Research Article

Sparse coding-based feature extraction for biometric remote authentication in Internet of Things

Haleh Amintoosi[1] · Ali Jaber Taresh[1]

Abstract
In recent years, Internet of Things (IoT) has attracted lots of attention. However, the security related issues such as authentication remain a challenge. The heterogeneity of IoT in terms of devices and communication makes most existing authentication mechanisms inapplicable. So, there is a need for a two-factor authentication mechanism to obtain an end-to-end authentication between IoT devices/applications. In this paper, we propose a sparse coding based feature extraction for biometric remote user authentication. The proposed scheme makes use of sparse codes hash operations and overcomplete dictionary to store/retrieve the biometric data efficiently. The performance analysis proves that the proposed method is robust against noise and able to obtain the accuracy of 0.97.

Keywords Internet of Things · Security · Authentication · Biometric · Sparse coding

1 Introduction

The basic idea of the Internet of Things is attaching embedded devices to everyday objects to make them smart objects/devices. These smart devices will be uniquely identified and able to communicate with each other to perform complex tasks for humanity.

In an unattended IoT environment, remote users can access IoT services through smart device applications in order to connect to any node. Once connected, the user can access desired information from specific nodes [1, 2]. The IoT node and the gateway node need to be totally assured of the legitimacy of the user, since he/she can access nodes data or might send commands to the nodes. Also, user needs that both the IoT node and gateway node are legitimate; so that the attacker cannot masquerade as node and transmit wrong data. This makes remote user authentication very crucial in IoT networks.

Traditional authentication schemes are normally dependent on single factors such as passwords. However, since IoT devices are vulnerable against physical attacks, and due to the insufficient security guarantee of single factor authentication approaches (e.g., passwords), it is reasonable to consider a second factor based on user's personal biometric characteristics such as fingerprint, iris scan, etc., which are hard to copy, and are more scalable.

The first step in authentication via biometric data is to extract the basic and dominant features from the data to be further used for authentication. For biometric data that changes with time and environment, a challenging problem is to perform the feature extraction in a way that it is resistant against noise. Moreover, since IoT devices have storage capacity constraints, finding a 'sparse' representation is also a challenge.

In this article, we present a sparse coding based feature extraction method for biometric remote user identification strategy with the aim of correctly authenticating users in the presence of noise and find an accurate sparse representation of dominant features of biometric data. The proposed scheme accurately extracts key biometric features based on an overcomplete dictionary and is robust against different noise levels.

✉ Haleh Amintoosi, amintoosi@um.ac.ir; Ali Jaber Taresh, ali1983jaber@gmail.com | [1]Computer Engineering Department, Faculty of Engineering, Ferdowsi University of Mashhad, Mashhad, Iran.

SN Applied Sciences (2019) 1:1098 | https://doi.org/10.1007/s42452-019-1135-7

Received: 21 May 2019 / Accepted: 20 August 2019

Published online: 26 August 2019

2 Related work

In recent years, many studies and surveys have been published to investigate the security issues related to IoT. Teh et al. [3] developed a prototype that utilizes an identity-based identification scheme to provide access control for incoming and outgoing personnel from the building. The novelty lies in the application of an identity-based identification scheme to ensure that no third-party malicious observers trying to break the system. Usman et al. [4] proposed a lightweight security algorithm which is a 64-bit block cipher with a 64-bit key. The algorithm architecture consists of Feistel and a uniform substitution-permutation network which was proved to provide the required security in just five encryption rounds. Li et al. [5] proposed a secure authentication scheme for IoT-based healthcare systems, able to guarantee user anonymity and avoid attacks such as password disclosure. Dhillon and Kalra [2] proposed a lightweight biometric based remote authentication scheme using Perceptual hash [6] function and XOR operation. Their method consists of four phases: user registration, login, authentication and password change.

In this paper, we propose a new feature extraction strategy based on sparse coding to extract powerful and robust features for biometric authentication. Our proposed method is in fact an extension to the work presented in [2] with the aim of increasing its resistance against inherent noise in biometric data. In particular, we incorporate K-Singular Value Decomposition (K-SVD) method to extract features from biometric data due to the fact that K-SVD is successful in providing a sparse representation of the data, which is essential in IoT environment. We also made use of Orthogonal Matching Pursuit (OMP) for feature extraction with the aim of improving the effectiveness of K-SVD and speeding up the process.

3 Feature extraction for biometric authentication

3.1 Network model

The proposed network model has 4 components: (1) IoT devices, able to collect sensitive information and communicate with other connected nodes, (2) a gateway node, connected to one or more IoT devices, which translates proprietary communication protocols to Internet Protocol. (3) The Internet and (4) PCs and mobile devices. By using the smart device, the user is able to access the IoT service through an application via Internet. We assume a network scenario for the user authentication in which, the communication is done between users and nodes in order to exchange information.

