Computationally Efficient Face Spoofing Detection with Motion Magnification

Samarth Bharadwaj, Tejas I. Dhamecha, Mayank Vatsa and Richa Singh
IIIT-Delhi, India
{samarthb, tejasd, mayank, rsingh}@iiitd.ac.in

Abstract

For a robust face biometric system, a reliable anti-spoofing approach must be deployed to circumvent the print and replay attacks. Several techniques have been proposed to counter face spoofing, however a robust solution that is computationally efficient is still unavailable. This paper presents a new approach for spoofing detection in face videos using motion magnification. Eulerian motion magnification approach is used to enhance the facial expressions commonly exhibited by subjects in a captured video. Next, two types of feature extraction algorithms are proposed: (i) a configuration of LBP that provides improved performance compared to other computationally expensive texture based approaches and (ii) motion estimation approach using HOOF descriptor. On the Print Attack and Replay Attack spoofing datasets, the proposed framework improves the state-of-art performance; especially HOOF descriptor yielding a near perfect half total error rate of 0% and 1.25% respectively.

1. Introduction

Face recognition systems are vulnerable to spoofing attacks with printed photos or replayed videos. Robust performance of existing face detection techniques has contributed to the ease of spoofing attacks on face biometric systems. Further, the wide availability of portable display units with high resolution has brought video replay attacks into the purview as well. The problem of spoofing is particularly compounded in mobile devices enabled with face recognition. For instance, mobile phone feature, *Face Unlock*, that uses face recognition to unlock a phone, has received criticism for being vulnerable to spoofing attacks [6], despite a blinking based liveness detection feature.

The literature on spoofing detection discuss two types of spoofing attacks, namely print and replay. Print attack uses printed photographs of a subject to spoof 2D face recognition systems, while replay attack presents a video of a live person to evade liveness detection. Further, a replay attack video could be of a digital photograph or a digital video replayed on a screen that is either fixed or hand-held. Few frames of real and spoofed videos are illustrated in Figure 1. Several techniques in literature on spoofing detection are based on the observation that face frames of a real person exhibit some unique texture properties in the image when compared to spoofed images.

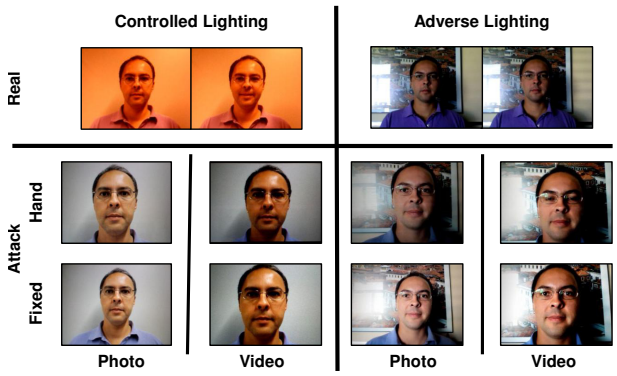

Figure 1. Few video frames from the Replay Attack database [4] illustrating photo (print) and video (replay) attacks.

On the NUAA dataset [16] of 15 subjects, Määttä *et al*. [10] found that for spoofing detection, Local Binary Patterns (LBP) were more efficient than local phase quantization as well as Gabor wavelet based descriptor. Further, they reported that concatenation of three LBP descriptors of different configuration was more efficient than using LBP with single configuration. Chakka *et al*. [2] evaluated the performance of six spoofing detection algorithms on the Print Attack database [1] in the IJCB 2011 counter measure to 2D facial spoofing competition. These algorithms primarily utilized texture and motion based approaches. Facial motions such as eye blinks and motion of head with respect to the background were also used to determine liveness. It was observed that texture based approaches resulted in 0% Half Total Error Rate (HTER). Määttä *et al*. [11] 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 approxima-

tion of each 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% HTER on the Print Attack dataset. In other research [7], power spectrum and LBP features were used in a fusion approach on a print attack database collected using a camera of an automated teller machine.

