Detecting Silicone Mask-Based Presentation Attack via Deep Dictionary Learning

Ishan Manjani, Snigdha Tariyal, Mayank Vatsa, *Senior Member, IEEE*, Richa Singh, *Senior Member, IEEE*, and Angshul Majumdar, *Senior Member, IEEE*

***Abstract*— In movies, film stars portray another identity or obfuscate their identity with the help of silicone/latex masks. Such realistic masks are now easily available and are used for entertainment purposes. However, their usage in criminal activities to deceive law enforcement and automatic face recognition systems is also plausible. Therefore, it is important to guard biometrics systems against such realistic presentation attacks. This paper introduces the first-of-its-kind silicone mask attack database which contains 130 real and attacked videos to facilitate research in developing presentation attack detection algorithms for this challenging scenario. Along with silicone mask, there are several other presentation attack instruments that are explored in literature. The next contribution of this research is a novel multilevel deep dictionary learning-based presentation attack detection algorithm that can discern different kinds of attacks. An efficient greedy layer by layer training approach is formulated to learn the deep dictionaries followed by SVM to classify an input sample as genuine or attacked. Experimental are performed on the proposed SMAD database, some samples with real world silicone mask attacks, and four existing presentation attack databases, namely, replay-attack, CASIA-FASD, 3DMAD, and UVAD. The results show that the proposed algorithm yields better performance compared with state-of-the-art algorithms, in both intra-database and cross-database experiments.**

***Index Terms*— Face recognition, silicone mask, presentation attack detection, deep dictionary.**

I. Introduction

IN the Hollywood movie *Mission Impossible*, Ethan Hunt wears silicone/latex mask to impersonate someone else's identity. Similarly, as shown in Fig. 1, the images from the movie *Mrs. Doubtfire* showcase the problem of concealing one's identity using realistic masks [1]. These masks are akin to real human faces, i.e. shape, texture, and appearance of these masks are similar to a human face (Fig. 2). The ease of availability of the silicone masks (for around $500), have led people to use them for recreational purposes. However, such masks can also be used for crime and wrongdoing. As early as 2010, cases of bank robberies have been reported where robbers conceal their identities using silicone masks, thereby leading the police to search for the wrong person [2]–[6]. Fig. 3 shows two examples of actual images of subjects caught on the camera during the robberies. These instances show that silicone facial masks can be used for presentation attack [7] which includes both, concealing one's identity and impersonating someone else's identity.

Unlike already studied face presentation attack instruments such as print and replay devices [8]–[10], and hard 3D face masks [11]–[13], it is challenging to distinguish between a real

Fig. 1. Robin Williams as *Mrs. Doubtfire* showcasing the effect of realistic masks. Images are obtained from Internet.

Fig. 2. A sample of silicone mask and a person wearing it. Images are obtained from Internet.

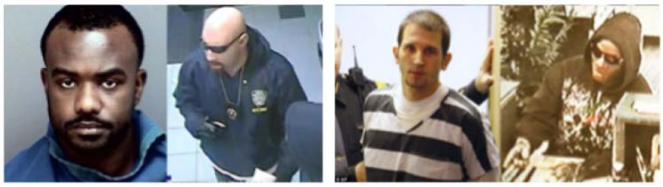

Fig. 3. Images of robbers without (first, third) and with (second, fourth) using masks to conceal their identity. Images are obtained from Internet.

Manuscript received July 18, 2016; revised December 28, 2016; accepted February 1, 2017. Date of publication March 1, 2017; date of current version May 3, 2017. This work was supported by the Ministry of Electronics and Information Technology, India and Infosys Center for Artificial Intelligence at IIIT Delhi. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Yunhong Wang.

I. Manjani was with the Indraprastha Institute of Information Technology Delhi, New Delhi 110020, India. He is now with Adobe Systems, Noida 201304, India (e-mail: ishan12041@iiitd.ac.in).

S. Tariyal, M. Vatsa, R. Singh, and A. Majumdar are with the Indraprastha Institute of Information Technology Delhi, New Delhi 110020, India (e-mail: snigdha1491@iiitd.ac.in; mayank@iiitd.ac.in; rsingh@iiitd.ac.in; angshul@iiitd.ac.in).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIFS.2017.2676720

Fig. 4. Sample frames of videos from the proposed database. The first four columns illustrate frames with silicone masks and hence belonging to the attacked class, while the frames of the next four columns belong to the genuine class.

sample and a sample where a person wears silicone masks. The algorithms used for presentation attack detection (PAD) generally rely on domain knowledge. Many algorithms utilize features which may be heavily pertinent to the kind of attack being detected, for instance, motion [14]–[20], texture [21]–[24], reflectance properties [25], [26], or image quality [27]. As much as a counter-measure algorithm depends on particular characteristics for its operation, the less is its generalizability across different attacks. Furthermore, the existing publicly available datasets are either captured in constrained sensing environments [8], [9] or semi-constrained sensing environments [10], [28]. Given the ultimate application of face biometrics in unconstrained environments, it is important to design presentation attack detection algorithms that are independent of the attack and environment.

With this motivation, we present an automatic feature learning technique through multilevel deep dictionaries for detecting face presentation attacks. The contributions of this paper can be summarized as follows:

- A Silicone Mask Attack Database (SMAD) is prepared which will be shared with the research community. The database has been compiled from online resources and consists of videos of people wearing silicone masks, along with genuine access videos. The videos have been captured in varied environments including unconstrained settings.
- A multilevel deep dictionary based face presentation attack detection algorithm is proposed. The feature learning is independent of any knowledge of attack types and does not seek to exploit any particular distinguishing attributes. In different layers, it encodes low to high level features of the presented samples and classification is performed using Support Vector Machine (SVM) classifier.
- The performance of the proposed algorithm is demonstrated on the SMAD, 3DMAD [11], CASIA-FASD [10], UVAD [28], and replay-attack database [9]. Both intra-database and cross-database experiments are performed to demonstrate the effectiveness and generalizability of the proposed algorithm across different kinds of attacks. We also compare the performance with existing state-of-the-art presentation attack detection algorithms.

II. Silicone Mask Attack Database

The Silicone Mask Attack Database consists of face videos with and without masks. The attack samples consist of a person wearing a real life silicone mask. The mask is a complete 3D structure to be worn around the head which fits well with proper holes for the eyes and mouth avoiding irregularities at these regions, such as in the print attacks. Some silicone masks also include hair, mustache, and beard for life-like impression. The silicone material is stretchable allowing the individual to speak and perform blink movements near the mouth and eye regions. The genuine videos are of people auditioning, interviewing, or hosting shows. Since the genuine videos in the database are also compiled from multiple sources, they introduce several important challenges such as illumination, background, distance from the camera, and capture quality. Fig. 4 shows sample frames of both classes - *genuine* and *attack*. Existing presentation attack databases generally contain both genuine and attacked samples of the person impersonating and the individual being impersonated. This allows the researchers to evaluate the effect of presentation attack on face recognition - both for evading and impersonating identity - along with designing attack detection algorithms. Since tailor made silicone masks are very expensive, the proposed SMAD does not contain the videos of the person whose identity is being impersonated.

