applications and Systems*, 2013, pp. 1–7.