Face Presentation Attack with Latex Masks in Multispectral Videos

Akshay Agarwal*, Daksha Yadav+, Naman Kohli+, Richa Singh*, Mayank Vatsa*, Afzel Noore+
*IIIT-Delhi, +West Virginia University
*{akshaya, rsingh, mayank}@iiitd.ac.in,
+{daksha.yadav, naman.kohli, afzel.noore}@mail.wvu.edu

Abstract

*Face recognition systems are susceptible to presentation attacks such as printed photo attacks, replay attacks, and 3D mask attacks. These attacks, primarily studied in visible spectrum, aim to obfuscate or impersonate a person's identity. This paper presents a unique multispectral video face database for face presentation attack using latex and paper masks. The proposed **M**ultispectral **L**atex Mask based Video **F**ace **P**resentation Attack (MLFP) database contains 1350 videos in visible, near infrared, and thermal spectrums. Since the database consists of videos of subjects without any mask as well as wearing ten different masks, the effect of identity concealment is analyzed in each spectrum using face recognition algorithms. We also present the performance of existing presentation attack detection algorithms on the proposed MLFP database. It is observed that the thermal imaging spectrum is most effective in detecting face presentation attacks.*

1. Introduction

The ability to correctly identify an individual based on their facial features has led to its usage in diverse applications ranging from border control to banking. Over the last decade, face recognition has gained widespread attention due to extensive usage in mobile phone authentication and surveillance applications. The advent of deep learning algorithms has further increased the performance of face recognition systems, which in turn, has led to its increased usage in commercial applications and access control environments.

With increasing usage of face authentication systems, *presentation attacks* are becoming a serious point of concern, particularly for unmanned applications such as ATM machines. As faces are easy to acquire without the subject's consent or awareness, impersonating someone's identity is easy [14]. Similarly, identity hiding is also less challenging with the usage of 3D hard masks [8] or more sophisticated silicone masks [13]. In 2016, ISO/IEC 30107 standards for

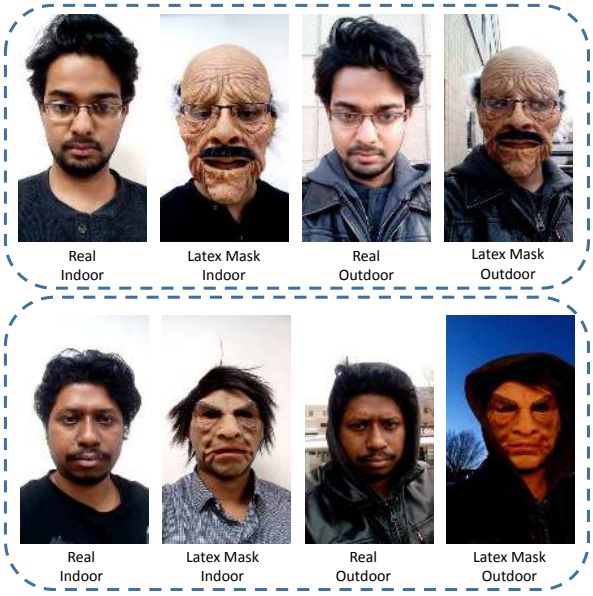


Figure 1: Showcasing the problem of unconstrained face presentation attack using 3D latex masks.

presentation attack detection (PAD) [10] also highlighted this aspect:

"*Since the beginning of these technologies, the possibility of subversion of recognition by determined adversaries has been widely acknowledged, as has the need for countermeasures to detect and defeat subversive recognition attempts, or presentation attacks.*"

Face presentation attacks, particularly with 3D masks, can be classified into two broad categories: (i) evasion, where the key aim is face obfuscation or identity concealment and (ii) impersonation, where the objective is to impersonate the identity of another individual. Most of the existing databases used in PAD are prepared in visible spectrum under controlled environment, except a few recent databases. Table 1 presents a summary of recent 3D mask databases used in face PAD literature along with their characteristics.

Table 1: Existing datasets available in the literature for face presentation attack detection. * denotes the number of subjects utilized for the presentation attack detection experiments.

