IET Biometrics

Research Article

IET Journals
The Institution of Engineering and Technology

Upgrading security and protection in ear biometrics

ISSN 2047-4938
Received on 10th April 2018
Revised 13th December 2018
Accepted on 29th January 2019
E-First on 20th February 2019
doi: 10.1049/iet-bmt.2018.5081
www.ietdl.org

Harsh Sinha[1], Pawan K. Ajmera[1] ✉

[1]*Department of Electrical and Electronics Engineering, Birla Institute of Technology and Science, Pilani, India*
✉ *E-mail: pawan.ajmera@pilani.bits-pilani.ac.in*

Abstract: Biometrics is being widely accepted for user authentication across the globe. Integration of biometrics in the daily life provokes the need to design secure authentication systems. This study proposes the use of outer ear images as a biometric modality. The comparable complexity between the human outer ear and face in terms of its uniqueness and permanence has increased interest in the use of ear as a biometric. However, similar to face recognition, it poses challenges of variation in illumination, contrast, rotation, scale and pose. Owing to the extensive work in the field of computer vision using convolutional neural networks (CNNs), its feasibility in the field of ear biometrics has been presented in this work. The proposed technique uses a CNN as a feature extractor and a support vector machine (SVM) for the classification task. The joint CNN-SVM framework is used for mapping ear images to random base-*n* codes. The codes are further hashed using the secure hash algorithm SHA-3 to generate secure ear templates. The feasibility of the proposed technique has been evaluated on annotated web ears dataset. This work demonstrates 12.52% average equal error rate without any image pre-processing, which shows that the proposed approach is promising in the field of secure ear biometrics.

1 Introduction

Biometrics is ingrained in recognising a user. The basic idea is pivoted on quantifying the behavioural and physiological characteristics for accurate user verification. The present biometric systems have some drawbacks. First of all, multiple readings of a biometric template are prone to variation in illumination, maquillage, pose, resolution and expressions. Human traits can be changed in accidents, surgical alterations and with old age. Hence, they rarely match unless biometric templates are acquired in a controlled environment. Second, biometric traits are non-revocable. When a password for ATM access is compromised, the bank can revoke old PIN and issue a new PIN. However, in the case of a biometric sample, replacement is not possible. In addition, different passwords can be chosen for different financial institutions. However, human traits once compromised can undermine security across different institutions.

Clearly, biometric templates were not designed for a secure storage. Cappelli *et al.* [1] showed that it is possible to perform masquerade attacks by reconstructing fingerprint images from templates. Therefore, biometric template protection becomes extremely important and there is an ever-growing need to look into newer avenues for better security.

This paper addresses the need for a secure and cancelable biometric template generation as an illustration to ear biometrics. The human outer ear as a biometric has gained popularity among the research community [2] owing to its comparable complexity with face in terms of robustness, distinctiveness, availability, accessibility and acceptability [3]. The ear is also being used in Android [4] and iOS applications [5].

This paper presents a secure and revocable authentication scheme in terms of sensitivity and specificity. The proposed methodology is evaluated on ear database, i.e. annotated web ears (AWE) database. A cancelable and tunable security is proposed by using random base-*n* codes to protect the authentication system from brute-force attacks.

The paper is organised as follows. The literature corresponding to secure biometrics and ear biometrics is described in Section 2. Section 3 discusses the proposed approach for secure ear recognition. The performance evaluation and the security for the proposed methodology are presented in Sections 4 and 5, respectively. Section 6 reports the performance of the proposed approach with some of the other prominent techniques. Section 7 summarises and concludes the paper.

2 History

2.1 Ear biometrics

Using a manual technique of 12 measurements from the centre of the ear, Iannarelli [6] built up the first ear recognition system. Recently, a standout amongst the most conspicuous ear biometric strategies was developed by Burge and Burger [7]. They localised the ear using deformable shapes and achieved identification utilising a Gaussian pyramid transformation of a human ear. Moreno *et al.* [8] developed the first fully automated system for ear recognition using an ensemble of neural classifiers with a compression network to extract macro-features. Ghoualmi *et al.* [9] expressed that the vast majority of biometric strategies give extremely poor precision when images are used without any pre-processing. Detailed works by Omara *et al.* [10] and Anwar *et al.* [11] have suggested better image pre-processing pipelines for recognition of variance in scale, pose and illumination.

In the purview of ear biometrics, researchers have primarily focused on devising new feature extraction techniques for ear images such as wavelet-based [12, 13] and filter-based [14, 15] techniques. One of the most efficient feature extraction techniques is force-field transformation that shows 99.2% recognition accuracy on the XM2VTS dataset [16].

As suggested by LeCun *et al.* [17], selecting features on the basis of classifier performance turns out to be a vastly improved approach. Taking a crude image as input without any pre-processing, a convolutional neural network (CNN) is able to generate feature vectors to train a posterior classifier pivoting on the back-propagation algorithm. This dispenses any explicit feature extraction pipeline. Therefore, CNNs perform better as compared to the traditional classification techniques that are based on hand-crafted features.

2.2 Cancelable biometrics

Secure biometric authentication has been studied widely by researchers either by (i) employing fuzzy methods to generate

biometric-based cryptosystems (biometric salting) or by (ii) applying a non-invertible transform to acquired biometric data [18].

2.2.1 Biometric cryptosystems: In the literature, biometric cryptosystems have been implemented by either combining a user-specific key with a biometric sample or directly generating a key from the biometric template to achieve a random distorted template [19]. Some of the common schemes are fuzzy commitment, fuzzy vault and fuzzy extractor. Apart from these, Nagar *et al.* [20] proposed a hybrid scheme for securing fingerprint templates.

The fuzzy commitment was proposed by Juels and Wattenberg [21]. It is based on de-committing of a key using a biometric sample by creating a hybrid between error correcting codes and cryptography. Fuzzy commitment schemes have been applied on the iris [22–24] and fingerprint [25]. However, such systems suffer from shortcomings such as infeasible assumptions, limited error correcting capacity and limited length of keys.

Fuzzy vault schemes [26] can be viewed as an extension to fuzzy commitment schemes. Such techniques have been employed as an application to smart cards [27], fingerprint [28–30], handwritten signature [31, 32], face recognition [31] and iris images [33]. Fuzzy vault schemes have been proved to be reliable, secure and revocable. However, they can be compromised if their parameters such as polynomials and chaff points are not protected. Moreover, original biometric data can be easily identified by correlating values from different vault systems. Chang *et al.* [34] proposed that it is possible to exploit the non-uniformity of biometric modalities from a set of chaff points using statistical analysis.

The fuzzy extractor scheme [35] is used to authenticate a user by extracting a string from the biometric sample as a key in such a way that extraction is invariant to noise. In order to enhance security, a helper key is generated at the time of the enrollment phase. The efficiency of the system depends upon the extractor and its ability to reproduce the same string or a string close enough (according to a pre-defined distance measure) in an uncontrolled environment.

2.2.2 Non-invertible transformation: It is possible to exploit biometric cryptosystems to extract the original biometric sample [1]. However, in the feature transformation approach, the biometric sample is transformed in such a way that it is non-invertible. Thus, it ensures that the original biometric sample is non-recoverable. The matching of biometric templates is performed in the transformed subspace. Non-invertible transformations were proposed by Ratha *et al.* [36]. They transformed fingerprint data using Cartesian, polar and functional transformations. A bi-directional 2D random projection was proposed by Leng *et al.* [37] for cancellable face and palmprint recognition. An extensive survey on cancellable biometrics has been presented by Rathgeb and Uhl [19].

