Edge and Cloud-aided Secure Sparse Representation for Face Recognition

Yitu Wang† Takayuki Nakachi† Hiroyuki Ishihara†
†NTT Network Innovation Laboratory, NTT Corporation
Yokosuka, Kanagawa, 239-0847 Japan
Email: {yitu.wang.dp, takayuki.nakachi.pu, hiroyuki.ishihara.my}@hco.ntt.co.jp

***Abstract*—Edge and cloud computing has recently emerged not only to meet the ever-increasing computation demands, but also to provide extra degree of diversity by collecting data from the mobile devices in service. However, this, in turn, has raised new technical challenges on the security issue, and calls for the design of new frameworks to exploit multi-device diversity. In this paper, we take the advantage of this benefit while preserving the privacy. Specifically, 1). To address the privacy issue, we develop a low-complexity encrypting algorithm based on random unitary transform, where it is proved both theoretically and through simulation that such encryption will not affect the result of face recognition. 2). To exploit multi-device diversity, we integrate the recognition results based on the dictionaries of each device into an aggregated output through ensemble learning, which has shown higher correctness of predictability than any individual methods. The designed framework not only contributes to the reduction of computation complexity at each device, but also proves to be effective and robust through simulation results.**

I. Introduction

Face recognition has long been an active area of research due to both the scientific challenge and its practical significance in a wide range of practical applications and future network paradigms, e.g., virtual reality applications and the Internet of Things (IoT) network. Significant theoretical and experimental research has been done to address this issue. With inspiration from the sparsity mechanism of the human vision system, the sparse representation based classification algorithms have received large attention [1]. [2] adopts K-Singular Value Decomposition (K-SVD) algorithm to learn a feature dictionary, next applies Orthogonal Matching Pursuit (OMP) to find the sparse representation of the testing image, then uses Support-Vector Machines (SVM) for face recognition. Another commonly adopted technique for face recognition is deep learning, which has been proved to be effective in extracting deep hierarchical features [3]. [4] adopts deep model, i.e., ConvNets, to extract high-level visual features, which can be used for improving the performance of face recognition. However, these techniques pose exigent requirements on computation capability, which cannot be easily satisfied by solely relying on mobile devices due to their limited computation resource.

Edge and cloud computing is emerged as a promising technology to provide cloud-computing capabilities at the edge of pervasive radio access networks in close proximity to mobile users, while reducing the traffic bottlenecks between edge and cloud in the core and backhaul networks [5]. In the literature, [6], [7] improve the computation efficiency of face recognition by offloading a part of the computation tasks to the edge and cloud. However, such strategies suffer from one major drawback, i.e., they use the edge and cloud simply to accelerate the computation, while neglecting the potential diversity provided by multiple devices. To exploit more dimensions of the network resources for not only satisfying the computation demands, but also improving the performance of face recognition, we allow the cloud to produce a joint face recognition result through combining the results based on the dictionaries from each device. We take one step further and try to study *What is the fundamental benefit of exploiting the multi-device diversity?* In the meantime, it is of great significance to prevent privacy leakage, especially when we allow the sharing of computing results by the cloud.

In this paper, we develop a framework for edge and cloud-aided face recognition based on privacy preserving sparse representation. The motivation and main contributions are summarized as follows,

1) **Preserve the privacy by random unitary transform:** Involved encrypting algorithms should ensure that dictionaries/recognition results can be trained/drawn from the encrypted images. Commonly adopted method that allows computation on ciphertexts, such as Homomorphic Encryption (HE) and secure Multi-Party Computation (MPC), is faced with the curse of dimensionality. To address this problem, we develop a low-complexity encrypting algorithm based on random unitary transform, where it is proved both theoretically and through simulation that such encryption will not affect the result of face recognition.
2) **Exploit multi-device diversity by ensemble learning:** The performance of the dictionary-based face recognition algorithms relies heavily on the number of training samples, where the excessive cost of bandwidth and storage makes it difficult to gather all the training samples at the cloud. Alternatively, with the diversity provided by the cloud, we integrate only the recognition results based on the dictionaries from each device into an aggregated output through ensemble learning, which proves to be effective and robust through simulation results.