The database comprises 130 videos, 65 genuine and 65 with mask. In most of the cases, the original videos are quite lengthy. However, in real world applications, the available

TABLE I
COMPARISON OF THE FACE PRESENTATION ATTACK DATABASES USED IN THIS RESEARCH

	Replay Attack [9]	**3DMAD [11]**	**CASIA-FASD [10]**	**UVAD [28]**	**SMAD**
Subjects	50	17	50	404	-
Videos	1200	255	650	17,076	130
Protocol	Train, development, and test sets	Leave one subject out	Train and test sets	Train and test sets	5-fold cross-validation
Characteristics	2D print, replay; controlled environment	Hard resin masks; controlled environment	Quality; warped video and cut-photo	2D replay; indoor and outdoor	Real life silicone masks; varying lighting and background

TABLE II
DETAILS OF THE SILICONE MASK ATTACK DATABASE

	Genuine	Attack	Total
Male	43	59	102
Female	22	6	28
Frames	15,901	11,996	27,897
Videos	**65**	**65**	**130**

videos may be of a very small duration. Therefore, sample videos in the database have been clipped to $3-10$ seconds to present enough data for designing attack detection algorithms while ensuring near frontal head pose of the subjects. The database consists of total $27,897$ frames with an average of 214 frames per video. Table I compares the characteristics of SMAD with four existing databases, namely Replay attack [9], 3DMAD [11], CASIA-FASD [10], and UVAD [28]. Table II provides the finer statistics of the proposed database in terms of frames, videos, male and female.

We postulate that the algorithms designed to detect presentation attack attempts in videos captured in such uncontrolled settings would significantly benefit the face recognition applications. Additionally, attack attempts through the silicone masks dismiss the possibilities of PAD algorithms seeking to exploit any particular distinguishing characteristic such as flat structure, visible borders, movements, or illumination properties.

The database will be made publicly available via http://iab-rubric.org/resources.html. We next present the experimental protocol and the performance measurement metrics for benchmarking and reporting results on the Silicone Mask Attack Database.

A. Experimental Protocol on SMAD

Videos from each class are equally distributed in five non-overlapping folds. In each iteration, three folds are used for training the PAD algorithm while the other two are used as the test set. The training folds can be used for parameter optimization and as validation or development set. This train-test split is randomly repeated five times for cross validation and performance evaluation. Along with the database, these five splits will also be released to the research community.

To demonstrate the performance of presentation attack detection algorithms, this study proposes two protocols: frame-based and video-based. The performance obtained by classifying all the individual frames as genuine or attacked is referred to as the *frame based approach*, while the *video based approach* classifies entire video samples into two classes.

B. Performance Measure

A presentation attack detection algorithm provides a measure of how likely a biometric sample is genuine or attacked. Depending on the type of application, the performance criteria and thresholds may change. For a very secure facility, the users may not want to allow any attacked sample whereas, for time and attendance systems, some level of attacks could be acceptable in lieu of less discomfort to the users. Therefore, we present the baseline results in terms of Equal Error Rate (EER), Half Total Error Rate (HTER) and Receiver Operating Characteristic (ROC) curves. Number of false accepts (accepting an attacked sample as genuine) and false rejects (rejecting a genuine sample as attacked) are calculated for different thresholds and the variation of false reject rate (FRR) with false accept rate (FAR) is visualized using ROC curves.[1] The EER represents the point along the ROC curve where the FAR equals the FRR. HTER, a popularly used metric in the PAD literature, combines both FAR and FRR by computing their average at a threshold obtained using the training and/or validation set, and the results are reported on the test set.

III. PROPOSED DEEP DICTIONARY VIA GREEDY LEARNING FOR PRESENTATION ATTACK DETECTION

Existing presentation attack detection algorithms for face are generally based on either hand-crafted features or deep neural network architectures. The challenge with hand-crafted texture features is that it is difficult for one feature to encode the variations across multiple kinds of presentation attacks. On the other hand, deep network based learnt features provide good generalization, but they require significantly large and representative training database. Therefore, in this research, we propose deep dictionary via greedy learning algorithm (DDGL) for face presentation attack detection. We first briefly provide the preliminaries and review of dictionary learning algorithms followed by the formulation of DDGL algorithm.

[1]ROC curves for cross validation are computed with threshold averaging [29].

A. Dictionary Learning: Preliminaries

Dictionary learning has been well studied for both compressive sensing and feature representation [30]–[32]. The traditional interpretation of dictionary learning is as follows: It learns a basis (D) and coefficients (Z) for representing the data (X). The columns of D are called 'atoms'. The basic formulation for dictionary is shown in Equation (1). The dictionary is learnt so that the coefficients (features) - along with the dictionary - can synthesize/generate the data. The network directed from representation to the input is called *synthesis learning* in signal processing. Dictionary learning employs an Euclidean cost function [33], given by

$$\min_{D,Z} ||X - DZ||_F^2 \tag{1}$$

Earlier, it was termed as 'matrix factorization' as dictionary learning represents the data (X) as a product of two matrices (D and Z). Equation (1) is solved using the method of optimal directions [34] which is basically an alternating minimization algorithm where the representation (Z) is updated assuming that the dictionary (D) is given (Equation 2); and then solves for D assuming Z is fixed (Equation 3).

$$Z_k \leftarrow \min_{Z} ||X - D_{k-1}Z||_F^2 \tag{2}$$

$$D_k \leftarrow \min_{D} ||X - DZ_k||_F^2 \tag{3}$$

Dictionary learning/matrix factorization is a bi-linear non-convex problem. However, each of the sub-problems Equations 2 and 3 are convex. The sub-problems are solved iteratively until local convergence.

Since the advent of compressed sensing [30], [35], researchers in signal processing and machine learning are interested in solving the dictionary for sparse coefficients. Most studies now impose an additional sparsity constraint on the representation (Z) [36], but it is not mandatory. The K-SVD [31] is probably the most well known technique for solving this problem; it is formulated as,

$$\min_{D,Z} ||X - DZ||_F^2 \quad \text{such that } ||Z||_0 \leq \tau \tag{4}$$

K-SVD proceeds in two stages: in the first stage it learns the dictionary and in the next stage it uses the learned dictionary to sparsely represent the data. Solving the l_0-norm minimization problem is NP hard [37]. K-SVD employs the greedy (sub-optimal) orthogonal matching pursuit (OMP) [38] to solve the l_0-norm minimization problem approximately. In the dictionary learning stage, K-SVD proposes an efficient technique to estimate the atoms one at a time using a rank one update. The major disadvantage of K-SVD is that it is a relatively slow technique owing to its requirement of computing the SVD (singular value decomposition) in every iteration. There are other efficient optimization based approaches for dictionary learning [32], [39] - these learn the full dictionary instead of updating the atoms separately.