Recently, there has been focus on employing deep learning methods for secure biometric authentication. Deep learning approaches are based on generating a transformed feature vector that uniquely represents biometric data yet reduces any correlation between original biometric data and the transformed feature vector. Such techniques have been applied to the face [38, 39] and palmprint [40]. In the field of secure biometrics, the ear sample is used in a multi-modal fashion either with the face or other soft biometrics. Paul and Gavrilova present a multi-modal recognition by fusion of features extracted from face and ear to generate the cancellable template using random projection and the fuzzy communication scheme [41].

The proposed methodology uses a CNN to obtain noise invariant feature vectors for classification. Support vector machine (SVM) is used to train a discriminant function that learns mapping of extracted features (obtained from CNN) to random base-*n* codes. To prevent violation of data, the codes are hashed using a cryptographic hash function, the secure hash algorithm (SHA3-256) [42]. Thus, the proposed methodology overcomes vulnerabilities such as infeasible assumptions, limited key length, error correction, non-uniformity of biometric data and data acquisition from uncontrolled environments.

3 Proposed methodology

This paper proposes a recognition methodology for ear biometrics that achieves high accuracy while maintaining a high level of security with no pre-assumptions in terms of variations in pose, illumination and the type of security attack.

As the human outer ear has been shown to be unique, invariant and permanent [6], the first step in the enrollment process is capturing an ear image. Similar to face recognition, the image can be captured from a digital camera. CNNs have proved their might in image classification and recognition. They can be trained to be invariant to pose, illumination and noise.

The proposed methodology uses a CNN as a feature extraction module that takes an image of the outer ear as input and generates a feature vector. This is achieved by removing the fully connected layers responsible for classification.

A typical CNN consists of several convolutional layers as well as fully connected layers. The fully connected layers classify extracted features into classes. The last layer prior to the fully connected layers (known as bottle neck features or BNFs) can be used as features with any generic classifier [43]. Researchers have shown that generic descriptors extracted from the penultimate layer of CNN are very efficient for classification [44, 45]. In this work, SVM is used as to classify BNFs obtained from CNN.

Traditional biometric authentication systems store original biometric samples. However, such personal identifiable information (PII) or sensitive personal information (SPI) can become an issue in terms of security as it is prone to data theft and information extortion. Therefore, to ensure security, random base-*n* codes are used as output labels for classification such that they bear no correlation with the original biometric sample.

These codes are hashed using secure hash algorithm (SHA-3) [42] and stored as a template for verification. Hashing serves as a non-invertible transformation that enables secure storage of codes (being used as classification labels).

The proposed joint CNN-SVM framework is illustrated in Fig. 1. The proposed methodology aims to exploit the advantages of CNN, SVM and SHA3-256 in a single mechanism. Therefore, the framework has been illustrated in three major components: a CNN serving as a feature extractor, an SVM acting as a posterior handling module for classification task and SHA3-256 for secure template generation.

During an attempt, a test sample is fed as an input to the trained model that computes a hash code and compares it with the codes stored in a database to authenticate the user. Hash codes are non-invertible; they eliminate any possibility of extracting the original biometric sample. In case, hash codes get compromised, the institution can simply use a different set of random codes as labels, thus introducing cancellability in the proposed methodology.

The subsequent sections describe the various components of proposed methodology.

3.1 Convolutional neural network

CNNs are a class of neural networks. Similar to traditional neural networks, they are composed of several weights and biases that are learned as per the desired mapping of inputs and outputs. The ability of CNNs in the field of large-scale image recognition [46, 47] was realised in the ImageNet Large Scale Visual Recognition Competition (ILSVRC) challenge [48].

A CNN is an end-to-end non-linear system that can be trained to learn high-level representations directly from raw images. The main components of the CNN architecture are convolution, pooling and fully connected layers. An input signal is a greyscale image of an ear as a matrix $\boldsymbol{I} \epsilon \mathbb{R}^{w \times h \times c}$, where w, h and c are input width, height and the number of channels, respectively. For convolution, a weight matrix $\boldsymbol{W} \epsilon \mathbb{R}^{p \times p \times c \times k}$ is convolved with input $\boldsymbol{I}$. The weight matrix spans across a small patch of size $p \times p$ with a stride s, where $p \leq \min(w, h)$. The weight sharing across patches helps to model local correlations in the input image. The weight matrix is

