Pattern Recognition 76 (2018) 273–287

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/patcog

Cancellable speech template via random binary orthogonal matrices projection hashing

Kong-Yik Chee[a], Zhe Jin[b], Danwei Cai[c,d], Ming Li[c,d], Wun-She Yap[a,*], Yen-Lung Lai[a], Bok-Min Goi[a]

[a] Lee Kong Chian Faculty of Engineering and Science, Universiti Tunku Abdul Rahman, Malaysia
[b] School of Information Technology, Monash University Malaysia, Malaysia
[c] SYSU-CMU Joint Institute of Engineering, School of Electronics and Information Technology, Sun Yat-Sen University, China
[d] SYSU-CMU Shunde International Joint Research Institute, China

ARTICLE INFO

Article history:
Received 23 April 2017
Revised 21 September 2017
Accepted 30 October 2017
Available online 1 November 2017

Keywords:
Cancellable biometrics
Speaker recognition
RBOMP hashing
PF function
Security & privacy

ABSTRACT

The increasing advancement of mobile technology explosively popularizes the mobile devices (e.g. iPhone, iPad). A large number of mobile devices provide great convenience and cost effectiveness for the speaker recognition based applications. However, the compromise of speech template stored in mobile devices highly likely lead to the severe security and privacy breaches while the existing proposals for speech template protection do not completely guarantee the required properties such as unlinkability and non-invertibility. In this paper, we propose a cancellable transform, namely Random Binary Orthogonal Matrices Projection (RBOMP) hashing, to protect a well-known speech representation (i.e. i-vector). RBOMP hashing is inspired from Winner-Takes-All hash and further strengthened by the integration of the prime factorization (PF) function. Briefly, RBOMP hashing projects the i-vector using random binary orthogonal matrices and records the discrete value. Due to the strong non-linearity of RBOMP, the resultant hashed code withstands the template invertibility attack. Further, the experimental results suggest that the speech template generated using RBOMP hashing can still be verified with reasonable accuracy. Besides that, rigorous analysis shows that the proposed cancellable technique for speech resists several major attacks while the other criteria of biometric template protection can be justified simultaneously.

1. Introduction

Given the advancement of technologies and the increase in the popularity of mobile devices, speaker recognition system is emerging into a rapid growing field of research. In [1], Unar et al. stated the possibilities of using voice biometric modalities in different applications involving mobile commerce and transactions. Voice, consisting of unique features of different speakers, is often used to identify and verify the legitimate user in numerous applications. Typically, speaker recognition can be categorized as speaker identification and speaker verification. Speaker identification classifies a given voice to a specific speaker, while speaker verification decides a pair of voices as from the same speaker. State-of-the-art speaker recognition systems widely use i-vector modeling as a frontend technique to jointly model speaker and channel variabilities in a speech utterance due to its favorable performance as well as its condensed representation [2]. Moreover, Probabilistic Linear Discriminative Analysis (PLDA) is commonly adopted as a supervised backend modeling approach to strengthen speaker information while restraining channel variability and other sources of undesired variabilities [3–5]. Instances of speaker recognition systems that use both i-vector and PLDA can refer to [6,7]. It is worth mentioning that two general methods applying Deep Neural Network (DNN) to speaker recognition system brought impressive gains in performance. The first method trained a DNN acoustic model to produce frame alignments by the standard Gaussian Mixture Model (GMM) in the conventional framework [8]. The second method used the DNN acoustic model to extract phonetic features [9,10]. The phonetic features are the outputs of the bottleneck layer of a DNN or the low dimensional features after applying PCA to DNN's outputs of tied triphone state phoneme posterior probabilities. The phonetic features were then concatenated to Mel Frequency Cepstral Coefficient (MFCC) to generate tandem feature.

In i-vector/PLDA framework, a speaker recognition system can determine the authenticity of a user by matching the voice reference (i.e. i-vector) stored in the database. However, this raises the concern on the protection of the voice reference (also known as

* Corresponding author.
E-mail address: yapws@utar.edu.my (W.-S. Yap).

https://doi.org/10.1016/j.patcog.2017.10.041
0031-3203/

template) stored in the database to prevent security and privacy threats. In [11,12], it has shown that biometric template leakage is considered as one of the most harmful attacks in the biometric security system. The compromised biometric template can lead the impostor to create physical spoof from the stolen template, replace the template and gain illegitimate access to the system [12–14]. It is further complicated by the fact that biometric traits are irreplaceable once compromised. Therefore, a biometric-based application equipped with template protection capability is urgently needed.