3.2 The proposed sparse coding-based feature extraction

We propose a new feature extraction strategy based on sparse coding to extract powerful and robust features from the biometric data. This strategy uses an overcomplete dictionary to match features and find dominant information, which leads to a sparse feature vector that is robust with respect to noise. We use K-SVD method to extract features from biometric data since K-SVD is able to provide a sparse representation, which is essential in IoT environment. We also leveraged OMP for feature extraction to improve the effectiveness of KSVD and speed up the process. In the following, we first describe these methods with more detail and then, explain the proposed method.

3.2.1 Basic concepts

Sparse representation modeling or sparse coding is a learning method which is focused at obtaining a sparse representation of the input data by keeping the combination of basic elements (i.e., atoms) of the data as well as those basic elements themselves. These atoms constitute a dictionary.

*K-Singular Value Decomposition (K-SVD) i*s a dictionary learning algorithm that utilizes a singular value decomposition approach to create a dictionary for sparse representations [7]. K-SVD works by iteratively alternating between sparse coding the input data, and updating the atoms to better fit the data [7].

Orthogonal Matching Pursuit (OMP) is an approximation algorithm that is leveraged to address the sparse coding problem. The basis of OMP is iteratively choosing the most correlated–with-the-residual columns.

3.2.2 The proposed method

Feature extraction and its role in biometric user authentication has been shown in Fig. 1. As shown in this figure, once biometric data is collected, its basic features are extracted via our proposed sparse coding-based feature extraction method. The result will be a feature vector which is then fed to the perceptual hashing process [2]. The output is a hash value representing the biometric data, saved for further authentication. As mentioned above, perceptual hashing has been used in order to produce a compact representation of the biometric data. The reason, as mentioned in [2] is the potential drawback in

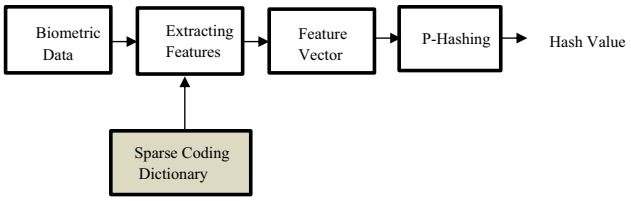

Fig. 1 Feature extraction role in biometric authentication

one-way common hash functions, which is the sensitivity to small changes of input, as for biometric data.

As mentioned above, an important issue in the sparsity of the solution and the accuracy of such representation is the selection of proper dictionary which can be done either by building a dictionary by utilizing a mathematical model of the data or by learning a dictionary to perform best on a training set [8]. Here, we choose the second method and utilize K-SVD to train the overcomplete dictionary to achieve the best representation for the data with strict sparsity constraints. We also use OMP for efficient approximation of sparse coding problem, which proves to be effective, especially from a run-time point of view [9].

In the following, we describe the dictionary learning process using K-SVD algorithm with more detail.

3.2.2.1 Construction of the dictionary The first step is the initialization of the dictionary. As previously mentioned, we select the training approach for the selection the dictionary and utilize K-SVD algorithm to train the overcomplete dictionary in order to achieve the best representation for the biometric data. As shown in Fig. 2, biometric data can be stored by matrix A in which, each column represents an image. In the first step, we create a dictionary D with the aim of presenting each image sparsely. D is first fixed and we consider the above optimization problem in order to find sparse representations with atoms summarized in matrix X.

3.2.2.2 Sparse coding In this step, we use an approximation algorithm to find a proper solution for the sparse representation. Matching Pursuit (MP) is one of the greedy algorithms that finds one atom at a time. It examines the dictionary to find the closest atom to the signal that best reduces Mean Square Error (MSE). The algorithm stops when MSE is below the distinction threshold. The orthogonal MP (OMP) is an improved version that re-evaluates the coefficients by least-square after each round. In each step, the column of dictionary that has the most correlation with the residual columns is selected and added into the already selected columns.

3.2.2.3 Dictionary update After the sparse coding step, the next is to find a better dictionary. This process is done by updating one column of the dictionary D at a time, while fixing X.

At the end, the result will be an efficient dictionary that stores the biometric data sparsely and can represent the data with high efficiency.