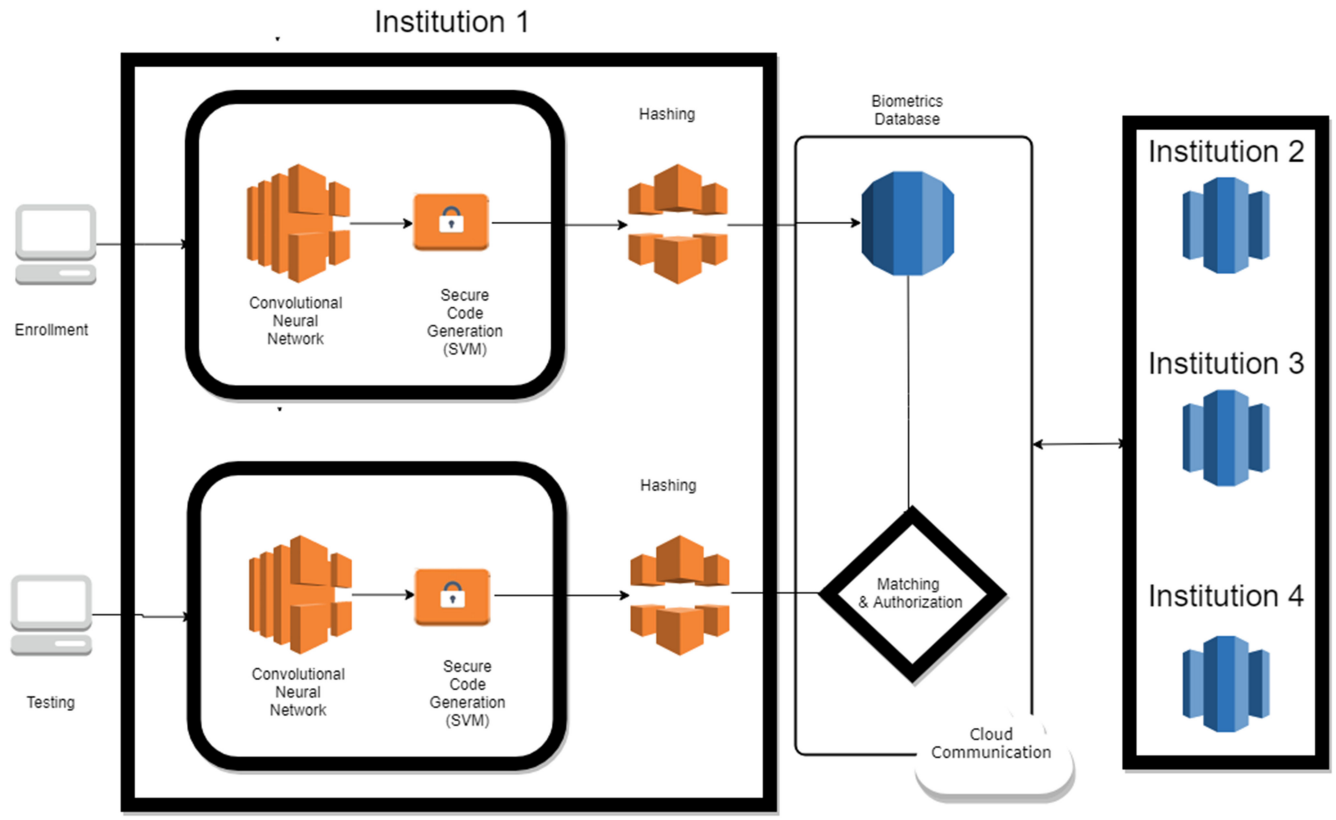

Fig. 1 *Proposed authentication system based on secure cancellable biometrics*

used to generate k feature maps. A convolution operation on the layer $\boldsymbol{I}$ can be summarised as follows:

$$O = \sigma\left(\sum_{c} \boldsymbol{W} \times \boldsymbol{I} + b\right) \tag{1}$$

where $\boldsymbol{O} \epsilon \mathbb{R}^{((w-p)/s)\times((h-p)/s)\times k}$ is the output matrix, b denotes bias and σ is an elementwise non-linearity operation (e.g. RELU).

After performing convolution, a pooling operation sub-samples the input matrix to summarise information, i.e. retains important information while reducing spatial resolution. In max-pooling, only the maximum value of spatial neighbourhood (say 2×2 window) is preserved. Thus, pooling helps in removing variability that exists due to illumination, pose, rotation and noise. It also helps to reduce the computation for later layers by reducing the size of matrices.

The proposed CNN is composed of four stacks of convolution and pooling layers followed by a fully connected layer. During training, the last layer is associated with a multiclass cross-entropy loss function as presented in the following equation:

$$\text{loss} = -\sum_{n=1}^{N} y_{\text{pred},t} \log(p_{\text{pred},t}) \tag{2}$$

where N = number of images (training samples), pred = predicted user id, t = actual target user id, p = predicted probability and y = a binary indicator 0 or 1, determining whether prediction is the same as target.

The parameters of CNN are trained using Adam optimiser [49] that takes into account benefits of Adagrad [50] by computing adaptive learning rates and RMSprop optimiser [51] by calculating decaying average of past squared gradients

$$\theta_{t+1} = \theta_t - \alpha \frac{m_t}{\sqrt{v_t + \epsilon}} \tag{3}$$

where θ_{t+1} = updated value of parameter, θ_t = previous value of parameter, α = step size, m_t = first-order moment (mean), v_t = second-order moment (variance), ϵ = small number (say 10^{-7} to prevent division-by-zero). The algorithm has been efficient across deep learning tasks as it prefers flat minima in error hyperplane avoiding local minima and thus achieving better generalisation [52, 53].

To avoid overfitting, dropout [54] and L2 regularisation are applied to both convolutional and fully connected layers. As a result, co-adaptation of nodes and over-dependence on large weights is prevented. In addition, employing batch normalisation [55] ensures that covariance shift is minimal, improving consistency and reproducibility of experiments.

Table 1 summarises the proposed CNN architecture as illustrated in Fig. 2.

3.2 Biometric template protection

In a traditional biometric authentication system, features from a biometric sample are stored in a database as a template. Such systems are prone to data leakage leading to identity theft. Therefore, traditional biometrics are dangerous for user privacy. It is important to store the template in such a way that it cannot be correlated with the subject's original biometric sample. This ensures that even if the template is compromised, the user cannot be impersonated.

This problem is addressed by mapping extracted features to random n-ary codes. The benefits of using random n-ary codes are twofold. First, it minimises any correlation between the biometric sample and the stored template. Second, if the template is compromised, there is flexibility to reassign another random code as a new user id, thus introducing cancellability.

In the proposed methodology, the original biometric template is not stored. Moreover, the random codes are allocated securely only during training to the client not stored in database. During an attempt, features are extracted from the ear image sample using CNN, which is fed to an SVM for classification. The SHA3-256 hash digest of the predicted label for the test sample (by SVM) is compared with stored hash digests for verification of the subject.

So, neither the original biometric sample nor the assigned random code is required to be stored in database eliminating any scope for data leakage. Rather, non-invertible hash digests are stored for matching during evaluation. These hash digests are non-

invertible, and hence, they cannot be exploited to retrieve the original biometric sample. Thus, user privacy is preserved.

3.2.1 n-ary code generation: Randomly generated base-n codes of length m (e.g. a binary code of bit length 256) are used as labels for different users. For example, binary (referred to as base-2) uses only two symbols 0 and 1, ternary (base-3) use three symbols 0, 1 and 2 and so on. Random generation of codes ensures no resemblance to the input biometric sample. Therefore, an intruder would have to brute-force all possible codes i.e. m^n attacks, which is computationally impossible provided $m > t$, a manually chosen threshold.

Entropy plays an important role in deciding whether a code is likely to be breached. For an n-ary code

$$H = -\sum_{i}^{n} p_i \cdot \log_n p_i \tag{4}$$

where H = entropy and p_i = occurrence probability of symbol i, assuming $\forall\ i,\ p_i > 0$.

According to (4), for maximum entropy of an n-ary code of length k, each symbol i must have an occurrence probability of $1/n$. To evaluate the performance of the proposed methodology, different base-n codes are used as classification labels (as shown in Figs. 3 and 4). Further, the methodology is also assessed for different code lengths. To evaluate the impact of code length on recognition accuracy, the following range was chosen for experimentation: $n\epsilon(2, 9)$ and $m\epsilon 2^{(6,11)}$.

Table 1 Summarised CNN architecture

Layer	Parameters	
convolution	patch size: 7 × 7	depth: 16
batch normalisation > ReLU activation	momentum: 0.9	epsilon: 0.001
maxpooling	patch size: 2 × 2	depth: 16
regularisation	dropout: 0.2	L2 beta: 0.5
convolution	patch size: 5 × 5	depth: 32
batch normalisation > ReLU activation	momentum: 0.9	epsilon: 0.001
max pooling	patch size: 2 × 2	depth: 32
regularisation	dropout: 0.2	L2 beta: 0.5
convolution	patch size: 3 × 3	depth: 64
batch normalisation > ReLU activation	momentum: 0.9	epsilon: 0.001
max Pooling	patch size: 2 × 2	depth: 64
regularisation	dropout: 0.2	L2 beta: 0.5
convolution	patch size: 1 × 1	depth: 128
batch normalisation > ReLU activation	momentum: 0.9	epsilon: 0.001
max pooling	patch size: 2 × 2	depth: 256
regularisation	dropout: 0.2	L2 beta: 0.5
fully connected layer	number of neurons: 512	—
fully connected layer	number of neurons: 80	—
regularisation	dropout: 0.2	L2 beta: 0.5
fully connected layer	number of neurons: 100	—

3.2.2 Support vector machines: The proposed technique uses a CNN as a feature extractor and an SVM for the classification. The joint CNN-SVM framework is used for mapping ear images to random base-n codes.