In the literature, a number of proposals have been reported to secure the biometric templates. The existing proposals in protecting biometric template can be divided into three types: biometric cryptosystems (or helper data methods), feature transformation (or cancellable biometrics) and hybrid biometric cryptosystem [12]. Biometric cryptosystems require the usage of helper data, a biometric-dependent public information which does not reveal the original biometric template, to retrieve or generate keys. Instance of biometric cryptosystem can refer to [15]. The authentication process for this approach is to perform biometric comparison to determine the validity of the key retrieved or generated. Depending on how the helper data is derived, this approach can further be divided into key-binding or key-generation systems [16]. On the other hand, cancellable biometrics transforms the original biometric feature in such a way that it is computationally difficult to reconstruct the original biometric feature [16,17]. The advantages of using this approach is that the adversary is computationally hard to recover the original biometric feature even if the transformed feature vector had been compromised. However, the transformation of feature often leads to the loss of accuracy and this will likely degrade the performance of the biometric recognition system [17]. Instances of cancellable biometric can refer to [18,19]. Lastly, the hybrid biometric cryptosystem is the combination of biometric cryptosystems and cancellable biometrics to enjoy the strength from each type of method. An ideal template protection scheme is required and must fulfill all of the following requirements [20]:

1. Irreversibility. It should always be computationally hard for the adversary to invert the protected biometric template.
2. Unlinkability. It should always be computationally hard for the adversary to distinguish whether multiple protected biometric templates were generated using the same biometric trait of a user.
3. Revocability. The protected biometric template should be able to be revoked or renewed to replace the old template while the original template should be computationally hard to be inverted from multiple protected biometric templates derived from the same biometric trait of a user.
4. Performance. The performance of the biometric recognition rate should not be seriously degraded.

1.1. Related works

In this section, the previous works on the speech template protection are discussed and summarized. Generally, the revisit of the speech template protection schemes follows the categories of biometric template protection, i.e. cancellable biometrics, biometric cryptosystems and hybrid biometric cryptosystem [12].

1.1.1. Cancellable biometrics

Cancellable biometrics, the intentional distortion of the biometric feature, was formalized by Ratha et al. [21] to protect the privacy of the user. In the event that the cancellable feature is compromised, the same biometric feature can be mapped into another new distinct template using the pre-designed distortion characteristics. Cancellable biometrics can further be divided into biometric salting and non-invertible transformation.

Biometric salting [22] blends an auxiliary data (e.g. a user specific key or password) with the biometric feature. A concrete example of biometric salting for speech template protection is probabilistic random projection proposed by Chong and Teoh [23]. Two-dimensional principal component analysis was applied on the feature matrix before going through a random projection process via an externally derived pseudo random-number. The projected matrix was then fed into a Gaussian Mixture Model (GMM) to obtain probabilistic speaker models. The presented scheme was shown to be resisted from the stolen-token attacks where even if the token had been compromised, the recognition performance of the system was still able to retain at the feature vector level. However, the scheme was vulnerable to attack via record multiplicity (ARM) as the adversary can recover the original feature template by exploiting multiple templates generated using different random projection matrices [24].

Cancellable biometrics also often refers to the use of one-way transformation function that converts the voice feature to a protected template that is computationally hard to be inverted [22]. In 2008, Xu and Cheng [25] proposed a cancellable voice template protection method based on fuzzy vault scheme [26]. Chaff points were added to the unordered Mel-Frequency Cepstral Coefficient matrix to create a vault and a prime accumulator was used to separate the genuine points from chaff points. Besides, a non-invertible function was used to conceal the raw features while polynomial reconstruction was used for authentication. However, Chang et al. [27] revealed that the selection of the chaff points is not independent as the selection of new chaff point depends on the location of the previous selected point. It was observed that the latecomers, referring to the points added later, will likely to have more nearby points. Hence, increasing the number of chaff points will likely lead the adversary to correctly guess the genuine points. In addition to that, if the prime accumulator had been compromised, the adversary will be able to easily determine the genuine points.

Recently, Pandev et al. [28,29] proposed a new technique called deep secure encoding for protecting face template. The face features were first extracted and trained using deep convolutional neural networks to generate an unprotected binary template. The unprotected binary template was divided into n k-bit blocks. Each k-bit block was then fed as an input of a cryptographic hash function (e.g. SHA-256). Finally, the n outputs of hash function were stored in the database for matching purposes. During the matching phase, the face image is first queried. Subsequently, similar training and feature extraction processes will be carried out using the queried face image to generate an unprotected binary template. The unprotected template is then divided into n k-bit blocks as the inputs of the underlying hash function. The n outputs of the hash will then be compared with the hashed codes stored in the database. The matching is successful if i out of n outputs of the hash are matched where i must be greater than the predefined threshold value. The proposed scheme is interesting as a random key is chosen and is embedded during face extraction and training processes to generate an unprotected binary template while no key is needed to secure the unprotected binary template. If the template is compromised, a new key will be selected and the training process must be carried out again to re-generate a new unprotected binary template. Thus, Pandev et al. claimed that their scheme offers the property of cancellability without using key (where no key is needed after the feature extraction and training processes). This idea is different with typical template protection schemes where key is needed in securing the unprotected template to offer the property of cancellability. The size of protected template is of $n \times k$ bits. In the experiment performed by Pandev et al. using two different datasets (i.e. CMU PIE and Extended Yale B), the size of a protected template is of $64 \times 1024 = 65536$ bits.