Database (Year)	**No. of Subjects**	**No. of Samples**	**2D Masks**	**3D Masks (Hard/Latex/ Silicone)**	**Varying Acquisition Environment**	**Videos**	**Multispectral (Wavelength in nm)**
3DMAD (2013) [8]	17	255	×	Hard	×	✓	×
3D Mask (2016) [12]	12	1,008	×	Hard	✓	✓	×
msspoof (2016) [4]	21	4,494	× (Print)	×	✓	×	✓(VIS and NIR)
SWIR (2016) [21]	5*	141	× (Print)	Silicone and Latex	×	×	✓(935, 1060, 1300, and 1550)
EMSPAD (2017) [19]	50	14,000	× (Print)	×	✓	×	✓(7 bands: 425-930)
SMAD (2017) [13]	–	130	×	Silicone	✓	✓	×
Proposed MLFP (2017)	10	1,350	✓	Latex	✓	✓	✓(VIS, NIR, and Thermal)

In the literature, thermal imagery has shown promising results for detecting disguise variations [7], [18]. Further, Pavlidis and Symosek [18] suggest that near-infrared spectrum can also help in detecting makeup and disguise accessories. Inspired from these existing research on facial disguise, it is our hypothesis that near-infrared and thermal spectrum can help in detecting 3D mask based face presentation attacks. In order to understand the impact of *spectrums* in detecting mask based presentation attacks, we prepare a novel database, named as **M**ultispectral **L**atex Mask based Video **F**ace **P**representation Attack (MLFP) database in this research. Samples of the database captured in different scenarios are shown in Figure 1. To the best of our knowledge, the proposed database is the first public face presentation attack database which contains videos in three different spectrums: visible (VIS), near-infrared (NIR), and thermal. The database contains 1,350 videos with presentation attacks using latex and paper masks (evasion/obfuscation scenarios) in both indoor and outdoor environments. We investigate the efficacy of existing PAD approaches in varying spectrums. The database will be made publicly available to the research community[1].

Next, we present the details of the MLFP database followed by the experimental analysis of mask based obfuscation on face recognition algorithms in Section 3. Section 4 presents the baseline results of existing presentation attack detection algorithms followed by concluding remarks.

[1] http://iab-rubric.org/resources.html

Table 2: Characteristics of the Multispectral Latex Mask based Video Face Presentation Attack (MLFP) database.

No. of Spectrums	3 (VIS, NIR, & Thermal)
Types of Masks	Latex (7) and Paper (3)
No. of Videos	1,350
Types of Videos	Real (150) and Attack (1,200)
No. of Subjects	10 (4 Females, 6 Males)
Video Duration	Minimum 10 seconds and Maximum 15 seconds
Environmental Variations	Indoor and Outdoor with Fixed and Random Backgrounds

2. Multispectral Latex Mask based Video Face Presentation Attack Database

The proposed MLFP database consists of 1,350 videos of 10 subjects with and without wearing the face masks. Videos are captured at different locations (indoor and outdoor) in unconstrained environment. Out of 1,350 videos in the database, 1,200 videos are with the subjects wearing mask (*attack videos*) and the remaining 150 videos are without the mask (*real videos*). The database consists of 10 subjects (4 females and 6 males) between the age of 23-38 years. Minimum video duration in the database is 10 seconds. The database contains more than 200,000 frames. It has been collected over a period of 3 months with the environmental temperature ranging from -15 °C to 15 °C. Figure 2 shows sample images and Table 2 summarizes the characteristics of the proposed MLFP database.

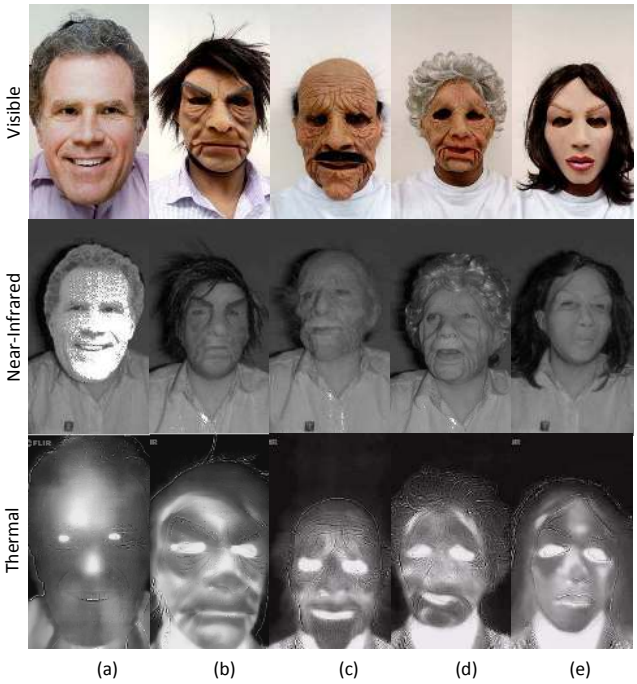


Figure 2: Sample images from the proposed MLFP database showcasing the masks used for obfuscation and its effects in the three spectrums. (a) is a 2D paper mask while masks (b)-(e) are 3D latex masks.

In the MLFP database, two types of face masks are utilized:

- 3D Latex Masks: Seven different latex masks are used for database collection. The masks are soft and hence, they conform to the subject's face shape. These masks allow life-like movement of mouth as well as face. While six masks cover the entire face, the seventh mask is a half-mask which covers the face region below the eyes.
- 2D Paper Masks: Three paper masks are utilized with cutouts for eyes. These are created using high resolution images on high quality card paper.