4 Experimental evaluation

4.1 Simulation setup

In order to evaluate the performance of our feature extraction method, we make use of extended Yale-B face database [10] that includes still images of 38 persons, each having 64 images taken under different illumination conditions. The simulation was performed on a laptop with Windows 7 Professional 64-bit operating system, Intel(R) Corei7-3610QM @2.30 GHz CPU and 4096 MB RAM in MATLAB2012. We used the K-SVD toolbox for dictionary training. The dictionary used was of size 100×625 and designed to handle image patches of size 8×8 pixels. The coefficients (e.g., atoms) were computed using OMP. The number of iterations was set to 20 and the average of results have been considered for evaluation.

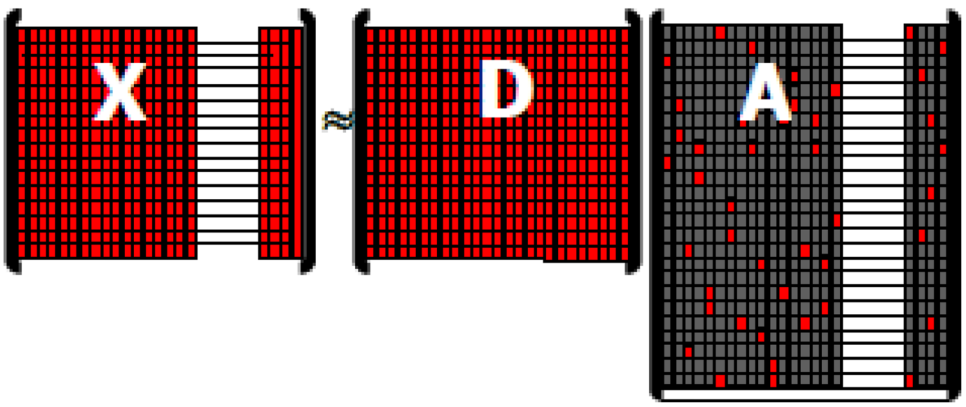

Fig. 2 Sparse coding using K-SVD

4.2 Simulation results

The quality of the trained dictionary can be evaluated computing the mean reconstruction error or Root Mean Square Error (RMSE). We expect that the quality of dictionary improves when dictionary is updated several times by K-SVD, resulting in a monotonic decrease in RMSE until reaching to a local minimum convergence. In order to figure out the performance of our proposed method, the evolution of RMSE was calculated at the end of each iteration, as shown in Fig. 3. As can be seen, RMSE is decreased as the dictionary becomes updated in subsequent iterations, leading to more accurate sparse representation.

In order to evaluate the performance in the presence of noise and for lower Signal to Noise Ratio (SNR), we selected a random face image from the dataset and added White Gaussian noise with varying SNR. We then applied the proposed method to observe the performance of sparse coding technique in denoising each patch of size 8×8 pixels from the noisy image. Figure 4 shows the results based on Peak SNR (PSNR) for different noise levels. As shown in this figure, the proposed method is robust against different levels of noise.

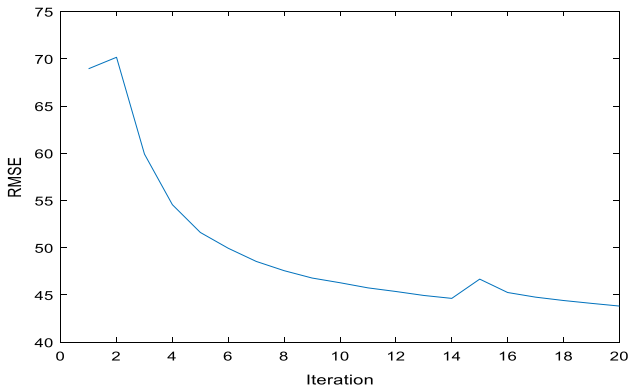

Fig. 3 Root Mean Square Error based on K-SVD algorithm

4.3 Performance analysis

The following provides the evaluation of the proposed strategy based on performance metrics. In order to calculate these metrics, we selected x = 2536 biometric data (e.g., face images) from the selected dataset and applied our proposed method to observe its performance. The results show that 2354 faces are authenticated correctly while the number of incorrect authentications is 182.

The overall accuracy of the proposed method is: $\frac{TP+TN}{x} = \frac{179+2303}{2536} = 0.97$.

The sensitivity or recall of the proposed feature is $\frac{TP}{TP+FN} = \frac{179}{179+3} = 0.983$.

The specificity of the proposed feature in accurately identifying is: $\frac{TN}{TN+FP} = \frac{2303}{2303+51} = 0.978$.

The precision of the proposed feature is raising alerts upon findings incorrect user is: $\frac{TP}{TP+FP} = \frac{179}{179+51} = 0.78$.

The prevalence of the proposed feature is: $\frac{TP+FN}{x} = \frac{179+3}{2536} = 0.072$.