B. Deep Dictionary via Greedy Learning

The idea of extending the shallow dictionary learning formulation has attracted attention in the recent years. Shen *et al.* [40] propose a hierarchical discriminative dictionary learning approach for visual categorization. They learn multiple dictionaries at different layers to capture varying scale information, which also includes encoding information from previous layers. Ophir *et al.* [41] apply dictionary learning in the analysis domain of wavelet transform to learn sub-dictionaries at different data scales, which are further used for sparse coding. Thiagarajan *et al.* [42] develop a multi-level approach where dictionaries after the first level are learnt on the residual representation of the previous level. They use the K-hyperline clustering algorithm for learning atoms of a single dictionary. Zheng and Jiang [43] for tag-taxonomy produce a dictionary for each node and concatenate them to produce level specific dictionaries. We present our deep dictionary learning paradigm as introduced in [44], and formulate a greedy learning algorithm for the optimization problem.

We interpret the columns of a simple dictionary not as atoms but as connections between the input and the representation layer, and extend it into a deep architecture. For the first layer, a dictionary is learnt to represent the data. In the second layer, the representation from the first layer acts as input; it learns a second dictionary to represent the features from first level. This concept can be extended to deeper layers. Further, we present how an approach of greedy level wise training can be used for solving the deep dictionary learning problem.

A single/shallow level of dictionary learning yields a latent representation of data and the dictionary atoms. Here, we propose to learn latent representation of data by learning multi-level dictionaries. The idea of learning deeper levels of dictionaries stems from the recent success of deep learning in various areas of machine learning [45]. As mentioned earlier, a single layer dictionary learning generally follows a synthesis framework, i.e. the dictionary (D_1) is learnt such that the features (Z) synthesize the data (X) along with the dictionary. This is expressed as,

$$X = D_1 Z \tag{5}$$

We propose to extend the shallow dictionary learning to multiple layers i.e., deep dictionary learning. Mathematically, the representation at the second layer can be written as:

$$X = D_1 \varphi(D_2 Z) \tag{6}$$

where, φ represents the activation function. The activation function is absent in the first layer because X can take any real value. With this, we can go deeper and deep dictionary learning can be expressed as (for N layers),

$$X = D_1 \varphi\big(D_2 \varphi(\ldots \varphi(D_N Z))\big) \tag{7}$$

The full optimization problem is,

$$\min_{D_1, D_2, \ldots, D_N, Z} \|X - D_1 \varphi\big(D_2 \varphi(\ldots \varphi(D_N Z))\big)\|_F^2 \tag{8}$$

Since this is a difficult problem to solve, inspired by the deep neural network architectures, we solve it via the greedy approach [46]. For the first layer, Z_1 is expressed as $\varphi\big(D_2 \varphi(\ldots \varphi(D_N Z))\big)$; so that the problem can be formulated as,

$$\min_{D_1, Z_1} ||X - D_1 Z_1||_F^2 \tag{9}$$

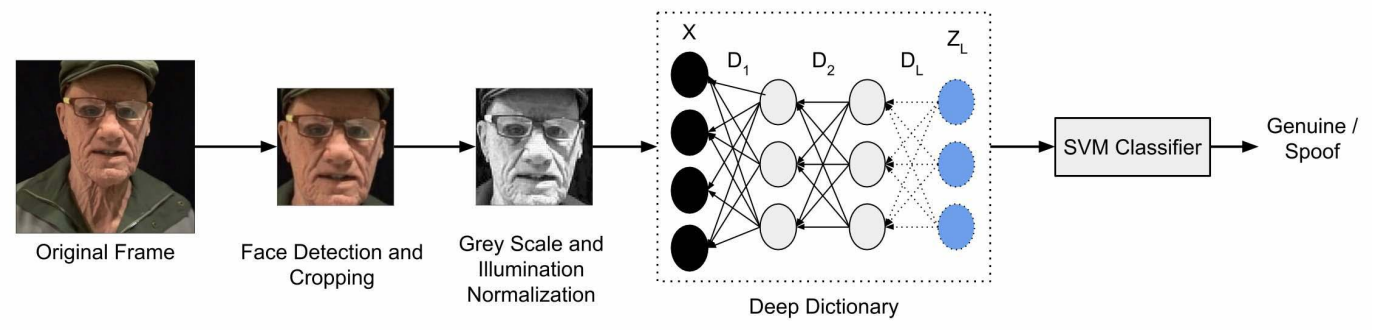

Fig. 5. Illustrating the steps involved in the proposed deep dictionary via greedy learning based face presentation attack detection algorithm.

This can be solved by the method of alternating directions (Equations 2 and 3). Once the coefficients for the first layer are learnt, we formulate learning of the second layer as,

$$\varphi^{-1}(Z_1) = D_2 Z_2, \text{ where } Z_2 = \varphi(D_3 \dots \varphi(D_N Z)) \quad (10)$$

MOD (Method of Optimizing Directions) is invoked once again and the formulation is,

$$\min_{D_2, Z_2} ||\varphi^{-1}(Z_1) - D_2 Z_2||_F^2 \quad (11)$$

The same greedy process is followed for further deeper layers as well. Following recent studies in dictionary learning, we next impose sparsity constraint on the representation of the final layer. The problem that needs to be solved is,

$$\min_{D_1, D_2, \dots, D_N, Z} \|X - D_1 \varphi\big(D_2 \varphi(\dots \varphi(D_N Z))\big)\|_F^2 + \lambda \|Z\|_1 \quad (12)$$

where, λ is the regularization parameter. Till the penultimate level, the intermediate representations are dense and hence can be solved as explained before. The sparsity is imposed at the final layer and hence the optimization problem at the final layer can therefore be expressed as,

$$\min_{D_N, Z} ||\varphi^{-1}(Z_{N-1}) - D_N Z||_F^2 + \lambda \|Z\|_1 \quad (13)$$

This can also be solved using alternating minimization.

$$Z_k \leftarrow \min_{Z} ||\varphi^{-1}(Z_{N-1}) - D_N Z||_F^2 + \lambda \|Z\|_1 \quad (14)$$

$$D_k \leftarrow \min_{Z} ||\varphi^{-1}(Z_{N-1}) - D_N Z||_F^2 \quad (15)$$

The dictionary update stage remains the same as before. However, for the coefficients, we need to solve an l_1-minimization problem. This can be efficiently solved using Iterative Soft Thresholding Algorithm [47]. Due to soft-thresholding, the l_1-norm does not yield exactly sparse coefficients. To get an exactly sparse representation one needs to minimize the l_0-norm. This too can be solved efficiently using the Iterative Hard Thresholding Algorithm [48].

C. Deep Dictionary Learning Based Algorithm for Face Presentation Attack Detection

Fig. 5 shows the steps involved in the proposed algorithm. The proposed presentation attack detection algorithm consists of an unsupervised feature learning step through deep dictionaries and a supervised classification mechanism. The deep dictionary architecture is learned using all the individual frames of the video samples. The classifier is trained on features of the video frames along with class labels. First, for all the video frames, facial region is detected using the Viola Jones [49] algorithm and segmented to a fixed size of $m \times n$. Next, a wavelet based illumination normalization technique [50] is applied to the facial regions and resized to 64×64.