A more challenging Replay Attack database involving spoofing attack by playing video or by displaying digital photo of the subject on an electronic device kept in front of the camera, was introduced in [4]. Baseline experiments were performed using different variants of LBP and three different classifiers. It was observed that SVM outperformed both Linear Discriminant Analysis (LDA) and χ^2 distance based classification. A recent research [13] used LBP from Three Orthogonal Planes (LBP-TOP) for spoofing detection in the Replay Attack database. LBP-TOP explicitly utilized 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$_{8,8,8,1,1,[1-2]}$ with SVM as the classifier achieved best HTER of 7.6% on the Replay Attack dataset. However, LBP-TOP is computationally expensive and may not scale to realtime applications. In a recent research, a geometric approach to replay attacks was proposed by [5] using two video databases.

Existing approaches to spoof detection widely use texture analysis with complex configurations to achieve better performance. However, a spoofing detection technique must not only be robust but also computationally efficient. In this regard, motion analysis based approach to spoofing detection are relatively less explored. In this research, we present a computationally efficient framework that utilizes motion magnification and texture/motion features for spoof detection. The key contributions of this paper can be summarized as:

- We propose a pre-processing approach to spoofing detection using motion magnification that substantially enhances the micro- and macro- facial motion usually exhibited by a subject. Our experiments indicate that appropriately magnified motion improves the performance of spoofing detection techniques, especially texture based approaches.

- We present a computationally efficient multiscale configuration of LBP that provides, along with motion magnification, improved performance as compared to existing LBP based approaches.

- A novel spoofing detection technique based on motion estimation using optical flow that is encoded with a sparsely pooled Histogram of Oriented Optical Flow (HOOF) [3] is also proposed. Evaluation on two spoofing databases show state-of-the-art performance of the proposed approach along with lower computation time.

Figure 2. Motion magnification of input video may accentuate facial expressions thereby aiding spoofing detection techniques.

2. Proposed Framework

It is our assertion that the performance of spoofing detection techniques can be improved with motion magnification as it might enhance the *liveness* nature of the face video. As shown in Figure 2, the proposed framework to spoofing detection first performs motion magnification. To automatically classify these enhanced videos for spoofing detection, we explore (1) texture features using LBP and (2) motion features using HOOF [3].

2.1. Motion Magnification

Motion magnification techniques in videos based on explicitly tracking a pixel's trajectory over time (Lagrangian approach) are computationally expensive and difficult to compute around occlusion boundaries thus resulting in artifacts. On the other hand, Eulerian approach to motion magnification directly amplifies temporal intensity changes at a given position without the need for explicit estimation [17]. Using appropriate temporal and spatial filtering, the desired motion is localized and then magnified under Taylor expansion assumption.

At first, each frame is decomposed into spatial Laplacian bands. Next, an ideal temporal bandpass filter is applied to each Laplacian band to isolate the desired temporal motion in each band. For instance, when a frequency band of 0.2-0.5 Hz which represents eye-lid movements [12] is applied, blinking motion of the subject is isolated. The isolated bandpassed signal is then multiplied by an amplification factor α and added to the original signal, as shown in Eq. 1.

$$\hat{I}(x, y, t) = I(x, y, t) + \alpha B(x, y, t) \tag{1}$$

where $B(x, y, t)$ is the output of a bandpassed filter for video $I(x, y, t)$, at positions x, y, t. Finally, the decomposed Laplacian bands are reconstructed to form the output video.

The magnification factor α is suitably attenuated with respect to a spatial cut-off frequency (λ_c), so as to reduce α for bands of higher frequencies. This minimizes the artifacts in the resultant video. It must be noted that the effect

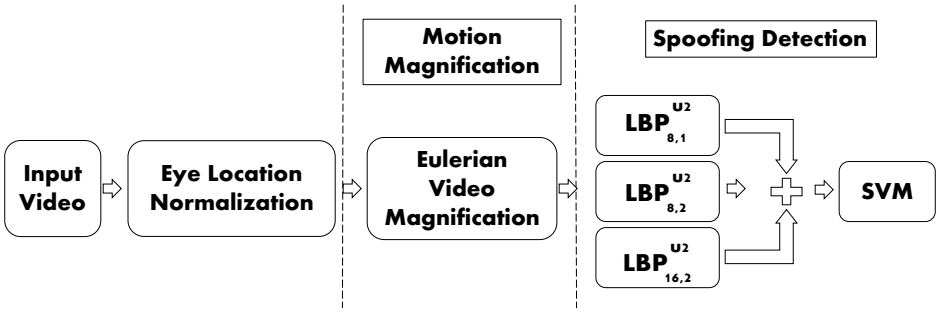

Figure 3. Illustrating the proposed texture based spoofing detection approach with motion magnification.

of magnification is dependent on the filter and the magnification factor α used. An optimal value of α is chosen by visual inspection of processed videos from the training set. The approach enhances facial movements including subtle motion such as blinking, saccadic and conjugate eye motion that may otherwise only be visible on close inspection of the video. It is our assertion that the enhanced motion may provide improved evidence of liveness of face video.