To summarize, results show that the authentication method is successful in achieving a high accuracy rate of 0.97. Moreover, it is successful in denoising the data, thus making it robust against biometric data replication and manipulation attacks.

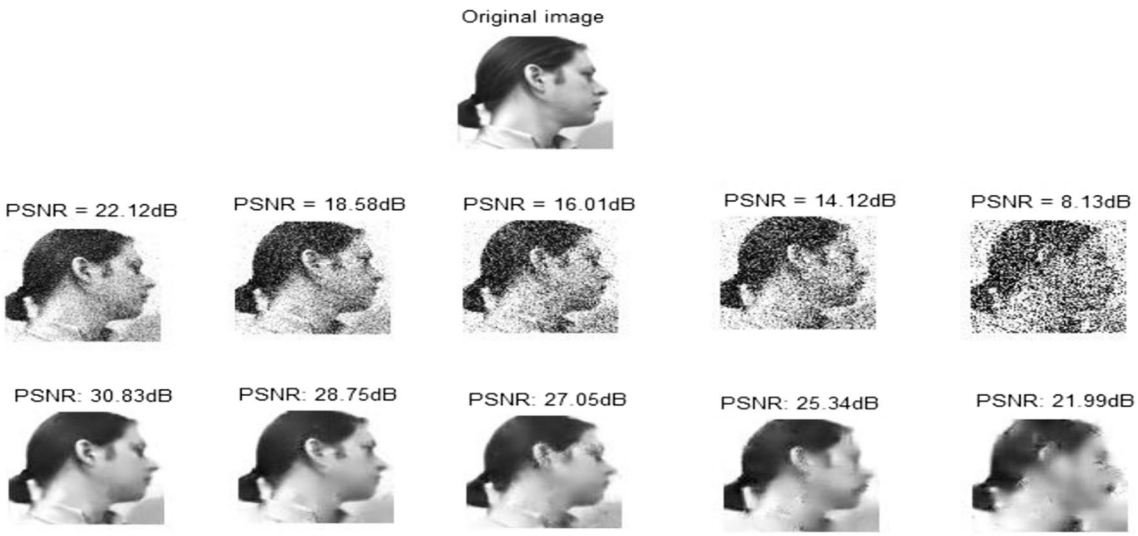

Fig. 4 Robustness of the proposed method based on K-SVD in facing different noise level (top row: noisy data, bottom row: reconstructed data based on K-SVD)

5 Conclusion

Security and privacy preservation for IoT devices is essential due to the connectivity of IoT devices and the sensitivity of the exchanged data. There seems to be the need for new authentication techniques to incorporate users' personal biometrics in order to improve the security of IoT applications. In this article, a new feature extraction scheme for sparse biometric user authentication has been proposed. The proposed scheme is based on sparse coding to extract powerful and robust features from the biometric data. The proposed method makes use of K-SVD to learn the dictionary and use OMP for feature extraction. Simulation results demonstrate that our proposed method is robust respect to the noise and is able to achieve the high accuracy rate of 0.97.

Compliance with ethical standards

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

1. Sfar AR et al (2018) A roadmap for security challenges in the Internet of Things. Digit Commun Netw 4(2):118–137
2. Dhillon PK, Kalra S (2017) A lightweight biometrics based remote user authentication scheme for IoT services. J Inf Secur Appl 34:255–270
3. Teh TY, Lee YS, Cheah ZY, Chin JJ (2017) IBI-mobile authentication: a prototype to facilitate access control using identity-based identification on mobile smart devices. Wirel Pers Commun 94(1):127–144
4. Usman M, et al (2017) A lightweight encryption algorithm for secure Internet of Things. arXiv:1704.08688
5. Li CT, Wu TY, Chen CL, Lee CC, Chen CM (2017) An efficient user authentication and user anonymity scheme with provably security for IoT-based medical care system. Sensors 17(7):1482
6. Niu XM, Jiao YH (2008) An overview of perceptual hashing. Acta Electron Sin 36(7):1405–1411
7. Aharon M, Elad M, Bruckstein A (2006) K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation. IEEE Trans Signal Process 54:34311–34322
8. Elad M, Aharon M (2006) Image denoising via sparse and redundant representations over learned dictionaries. IEEE Trans Image Process 15(12):3736–3745
9. Rubinstein R, Zibulevsky M, Elad M (2008) Efficient implementation of the K-SVD algorithm using batch orthogonal matching pursuit (technical report CS-2008-08). Computer Science Department, Technion
10. Georghiades AS, Belhumeur PN, Kriegman DJ (2001) From few to many: illumination cone models for face recognition under variable lighting and pose. IEEE Trans Pattern Anal Mach Intell 23(6):643–660