1) Training Phase: Let v_j be the set of all the frames of the j^{th} video sample. Each frame of the video sample v is pre-processed and reshaped to a $mn \times 1$ vector. To train the deep dictionary, a set $X = [v_1, v_2, \dots, v_N]$ is created where X is the set of all the frames of all the training video samples. The deep dictionary is learned through the greedy level wise algorithm on the data X.

Initially, level one dictionary coefficients D_1 are alternatively minimized with X to learn the pair $\{D_1, Z_1\}$. $\varphi^{-1}(Z_1)$ is minimized with D_2 to learn $\{D_2, Z_2\}$. Similarly $\{D_L, Z_L\}$ are learned at level L. $[D_1, D_2, \dots, D_L]$ are the learned L-level dictionary and Z_L denotes the data X in the final projection subspace of the deep dictionary. Z_L is the final layer representation of each frame of the N video samples. The data $\{Z_L, y\}$ where, y denotes the class label (genuine or attacked) for each frame, is used for supervised training. A linear SVM classifier [51] is learnt over $\{Z_L, y\}$ to obtain the distinguishing hyperplane between the two classes.

2) Testing Phase: For a given test video sample v_t (a set of frames), the aim is to classify the entire video and each of the individual frames as either genuine or attacked. The objective is achieved as follows: the set of frames v_t is projected through dictionary coefficients $[D_1, D_2, \dots, D_L]$ one level after the other to obtain $Z_L^{v_t}$. Input to the classifier is $Z_L^{v_t}$ and it provides a score for each frame in v_t. These frame scores are used to measure the frame classification performance. Normalized scores from all the frames are combined using sum rule to determine the classification output of the test video.

IV. Experiments and Results

This section summarizes the experiments performed and the results obtained to demonstrate the efficacy of the proposed

TABLE III

EER (%) ON THE TEST SET OF THE SILICONE MASK ATTACK DATABASE USING DIFFERENT FEATURE LEARNING TECHNIQUES

Feature Learning	Frames	Videos
Deep Belief Network	16.2	16.9
Shallow Dictionary	14.9	14.6
Deep Dictionary: Under-Complete	14.7	13.9
Deep Dictionary: Over-Complete	**14.7**	**12.3**

TABLE IV

THE EFFECT OF DICTIONARY INITIALIZATION TECHNIQUES ON THE EER (%) ON THE SILICONE MASK DATABASE

Initialization Techniques	Frames	Videos
Random Frames	14.9	14.6
PCA	**14.8**	14.6
QR Decomposition	16.2	16.2
Random	15.8	15.4
Ones	14.9	**14.6**

algorithm. We first evaluate the performance of DDGL algorithm on SMAD followed by showcasing the results on data collected from some real world incidents. We next demonstrate the results on four existing databases and compare the performance with state-of-the-art algorithms in both intra-database and cross-database scenarios.

A. Results on Silicone Mask Attack Database

The protocol accompanying the proposed database is followed to obtain the train, development, and test sets. The training data is used to learn the dictionary architecture and classifier, while parameters are optimized using the development set. The training and development data together are used to learn the final classifier and samples of the test data are predicted as genuine or attacked/fake. Five iterations of cross validation are performed and the EER (%) obtained on the test set is shown in Table III. Since the proposed DDGL features are inspired from representation learning via deep learning, the performance is also compared with Deep Belief Network (DBN) [52], which is a deep learning based representation learning technique. The results of DBN are computed with the same pipeline as the proposed DDGL based presentation attack detection. We next analyze the results of the proposed PAD algorithm applied on the SMAD in terms of dictionary initialization techniques and levels of dictionary for both frame-based protocol and video-based protocol.

1) Dictionary Initialization Technique: The first step in dictionary learning is initializing to the initial configuration of the dictionary projection subspace. The initialization is crucial to learning a good transformation domain D, learned representation Z, and avoiding settling to local minima. The dictionary may be initialized with either a matrix of random real values, ones, the basis learned via PCA, Q from the QR decomposition of training data, or randomly selected face frames vectorized to represent a principal axis of the transformation domain. In order to understand the effect, experiments are performed on shallow dictionary formulation. Table IV summarizes the results with varying initialization techniques. A difference of 1.39% and 1.53% is observed in the equal error rates with different initialization techniques for both frame and video based protocols, respectively. In this research, initialization with a matrix of ones, real values leads to better performance and hence they are used for experiments with multilevel deep dictionaries.

2) Dictionary Levels: We train a 1-level shallow dictionary, 2-level under-complete deep dictionary, and a 3-level over-complete deep dictionary. The atom architectures are shallow: [4096 512], under-complete: [4096 2300 512], and over-complete: [4096 8000 4096 512], respectively. The 1-level dictionary is the traditional dictionary learning formulation which serves as a baseline to the proposed deep architectures. As shown in Table III, shallow dictionary performs better than the DBN feature learning. Results indicate improvement beyond the shallow dictionary baseline with the multilevel deep dictionaries for the video based protocol. The 2-level architecture performs better than the 1-level, while the 3-level performs better than all other approaches with an EER of 12.3%. We have also performed t-test to understand the statistical significance and the tests showed that, at 99% confidence interval, the results of shallow and deep dictionaries are statistically different.

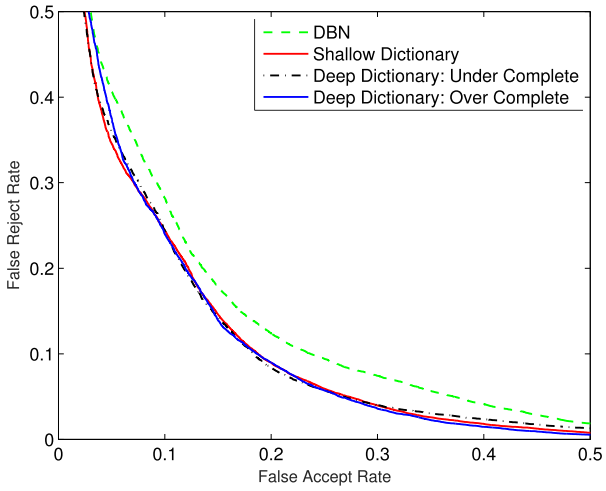

Fig. 6. ROC curves of the frame based protocol with different feature learning techniques on SMAD.

3) Frame Based Protocol: Given a single frame of the face of a person, the aim is to classify it as a genuine sample or an attack attempt. A good prediction algorithm using a single frame may be very useful in cases similar to the two cited robbery incidents. The results of the frame based protocol are shown in ROC curves of Fig. 6 and Table III. The dictionary based techniques outperform the DBN approach by 2-3% for both frame and video based protocols. Fig. 7 summarizes the HTER of DBN, shallow dictionary and deep dictionary along with standard deviation on the SMAD. The HTERs are in the range of $14.7 \pm 3.2\%$ to $16.0 \pm 3.8\%$.