A CNN is trained to classify, but once trained, a CNN becomes inclusive to classes. Re-training CNN for the addition of new classes is a cumbersome task as it requires collecting images from all of the users again. The ability of CNNs to generate noise-invariant feature vectors can be exploited by extracting generic descriptors from activations of the penultimate layer of CNN (bottle neck features or BNFs) [43–45]. These descriptors can be treated as features of a biometric sample.

The proposed methodology uses a linear SVM for classification of BNF extracted using CNN. Moreover, the linear SVM is trained in an incremental fashion that enables the addition of new users at any point in time.

The last convolution layer is ravelled into a single vector of size 131,072, which poses a hindrance for training SVM. SVMs scale super-linearly [56]. Therefore, principal components analysis is employed to reduce the vector size to 1024. SVMs are based on the quadratic programming problem and their complexity scales between $O(n_{\text{features}} \times n_{\text{samples}}^2)$ and $O(n_{\text{features}} \times n_{\text{samples}}^3)$ [57]. There is a chance of overfitting if n_{features} is much greater than n_{samples}. As the number of ear image samples in AWE dataset is 1000, a feature vector of size 1024 is an appropriate choice.

For an efficient calculation of principal components, randomised singular value decomposition [58] was used. The final classification is attained using a linear SVM with $\gamma = 0.001$, which is inversely related to the radius of influence and the regularisation parameter, $C = 10$. The linear SVM is trained using one versus all multi-class strategy.

3.3 Secure storage of assigned codes

This section explains the strategy for generating secure biometric templates. The security concerns in a traditional biometric system can be alleviated using cryptographic hash functions. Such functions transform an arbitrary block of data to a fixed size, non-invertible string. This removes any kind of resemblance between the biometric sample and stored template.

Cryptographic hash functions are used to generate hash digests of pre-defined length from input data. The utility of a hash function is characterised by its time complexity, diversity, and non-invertibility. Therefore, the security of a hash function can be assessed by its speed, distinctness in different hash digests and the difficulty in decrypting the generated hash digest.

NIST recently released SHA-3 [42], the latest companion of secure hash algorithms. With successful attacks on previously accepted standard hash function such as MD5 [59] and SHA-1 [60], there has been dependence on SHA-2. However, both SHA-1 and SHA-2 come under a common class of algorithm called Merkle Damgård construction [61]. Hence, it is likely that even SHA-2 might have similar weaknesses. Therefore, NIST launched SHA-3 in 2007 [62] that was accepted as the new standard in 2012.

In this work, the extracted features from ear images (by CNN) are mapped to random codes (as class labels) using SVM (as explained in Section 3.2.2). These random codes are hashed using SHA-3 [42] at the last stage of enrollment for secure storage of random codes. In this paper, SHA-3 is used because it is the new standard for robust security. A user is verified by matching hash

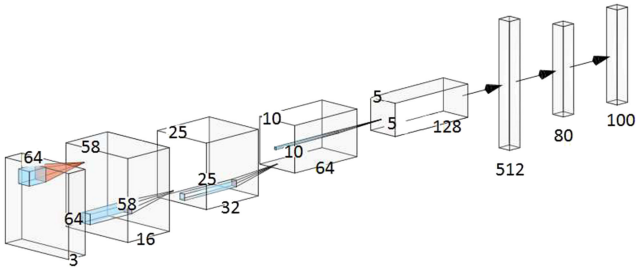

Fig. 2 *CNN used in proposed methodology [the max-pool layers are skipped for brevity]*

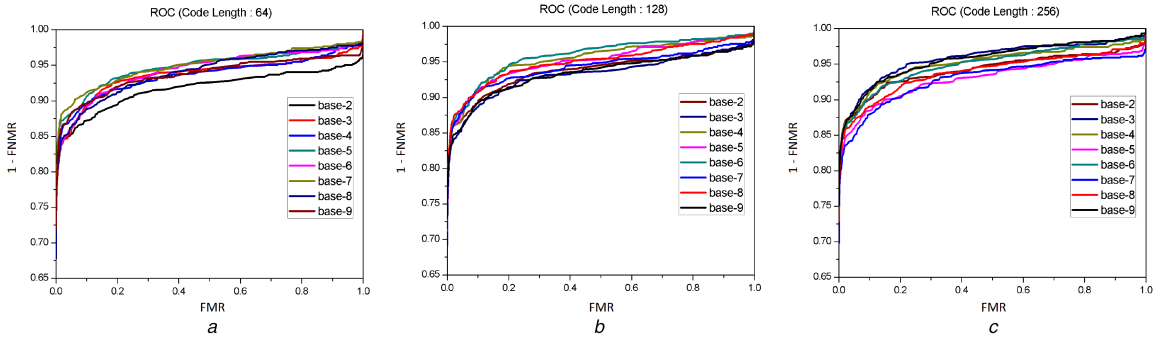

Fig. 3 *Impact of nature and length of random codes (used as classification labels) on recognition performance of the proposed methodology. base-n denotes number of symbols used for code generation. base-2 (two symbols: 0 and 1), base-3 (three symbols: 0, 1 and 2), base-4 (four symbols: 0–3) base-5 (five symbols: 0–4) base-6 (six symbols: 0–5) base-7 (seven symbols: 0–6) base-8 (eight symbols: 0–7) base-9 (nine symbols: 0–8) ROC curves for codes of lengths 64*

(a), 128 ***(b)*** and 256 ***(c)*** with different number of symbols used for code generation

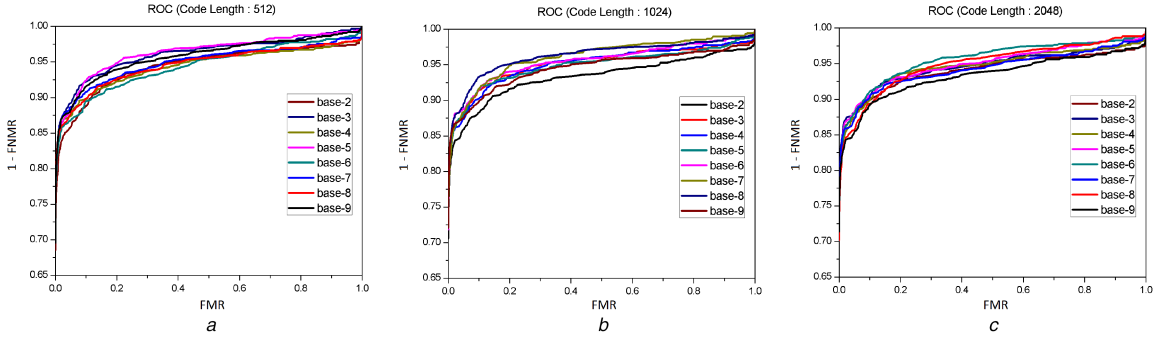

Fig. 4 *Impact of nature and length of random codes (used as classification labels) on recognition performance of the proposed methodology. base-n denotes number of symbols used for code generation. base-2 (two symbols: 0 and 1), base-3 (three symbols: 0, 1 and 2), base-4 (four symbols: 0–3) base-5 (five symbols: 0–4) base-6 (six symbols: 0–5) base-7 (seven symbols: 0–6) base-8 (eight symbols: 0–7) base-9 (nine symbols: 0–8) ROC curves for codes of lengths 512*

(a), 1024 ***(b)***, and 2048 ***(c)*** with different numbers of symbols used for code generation

digest of his test biometric sample (ear image) with the hash digest template. The proposed methodology uses SHA3-256 with the permutation function of the sponge construction [63] characterised by bitrate = 1088, capacity = 512 and output size = 256.