The 2D paper masks are smooth and reflect more light as compared to 3D latex masks. On the other hand, 3D latex masks have textural features such as wrinkles and in some cases, facial hair. Both latex and 2D paper masks have variations in terms of gender and age.

For each subject, videos are collected in two different acquisition environments: indoor and outdoor as well as two different backgrounds: fixed and random in visible and thermal spectrums. Due to the inability of existing systems to acquire good quality data in NIR spectrum in daylight, NIR data is collected in indoor environment only. Figure 3 shows

Figure 3: Illustrating low quality of the data captured in NIR spectrum in daylight. Image on the left is a good quality image captured indoors and image on the right is captured in daytime outdoor environment.

a sample image captured in NIR in daylight. In all the above mentioned conditions, data is collected from both real and mask attack videos with all ten masks.

2.1. Database Acquisition

Figure 4 shows the database collection process in which three devices are utilized for acquisition under different spectrums:

- Visible spectrum videos are collected using Android smartphones with 8 megapixels camera at frame resolution of $1,280 \times 720$ pixels.
- FLIR ONE thermal imager for Android[2] is used for thermal data collection. This imager has operating temperature range of 32°F to 95°F and can be used in conjunction with an Android phone to capture images and videos in thermal spectrum at 640×480 video resolution.
- Videos in NIR spectrum are captured using Microsoft Kinect for Windows V2[3] with output video resolution of 424×512.

2.2. Comparison with Existing Databases

As shown in Table 1, existing publicly available databases for face presentation attack are classified and compared with the proposed MLFP database based on the following categories:

Single Spectrum vs Multispectral: Existing 3D mask databases such as 3D Mask Attack Database (3DMAD) [8] and Silicone Mask Attack Database (SMAD) [13] are captured only in a single spectrum. However, due to the widespread usage of multispectral face recognition systems,

[2] http://www.flir.com/flirone/products/?id=81752
[3] http://www.xbox.com/en-US/xbox-one/accessories/kinect

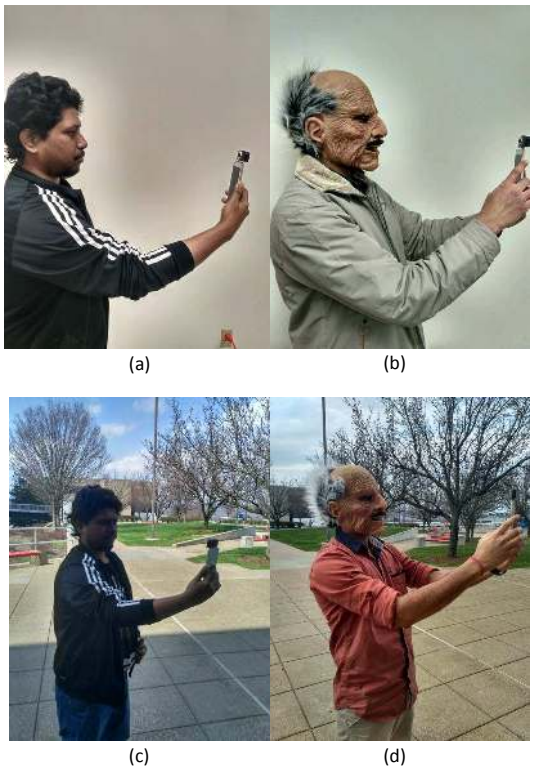


Figure 4: Illustration of data acquisition process in terms of varying environment where (a) real video data collection with fixed background indoors, (b) mask attack video data collection with fixed background indoors, (c) real video data collection with random background outdoors, and (d) mask attack video data collection with random background outdoors.

it is crucial to investigate the effect of multispectral face presentation attacks on recognition systems and develop multispectral face PAD algorithms. Recently, some databases have been proposed with multispectral imaging data acquisition [4, 21]. Compared to existing databases, the proposed MLFP database is the only database containing videos captured in three different imaging spectrums (VIS, NIR, and thermal).

Hard Masks vs Soft Masks: Most of the publicly available 3D mask databases [8, 12] are captured by wearing the hard resin mask which are easier to detect as compared to latex or silicone mask. Latex masks are comparatively softer and hence, can adjust to different facial sizes, shapes, and movements. The proposed MLFP database has been created by utilizing soft latex masks which makes the problem of face presentation attack more challenging.

Images vs Videos: Existing multispectral databases such as Multispectral Spoof Database (msspoof) [4] and Extended Multispectral Presentation Attack Database (EMSPAD) [19] consist of print attacks. Several algorithms have been proposed for face PAD where such attacks are easily detected [17, 22, 23]. Also, in many instances, the primary acquisition mode of face recognition pipeline is videos. Due to this, the proposed MLFP database comprises videos of face presentation attacks with minimum video duration of 10 seconds.