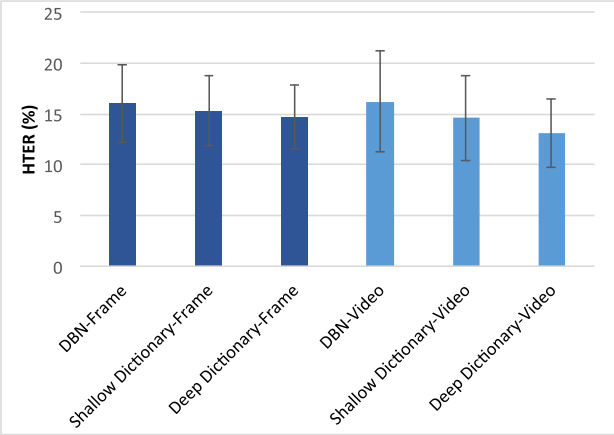

Fig. 7. HTER (%) of attack detection algorithm on frame and video based protocols of the silicone mask attack database.

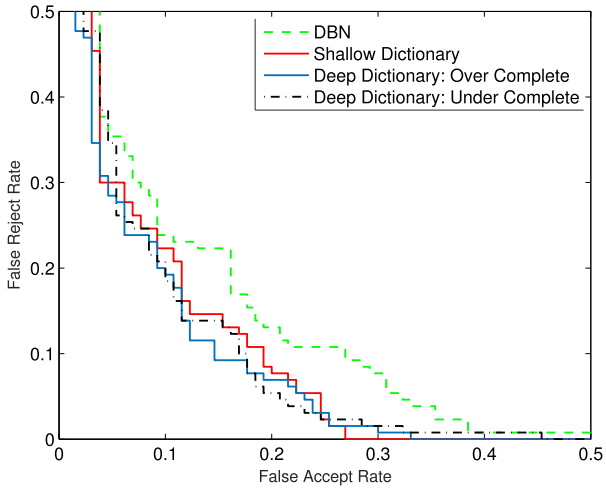

Fig. 8. ROC curves of the video based protocol with different feature learning techniques on SMAD.

4) Video Based Protocol: The video based protocol measures the capability of an algorithm to classify the video samples as genuine or attack. The normalized scores generated by the classifier for all the frames of a video are used for classification. The video based approach uses more data which is generally feasible in most situations with the same camera capturing multiple frames instead of one. The ROC curves for this protocol are shown in Fig. 8. The results show that all the algorithms are better at classifying videos as genuine or attacked as compared to their performance on individual frames except the Deep Belief Network based feature learning. The maximum improvement when classifying videos instead of frames is in the case of the over complete deep dictionary learning (2.36%). We have also performed classification experiments with the first 25 frames, first 50 frames, and all frames to determine the minimum number of frames required for reliable video classification. We observe that the HTER with 25 frames is 1.1% higher than all frames, whereas with 50 frames, slightly higher HTER of 13.5% is achieved (HTER of all frames is 13.1%). It can be inferred that a video of two seconds or more (with 30 frames per second capture) can yield good results.

TABLE V

COMPUTATIONAL TIME OF PROPOSED DEEP DICTIONARY LEARNING BASED ANTI-SPOOFING ALGORITHM

Frames	Operation	Time (s)
13,000	Dictionary Learning (per epoch)	8.33
Per Frame	Illumination Normalization	0.025
	Feature Extraction	0.277
	Classification	0.045
	Total Test	**0.347**

Fig. 9. New identities assumed through silicone masks as reported in [2]–[6].

5) Computational Requirements: Table V lists the time required for dictionary learning and frame classification. The computational time requirements are for Matlab 2014b implementation on the Intel(R) Xeon(R) E5-2695 CPU@2.40 GHz and 128GB RAM. The training time for a dictionary per epoch is 8.33 seconds and the time required to determine whether the given frame belongs to a genuine or attacked sample is 0.34 seconds including computations for illumination normalization, feature extraction, and classification.

B. Results on Real World Incidents

As mentioned in the Introduction section, there have been incidents where people have used silicone masks to assume new identities while they indulged in unlawful activities such as bank robberies [2]–[6], We evaluate the performance of the proposed deep dictionary algorithm for such situations. It allows a platform to explore the usefulness of the algorithm outside the scope of collected databases such as the SMAD and test its efficacy in real deployment scenarios. The trained over-complete deep dictionary and SVM classifier models learnt from the SMAD are used for this purpose. The algorithm is tested on images of people wearing masks in the reported four incidents, shown in Fig. 9. It is to be noted that these samples are not part of the SMAD. The algorithm correctly detects three cases as attacked whereas one case is misclassified as a real image. In three out of the four given cases, the algorithm would have helped the police know that the identities were assumed rather than real, even with only a single image.

C. Results on Existing Presentation Attack Databases

We have also compared the performance of the proposed algorithm with existing algorithms on four publicly available databases along with SMAD.

TABLE VI

SUMMARIZING THE HTERS (%) ON FACE PRESENTATION ATTACK DATABASES. RESULTS ARE OBTAINED FROM RESPECTIVE RESEARCH PAPERS AND VALUES ARE ROUNDED TO ONE DECIMAL PLACE. '-' REPRESENTS THAT HTER VALUES ARE NOT AVAILABLE. *RESULTS ARE COMPUTED BY THE AUTHORS

Algorithm	Replay-Attack	CASIA-FASD	3DMAD	UVAD	SMAD
$LBP-TOP^{\mu 2}_{8,8,8,1,1,1}$					21.5
Erdogmus and Marcel (2014) [53]	-	-	0.1	-	20.8*
Gragnaniello et al. (2015) [12]	9.4	-	0.0	-	-
Wen et al. (2015) [54]	7.4	-	-	-	-
Tirunagari et al. (2015) [55]	0.0	21.8	-	-	-
Pinto et al. (2015) [56]	2.8	14.3	8.0	29.9	-
Arashloo et al. (2015) [57]	0.0	-	-	-	-
Boulkenafet et al. (2016) [58]	3.5	-	-	-	-
Siddiqui et al.* (2016) [24]	0.0	3.8	0.0	27.6	20.4
Deep Belief Network*	1.4	10.8	0.5	30.7	19.2
Proposed DDGL	**0.0**	**1.3**	**0.0**	**16.5**	**13.1**

- For Replay Attack database, we have followed the protocol presented in [9].
- The 3DMAD database [11] consists of biometric samples of 17 subjects captured across 3 sessions. The biometric samples with resolution 640×480 are captured using the Microsoft Kinect for both RGB and depth data, however we use only RGB data and performed experiments according to the protocol defined in [53].
- The experiments on the CASIA-FASD [10] are performed according to the protocol related to the overall database.
- For UVAD database [28], we have followed the protocol used in [56].
- The results of SMAD are reported on the video protocol.