The rest of this paper is organized as follows. Section II

978-9-0827-9703-9/19/$31.00 ©2019 IEEE

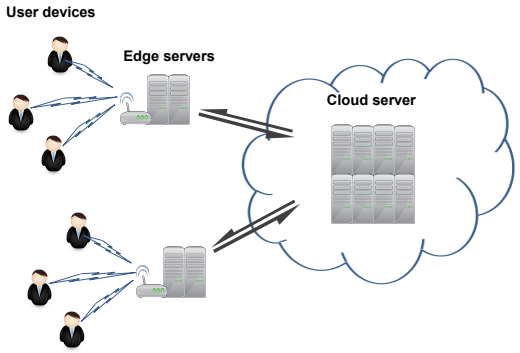


Fig. 1. Architecture of the edge and cloud-aided system

presents the system model. In Section III, we propose the edge and cloud-aided face recognition framework based on secure sparse representation. Following this, the performance of the proposed algorithm is evaluated in Sections IV. Finally, this paper concludes with Section V.

II. System Model

In this section, we first introduce the architecture of the edge and cloud-aided system. After discussing the method for face recognition based on sparse representation, we formulate the optimization problem under privacy preserving constraints.

A. Edge and Cloud-aided System

Consider an edge and cloud system as shown in Fig. 1, where N single core mobile devices, denoted as set $\mathcal{N}$, are assisted by M edge servers, denoted as set $\mathcal{M}$, and one remote cloud server C. The mobile devices are running applications involving face recognition[1], such as interactive gaming and virtual reality applications [5]. Among mobile devices in $\mathcal{N}$, L classes of person are to be recognized, denoted as $\mathcal{L}$, and each mobile device $j \in \mathcal{N}$ has b_i^j training samples for class $i \in \mathcal{L}$, denoted as $\mathcal{B}_i^j$ [8]. Each edge server in $\mathcal{M}$ is a light computing center deployed at a wireless access point, while a remote cloud is equipped with a stronger processor and connects with edge servers using the backbone network [9].

In the edge and cloud-aided computing scenario, a mobile device will offload its computation tasks to the edge server in close proximity via wireless channels[2]. The edge servers together with the cloud server will execute the computation tasks on behalf of the mobile device.

B. Sparse Representation for Face Images

The face recognition problem is defined as, using labeled training samples from L distinct classes to determine the class to which a new testing sample belongs. In order to achieve this goal, we adopt the *face subspace model* [8], which is formally defined as follows,

Definition 1 (Face Subspace Model)**.** *Given the training samples $\mathcal{B}_i^j, i \in \mathcal{L}, j \in \mathcal{N}$, each of which can be expressed as a column vector $\boldsymbol{d}_{(i,n)}^j \in \mathbb{R}^{m\times 1}, n \in \{1, 2, \cdots, b_i^j\}$ by stacking its columns. A dictionary $\boldsymbol{D}_i^j$ can be formulated accordingly as $\boldsymbol{D}_i^j = \left[\boldsymbol{d}_{(i,1)}^j, \boldsymbol{d}_{(i,2)}^j, \cdots, \boldsymbol{d}_{(i,b_i^j)}^j\right] \in \mathbb{R}^{m\times b_i^j}$. Any testing sample from the same class $\boldsymbol{y}_i^j \in \mathbb{R}^{m\times 1}$ approximately lies in the subspace spanned by $\mathcal{B}_i^j$, which can be expressed as follows*

$$\boldsymbol{y}_i^j = \boldsymbol{D}_i^j \boldsymbol{X}_i^j, \tag{1}$$

where $\boldsymbol{X}_i^j = \left[x_{(i,1)}^j, x_{(i,2)}^j, \cdots, x_{(i,b_i^j)}^j\right]^T \in \mathbb{R}^{b_i^j\times 1}$ is the weight of each component. ■

Note that the face subspace model is flexible enough to capture much of the variation in real data sets, and proved to be effective in the context of face recognition.