Constrained vs Unconstrained: Existing face presentation attack databases such as Short Wave Infrared (SWIR) [21] and 3DMAD [8] have been captured in constrained indoor environment. To simulate real world scenarios with unrestricted backgrounds and illumination, the proposed MLFP database has been acquired both indoors and outdoors with fixed as well as random backgrounds.

3. Effect of Face Obfuscation on Face Verification Algorithms across Multiple Spectrums

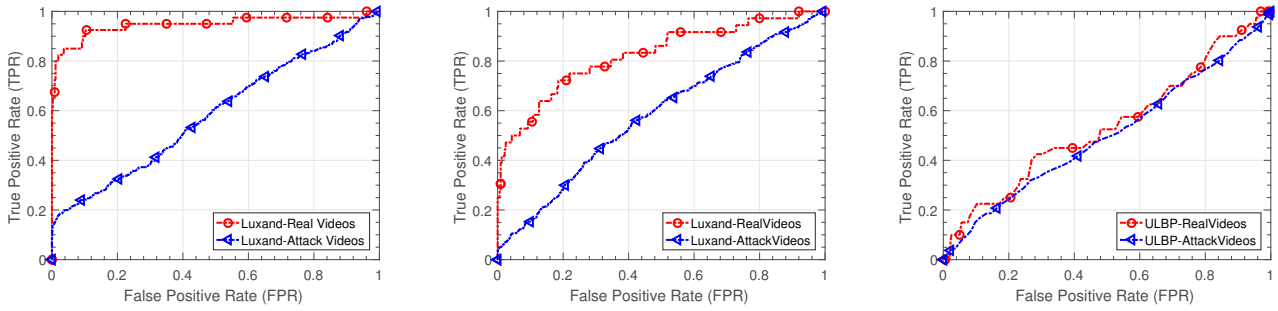
In this section, we investigate the effectiveness of presentation attacks for the purpose of identity evasion. Face obfuscation refers to concealment of one's identity to avoid recognition by wearing a mask or any other mechanism. Next, we examine the performance of face recognition algorithms to determine the effectiveness of these masks in obfuscating one's own identity.

3.1. Face Verification Algorithms for Evaluation

For VIS and NIR spectrums, a commercial off-the-shelf system (COTS), Luxand [1] is utilized for face recognition. To the best of our knowledge, there is no commercial face recognition system for thermal spectrum. Therefore, in place of COTS, Uniform Local Binary Patterns (ULBP) [15] is utilized for the thermal spectrum. ULBP features are extracted separately from the gallery and probe videos of the thermal spectrum and matching is performed by computing χ^2 distance.

3.2. Experimental Protocol

The proposed MLFP database consists of videos of each subject wearing all the ten masks. In order to analyze how effectively subjects are able to obfuscate their identity, a face verification experiment is designed. The gallery of the experiment is fixed, containing 10 real videos (one for each subject). The real enrollment video is captured in well-illuminated indoor environment for the three spectrums. Two probe sets are created; the first probe set, called *Real Videos*, is created with 40 real videos (four for each subject). Using this probe set, the goal is to illustrate the performance of face verification under regular/normal conditions. The second probe set, called *Attack Videos*, comprises of 400 presentation attack videos (40 of each subject). This probe set is designed to examine if an individual



(a) COTS in VIS (b) COTS in NIR (c) ULBP in Thermal

Figure 5: Illustrating the effect of presentation attacks on existing face recognition systems in the three spectrums.

is able to conceal his/her identity using face masks. After fixing the gallery and probe partitions, video-to-video face recognition experiment is performed.

3.3. Experimental Results

The results of the face obfuscation experiment are presented using Receiver Operating Characteristic (ROC) curves (Figure 5) and Equal Error Rate (EER). EER is defined as the threshold point where false positive rate is equal to false reject rate. It is observed that the performance of the COTS is significantly high in the visible spectrum for Real Videos with an EER of 10.0%. However, for presentation attack videos, the face recognition performance dramatically decreases with an EER of 44.9%. This decline in the accuracy of face recognition indicates that the individuals in the probe videos are able to effectively conceal their identity when they are wearing masks. A similar trend is observed in the NIR spectrum where COTS yields an EER of 26.7% with real videos and 44.5% for attack videos. Likewise, for ULBP in thermal spectrum, an EER of 48% is observed for real videos probe which increases to 51% for attack video probe. This experiment demonstrates that the problem of face presentation attack influences all the imaging spectrums and establishes the requirement of efficient PAD algorithms to detect these attacks.

4. Evaluating Face Presentation Attack Detection Algorithms across Multiple Spectrums

The previous section demonstrates the effect of face obfuscation performed by wearing different masks on the performance of commercial facial recognition systems. In this section, we analyze the performance of existing face PAD algorithms to classify an input video as real or attacked.