Table VI summarizes results of the proposed algorithm in terms of the HTER. For comparison, eight recent and state-of-the-art algorithms ([12], [24], [53], [54], [55], [56], [57], [58]) are selected and the results are reported directly from the published papers, except in few cases as marked in Table VI. On the 3DMAD and Replay-attack database, the proposed algorithm and some existing algorithms are able to achieve 0% HTER. This suggests that samples in these two databases can be easily classified as spoof or non-spoof. On the other hand, CASIA-FASD and UVAD are more challenging, and existing algorithms have shown higher HTER values, particularly on UVAD. On CASIA-FASD, the proposed algorithm yields 1.3% HTER which is about three times better than the existing literature. On UVAD, the reported state-of-the-art results is 29.9% [56], and the proposed algorithm improves it to 16.5%. For completeness purposes, we have also performed experiments with DBN and the results are reported in Table VI. The DBN approach yields lower error on replay-attack and 3DMAD but it does not perform well on CASIA-FASD and UVAD. It is our assertion that, in order to achieve lower errors for complex environment, DBN requires large amount of training data (which also means excessive training time and computational requirements). In contrast, DDGL efficiently learns the dictionaries with the data available at hand, with faster learning. Finally, we have also computed HTER values for the SMAD and performance is compared with two existing algorithms [24], [53]. The results in Table VI show that the proposed algorithm yields lower errors compared to existing approaches.

D. Cross-Database Experiments

To evaluate the generalizability of the proposed algorithm, we have performed cross-database experiments with all five databases. The experiments are performed according to the protocol proposed by Pinto *et al.* [56]. With five databases, there are five sets of cross-database experiments. For example, the first set of results are reported when the model is trained on the Replay-attack database and tested on other four databases. We have implemented an existing algorithm [24] for performance comparison and Table VII shows the HTER values of the proposed and existing algorithms. Since the experimental protocol is same, we can directly compare the results with Pinto *et al.* [56] for selected experiments. Similarly, for two experiments, direct comparison can be performed with [58].

As shown in Table VII, the proposed algorithm consistently outperforms other face presentation attack detection approaches and is more generalizable. It is observed that training on CASIA-FASD or UVAD databases generally yields lower error on other test databases. This is primary due to the fact that these two databases have multiple attack variations and training on them helps in achieving better features and classification decision boundaries. Existing algorithms also show similar trend; however, they yield two times more error compared to the proposed algorithm. On using SMAD for training and 3DMAD for testing, HTER of 14.1% is observed. We believe that this is due to the fact that SMAD has more variations than 3DMAD and training on SMAD helps in achieving better representation and improved classification results on 3DMAD. On training with 3DMAD, the proposed algorithm shows best cross-database test results on SMAD; however, it is not able to generalize as well in the reverse case. On training with either 3DMAD or SMAD, higher errors are observed on Replay, CASIA-FASD, and UVAD. This is due to two reasons: (i) availability of limited training data and (ii) very large gap between types of attacks in these cross-database settings.

TABLE VII
SUMMARIZING THE HTER (%) FOR CROSS DATABASE EXPERIMENTS

Train Database	Test Database	Implemented by Authors		Reported Results	
		Proposed DDGL	Siddiqui et al. [24]	Pinto et al. [56]	Boulkenafet et al. [58]
Replay-Attack	CASIA-FASD	**27.4**	44.6	50.0	37.7
	3DMAD	**21.6**	40.0	48.0	-
	UVAD	**30.9**	44.8	44.5	-
	SMAD	**32.0**	50.0	-	-
CASIA-FASD	Replay-Attack	**22.8**	35.4	34.4	30.3
	3DMAD	**30.2**	46.4	46.0	-
	UVAD	**32.5**	42.7	40.1	-
	SMAD	**31.2**	48.0	-	-
UVAD	Replay-Attack	**26.8**	40.6	42.8	-
	3DMAD	**28.1**	46.2	44.0	-
	CASIA-FASD	**20.9**	40.2	38.5	-
	SMAD	**30.0**	46.8	-	-
3DMAD	Replay-Attack	**40.9**	49.4	-	-
	UVAD	**44.2**	50.0	-	-
	CASIA-FASD	**42.8**	48.1	-	-
	SMAD	**29.9**	44.8	-	-
SMAD	Replay-Attack	**42.6**	50.1	-	-
	UVAD	**44.8**	47.8	-	-
	CASIA-FASD	**44.6**	48.1	-	-
	3DMAD	**14.1**	28.6	-	-

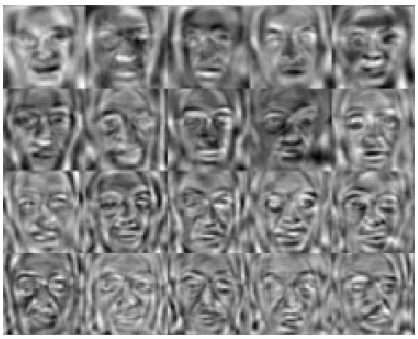

Fig. 10. Visualization of some dictionary atoms showcasing that the proposed algorithm is able to learn robust features.

Fig. 11. An illustration of misclassifications by the Deep Dictionary based PAD algorithm.

E. Discussion

In the proposed deep dictionary learning based approach, we are learning full dictionary at each level and as shown in Fig. 10, robust features are extracted from multiple levels. Using this approach, abstract representation is learnt at deeper levels which helps in better discrimination using SVM classifier. Moreover, layer-wise learning aids in achieving convergence at each layer. Due to these features of the proposed algorithm, experimental results showcase the superior performance of the proposed feature representation algorithm for face presentation attack detection. We experimentally observe that unlike traditional deep learning approaches, the proposed formulation does not require very large amount of training data to achieve higher performance and therefore lower HTERs are obtained on smaller databases as well. Computationally, the proposed algorithm is fast at both training and testing stages which also showcases the usefulness in real world cases. While the results are highly encouraging, there are some misclassified cases as well. As shown in Fig. 11, the silicon mask attack database contains images which have combined variations due to pose, expression, and illumination. These confounding variations sometimes cause errors in correctly discriminating between a mask attack and genuine sample. Finally, the experiments also illustrate that the proposed algorithm is also robust to cross-database variations and hence generalizable.

V. CONCLUSION AND FUTURE WORK

Advancements and popularity of biometric systems have instigated widespread usage in civil and law enforcement applications. However, this also instigates attempts to deceive the biometric systems by presentation attacks which pose a significant threat to successful implementation of the technology. This paper introduces the problem of silicone mask based face attack and presents one-of-a-kind Silicone Mask Attack Database. In order to address this arduous research

challenge, the paper also presents a presentation attack detection algorithm using a novel formulation of multilevel deep dictionary via greedy learning. Experimental results on the proposed silicone mask database and four existing databases show that the proposed algorithm outperforms several state-of-the-art attack detection algorithms. It is also observed that the proposed method performs well even in the presence of confounding unconstrained environment and detects attacks from real world silicone mask attack samples along with replay-attack, 3D mask attack, CASIA, and UVA databases. It is our assertion that the availability of silicone mask attack database can initiate research efforts in building algorithms for unconstrained face presentation attack detection. As a future research direction, we are currently working towards improving the cross-database performance of the proposed presentation attack detection algorithm. Moreover, the proposed algorithm can also be extended towards presentation attack detection in other biometric modalities such as detecting contact lens and print attack in iris [59]–[61].