By grouping samples from all the classes $\mathcal{L}$ for device $j \in \mathcal{N}$, a dictionary $\boldsymbol{D}^j$ is formulated as follows,

$$\boldsymbol{D}^j = \left[\boldsymbol{D}_1^j, \boldsymbol{D}_2^j, \cdots, \boldsymbol{D}_L^j\right]. \tag{2}$$

According to Definition. 1, any testing image $\boldsymbol{y}^j, j \in \mathcal{N}$ can be sparsely represented over the dictionary $\boldsymbol{D}^j$,

$$\boldsymbol{y}^j = \boldsymbol{D}^j \boldsymbol{X}^j, \tag{3}$$

where $\boldsymbol{X}^j = \left[\boldsymbol{X}_1^j; \boldsymbol{X}_2^j; \cdots; \boldsymbol{X}_L^j\right] \in \mathbb{R}^{\sum_i b_i^j \times 1}$ is the sparse coefficient.

If $\sum_{i\in\mathcal{L}} b_i^j > m$ and $\boldsymbol{D}^j$ is a full-rank matrix, then Eq. (3) is under-determined, such that its solution is not unique. Alternatively, this difficulty can be resolved by solving the following l^1-minimization problem,

$$(P_0) \quad \hat{\boldsymbol{X}}^j = \arg\min_{\boldsymbol{X}^j} \left|\left|\boldsymbol{D}^j\boldsymbol{X}^j - \boldsymbol{y}^j\right|\right|_2 \quad \text{s.t.} \quad ||\boldsymbol{X}^j||_1 \leq \epsilon, \tag{4}$$

where ϵ represents the sparsity constraint. The above optimization problem can be efficiently solved using Matching Pursuit (MP) or Orthogonal Matching Pursuit (OMP).

C. Problem Formulation

Given a testing sample $\boldsymbol{y}^j, j \in \mathcal{N}$, we first compute its sparse representation $\hat{\boldsymbol{X}}^j$ by solving P_0 in Eq. (4). In the ideal case, the nonzero entries in $\hat{\boldsymbol{X}}^j$ should be associated with the columns of $\boldsymbol{D}^j$ from a single class. For example, if $\hat{\boldsymbol{X}}_l^j = \boldsymbol{0}, \forall l \neq i$, then we can easily assign $\boldsymbol{y}^j$ to class i. However, due to noise and modeling error, there exists small nonzero entries associated with other classes. To address this problem, define $\boldsymbol{\delta}_l^j = [0, \cdots, 0, 1, \cdots, 1, 0, \cdots, 0], j \in \mathcal{N}, l \in \mathcal{L}$, whose nonzero entries not only correspond to the entries in $\boldsymbol{X}^j$, but also associate with the l-th class only. Using merely the coefficients associated with the l-th class, we can approximate the testing sample $\boldsymbol{y}^j$ by $\boldsymbol{D}^j\boldsymbol{\delta}_l^j\hat{\boldsymbol{X}}^j$, then classify $\boldsymbol{y}^j$ according to the following optimization problem,

$$(P_1) \quad \min_l r_l^j(\boldsymbol{y}^j) = \left|\left|\boldsymbol{y}^j - \boldsymbol{D}^j\boldsymbol{\delta}_l^j\hat{\boldsymbol{X}}^j\right|\right|_2^2, \tag{5}$$

[1]All the captured images are pre-processed by the mobile devices, i.e., images are segmented into fine-grained head pictures [8].

[2]Some physical layer access scheme, e.g., Code Division Multiple Access (CDMA), is adopted to allow multiple devices to share the same edge server simultaneously and efficiently. There are also existing algorithms for the matching between mobile devices and edge servers based on their channel quality [9]. In this paper, we assume such matching is accomplished during network setup period.

where $r_l^j(\boldsymbol{y}^j)$ represents the class specific approximation error.

While such a scheme has been shown to be effective in face recognition [8], the classifying decision is made based on local dictionary only, which makes this method vulnerable to noise and modeling error. In the edge and cloud-aided system, our objective is to construct a framework to further minimize the reconstruction error by exploiting the multi-device diversity, while ensuring the security of the information passing during the whole process, which can be formally formulated as

$$\begin{aligned}(P_2)\quad \min_{l,j}\ & \overline{r}_l^j(\overline{\boldsymbol{y}}^k) = \left|\left|\overline{\boldsymbol{y}}^k - \overline{\boldsymbol{D}}^j\boldsymbol{\delta}_l^j\tilde{\boldsymbol{X}}^{(j,k)}\right|\right|_2^2 \\ s.t.\ \ & \overline{\boldsymbol{y}}^k = f(p,\boldsymbol{y}^k) \\ & \overline{\boldsymbol{D}}^j = f(p,\boldsymbol{D}^j) \\ & r_l^j(\boldsymbol{y}^k) = \overline{r}_l^j(\overline{\boldsymbol{y}}^k),\end{aligned} \tag{6}$$