4.1. Experimental Setup

This experiment is performed on the proposed MLFP database to evaluate the efficacy of baseline face PAD algorithms. As the database consists of all the subjects wearing all the masks, the following partitions have been created to ensure balanced unseen split of data in testing and training partitions as well as evaluating the generalizability of the algorithm with respect to the masks:

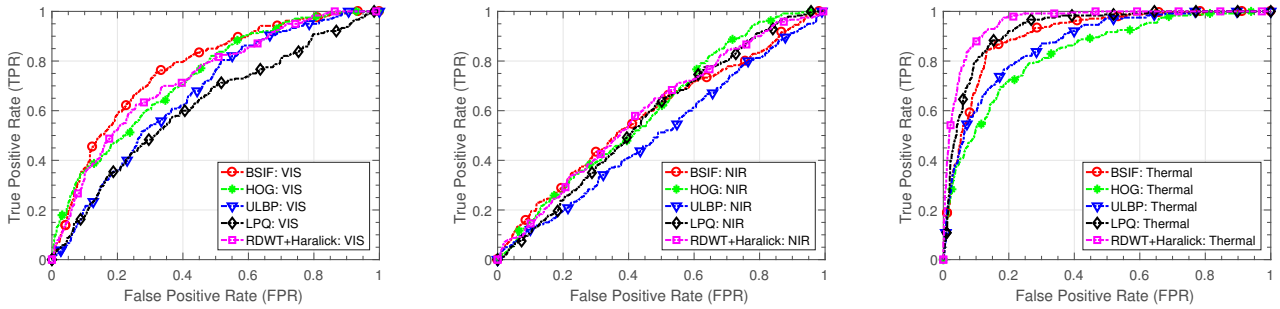
- **Partition based on Subjects:** While analyzing the PAD performance, the subjects are divided into three disjoint partitions containing three, three, and four randomly chosen subjects. For each cross-validation fold, one partition is chosen for training while the remaining folds are assigned to the testing set for evaluating the algorithm. Subjects in the training and testing sets are non-overlapping so that the PAD algorithm does not encode any subject-specific information which is essential for real-world scenarios.

- **Partition based on Masks:** As described earlier, MLFP database has been created by utilizing three paper and seven latex masks. The performance is evaluated on unseen masks by ensuring that the masks which are utilized for training are not used in testing. The masks are divided into three partitions where each partition contains randomly chosen one paper mask and two latex masks in the training fold while the remaining unseen masks are utilized for testing. The tenth mask which is unique as it is a half mask (containing both real and mask areas), is always utilized in the testing set.

4.2. Algorithms Utilized for Face Presentation Attack Detection

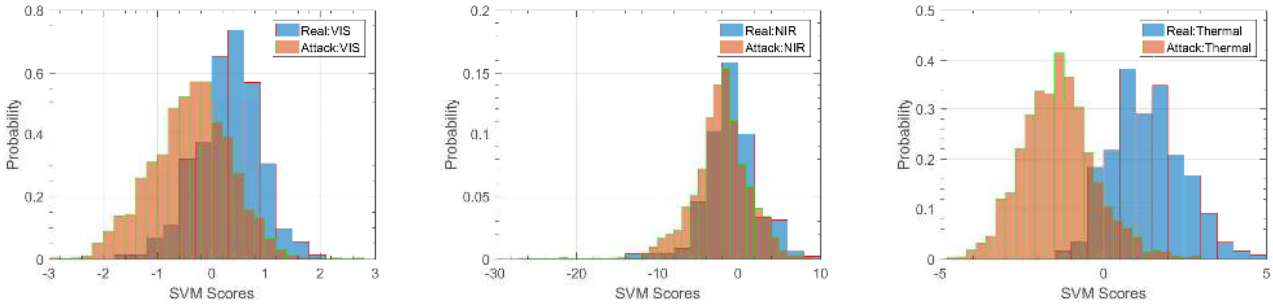
For experimental evaluation, the following five existing algorithms are utilized:

1. RDWT+Haralick Features [2]: The algorithm proposed by Agarwal et al. [2] utilizes a combination of Redundant Discrete Wavelet Transform (RDWT) and Haralick features. In recent literature, this approach



(a) Video-based PAD in VIS (b) Video-based PAD in NIR (c) Video-based PAD in Thermal

Figure 6: ROC curves showing the performance of video-based face presentation attack detection (PAD) algorithms on the proposed MLFP database.



(a) BSIF scores in VIS (b) RDWT+Haralick scores in NIR (c) RDWT+Haralick scores in Thermal

Figure 7: SVM score distribution of real and attack videos of best performing presentation attack detection algorithms in the three spectrums.

has shown very high performance on non-mask based presentation attack databases.