4 Performance evaluation

The following section provides an overview of datasets, evaluation protocols and specifications of parameters used for performance evaluation.

4.1 Dataset

The AWE dataset [64] contains 1000 cropped ear images of 100 distinct subjects that were collected from the web with the goal of studying unconstrained ear recognition. The publicly available USTB-III dataset [65] consists of 786 face profile images of 79 subjects under different illumination from a distance of 1.5 m. Every person has 10 images including a profile image and images turning to left 5°, 10°, 15° and 20°.

The performance measures for the proposed methodology are equal error rate (EER), false match rate (FMR) and false non-match rate (FNMR). These standard measures were realised by taking the average of receiver operating characteristic (ROC) across all classes.

4.2 Experimental setup

First of all, ear images are extracted from profile images as ear detection is an indispensable part of ear recognition. A histogram of oriented gradients (HOG) with an SVM framework has been employed to crop ear images from face profile images [66]. Furthermore, the cropped ear image is grey-scaled, resized (64×64) and normalised before feeding them to CNN for feature extraction. Traditional image pre-processing algorithms (such as histogram equalisation) have a tendency to distort the image based on outliers, which results in loss of information. Hence, no explicit image pre-processing algorithms were used.

From the available 10 images for every user, 6 images are used for training, 2 images for validation and 2 images for testing. The proposed CNN is trained on 100 users (AWE dataset). The 79 users from USTB III dataset are used to simulate impostors during the training phase. The weight of the proposed CNN model is initialised using a normal distribution of $\mu = 0$, $\sigma = 0.01$. The architecture consists of ∼90,000 parameters that were sufficient for a database consisting of 10 images per user. Deeper architectures tend to converge to a local minimum due to limited training data, thereby over-fitting and yielding less accurate results.

4.2.1 Computational complexity: The following section presents computational efficiency for the proposed methodology, which is a key aspect in biometric authentication systems.

A simple two-layer three-node neural network with threshold activation functions is NP-complete [67]. Recently, a general formula for the total time complexity of all convolutional layers in a CNN was presented by He and Sun [68].

CNNs use a massive network of shared weights and biases that enables automatic feature extraction. However, it increases the computational complexity of the CNN. Despite high computational complexity of CNN, several tricks that enable *improper learning* are ReLU activation function, overspecification and regularisation [69]. The evaluated time highly depends on the specific hardware configuration and the platform used for evaluating the CNN architecture.

All of the experiments in this paper are performed on Dell Precision Tower 5810 with CPU as Intel Xeon Processor and two 2-GB Nvidia Quadro K620 GPUs. The scripts are written in python based on TensorFlow and scikit-learn. The running time for a single image is ∼1 s, which includes reading input image, inference and visualisation.

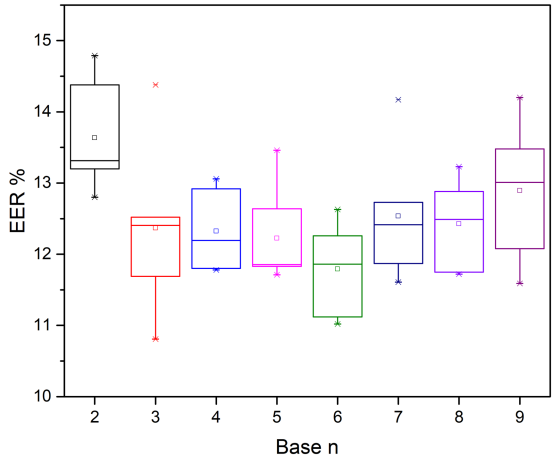

Fig. 5 *Graph visualising EER values across different base-n codes*

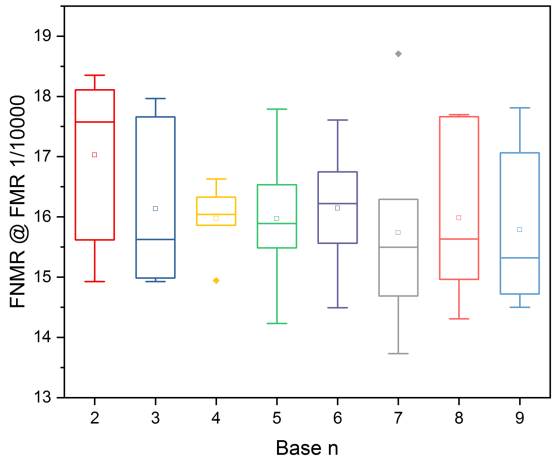

Fig. 6 *Graph visualising FNMR@FMR 1/10,000 values across different base-n codes*

4.3 Results

This section discusses performance evaluation of proposed methodology evaluated on the fused dataset, as described in Section 4.1. Figs. 3 and 4 show ROC curves. Figs. 5 and 6 show EER and FNMR at FMR 1/10,000 values for different code lengths, respectively. FNMR and FMR are defined as shown in the following equations:

$$\text{FMR} = \frac{\text{FP}}{\text{FP} + \text{TN}} \tag{5}$$

$$\text{FNMR} = \frac{\text{FN}}{\text{FN} + \text{TP}} \tag{6}$$

where FMR = false match rate, FNMR = false non-match rate, FP = number of false positives, TN = number of true negatives, TN = number of true negatives, FN = number of false negatives and TN = number of true positives. EER is defined as the point at which FMR equals FNMR.

ROC curves are shown in Figs. 3 and 4 with different sub-figures (*a*–*c*), exhibiting performance of the proposed method corresponding to the various lengths of random codes (64, 128, 256, 512, 1024 and 2048). Each curve in a sub-figure corresponds to a ROC curve for a different length of the random code. For example, Fig. 3*a* shows ROC curves for codes of length 64 with different numeral systems such as binary and ternary that are used for random code generation. The ROC curves demonstrate the discriminating ability of a classifier based on the sensitivity (1 − FNMR) and specificity (FMR).

The proposed methodology achieves up to 12.52% average EER on the AWE dataset consisting of 100 users. The distribution of EER values and FNMR@FMR 1/10,000 values is illustrated as box plots in Figs. 5 and 6. Box plots characterise a distribution using its minimum, lower quartile, median, upper quartile and maximum values. Both the figures have similar interquartile regions across all code lengths, which show that the EER values and FNMR@FMR 1/10,000 values are stable with respect to code length m and base n ($\mu_{\text{EER}} = 12.52\%$, $\mu_{\text{FNMR @ FMR 1/10,000}} = 16.1\%$). This allows the authentication system to flexibly choose a security level.

5 Security analysis

The following section analyses the security of the proposed biometric authentication in defence to possible attacks such as false acceptance rate (FAR) attacks, linkage attacks and hill climbing attacks.

5.1 FAR attacks

There is a considerable chance of false acceptance in biometric authentication systems. Such systems often face false rejects due to variations in illumination, contrast, rotation, scale and pose in the same class. Sometimes, there can be false acceptances due to high correlation between different classes.