where $f(\cdot)$ is the encrypting function, p is the key for encryption, and $\tilde{\boldsymbol{X}}^{(j,k)}, j,k \in \mathcal{N}$ represents the sparse representation of $\overline{\boldsymbol{y}}^k$ under $\overline{\boldsymbol{D}}^j$. The first two constraints guarantee the security of the system, and the last one ensures the algorithm operating on secured plane without performance loss.

III. Edge and Cloud-aided Secure Sparse Representation for Face Recognition

In this section, we propose the framework for edge and cloud-aided secure sparse representation for face recognition. To satisfy the privacy preserving constraints in Eq. (6), we briefly introduce random unitary transform and outline three important properties, based on which we prove that the result of face recognition will not be influenced. To utilize the multi-device diversity, we propose a two-stage ensemble learning framework, 1). The sparse representation and the associated reconstruction error is calculated according to each dictionary at the cloud, which serves as a member classifier. 2). These member classifiers are combined into an aggregated classifier by solving P_2 in Eq. (6) to obtain a refined solution.

A. Random Unitary Transform

In order to not only preserve the privacy of the system, but also enable algorithms operating on secure plane, the random unitary transform is one promising method, which is proved to be effective for biometric template protection and network Brain Machine Interface (BMI) coding [10], [11].

Any vector $\boldsymbol{v} \in \mathbb{R}^{m\times 1}$ encrypted by random unitary matrix $\boldsymbol{Q}_p \in \mathbb{C}^{m\times m}$ with private key p can be expressed as follows,

$$\overline{\boldsymbol{v}} = f(p,\boldsymbol{v}) = \boldsymbol{Q}_p\boldsymbol{v}, \tag{7}$$

where $\overline{\boldsymbol{v}}$ is the encrypted vector, and the unitary matrix $\boldsymbol{Q}_p$ satisfies

$$\boldsymbol{Q}_p^*\boldsymbol{Q}_p = \boldsymbol{I}, \tag{8}$$

where $[\cdot]^*$ and $\boldsymbol{I}$ represents the Hermitian transpose and identity matrix, respectively. Gram-Schmidt orthogonalization can be adopted for generating $\boldsymbol{Q}_p$[3]. The encrypted vector has three properties [12] as follows,

[3]Such encrypting technique has been proved to be robust in terms of brute-face attack, diversity and irreversibility [13]. The security can be further enhanced by updating the private key periodically.

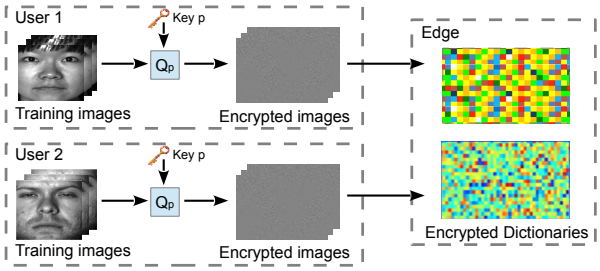


Fig. 2. Training: Generating encrypted dictionaries at each edge server

- Conservation of the Euclidean distances

$$||\boldsymbol{v}_i - \boldsymbol{v}_j||_2^2 = ||\overline{\boldsymbol{v}}_i - \overline{\boldsymbol{v}}_j||_2^2, \tag{9}$$

- Norm isometry

$$||\boldsymbol{v}||_2^2 = ||\overline{\boldsymbol{v}}||_2^2, \tag{10}$$

- Conservation of inner products

$$\boldsymbol{v}_i \times \boldsymbol{v}_j^T = \overline{\boldsymbol{v}}_i \times \overline{\boldsymbol{v}}_j^T. \tag{11}$$