2. Binarized Statistical Image Features (BSIF) [11]: BSIF are calculated by learning a subspace from independent component analysis of natural images in a patch-wise manner. This is followed by binarization of the coordinates in this new basis by thresholding to obtain the final binary codes for each pixel. In this research 8bit filter of size 11×11 is used.

3. Uniform Local Binary Patterns (ULBP) [15]: Traditional textural feature descriptor ULBP is utilized to encode the texture variations between real samples and attacked samples.

4. Local Phase Quantization (LPQ) [16]: LPQ is computed by utilizing short-term Fourier transform on local windows which is followed by computing local Fourier coefficients at four frequency points. These coefficients are quantized for producing the final phase information. ULBP, LPQ, and BSIF have been utilized by Boulkenafet et al. [3] for color texture analysis based face PAD.

5. Histogram of Oriented Gradients (HOG) [6]: HOG features are computed by aggregating edge gradients across local cells of an image. HOG is selected for evaluation purposes as the edge information in masks can be encoded by gradients in this feature descriptor.

Feng et al. [9] and Siddiqui et al. [20] have demonstrated that applying PAD algorithm on the input face images yield better results than detected face images only. Therefore in this research, the features are computed on the input image without performing face detection. This also reduces the computation time which is important for face PAD as it is the *preprocessing* step in the face recognition pipeline. Each feature is computed by dividing an input frame into 4×4 blocks. The extracted features are used in conjunction with linear Support Vector Machine (SVM) [5] for real vs attack classification.

4.3. Experimental Results

In this section, we investigate the effect of multispectral PAD on videos and individual frames. Frame-based attack detection results are computed by dividing the input videos into its constituent frames and applying face PAD algorithm

Table 3: EER (%) of video-based face presentation attack detection on the proposed MLFP database.

Features	VIS	NIR	Thermal
HOG [6]	34.9	45.5	24.5
ULBP [15]	38.7	49.0	21.3
BSIF [11]	**29.2**	43.3	15.0
LPQ [16]	40.0	43.8	13.7
RDWT + Haralick [2]	32.9	**42.0**	**10.8**

Table 4: EER (%) of frame-based face presentation attack detection on the proposed MLFP database.

Features	VIS	NIR	Thermal
HOG [6]	36.9	45.7	29.6
ULBP [15]	40.2	49.5	28.2
BSIF [11]	**32.5**	46.7	25.0
LPQ [16]	43.7	45.5	22.1
RDWT + Haralick [2]	36.2	**44.4**	**15.4**

on frame-wise basis. The PAD problem is a binary classification problem where a frame/video is classified as real or attack. For computing video-based results, the frame-based scores are combined by computing the average of all frames to produce the final video classification result.

4.3.1 Video-based Face Presentation Attack Detection

The results of video-based face PAD algorithms are summarized in Table 3 and Figure 6.

PAD performance in VIS spectrum: We observe that BSIF features yield the best face PAD performance and it outperforms the other approaches by 3-10% in terms of EER. The performance of BSIF can be attributed to its ability to encode micro-level variations in smoothness and texture between real and attack samples in the visible spectrum.

PAD performance in NIR and thermal spectrums: As seen in Table 3 and Figure 6, RDWT+Haralick [2] framework yields the best performance for video-based face PAD in NIR and thermal spectrum. This result indicates that the textural changes occurring when an individual is wearing a mask has been effectively captured by the combination of RDWT and Haralick features.

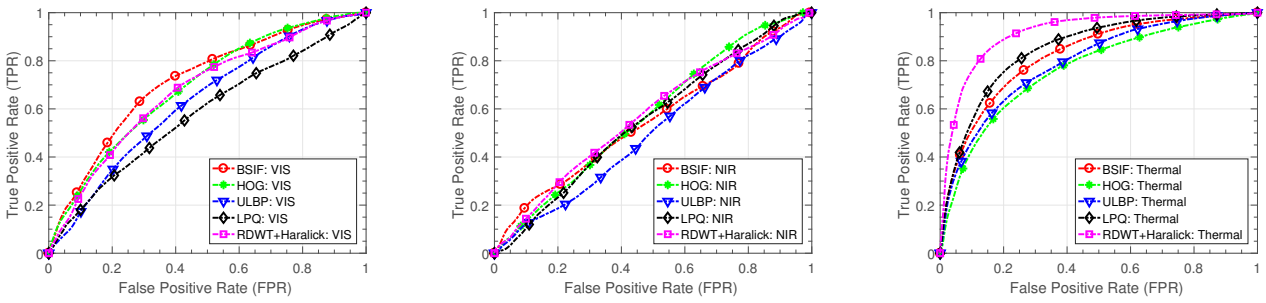
Comparing PAD performance across multiple spectrums: Upon comparing the effectiveness of different spectrums for face PAD, it is observed that thermal spectrum demonstrates the minimum EER of 10.8%. It is to be noted that this EER is at least 19-32% better as compared to NIR and VIS spectrums. The match score distribution plots shown in Figure 7 also support this observation. Upon analyzing real and attack video classification scores of thermal spectrum (Figure 7c), it is seen that there is less overlap between the scores of the two classes as compared to VIS and NIR. Therefore, it can be inferred that the heat patterns in thermal spectrum are more discriminatory with respect to real vs attack video classification. Furthermore, analysis of Figure 7b reveals a notable overlap between real and attack scores for the NIR spectrum which is also reflected in the highest EER of all the features in the NIR spectrum. This result accentuates the need for developing an accurate face PAD algorithm for NIR spectrum due to increasing popularity of NIR-based face recognition systems.