2.2. Feature Extraction

Motion magnified video of a subject can be classified for spoofing detection using either texture or motion based features. As mentioned, texture features are widely explored in spoofing detection literature as compared to motion based features. In this research, we propose the following texture and motion based features for spoofing detection.

2.2.1 Multiscale LBP

Inspired from various texture based spoofing detection approaches [1, 2, 4, 8, 10, 11, 13], we explore the utility of LBP based features along with motion magnification. In literature it has been known that feature level concatenation of global LBP features are efficient for spoofing detection. It is our assertion that after motion magnification, comparatively coarser texture features should suffice for spoofing detection. To encode texture information at multiple scales, we propose to use feature concatenation of the three LBP configurations ($\text{LBP}^{u2}_{8,1}$, $\text{LBP}^{u2}_{8,2}$, and $\text{LBP}^{u2}_{16,2}$), as shown in Figure 3. For classification, as used in existing literature, SVM with Radial Basis Function (RBF) kernel is used. As opposed to [10], that computes overlapping local histograms of $\text{LBP}^{u2}_{8,1}$, resulting in the overall feature of size 833; we only compute global histograms at all three scales, resulting in a descriptor of size 361 (i.e. 59+59+243).

2.2.2 Histogram of Oriented Optical Flows

Since subtle facial motion is partially involuntary, it is our assertion that motion estimation using optical flow may aid as an anti-spoofing feature. Optical flow has previously been used in several applications including identification of facial micro-expressions in videos [15]. Optical flow is a dense motion estimation technique that computes the motion of each pixel by solving the optimization problem shown in Eq. 2.

$$\frac{\partial I}{\partial x}V_x + \frac{\partial I}{\partial y}V_y + \frac{\partial I}{\partial t} = 0 \quad (2)$$

The flow in horizontal (V_x) and vertical directions (V_y) are used to compute the orientation based flow vector. In this research, conjugate gradient approach [9] is used to solve the optimization problem due to its low computational complexity. However, raw optical flow per pixel may be too spatially constrained, and may encode redundant background or unwanted motion. As illustrated in Figure 4, the flow vectors are computed and pooled over local block regions. Specifically, optical flow from the face region is computed between frames at a fixed interval (k). A histogram of the optical flow orientation angle, weighted by the magnitude is computed over local blocks and concatenated to form a single vector. The vector thus obtained is termed as the Histogram of Oriented Optical Flows (HOOF) [3]. Further, a final high dimensional feature vector is obtained by concatenating all the sampled frames as shown in Eq. 3.

$$H_I = [HOOF(I_t, I_{t+k})\ HOOF(I_{t+k+1}, I_{t+2k})...] \quad (3)$$

The length of vector H_I is a function of the sampling interval k. The vectors of equal length are computed by appropriately truncating the input video. Finally, dimensionality reduction using PCA, at 95% Eigen energy, is applied to reduce the dimensionality of H_I. For classification, a two-class LDA is used to obtain a uni-dimensional projection of the reduced feature vector.

3. Experimental Evaluation

A spoofing detection technique must be robust across different types of attacks. Therefore, the experiments are performed on two publicly availably databases, namely (1) Print Attack database [1] and (2) Replay Attack database [4]. Both the databases are associated with a fixed experimental protocol. The details are described as follows.