ACKNOWLEDGMENT

The authors acknowledge the associate editor and reviewers for their thoughtful comments.

REFERENCES

[1] *Mrs. Doubtfire*, accessed on Mar. 20, 2017. [Online]. Available: https://en.wikipedia.org/wiki/Mrs._Doubtfire

[2] *The White Robber Who Carried out Six Raids Disguised as a Black man*, accessed on Mar. 20, 2017. [Online]. Available: http://alturl.com/rvmqa

[3] *Bank Robbers Stump the NYPD With Life-Like Masks That Made Them Look White Even Though They may Have Been Black or Hispanic*, accessed on Mar. 20, 2017. [Online]. Available: http://alturl.com/9ah43

[4] *North Carolina Bank Robber Caught After Using Latex Mask to Look Like old man: Cops*, accessed on Mar. 20, 2017. [Online]. Available: http://alturl.com/p8rbh

[5] *Man in Disguise Boards International Flight*, accessed on Mar. 20, 2017. [Online]. Available: http://alturl.com/z7vdd

[6] *The man in the Latex Mask: Black Serial Armed Robber Disguised Himself as a White man to rob Betting Shops*, accessed on Mar. 20, 2017. [Online]. Available: http://alturl.com/ymngx

[7] *ISO/IEC 30107-1: Information Technology—Biometric Presentation Attack Detection—Part I: Framework*, accessed on Mar. 20, 2017. [Online]. Available: http://www.iso.org/iso/catalogue_detail.htm?csnumber=53227

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

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

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

[11] N. Erdogmus and S. Marcel, "Spoofing in 2D face recognition with 3D masks and anti-spoofing with Kinect," in *Proc. IEEE BTAS*, Oct. 2013, pp. 1–6.

[12] D. Gragnaniello, G. Poggi, C. Sansone, and L. Verdoliva, "An investigation of local descriptors for biometric spoofing detection," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 4, pp. 849–863, Apr. 2015.

[13] D. Menotti *et al.*, "Deep representations for iris, face, and fingerprint spoofing detection," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 4, pp. 864–879, Apr. 2015.

[14] O. V. Komogortsev, A. Karpov, and C. D. Holland, "Attack of mechanical replicas: Liveness detection with eye movements," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 4, pp. 716–725, Apr. 2015.

[15] K. Kollreider, H. Fronthaler, and J. Bigun, "Verifying liveness by multiple experts in face biometrics," in *Proc. IEEE CVPR Workshops*, Jun. 2008, pp. 1–6.

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

[17] H.-K. Jee, S.-U. Jung, and J.-H. Yoo, "Liveness detection for embedded face recognition system," *Int. J. Biomed. Sci.*, vol. 1, no. 4, pp. 235–238, 2006.

[18] K. Kollreider, H. Fronthaler, M. I. Faraj, and J. Bigun, "Real-time face detection and motion analysis with application in 'liveness' assessment," *IEEE Trans. Inf. Forensics Security*, vol. 2, no. 3, pp. 548–558, Sep. 2007.

[19] J. Bigun, H. Fronthaler, and K. Kollreider, "Assuring liveness in biometric identity authentication by real-time face tracking," in *Proc. IEEE CIHSPS*, Jun. 2004, pp. 104–111.

[20] A. Ali, F. Deravi, and S. Hoque, "Liveness detection using gaze collinearity," in *Proc. IEEE EST*, Oct. 2012, pp. 62–65.

[21] N. Kose and J.-L. Dugelay, "Countermeasure for the protection of face recognition systems against mask attacks," in *Proc. IEEE FG*, Apr. 2013, pp. 1–6.

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

[23] T. de Freitas Pereira, A. Anjos, J. M. De Martino, and S. Marcel, "Lbp-top based countermeasure against face spoofing attacks," in *Proc. ACCV Workshops*, 2013, pp. 121–132.

[24] T. A. Siddiqui *et al.*, "Face anti-spoofing with multifeature videolet aggregation," in *Proc. IAPR Int. Conf. Pattern Recognit.*, Dec. 2016, pp. 1–5.

[25] D. F. Smith, A. Wiliem, and B. C. Lovell, "Face recognition on consumer devices: Reflections on replay attacks," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 4, pp. 736–745, Apr. 2015.

[26] W. Kim, S. Suh, and J.-J. Han, "Face liveness detection from a single image via diffusion speed model," *IEEE Trans. Image Process.*, vol. 24, no. 8, pp. 2456–2465, Aug. 2015.

[27] J. Galbally, S. Marcel, and J. Fierrez, "Image quality assessment for fake biometric detection: Application to iris, fingerprint, and face recognition," *IEEE Trans. Image Process.*, vol. 23, no. 2, pp. 710–724, Feb. 2014.

[28] A. Pinto, W. R. Schwartz, H. Pedrini, and A. Rocha, "Using visual rhythms for detecting video-based facial spoof attacks," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 5, pp. 1025–1038, May 2015.

[29] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognit. Lett.*, vol. 27, no. 8, pp. 861–874, Jun. 2006.

[30] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[31] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.

[32] M. Yaghoobi, T. Blumensath, and M. E. Davies, "Dictionary learning for sparse approximations with the majorization method," *IEEE Trans. Signal Process.*, vol. 57, no. 6, pp. 2178–2191, Jun. 2009.

[33] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, Oct. 1999.

[34] K. Engan, S. O. Aase, and J. Hakon Husoy, "Method of optimal directions for frame design," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 5. Mar. 1999, pp. 2443–2446.

[35] E. J. Candès and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?" *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5406–5425, Dec. 2006.

[36] R. Rubinstein, A. M. Bruckstein, and M. Elad, "Dictionaries for sparse representation modeling," *Proc. IEEE*, vol. 98, no. 6, pp. 1045–1057, Jun. 2010.

[37] B. K. Natarajan, "Sparse approximate solutions to linear systems," *SIAM J. Comput.*, vol. 24, no. 2, pp. 227–234, 1995.

[38] Y. C. Pati, R. Rezaiifar, and P. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Proc. SSC*, 1993, pp. 40–44.

[39] A. Rakotomamonjy, "Applying alternating direction method of multipliers for constrained dictionary learning," *Neurocomputing*, vol. 106, pp. 126–136, Apr. 2013.

[40] L. Shen, S. Wang, G. Sun, S. Jiang, and Q. Huang, "Multi-level discriminative dictionary learning towards hierarchical visual categorization," in *Proc. IEEE CVPR*, Jun. 2013, pp. 383–390.

[41] B. Ophir, M. Lustig, and M. Elad, "Multi-scale dictionary learning using wavelets," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 5, pp. 1014–1024, Sep. 2011.