Therefore, biometric authentication systems are developed assuming a very low probability of false acceptances (referred to as false acceptance rate, FAR). According to the UK Government's Biometrics Working Group (BWG) [70], typical systems are configured at FMR 1/10,000. This indicates that two identical subjects can be found in 10^4 trials, which require that the attacker must already have a large database.

The security analysis of the proposed system is shown in Fig. 6. The proposed methodology attains $\mu_{\text{FNMR @ FMR 1/10,000}} = 16.1\%$.The distribution of FNMR@FMR 1/10,000 values is illustrated as box plots in Fig. 6. The figure has similar interquartile regions across all code lengths, which show that the FNMR@FMR 1/10,000 values are stable with respect to code length m and base n.

5.2 Linkage attacks

If the same biometric characteristic is utilised in different institutions, as explained in Fig. 1, it results in similar or correlated identities being stored in different databases. Such an unexpected linkage of unrelated applications may be exploited by an adversary.

To overcome this drawback, the proposed methodology can generate distinct pseudo user identities from the same biometric features. In this work, an SVM is used to train a discriminant function that learns mapping of extracted features (obtained from CNN) to random base-n codes. Randomly generated base-n codes of length-m (for example, binary code of bit length 256) were used as labels for different users. As the codes are randomly generated, they bear no resemblance to the input image. Consequently, there is no possibility of correlated identities. Rather, the proposed methodology provides the flexibility that different institutions can use different random codes as user ids for the same person, thereby removing any unexpected linkages.

5.3 Hill climbing attacks

A hill climbing attack deals with the possibility of an attacker to generate synthetic examples to exploit false acceptance of a biometric authentication system. An attacker uses an application that generates biometric templates pivoting on match score of the biometric system until it overcomes the decision threshold.

This weakness of the traditional biometric system is addressed in this work by using randomly generated base-n codes of length-m as labels for different users. Further, the codes are hashed using SHA-3 for secured storage. As the stored hash digests are non-invertible and bear no resemblance to input biometric data, an intruder would have to brute-force all possible codes, i.e. m^n attacks, which is computationally impossible provided $m > t$, a manually chosen threshold. For example, if a binary code of length 512 is used for authentication, an attacker would have to brute force $2^5$12 codes, which is infeasible.

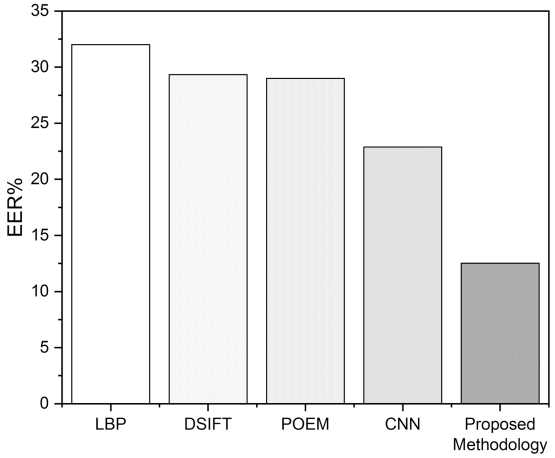

Fig. 7 *Comparative performance of the proposed methodology*

6 Comparative performance

This section focuses on assessing the performance with some of the other prominent techniques that have used the AWE dataset. Hansley *et al.* [71] and Emersic *et al.* [64] have evaluated several holistic, handcrafted and learned features to tackle the problem of unconstrained ear recognition. Some of feature extraction techniques such as local binary patterns (LBP) [71], dense scale invariant feature transform (DSIFT) [71], patterns of oriented edge magnitudes (POEM) [64] and CNN [71] are illustrated in Fig. 7. The results are evaluated on the AWE dataset. Fig. 7 also shows the performance of the proposed methodology that achieves the best EER of 12.52% on the AWE dataset.

The joint CNN-SVM framework outperforms traditional classifiers as it aims to exploit the strengths of CNN and SVM in a single mechanism. As suggested by Lecun *et al.* [17], selecting features on the basis of classifier performance turns out to be an improved approach. A CNN is able to generate discriminative feature vectors that dispense any explicit feature extraction pipeline. CNNs perform better as compared to hand-crafted features. Therefore, the proposed methodology outperforms feature extraction techniques such as LBP, DSIFT and POEM.

In comparison to CNN proposed by Hansley *et al.* [71], the proposed methodology uses a similar CNN architecture. However, the proposed architecture differs in the following aspects. First, this work uses CNN exclusively for feature extraction. A linear SVM is used for classifying extracted BNF features from CNN. A CNN is trained pivoting on the back-propagation algorithm that is based on empirical risk minimisation. The training procedure stops as soon as first the separating hyperplane is found irrespective of the nature of minima (local or global). On the other hand, SVM aims to minimise the generalisation errors using structural risk minimisation. SVM solves a quadratic programming problem to achieve a global optimum [72]. Therefore, SVM performs better as compared to CNN for the classification task.

7 Conclusions

This paper presents a secure and cancellable ear biometric authentication system. Integrating advantages of CNN (feature extraction from images), SVM (ability to rank different classes) and SHA-3 (non-invertible secure hash) paves the way for a secure ear biometric system. The evaluations and experiments for the proposed methodology demonstrate high EER of 12.52% irrespective of the nature (base-n) and length of labels (m), where $n\epsilon(2, 9)$ and $m\epsilon 2^{(6, 11)}$. To summarise, we have the following:

i. Outer ear image as a biometric is viable as experiments show a high verification rate with average EER = 12.52% ± 1.5%
ii. The proposed method uses SHA-3 for storage of templates that is non-invertible, and hence, there is no scope for an intrusion.
iii. The EER results show that the system is invariant to bit length as well base-n for secure code. Hence, any enterprise can choose the desired bit length for a tunable level of security.
iv. The proposed methodology is analysed to be competent against FAR attacks, linkage attacks and hill climbing attacks.

8 References

[1] Cappelli, R., Maio, D., Lumini, A., *et al.*: 'Fingerprint image reconstruction from standard templates', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2007, **29**, (9), pp. 1489–1503

[2] Pflug, A., Busch, C.: 'Ear biometrics: a survey of detection, feature extraction and recognition methods', *IET Biometrics*, 2012, **1**, (2), pp. 114–129

[3] Benarous, L., Kadri, B., Bouridane, A.: 'A survey on cyber security evolution and threats: biometric authentication solutions', in Jiang, R., Al-maadeed, S., Bouridane, A., *et al.* (Eds.): '*Biometric security and privacy*' (Springer, Cham, Switzerland, 2017), pp. 371–411

[4] Boczek, M.: 'Ear biometric capture, authentication, and identification method and system'. US Patent 9,613,200, 4 April 2017

[5] Bargal, S.A., Welles, A., Chan, C.R., *et al.*: 'Image-based ear biometric smartphone app for patient identification in field settings'. VISAPP (3), Berlin, Germany, 2015, pp. 171–179

[6] Iannarelli, A.V.: '*Ear identification*' (Paramont Publishing Company, Fremont, California, 1989)

[7] Burge, M., Burger, W.: 'Ear biometrics in computer vision'. Proc. 15th Int. Conf. Pattern Recognition, Barcelona, Spain, 2000, Art. no. 826830