Real vs Attack classification analysis: Face PAD performance is further analyzed with respect to percentage of real and attack videos classified correctly. In VIS, the proportion of correctly classified real videos (72.9%) is higher than correctly classified attack videos (69.7%). On the contrary, in both NIR and thermal, the probability of correctly detecting attack videos is higher than real videos. In scenarios where NIR is widely used such as surveillance, misclassification of real/genuine videos may prevent genuine users to gain access. Overall, thermal spectrum gives the best performance in classifying real videos and presentation attack videos.

Analysis of PAD in varying environments: Ideally, we expect an algorithm to detect attacks accurately even in outdoor/unconstrained environments. Therefore, we also perform post-hoc analysis to examine the effect of environment (indoor and outdoor) in detection of presentation attacks. This is performed by computing the percentage of videos correctly classified in each environmental setting using the best performing features in VIS and thermal spectrums. Using BSIF in VIS spectrum, videos captured indoors are detected with 71.0% accuracy with BSIF as compared to 69.1% accuracy for outdoor videos. Similarly, using RDWT+Haralick features in thermal spectrum, indoor videos are detected with 89.9% accuracy as compared to 88.8% accuracy for outdoor videos. These results indicate that across both the spectrums, more indoor videos are classified correctly as compared to outdoor videos. This can be attributed to the presence of uncontrolled background in outdoor videos, thus, leading to more misclassifications.

4.3.2 Frame-based Face Presentation Attack Detection

Following the same protocol as video-based face PAD, performance of existing algorithms is evaluated on frame-wise basis. Thus, in frame based PAD experiment, each frame of a video is classified as real or attack. The results of frame-based PAD algorithms are summarized in Table 4 and Figure 8.



(a) Frame-based PAD in VIS (b) Frame-based PAD in NIR (c) Frame-based PAD in Thermal

Figure 8: ROC curves showing the performance of frame-based face presentation attack detection algorithms on the proposed MLFP database in the three spectrums.

Frame-wise PAD performance: Firstly, similar to video-based results, RDWT+Haralick features yield the best results for detecting face presentation attacks with an EER of 44.4% and 15.4% respectively for NIR and thermal spectrums. In NIR, RDWT+Haralick features are 1-5% better and in thermal spectrum, RDWT+Haralick features are 6-14% better than other algorithms utilized for comparison. We also observe that BSIF features produce the highest accuracy for detecting a given frame as real or attack in the visible spectrum. As video-based face detection accuracy is computed by averaging frame-based scores, it is interesting to note that the performance of frame-based face PAD is lower as compared to video-based face PAD for all the features across all the spectrums. This result highlights the importance of videos for face PAD as they are a larger source of information which turns out to be beneficial for detecting instances of face presentation attacks.

Comparing frame-wise PAD performance across multiple spectrums: It is also observed that irrespective of the features utilized, the performance computed on the frames of the thermal spectrum is better as compared to visible and NIR spectrums for face PAD. For instance, RDWT+Haralick features of thermal spectrum frames demonstrate 21-29% higher accuracy as compared to the same features in the other two spectrums. One reason can be that there is a significant variation between masks and regular faces in thermal spectrum.

5. Conclusion

Developing accurate face presentation attack detection algorithms is a critical research area due to widespread usage of face recognition systems worldwide. However, limited attention has been paid to multi-spectral video-based face PAD. To promote further research in this area, we proposed a novel database: Multispectral Latex Mask based Video Face Presentation Attack (MLFP) database. In this database, face presentation attacks are replicated by utilizing 2D paper masks and 3D latex masks in three different spectrums: visible, near-infrared, and thermal imaging. Utilizing this database, we demonstrate the effect of these presentation attacks in different spectrums on the performance of existing face recognition systems. It is observed that in all three spectrums, face recognition performance drastically declines when an attack video is presented as probe. Further, we analyze the efficacy of baseline face PAD algorithms on the proposed MLFP database. It is observed that RDWT+Haralick features in thermal spectrum demonstrate the highest performance for classifying an input video as real or attack. It is crucial to understand that there is a significant scope of improvement in detecting face presentation attacks in videos. In future, we will investigate the efficacy of textural and structural features for designing an accurate multi-spectral video-based face PAD algorithm.