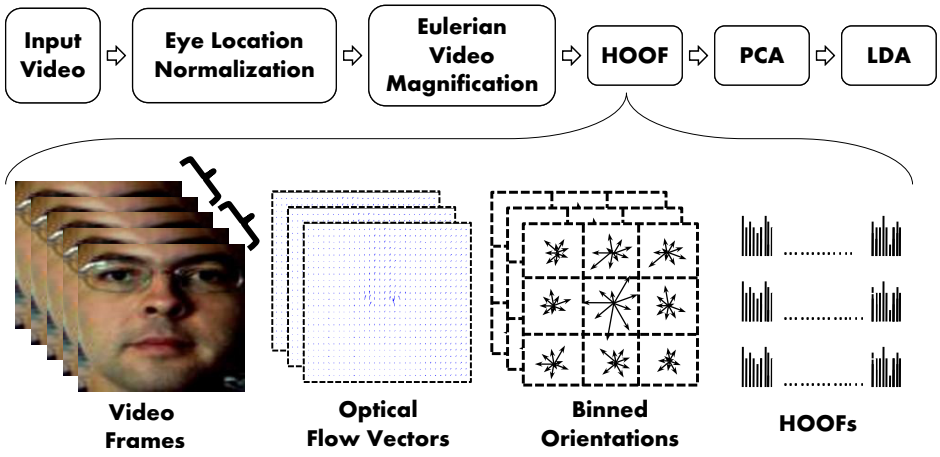

Figure 4. An illustration of the proposed approach with motion magnification and HOOF descriptor. HOOF descriptors obtained between pairs of frames at a fixed interval are concatenated to create a single feature vector.

3.1. Databases

The Print Attack database [1] consists of 200 real access and 200 printed-photo attack attempt videos of 50 subjects. Further, the dataset consists of training (120 videos), development (120 videos), and testing (160 videos) subgroups. The training and the 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 database [4] consists of 1200 videos that include 200 real access videos, 200 print attack videos, 400 phone attack videos, and 400 tablet attack videos. The dataset consists of training (360 videos), development (360 videos), and testing (480 videos) subgroups. The training and the development subgroups contain 60 real access videos and 300 attack videos each, whereas the testing subgroup contains 80 real access and 400 attack videos.

In the experiments on both Print Attack [1] and Replay Attack [4] databases, the standard predefined experimental protocols are followed, i.e., classifier model is learned on the training set, and the development set is used for parameter tuning. As shown in Figure 5, both the datasets are first pre-processed by cropping the face region based on eye coordinates obtained from a commercial face recognition system. In order to correct for small inconsistencies in eye detection, global image registration [14] is applied with the first frame as reference. This process effectively minimizes the motion in videos that are not facial motion. The normalization process is used in conjunction with existing literature and may also help reduce the effect of hand motion in spoof attacks. Not all videos contain the same number of frames, therefore only the first 230 frames from the videos are used. To ensure fair comparison across different spoofing detection algorithms, the same pre-processed images (eye-detection and frame registration) are used in all experiments.

3.2. Results and Analysis

To demonstrate the effectiveness of the proposed framework, the results of both texture and motion based feature extraction approaches are computed with and without motion magnified videos. The experimental results in Table 1 present the performance of spoofing detection in terms of HTER (%). For motion magnification, optimal parameters are selected as $\alpha = 50, \lambda_c = 10$, and an ideal bandpass filter with band $100 - 120$ Hz is used. Note that for computing texture based features, all the frames are first converted to gray scale. The parameter estimation for SVM is performed using grid search where the objective of grid search is defined in terms of optimizing the equal error rate on the development set.

For the proposed motion estimation approach, HOOF feature vectors of the test set (after dimensionality reduction) are projected using LDA to a single dimension and the classification is performed using (1) thresholding and (2) nearest neighbor. In the experiment HOOF + LDA (thresholding), a threshold is computed from the development set

Figure 5. The input video is normalized by first cropping and registering each frame, with the first frame as reference.

Table 1. Performance of various approaches in terms of HTER in percentage. ∗Result as reported in citation (under the same experimental protocol).