B. Secure Sparse Representation and Recognition

According to random unitary transform, the encrypted training samples $\overline{\boldsymbol{d}}_{(i,n)}^j \in \mathbb{R}^{m\times 1}, i \in \mathcal{L}, j \in \mathcal{N}, n \in \{1,2,\cdots,b_i^j\}$ and testing samples $\overline{\boldsymbol{y}}^j, j \in \mathcal{N}$ are generated as follows,

$$\begin{aligned}\overline{\boldsymbol{d}}_{(i,n)}^j &= f\left(p,\boldsymbol{d}_{(i,n)}^j\right) = \boldsymbol{Q}_p\boldsymbol{d}_{(i,n)}^j, \\ \overline{\boldsymbol{y}}^j &= f(p,\boldsymbol{y}^j) = \boldsymbol{Q}_p\boldsymbol{y}^j.\end{aligned} \tag{12}$$

According to Eq. (2), the encrypted dictionary $\overline{\boldsymbol{D}}^j, j \in \mathcal{N}$ is generated as follows,

$$\overline{\boldsymbol{D}}^{\boldsymbol{j}} = f\left(p,\boldsymbol{D}^{\boldsymbol{j}}\right) = \boldsymbol{Q}_p\boldsymbol{D}^{\boldsymbol{j}}. \tag{13}$$

To obtain the secure sparse representation, we consider the following optimization problem operating on secured plane,

$$(P_3)\quad \tilde{\boldsymbol{X}}^{(j,k)} = \arg\min_{\boldsymbol{X}^j}\left|\left|\overline{\boldsymbol{D}}^j\boldsymbol{X}^j - \overline{\boldsymbol{y}}^k\right|\right|_2 \text{s.t.}\quad ||\boldsymbol{X}^j||_0 \le \epsilon. \tag{14}$$

It is proved that $\tilde{\boldsymbol{X}}^{(j,k)}$ by solving Eq. (14) is exact the same as $\hat{\boldsymbol{X}}^{(j,k)}$ by solving Eq. (4) [14].

In the following theorem, we prove that the result of face recognition is not affected as well.

Theorem 1. *The result $\overline{r}_l^j(\overline{\boldsymbol{y}}^k)$ by solving Eq. (6) is exact the same as the result $r_l^j(\boldsymbol{y}^k)$ by solving*

$$(P_4)\quad \min_{l,j} r_l^j(\boldsymbol{y}^k) = \left|\left|\boldsymbol{y}^k - \boldsymbol{D}^j\boldsymbol{\delta}_l^j\hat{\boldsymbol{X}}^{(j,k)}\right|\right|_2^2. \tag{15}$$

■

Proof: Observing $\overline{r}_l^j(\overline{\boldsymbol{y}}^k)$ is usually small, we have

$$\overline{r}_l^j(\overline{\boldsymbol{y}}^k) = ||\overline{\boldsymbol{y}}^k||_2^2 - \frac{\overline{\boldsymbol{y}}^k\left(\overline{\boldsymbol{D}}^j\boldsymbol{\delta}_l^j\hat{\boldsymbol{X}}^{(j,k)}\right)^T}{||\overline{\boldsymbol{D}}^j\boldsymbol{\delta}_l^j\hat{\boldsymbol{X}}^{(j,k)}||_2^2}. \tag{16}$$

From the properties of the unitary transform, we have $||\overline{\boldsymbol{y}}^k||_2^2 = ||\boldsymbol{y}^k||_2^2, \forall j \in \mathcal{N}$ (Norm isometry), $\overline{\boldsymbol{y}}^k\overline{\boldsymbol{D}}^j\boldsymbol{\delta}_l^j\hat{\boldsymbol{X}}^{(j,k)} =$