[42] J. J. Thiagarajan, K. N. Ramamurthy, and A. Spanias, "Multilevel dictionary learning for sparse representation of images," in *Proc. DSPW*, Oct. 2011, pp. 271–276.
[43] J. Zheng and Z. Jiang, "Tag taxonomy aware dictionary learning for region tagging," in *Proc. IEEE CVPR*, Dec. 2013, pp. 369–376.
[44] S. Tariyal, A. Majumdar, R. Singh, and M. Vatsa, "Deep dictionary learning," *IEEE Access*, vol. 4, pp. 10096–10109, 2016.
[45] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.
[46] Y. Bengio, P. Lamblin, D. Popovici, and H. Larochelle, "Greedy layer-wise training of deep networks," in *Neural Information Processing Systems*. Cambridge, MA, USA: MIT Press, 2007, pp. 153–160.
[47] I. Daubechies, M. Defrise, and C. de Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Commun. Pure Appl. Math.*, vol. 57, no. 11, pp. 1413–1457, Nov. 2003.
[48] T. Blumensath and M. E. Davies, "Iterative thresholding for sparse approximations," *J. Fourier Anal. Appl.*, vol. 14, nos. 5–6, pp. 629–654, Dec. 2008.
[49] P. Viola and M. J. Jones, "Robust real-time face detection," *Int. J. Comput. Vis.*, vol. 57, no. 2, pp. 137–154, 2004.
[50] V. Štruc and N. Pavešic, "Photometric normalization techniques for illumination invariance," *Face Image Anal., Techn. Technol.*, vol. 52, pp. 279–300, Mar. 2011.
[51] C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining Knowl. Discovery*, vol. 2, no. 2, pp. 121–167, 1998.
[52] Y. Bengio, P. Lamblin, D. Popovici, and H. Larochelle, "Greedy layer-wise training of deep networks," in *Advances in Neural Information Processing Systems*, vol. 19. Cambridge, MA, USA: MIT Press, 2007, p. 153.
[53] N. Erdogmus and S. Marcel, "Spoofing face recognition with 3D masks," *IEEE Trans. Inf. Forensics Security*, vol. 9, no. 7, pp. 1084–1097, Jul. 2014.
[54] D. Wen, H. Han, and A. K. Jain, "Face spoof detection with image distortion analysis," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 4, pp. 746–761, Apr. 2015.
[55] S. Tirunagari, N. Poh, D. Windridge, A. Iorliam, N. Suki, and A. Ho, "Detection of face spoofing using visual dynamics," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 4, pp. 762–777, Apr. 2015.
[56] A. Pinto, H. Pedrini, W. R. Schwartz, and A. Rocha, "Face spoofing detection through visual codebooks of spectral temporal cubes," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 4726–4740, Dec. 2015.
[57] S. R. Arashloo, J. Kittler, and W. Christmas, "Face spoofing detection based on multiple descriptor fusion using multiscale dynamic binarized statistical image features," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 11, pp. 2396–2407, Nov. 2015.
[58] Z. Boulkenafet, J. Komulainen, and A. Hadid, "Face spoofing detection using colour texture analysis," *IEEE Trans. Inform. Forensics Security*, vol. 11, no. 8, pp. 1818–1830, 2016.
[59] N. Kohli, D. Yadav, M. Vatsa, and R. Singh, "Revisiting iris recognition with color cosmetic contact lenses," in *Proc. 6th IAPR Int. Conf. Biometrics (ICB)*, Jun. 2013, pp. 1–7.
[60] D. Yadav, N. Kohli, J. S. Doyle, R. Singh, M. Vatsa, and K. W. Bowyer, "Unraveling the effect of textured contact lenses on iris recognition," *IEEE Trans. Inf. Forensics Security*, vol. 9, no. 5, pp. 851–862, May 2014.
[61] P. Gupta, S. Behera, M. Vatsa, and R. Singh, "On iris spoofing using print attack," in *Proc. IEEE ICPR*, Oct. 2014, pp. 1681–1686.

Ishan Manjani received the B.Tech. degree in computer science and engineering from the Indraprastha Institute of Information Technology, Delhi, India, in 2016. He has served as an Undergraduate Researcher with the Computer Vision Research Laboratory, University of Notre Dame, USA. He is currently a member of Technical Staff with Adobe Systems, India. His research interests include computer vision, face biometrics, and deep learning. He is a recipient of the S.N. Bose Fellowship awarded by the Indo-US Science and Technology Forum.

Snigdha Tariyal received the master's degree from the Indraprastha Institute of Information Technology Delhi, New Delhi. She is currently a Teaching Fellow with the Indraprastha Institute of Information Technology Delhi. Her area of research includes signal processing and dictionary learning.

Mayank Vatsa (S'04–M'09–SM'14) received the M.S. and Ph.D. degrees in computer science from West Virginia University, Morgantown, USA, in 2005 and 2008, respectively. He is currently an Associate Professor with the Indraprastha Institute of Information Technology Delhi, India, and an Adjunct Associate Professor with West Virginia University, USA. He is also the Head of the Infosys Center for Artificial Intelligence at IIIT-Delhi. He has authored over 200 publications in refereed journals, book chapters, and conferences. His areas of interest are biometrics, image processing, computer vision, machine learning and information fusion. He is a recipient of the AR Krishnaswamy Faculty Research Fellowship, the FAST Award by DST, India, and several best paper and best poster awards in international conferences. He is also the Vice-President (Publications) of the IEEE Biometrics Council, an Associate Editor of IEEE ACCESS, and an Area Editor of *Information Fusion* (Elsevier). He served as the PC Co-Chair of ICB 2013, IJCB 2014, and ISBA 2017.

Richa Singh (S'04–M'09–SM'14) received the Ph.D. degree in computer science from West Virginia University, Morgantown, USA, in 2008. She is currently an Associate Professor with the Indraprastha Institute of Information Technology Delhi, New Delhi, India, and an Adjunct Associate Professor with West Virginia University, USA. She has authored over 200 publications in refereed journals, book chapters, and conferences. Her areas of interest are biometrics, pattern recognition, and machine learning. She is a recipient of the Kusum and Mohandas Pai Faculty Research Fellowship at the Indraprastha Institute of Information Technology, the FAST Award by DST, India, and several best paper and best poster awards in international conferences. She is also an Editorial Board Member of *Information Fusion* (Elsevier), and the Associate Editor of IEEE Access and *EURASIP Journal on Image and Video Processing* (Springer). She has served as the General Co-Chair of ISBA 2017 and a PC Co-Chair of BTAS 2016.

Angshul Majumdar (SM'15) received the bachelor's degree from Bengal Engineering College, Shibpur, and the master's and Ph.D. degrees from The University of British Columbia in 2009 and 2012, respectively. He is currently an Assistant Professor with the Indraprastha Institute of Information Technology Delhi, New Delhi. He has coauthored over 120 papers in journals and reputed conferences. He is the author of *Compressed Sensing for Magnetic Resonance Image Reconstruction* (Cambridge University Press) and a coeditor of the *MRI: Physics, Reconstruction and Analysis* (CRC Press). His research interests are broadly in the areas of signal processing and machine learning. He is currently serving as the Chair of the IEEE SPS Chapter's committee and the Chair of the IEEE SPS Delhi Chapter.