Approach	Print Attack				Replay Attack			
	Normal		Magnified		Normal		Magnified	
	Dev	Test	Dev	Test	Dev	Test	Dev	Test
$LBP^{u2}_{8,1}$+SVM [4]∗	-	-	-	-	14.84	15.16	-	-
$LBP^{u2}_{8,1}$+SVM (RBF)	5.00	3.12	1.66	0.63	10.00	14.87	6.60	10.20
$LBP^{u2}_{8,2}$+SVM (RBF)	5.00	2.50	1.66	1.88	11.66	14.37	6.66	**6.62**
$LBP^{u2}_{16,2}$+SVM (RBF)	5.00	3.12	1.66	1.87	8.50	12.87	6.50	8.75
$LBP^{u2}_{8,1}$+$LBP^{u2}_{8,2}$+$LBP^{u2}_{16,2}$ + SVM (RBF)	3.33	5.60	1.66	1.25	8.55	11.75	5.16	**6.62**
$LBP\text{-}TOP_{8,8,8,1,1,[1-2]}$+SVM [13]∗	-	-	-	-	7.88	7.60	-	-
HOOF+LDA (thresholding)	0.00	3.13	0.00	3.13	0.00	**3.75**	0.00	**4.38**
HOOF+LDA (NN)	0.00	0.62	0.00	**0.00**	0.00	**1.25**	0.00	**1.25**

for classification. On the other hand, for HOOF + LDA (NN), a nearest-neighbor is computed from the training set. Table 1 shows the results of texture and motion based approaches on both the datasets. The inferences drawn from the experimental results are as follows:

- The results show that only using the texture based multiscale LBP with SVM classification yields 5.6% HTER on the Print Attack database and 11.75% on the Replay Attack database. However, the HTER improves to 1.25% and 6.62% when the videos are preprocessed using the proposed motion magnification technique. These results are better than the state-of-the-art of 7.6% on the Replay Attack database provided by LBP-TOP [13]. The results of individual components of multiscale LBP also show that motion magnification improves the performance of individual components significantly. The improvement may be attributed to the exaggeration of *liveness* features of the face, such as blinking, twitching, and saccadic movement of eyes. Since motion magnification approach enhances the changes in intensity values in video, it may also enrich the texture of the magnified video.

- The results of HOOF feature extraction which is based on the motion estimation approach show that HOOF+LDA (NN) provides near-perfect classification performance on both the datasets (0% and 1.25% HTER on Print Attack and Replay Attack respectively), thereby enhancing the state-of-the-art by 6.35% on the Replay Attack. The distributions of uni-dimensional LDA projections of HOOF descriptors shown in Figure 6 indicate an improved separation between real and attack classes when using motion magnification. This illustrates that HOOF descriptor is able to correctly encode real facial movements.

- In the proposed HOOF+LDA (NN), only two samples from the test set of Replay Attack database are wrongly classified (real as attack). Further analysis of the two misclassifications revealed that both samples are incorrectly registered which may have contributed to the misclassification.

Table 2. Time taken in execution of various stages for one video (375 frames).

Stage	Time (seconds)
Registration	293.8
Motion magnification	28.4
HOOF feature extraction	15.2
$LBP^{u2}_{8,1}$+$LBP^{u2}_{8,2}$+$LBP^{u2}_{16,2}$ feature extraction	14.3
$LBPTOP_{8,8,8,1,1,[1-2]}$ feature extraction	734.0

- As shown in Table 2, the MATLAB implementations of the proposed techniques are computationally efficient. The execution time for complete pipeline (preprocessing, motion magnification and HOOF feature extraction) takes less time than LBP-TOP features extraction alone. It shows that the proposed approach outperforms existing approaches in terms of accuracy as well as computational time.

- The performance of our implementation of $LBP^{u2}_{8,1}$+SVM (RBF) outperforms the reported results in [4]. This may be attributed to the preprocessing stage (better eye-detection and frame registration).

4. Conclusion and Future Directions

Anti-spoofing in face recognition systems must quickly mature to provide a robust and computationally efficient solution to improve the practicality of face biometrics. This research presents a novel framework for spoofing detection

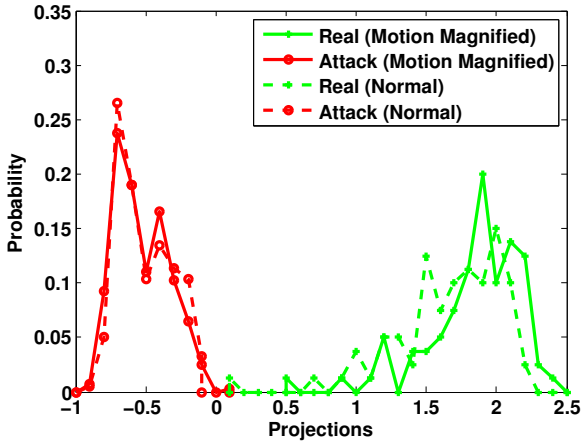

Figure 6. Histogram of the probe projections of HOOF descriptor using LDA.