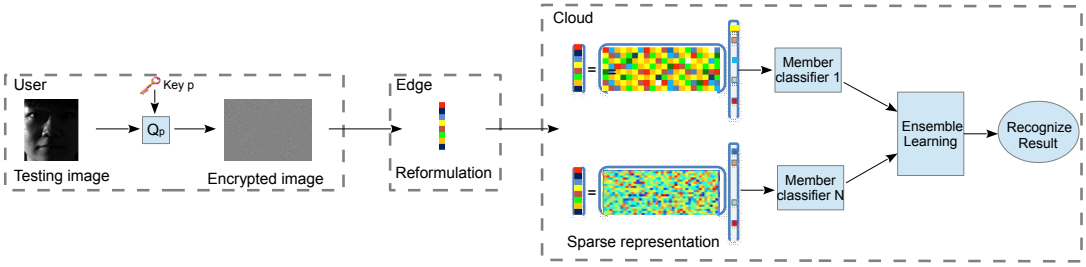


Fig. 3. Recognizing: Secure face recognition at the cloud

$\boldsymbol{y}^k \boldsymbol{D}^j \boldsymbol{\delta}_l^j \hat{\boldsymbol{X}}^{(j,k)}, \forall j \in \mathcal{N}, l \in \mathcal{L}$ (Conservation of inner products), and $\overline{\boldsymbol{D}}^j \boldsymbol{\delta}_l^j \hat{\boldsymbol{X}}^{(j,k)} = \boldsymbol{D}^j \boldsymbol{\delta}_l^j \hat{\boldsymbol{X}}^{(j,k)}, \forall j \in \mathcal{N}, l \in \mathcal{L}$ (Norm isometry). Therefore, Eq. (16) can be rewritten as follows,

$$\overline{r}_l^j(\overline{\boldsymbol{y}}^k) = ||\boldsymbol{y}^k||_2^2 - \frac{\boldsymbol{y}^k \left(\boldsymbol{D}^j \boldsymbol{\delta}_l^j \hat{\boldsymbol{X}}^{(j,k)}\right)^T}{||\boldsymbol{D}^j \boldsymbol{\delta}_l^j \hat{\boldsymbol{X}}^{(j,k)}||_2^2} = \left|\left|\boldsymbol{y}^k - \boldsymbol{D}^j \boldsymbol{\delta}_l^j \hat{\boldsymbol{X}}^{(j,k)}\right|\right|_2^2, \tag{17}$$

which is the same as $r_l^j(\boldsymbol{y}^k)$ in P_4 in Eq. (15). ■

Therefore, we can safely draw the conclusion that, by adopting unitary random transform, the designed algorithm operates on secured plane without any performance degradation.

C. Ensemble Learning Framework

Observing that the training samples are differently and independently chosen according to different devices, the relative uniqueness of the information available in each dictionary prompts the member classifiers to capture different patterns. In order to take such advantage, we propose an edge and cloud-aided ensemble learning framework, where the training stage and recognizing stage are involved.

1) Dictionary Training:

The dictionary training stage consists of three steps, as shown in Fig. 2. First, each device $j \in \mathcal{N}$ encrypts all the training images, and transmits to the designated edge server. Then, the encrypted dictionary $\overline{\boldsymbol{D}}^j$ is formulated according to Eq. (13). Finally, the encrypted dictionaries $\overline{\boldsymbol{D}}^j, \forall j \in \mathcal{N}$ are transmitted to the remote cloud server.

Since the training samples are not directly transmitted to the cloud, the amount of required network bandwidth between the edge servers and the cloud can be notably reduced. [4]

2) Face Recognizing:

The face recognizing stage is illustrated in Fig. 3. First, a device k encrypts the testing image $\boldsymbol{y}^k$, and transmits to the designated edge server. Next, the edge server transmits the reformulated encrypted testing image for recognition to the cloud. Then, each encrypted dictionary serves as a member classifier of the ensemble learning framework. Upon receiving a encrypted testing image $\overline{\boldsymbol{y}}^k$, the each member classifier $j \in \mathcal{N}$ on the cloud will calculate the sparse representation $\tilde{\boldsymbol{X}}^{(j,k)}$ by solving P_3 in Eq. (14). The reconstruction error $r_l^j(\overline{\boldsymbol{y}}^k)$ as well as the classification result are obtained by solving the secured version of P_1 in Eq. (5). Finally, we combine the results of member classifiers by solving P_2 in Eq. (6).

[4] [8] suggests that the dimension of dictionary for face recognition can be reduced significantly only with little influence on the performance.

Algorithm 1 Edge and Cloud-aided Secure Sparse Representation for Face Recognition

1: **[Training Stage]**
2: Training images $\boldsymbol{d}_{(i,n)}^j$ are encrypted according to Eq. (12), and transmitted to the designated edge server.
3: The edge servers formulate encrypted dictionaries $\overline{\boldsymbol{D}}^j$ according to Eq. (13), and upload to the cloud.
4: **[Testing Stage]**
5: Testing image $\boldsymbol{y}^k$ is encrypted according to Eq. (12), and transmitted to the designated edge server.
6: The edge servers upload the formulated encrypted testing vector $\overline{\boldsymbol{y}}^k$ to the cloud.
7: Each member classifier $j \in \mathcal{N}$ finds the sparse representation $\tilde{\boldsymbol{X}}^{(j,k)}$ by solving P_3 in Eq. (14) using OMP, and then calculates reconstruction error $\overline{r}_l^j(\overline{\boldsymbol{y}}^k)$ by solving the secured version of P_1 in Eq. (5) in a parallel manner.
8: The cloud combines the result and generates the recognition result by solving P_2 in Eq. (6).