[8] Moreno, B., Sanchez, A., Vélez, J.F.: 'On the use of outer ear images for personal identification in security applications'. Proc. IEEE 33rd Annual Int. Carnahan Conf. Security Technology, Madrid, Spain, 1999, pp. 469–476

[9] Ghoualmi, L., Draa, A., Chikhi, S.: 'An ear biometric system based on artificial bees and the scale invariant feature transform', *Expert Syst. Appl.*, 2016, **57**, pp. 49–61

[10] Omara, I., Li, F., Zhang, H., *et al.*: 'A novel geometric feature extraction method for ear recognition', *Expert Syst. Appl.*, 2016, **65**, pp. 127–135

[11] Anwar, A.S., Ghany, K.K.A., Elmahdy, H.: 'Human ear recognition using geometrical features extraction', *Procedia Comput. Sci.*, 2015, **65**, pp. 529–537

[12] Sana, A., Gupta, P., Purkait, R.: 'Ear biometrics: a new approach', in Pal, P. (Ed.): '*Advances in pattern recognition*' (World Scientific, India, 2007), pp. 46–50

[13] Wang, Y., Mu, Z.-C., Zeng, H.: 'Block-based and multi-resolution methods for ear recognition using wavelet transform and uniform local binary patterns'. Proc. 19th Int. Conf. Pattern Recognition, Tampa, Florida, USA, 2008, pp. 1–4

[14] Jamil, N., AlMisreb, A., Halin, A.A.: 'Illumination-invariant ear authentication', *Proc. Comput. Sci.*, 2014, **42**, pp. 271–278

[15] Meraoumia, A., Chitroub, S., Bouridane, A.: 'An automated ear identification system using Gabor filter responses'. Proc. IEEE 13th Int. New Circuits and Systems Conf., Grenoble, France, 2015, pp. 1–4

[16] Hurley, D.J., Nixon, M.S., Carter, J.N.: 'Force field feature extraction for ear biometrics', *Comput. Vis. Image Underst.*, 2005, **98**, (3), pp. 491–512

[17] LeCun, Y., Boser, B., Denker, J.S., *et al.*: 'Backpropagation applied to handwritten zip code recognition', *Neural Comput.*, 1989, **1**, (4), pp. 541–551

[18] Jain, A.K., Nandakumar, K., Nagar, A.: 'Biometric template security', *EURASIP J. Adv. Signal Process.*, 2008, **2008**, p. 113

[19] Rathgeb, C., Uhl, A.: 'A survey on biometric cryptosystems and cancelable biometrics', *EURASIP J. Inf. Secur.*, 2011, **2011**, (1), p. 3

[20] Nagar, A., Nandakumar, K., Jain, A.K.: 'A hybrid biometric cryptosystem for securing fingerprint minutiae templates', *Pattern Recognit. Lett.*, 2010, **31**, (8), pp. 733–741

[21] Juels, A., Wattenberg, M.: 'A fuzzy commitment scheme'. Proc. 6th ACM Conf. Computer and Communications Security, Singapore, 1999, pp. 28–36

[22] Rathgeb, C., Uhl, A.: 'Statistical attack against fuzzy commitment scheme', *IET Biometrics*, 2012, **1**, (2), pp. 94–104

[23] Hao, F., Anderson, R., Daugman, J.: 'Combining crypto with biometrics effectively', *IEEE Trans. Comput.*, 2006, **55**, (9), pp. 1081–1088

[24] Bringer, J., Chabanne, H., Cohen, G., *et al.*: 'Theoretical and practical boundaries of binary secure sketches', *IEEE Trans. Inf. Forensics Sec.*, 2008, **3**, (4), pp. 673–683

[25] Tong, V.V.T., Sibert, H., Lecoeur, J., *et al.*: 'Biometric fuzzy extractors made practical: a proposal based on fingercodes'. Proc. Int. Conf. Biometrics, Seoul, Korea, 2007, pp. 604–613

[26] Juels, A., Sudan, M.: 'A fuzzy vault scheme', *Des. Codes Cryptogr.*, 2006, **38**, (2), pp. 237–257

[27] Charles Clancy, T., Kiyavash, N., Lin, D.J.: 'Secure smartcard based fingerprint authentication'. Proc. ACM SIGMM Workshop Biometrics Methods and Applications, Berkeley, California, USA, 2003, pp. 45–52

[28] Uludag, U., Pankanti, S., Jain, A.K.: 'Fuzzy vault for fingerprints'. Proc. Int. Conf. Audio-and Video-Based Biometric Person Authentication, New York, USA, 2005, pp. 310–319

[29] Nandakumar, K., Jain, A.K., Pankanti, S.: 'Fingerprint-based fuzzy vault: implementation and performance'. *IEEE Trans. Inf. Forensics Sec.*, 2007, **2**, (4), pp. 744–757

[30] Tams, B.: 'Unlinkable minutiae-based fuzzy vault for multiple fingerprints', *IET Biometrics*, 2016, **5**, (3), pp. 170–180

[31] Dong, J., Tan, T.: 'Security enhancement of biometrics, cryptography and data hiding by their combinations', 5th International Conference on Visual Information Engineering (VIE 2008), Xian China, 2008, pp. 239–244

[32] Freire-Santos, M., Fierrez-Aguilar, J., Ortega-Garcia, J.: 'Cryptographic key generation using handwritten signature', in Flynn, P.J., Pankanti, S. (Eds.): '*Biometric technology for human identification III*', vol. 6202 (International Society for Optics and Photonics, SPIE, USA, 2006), p. 62020N