in face recognition systems. 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. Further, we present a motion estimation based technique using optical flow descriptor (HOOF). The HOOF descriptors obtained from motion magnified videos provide state-of-the-art performance on the Print Attack and Replay Attack datasets in terms of accuracy and computational efficiency. We are currently improving the approach in more challenging and adversarial conditions using a combination of motion and texture based techniques.

References

[1] A. Anjos and S. Marcel. Counter-measures to photo attacks in face recognition: a public database and a baseline. In *International Joint Conference on Biometrics*, 2011. 1, 3, 4

[2] 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 *International Joint Conference on Biometrics*, pages 1–6, 2011. 1, 3

[3] 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 *Conference on Computer Vision and Pattern Recognition*, pages 1932–1939, 2009. 2, 3

[4] I. Chingovska, A. Anjos, and S. Marcel. On the effectiveness of local binary patterns in face anti-spoofing. In *International Conference of the Biometrics Special Interest Group*, pages 1–7, 2012. 1, 2, 3, 4, 5

[5] M. De Marsico, M. Nappi, D. Riccio, and J. Dugelay. Moving face spoofing detection via 3D projective invariants. In *International Conference on Biometrics*, pages 73–78, 2012. 2

[6] R. D. Findling and R. Mayrhofer. Towards face unlock: on the difficulty of reliably detecting faces on mobile phones. In *International Conference on Advances in Mobile Computing & Multimedia*, pages 275–280, 2012. 1

[7] G. Kim, S. Eum, J. K. Suhr, D. I. Kim, K. R. Park, and J. Kim. Face liveness detection based on texture and frequency analyses. In *International Conference on Biometrics*, pages 67–72, 2012. 2

[8] J. Komulainen, A. Hadid, and M. Pietikainen. Face spoofing detection using dynamic texture. In *ACCV Workshop on Computer Vision with Local Binary Pattern Variants*, 2012. 3

[9] C. Liu. *Beyond Pixels: Exploring New Representations and Applications for Motion Analysis*. PhD thesis, Massachusetts Institute of Technology, 2009. 3

[10] J. Määttä, A. Hadid, and M. Pietikainen. Face spoofing detection from single images using micro-texture analysis. In *International Joint Conference on Biometrics*, pages 1–7, 2011. 1, 3

[11] J. Määttä, A. Hadid, and M. Pietikainen. Face spoofing detection from single images using texture and local shape analysis. *IET Biometrics*, 1(1):3–10, 2012. 1, 3

[12] G. Pan, L. Sun, Z. Wu, and S. Lao. Eyeblink-based anti-spoofing in face recognition from a generic webcamera. In *International Conference on Computer Vision*, pages 1–8, 2007. 2

[13] T. d. F. Pereira, A. Anjos, J. M. De Martino, and S. Marcel. LBP-TOP based countermeasure against facial spoofing attacks. In *ACCV Workshop on Computer Vision With Local Binary Pattern Variants*, 2012. 2, 3, 5

[14] S. Periaswamy and H. Farid. Elastic registration in the presence of intensity variations. *IEEE Transactions on Medical Imaging*, 22(7):865–874, 2003. 4

[15] M. Shreve, S. Godavarthy, D. Goldgof, and S. Sarkar. Macro- and micro-expression spotting in long videos using spatio-temporal strain. In *International Conference on Automatic Face Gesture Recognition and Workshops*, pages 51–56, 2011. 3

[16] X. Tan, Y. Li, J. Liu, and L. Jiang. Face liveness detection from a single image with sparse low rank bilinear discriminative model. In K. Daniilidis, P. Maragos, and N. Paragios, editors, *European Conference on Computer Vision*, volume 6316 of *Lecture Notes in Computer Science*, pages 504–517. Springer Berlin Heidelberg, 2010. 1

[17] 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*, 31(4), 2012. 2