6. Acknowledgement

This research is supported in part by the Ministry of Electronics and Information Technology, Government of India. Akshay Agarwal is supported through Visvesvaraya PhD Fellowship.

References

[1] Luxand. https://www.luxand.com/facesdk/. 4

[2] A. Agarwal, R. Singh, and M. Vatsa. Face anti-spoofing using Haralick features. In *IEEE International Conference on Biometrics Theory, Applications and Systems*, pages 1–6, 2016. 5, 7

[3] Z. Boulkenafet, J. Komulainen, and A. Hadid. Face spoofing detection using colour texture analysis. *IEEE Transactions on Information Forensics and Security*, 11(8):1818–1830, 2016. 6

[4] I. Chingovska, N. Erdogmus, A. Anjos, and S. Marcel. Face recognition systems under spoofing attacks. In *Face Recognition Across the Imaging Spectrum*, pages 165–194. Springer, 2016. 2, 4

[5] C. Cortes and V. Vapnik. Support-vector networks. *Machine Learning*, 20(3):273–297, 1995. 6

[6] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, pages 886–893, 2005. 6, 7

[7] T. I. Dhamecha, A. Nigam, R. Singh, and M. Vatsa. Disguise detection and face recognition in visible and thermal spectrums. In *International Conference on Biometrics*, pages 1–8, 2013. 2

[8] N. Erdogmus and S. Marcel. Spoofing in 2D face recognition with 3D masks and anti-spoofing with Kinect. In *IEEE International Conference on Biometrics: Theory, Applications and Systems*, pages 1–6, 2013. 1, 2, 3, 4

[9] L. Feng, L.-M. Po, Y. Li, X. Xu, F. Yuan, T. C.-H. Cheung, and K.-W. Cheung. Integration of image quality and motion cues for face anti-spoofing: A neural network approach. *Journal of Visual Communication and Image Representation*, 38:451–460, 2016. 6

[10] Information technology - biometric presentation attack detection. Standard ISO/IEC 30107-1:2016 - Part 1 - Framework. 1

[11] J. Kannala and E. Rahtu. BSIF: Binarized statistical image features. In *International Conference on Pattern Recognition*, pages 1363–1366, 2012. 6, 7

[12] S. Liu, B. Yang, P. C. Yuen, and G. Zhao. A 3D mask face anti-spoofing database with real world variations. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 1551–1557, 2016. 2, 4

[13] I. Manjani, S. Tariyal, M. Vatsa, R. Singh, and A. Majumdar. Detecting silicone mask-based presentation attack via deep dictionary learning. *IEEE Transactions on Information Forensics and Security*, 12(7):1713–1723, 2017. 1, 2, 3

[14] S. Marcel, M. S. Nixon, and S. Z. Li. *Handbook of Biometric Anti-Spoofing*. Springer, 2014. 1

[15] T. Ojala, M. Pietikainen, and T. Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7):971–987, 2002. 4, 6, 7

[16] V. Ojansivu and J. Heikkilä. Blur insensitive texture classification using local phase quantization. In *International Conference on Image and Signal Processing*, pages 236–243. Springer, 2008. 6, 7

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

[18] I. Pavlidis and P. Symosek. The imaging issue in an automatic face/disguise detection system. In *IEEE Workshop on Computer Vision Beyond the Visible Spectrum: Methods and Applications*, pages 15–24, 2000. 2

[19] R. Raghavendra, R. Kiran, V. Sushma, C. Faouzi, and C. Busch. On the vulnerability of extended multispectral face recognition systems towards presentation attacks. In *IEEE International Conference on Identity, Security and Behavior Analysis*, pages 1–8, 2017. 2, 4

[20] T. A. Siddiqui, S. Bharadwaj, T. I. Dhamecha, A. Agarwal, M. Vatsa, R. Singh, and N. Ratha. Face anti-spoofing with multifeature videolet aggregation. In *International Conference on Pattern Recognition*, pages 1–6, 2016. 6

[21] H. Steiner, A. Kolb, and N. Jung. Reliable face anti-spoofing using multispectral SWIR imaging. In *International Conference on Biometrics*, pages 1–8, 2016. 2, 4

[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 *European Conference on Computer Vision*, pages 504–517. Springer, 2010. 4

[23] Z. Zhang, D. Yi, Z. Lei, and S. Z. Li. Face liveness detection by learning multispectral reflectance distributions. In *IEEE International Conference on Automatic Face & Gesture Recognition and Workshops*, pages 436–441, 2011. 4