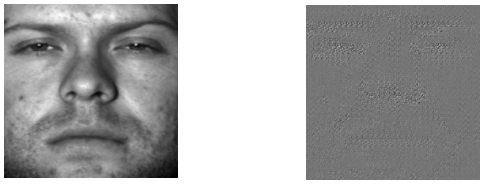
On the computational complexity of Algorithm 1, the most time-consuming part is the OMP algorithm in Line 7, in which the running time is $O(maz)$ for each member classifier, where m denotes the dimension of the testing image, $a = \max_{j \in \mathcal{N}} \sum_{i \in \mathcal{L}} b_i^j$ represents the number of rows in the dictionary, and z is the sparsity of the sparse representation.

IV. Simulation Results

In this section, we investigate the performance of the proposed framework by simulation. For performance comparison, we adopt three baseline algorithms,

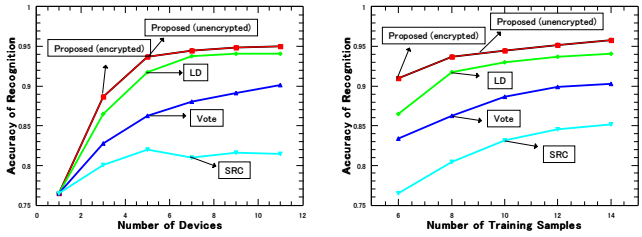
- Stable Sparse Representation-based Classification (SRC) by solving the l^1-minimization problem in Eq. (4) [8].
- Aggregating the training samples from all the devices in $\mathcal{N}$, and formulating a Large Dictionary (LD) at the cloud, and perform SRC for face recognition.
- Combining the results of member classifiers through Majority Voting (Vote) at the cloud [15].

The Extended YaleB database is one of the commonly used database for face recognition, the cropped and normalized face images are captured according to different angles and lighting conditions [16], [17]. We randomly select 32 images for each individual as the entire training set while the rest for testing. For each device, the training samples are randomly selected from the entire training set with replacement.



(a) Original image (b) Encrypted image

Fig. 4. Security verification



(a) 8 Training samples on average (b) 5 Devices

Fig. 5. Performance evaluation

Fig. 4 demonstrates the privacy-preserving property of the proposed framework. It is shown that the original image is difficult to be recognized from the encrypted one, and it would be computationally expensive to obtain the original image without the knowledge of the private key.

Fig. 5 provides the performance comparison among the proposed and baseline algorithms. The proposed algorithm achieves superior performance than all the baseline algorithms. First, majority voting, as an universal ensemble approach, does not perform well, which signifies the importance to design a good combiner that is related to the optimization problem. Second, as for the LD algorithm, 1). The computation complexity is $O(Nmaz)$, which is larger than that of the proposed algorithm. 2.) It requires large bandwidth between edge and cloud. Third, it is verified that by adopting random unitary transform, the proposed framework operates on secured plane without any performance degradation. Finally, as the number of devices grows, the performance improvement brought by ensemble learning is significant.

One interesting phenomenon is observed in Table. I. When we use ensemble learning in the case where one member classifier has full-knowledge of the entire training set, it is still possible to further enhance the performance. Moreover, observing that even though the performance of member classifier 2 is very weak, the result of ensemble learning is not influenced, which demonstrates the robustness of the proposed framework to the existence of a few weak member classifiers.

V. Conclusions

In this paper, we develop a framework for edge and cloud-aided face recognition based on secure sparse representation. To guarantee the privacy, we adopt random unitary transform, with which algorithms proves to be compatible with the secured plane. To reduce the computation demands at each device, the dictionary learning is conducted at each edge server, and the recognition is accomplished at the cloud. To exploit multi-device diversity, we combine only the computation results of each member classifier into an aggregated output. Finally, the simulation results verify the secure property as well as the superiority of the proposed framework.