[33] Lee, Y.J., Bae, K., Lee, S.J., *et al.*: 'Biometric key binding: fuzzy vault based on iris images'. Proc. Int. Conf. Biometrics, Seoul, Korea, 2007, pp. 800–808
[34] Chang, E.-C., Shen, R., Teo, F.W.: 'Finding the original point set hidden among chaff'. Proc. ACM Symp. Information, Computer and Communications Security, Taipei, Taiwan, 2006, pp. 182–188
[35] Li, Q., Guo, M., Chang, E.-C.: 'Fuzzy extractors for asymmetric biometric representations'. Proc. IEEE Computer Society Conf. Computer Vision and Pattern Recognition Workshops, Anchorage, Alaska, USA, 2008, pp. 1–6
[36] Ratha, N.K., Chikkerur, S., Connell, J.H., *et al.*: 'Generating cancelable fingerprint templates', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2007, **29**, (4), pp. 561–572
[37] Leng, L., Zhang, S., Bi, X., *et al.*: 'Two-dimensional cancelable biometric scheme'. Proc. Int. Conf. Wavelet Analysis and Pattern Recognition, Xian, China, 2012, pp. 164–169
[38] Pandey, R.K., Zhou, Y., Kota, B.U., *et al.*: 'Deep secure encoding for face template protection'. Proc. IEEE Conf. Computer Vision and Pattern Recognition Workshops, Las Vegas, NV, USA, 2016, pp. 77–83
[39] Pandey, R.K., Zhou, Y., Kota, B.U., *et al.*: 'Learning representations for cryptographic hash based face template protection', in Bhanu, B., Kumar, A. (Eds.): '*Deep learning for biometrics*' (Springer, Cham, Switzerland, 2017), pp. 259–285
[40] Meraoumia, A., Kadri, F., Bendjenna, H., *et al.*: 'Improving biometric identification performance using PCANet deep learning and multispectral palmprint', in Jiang, R., Al-maadeed, S., Bouridane, A., *et al.* (Eds.): '*Biometric security and privacy*' (Springer, Cham, Switzerland, 2017), pp. 51–69
[41] Paul, P.P., Gavrilova, M.: 'Multimodal biometric approach for cancelable face template generation', in Braun, J.J. (Ed.): '*Multisensor, multisource information fusion: architectures, algorithms, and applications*' vol. 8407, (International Society for Optics and Photonics, SPIE, USA, 2012), p. 84070H
[42] Dworkin, M.J.: 'SHA-3 standard: permutation-based hash and extendable-output functions'. Technical report, 2015
[43] Donahue, J., Jia, Y., Vinyals, O., *et al.*: 'DECAF: A deep convolutional activation feature for generic visual recognition'. Proc. Int. Conf. Machine Learning, Beijing, China, 2014, pp. 647–655
[44] Oquab, M., Bottou, L., Laptev, I., *et al.*: 'Learning and transferring mid-level image representations using convolutional neural networks'. Proc. IEEE Conf. Computer Vision and Pattern Recognition, Columbus, Ohio, USA, 2014, pp. 1717–1724
[45] Razavian, A.S., Azizpour, H., Sullivan, J., *et al.*: 'CNN features off-the-shelf: an astounding baseline for recognition'. Proc. IEEE Conf. Computer Vision and Pattern Recognition Workshops, Columbus, Ohio, USA, 2014, pp. 806–813
[46] Krizhevsky, A., Sutskever, I., Hinton, G.E.: 'ImageNet classification with deep convolutional neural networks'. Proc. Advances in Neural Information Processing Systems, Lake Tahoe, Nevada, USA, 2012, pp. 1097–1105
[47] Liu, S., Deng, W.: 'Very deep convolutional neural network based image classification using small training sample size', 3rd IAPR Asian Conference on Pattern Recognition (ACPR), Kuala Lumpur, 2015, pp. 730–734
[48] Russakovsky, O., Deng, J., Su, H., *et al.*: 'ImageNet large scale visual recognition challenge', *Int. J. Comput. Vis.*, 2015, **115**, (3), pp. 211–252
[49] Kingma, D., Ba, J.: 'Adam: a method for stochastic optimization', 3rd International Conference on Learning Representations (ICLR 2015), California, 2015
[50] Duchi, J., Hazan, E., Singer, Y.: 'Adaptive subgradient methods for online learning and stochastic optimization', *J. Mach. Learn. Res.*, 2011, **12**, pp. 2121–2159
[51] Tieleman, T., Hinton, G.: 'Lecture 6.5-rmsprop: divide the gradient by a running average of its recent magnitude', in '*COURSERA: neural networks for machine learning*' (2012), www.cs.toronto.edu/~tijmen/csc321/slides/lecture_slides_lec6.pdf
[52] Hochreiter, S., Schmidhuber, J.: 'Flat minima', *Neural Comput.*, 1997, **9**, (1), pp. 1–42
[53] Heusel, M., Ramsauer, H., Unterthiner, T., *et al.*: 'GANs trained by a two time-scale update rule converge to a local Nash equilibrium'. Proc. Advances in Neural Information Processing Systems, Long Beach, California, USA, 2017, pp. 6629–6640
[54] Srivastava, N., Hinton, G.E., Krizhevsky, A., *et al.*: 'Dropout: a simple way to prevent neural networks from overfitting', *J. Mach. Learn. Res.*, 2014, **15**, (1), pp. 1929–1958
[55] Ioffe, S., Szegedy, C.: 'Batch normalization: accelerating deep network training by reducing internal covariate shift', Proceedings of the 32nd International Conference on Machine Learning PMLR, 2015, pp. 448–456
[56] Menon, A.K.: '*Large-scale support vector machines: algorithms and theory*', vol. 117 (Research Exam, University of California, San Diego, 2009)
[57] Fan, R.-E., Chang, K.-W., Hsieh, C.-J., *et al.*: 'LIBLINEAR: a library for large linear classification', *J. Mach. Learn. Res.*, 2008, **9**, pp. 1871–1874
[58] Halko, N., Martinsson, P.-G., Tropp, J.A.: 'Finding structure with randomness: probabilistic algorithms for constructing approximate matrix decompositions', *SIAM Rev.*, 2011, **53**, (2), pp. 217–288
[59] Sotirov, A., Stevens, M., Appelbaum, J., *et al.*: 'Md5 considered harmful today, creating a rogue CA certificate'. Proc. 25th Annual Chaos Communication Congress, no. EPFL-CONF-164547, Berlin, Germany, 2008
[60] Schneier, B.: 'Schneier on security: cryptanalysis of SHA-1', 2005, Schneier.com
[61] Merkle, R.C.: 'Secrecy, authentication, and public key systems'. Ph.D. Dissertation, Stanford University, Stanford, CA, USA, AAI8001972, 1979
[62] Secure Hash Standard, Federal information processing standards publication 180-1, 1995
[63] Bertoni, G., Daemen, J., Peeters, M., *et al.*: 'Sponge functions'. Proc. ECRYPT Hash Workshop, Barcelona, Spain, vol. 2007, 2007
[64] Emersic, Z., Struc, V., Peer, P.: 'Ear recognition: more than a survey', *Neurocomputing*, 2017, **255**, pp. 26–39
[65] University of Science and Technology, Beijing, USTB database. http://www1.ustb.edu.cn/resb/en/index.htm, 2004
[66] Sinha, H., Manekar, R., Sinha, Y., *et al.*: 'Convolutional neural network-based human identification using outer ear images', in Bansal, J., Das, K., Nagar, A., *et al.* (Eds.): '*Soft computing for problem solving*' (Springer, Singapore, 2019), pp. 707–719
[67] Blum, A., Rivest, R.L.: 'Training a 3-node neural network is NP complete'. Proc. Advances in Neural Information Processing Systems, Denver, Colorado, USA, 1989, pp. 494–501
[68] He, K., Sun, J.: 'Convolutional neural networks at constrained time cost'. Proc. IEEE Conf. Computer Vision and Pattern Recognition, Boston, Massachusetts, USA, 2015, pp. 5353–5360
[69] Livni, R., Shalev-Shwartz, S., Shamir, O.: 'On the computational efficiency of training neural networks'. Proc. Advances in Neural Information Processing Systems, Montreal, Quebec, Canada, 2014, pp. 855–863
[70] UK Biometrics Working Group: 'Use of biometrics for identification and authentication: advice on product selection', Tech. report, UK, Government Office of the e-Envoy, 2002, www.cesg.gov.uk/Publications/Documents/biometricsadvice.pdf
[71] Hansley, E.E., Segundo, M.P., Sarkar, S.: 'Employing fusion of learned and handcrafted features for unconstrained ear recognition', *IET Biometrics*, 2018, **7**, (3), pp. 215–223
[72] Niu, X.-X., Suen, C.Y.: 'A novel hybrid CNN–SVM classifier for recognizing handwritten digits', *Pattern Recognit.*, 2012, **45**, (4), pp. 1318–1325