TABLE I
Recognition Accuracy

Device 1 (Entire training set)	0.9430
Device 2 (8 training samples per class)	0.7540
Device 3 (10 training samples per class)	0.8482
Device 4 (10 training samples per class)	0.8491
Device 5 (12 training samples per class)	0.9009
Cloud (Proposed framework)	0.9544

References

[1] Y. Xu, Z. Li, J. Yang, and D. Zhang, "A survey of dictionary learning algorithms for face recognition," *IEEE Access*, vol. 5, pp. 8502-8514, Apr. 2017.

[2] Z. Jiang, Z. Lin, and L. Davis, "Learning a discriminative dictionary for sparse coding via label consistent K-SVD," *Proc. of IEEE CVPR 2011*, pp. 1697-1704, Jun. 2011.

[3] B. Shickel, P. Tighe, A. Bihorac, and P. Rashidi, "Deep EHR: A survey of recent advances in deep learning techniques for electronic health record (EHR) analysis," *IEEE J. Biomed. Health Informat.*, vol. 22, no. 5, pp. 1589-1604, Sept. 2018.

[4] R. Ranjan, V. Patel, and R. Chellappa, "Hyperface: A deep multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 41, no. 1, pp. 121-135, Jan. 2019.

[5] T. Taleb, K. Samdanis, B. Mada, H. Flinck, S. Dutta, and D. Sabella, "On multi-access edge computing: A survey of the emerging 5G network edge cloud architecture and orchestration," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 3, pp. 1657-1681, Sept. 2017.

[6] P. Hu, H. Ning, T. Qiu, H. Song, Y. Wang, and X. Yao, "Security and privacy preservation scheme of face identification and resolution framework using fog computing in internet of things," *IEEE Internet Things J.*, vol. 4, no. 5, pp. 1143-1155, Oct. 2017.

[7] P. Hu, H. Ning, T. Qiu, Y. Zhang, and X. Luo, "Fog computing based face identification and resolution scheme in internet of things," *IEEE Trans. Ind. Informat.*, vol. 13, no. 4, pp. 1910-1920, Aug. 2017.

[8] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 31, no. 2, pp. 210-227, Feb. 2009.

[9] X. Chen, L. Jiao, W. Li, and X. Fu, "Efficient multi-user computation offloading for mobile-edge cloud computing," *IEEE/ACM Trans. Netw.*, vol. 24, no. 5, pp. 2795-2808, Oct. 2016.

[10] I. Nakamura, Y. Tonomura, and H. Kiya, "Unitary transform-based template protection and its application to l^2-norm minimization problems," *IEICE Trans. Inform. Sys.*, vol. E99-D, no.1, pp. 60-68, Jan. 2016.

[11] T. Nakachi, H. Ishihara, and H. Kiya, "Privacy-preserving network BMI decoding of covert spatial attention," *Proc. of IEEE ICSPCS 2018*, pp. 1-8, Dec. 2018.

[12] T. Maekawa, T. Nakachi, S. Shiota, and H. Kiya, "Privacy-preserving SVM computing by using random unitary transformation," to appear in *Proc. of IEEE ISPACS 2019*, arXiv:1809.07055.

[13] Y. Saito, I. Nakamura, S. Shiota, and H. Kiya, "An efficient random unitary matrix for biometric template protection," *Joint Proc. of SCIS 2016 and ISAIS 2016,* pp. 366-370, 2016.

[14] T. Nakachi, and H. Kiya, "Practical secure OMP computation and its application to image modeling," *Proc. of ACM ICIHIP 2018*, pp. 25-29, Sept. 2018.

[15] L. Yu, S. Wang, and K. Lai, "Credit risk assessment with a multistage neural network ensemble learning approach," *Expert syst. appl.*, vol. 34, no. 2, pp. 1434-1444, Feb. 2008.

[16] A. Georghiades, P. Belhumeur, and D. Kriegman, "From few to many: Illumination cone models for face recognition under variable lighting and pose," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 23, no. 6, pp. 643-660, Jun. 2001.

[17] K. Lee, J. Ho, and D. Kriegman, "Acquiring linear subspaces for face recognition under variable lighting," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 27, no. 5, pp. 684-698, May 2005.