Robust feature modeling for face authentication in smart device

Ao Li [1*], Xin Liu [1], Hailong Jiang [2]

1. School of Computer Science and Technology, Harbin University of Science and Technology, Harbin, China

2. Department of Computer Science, Kent State University, USA

Email: {dargonboy@126.com (Ao Li), 1615927278@qq.com (Xin Liu), hjiang13@kent.edu (Hailong Jiang)}

***Abstract*—As smart devices become widespread, security has emerged as a crucial problem for users. Biometrics is an effective means of user authentication that has been applied to smart devices. Face authentication has been widely studied in this context because of its short interaction time and quick authentication. In this paper, we propose robust feature modeling for face authentication in smart devices. Our approach uses coefficients of low-rank representation as a measure of similarity between samples. Label regression and reconstructed label information are then used to supervise feature subspace learning. We added three types of noise and interference to three public face datasets to simulate potential problems in face authentication on smart devices. The results of the experiment show that our proposed approach can achieve better authentication performance than conventional methods.**

Keywords—smart device; face authentication; feature subspace learning; low-rank representation

I. Introduction

Since the advent of mobile technology, cellphones have evolved from call-only implements to today's full-featured smart mobile devices. Due to the rapid development of mobile devices, smart devices have become essential tools in many people's daily lives. Not only are we using these devices to communicate with others and propagate news, we also regard them as data hubs, and this includes e-mails, photos, browsing history, and even passwords. Even though all these sensitive data are contained in a device, many users choose not to protect their devices. Thus, a successful attack on a smart mobile device or its loss can have dire consequences, such as loss or leakage of important or sensitive information.

User authentication on smart devices is a critical part of ensuring device security. Most smart devices today have authentication capabilities, and various measures for authenticating them have been proposed, such as multi-touch gestures on the screen [1], gait recognition [2], and device movement patterns [3]. Gestures on the touchscreen provide effective security, but if we check our devices 100 times a day, unlocking them takes approximately two seconds each time, which means that we spend over three minutes unlocking them every day.

To further improve efficiency, biometrics is widely used in smart devices. Biometric technology uses the inherent physiological characteristics of the human body, such as fingerprints, faces, and irises, and such behavioral features as handwriting, speech, and gait to identify individuals. Its applications include forensics, person authentication, and access control [4, 5, 6, 7, 8].

The use of biometric systems can be divided into the following three categories: the closed-set identification task, open-set identification task, and the authentication task. Authentication tasks are among the most commonly used tasks in biometric systems, where a sample to be identified is assigned to one of two classes by a classifier, indicating whether or not a given sample belongs to a particular person [9].

Fingerprints, speech, and face authentication are nowadays embedded into smart mobile device system. Fingerprints and speech authentication have been successfully applied, and although facial authentication is not yet mature—particularly in poor lighting, in cases of occlusion, and when images of only parts of the face are available—it has been used in the latest smartphones. The use of the face as a biometric feature has promising prospects because it can be optimized to use as small a number of human–computer interactions as possible for authentication [10].

In this paper, we focus on face authentication in smart devices. In most situations, using the face to identify someone is an easy task for humans, but this is far more challenging for computers. Improving the accuracy of face recognition has long been among the most active areas of research in computer vision and pattern recognition. Face authentication has a wide range of applications in daily life, such as to unlock smartphones, for online payment, and for access to places of employment.

Many methods have been proposed for face recognition. For example, the principal component analysis (PCA)-based technique in [11] is a feature-learning method that searches for a subspace that maximizes the variance of the projected sample. Pentland et al. proposed the modular PCA in [12], which can be used for variable face recognition with a view-based, multiple-observer eigenspace technique. Linear discriminant analysis (LDA) learns a projection matrix via the Fisher linear discriminant, and completes the classification and recognition tasks in a low-dimensional subspace with discriminative information [13]. Elastic graph matching (EGM) [14] has been applied to face authentication, and usually consists of two steps. The first is matching with a rigid grid, and the second a deformation of the grid guaranteed to match under the premise of some rotation, scaling, and even deformation of the object. Jonsson et al. proposed a support vector machine (SVM) model for face authentication in [15] that outperforms the eigenface technique in their experiments.

In the last decade, face recognition models based on low-rank representation have attracted considerable attention from researchers. Liu et al. proposed the latent low-rank representation model (latent LRR), where feature extraction and subspace segmentation were integrated into a framework to extract prominent features from data by exploiting the latent structural information hidden in low-dimensional data [16]. To eliminate the problem whereby the independent learning of two matrices in the latent LRR cannot ensure overall optimization, Fang et al. proposed a supervised feature extraction method using an approximate low-rank method (SFE-ALR) that treats the two matrices in the latent

The paper was supported in part by National Natural Science Foundation of China(Grant 61501147), China Postdoctoral Science Foundation(Grant 2016M601438), University Nursing Program for Young Scholars with Creative Talents in Heilongjiang Province(Grant UNPYSCT-2018203), Excellent Youth Scholar Fund of Heilongjiang Province(Grant JJ2019YX0116), The Foundation of Talents of Harbin University of Science and Technology, Talents Program of HRBUST(LGYC2018JQ013).

LRR as one, such that they promote each other to obtain better performance [17]. Meanwhile, Ma et al. combined dictionary learning and low-rank constraints to develop a discriminative low-rank dictionary learning algorithm for sparse representation (DLRDSR) [18] to solve the problem of face recognition. We think that these models based on low-rank representation can be applied to smart device systems.

Statistical models are also a popular class of methods for face authentication. In contrast to methods that use holistic features, statistical models typically use local features that exclusively describe partial facial features. For instance, the hidden Markov model (HMM) can successfully encode facial features by combining a priori structural information with statistical information [19]. In [20], Lucey et al. used the Gaussian mixture model (GMM) framework to simulate the conditional density function of a parts-based representation of the face.

To solve the problem of inaccurate face authentication under noise and interference, we propose robust feature modeling for face authentication to enhance the ability to distinguish between faces. We summarize the contributions of our approach as follows: 1) We develop a robust feature model. In the proposed objective function, the coefficients of low-rank representation represent the similarity weights by introducing the instances from the same class as positive and the ones from different classes as negative, which makes the coefficient matrix more diverse. 2) We introduce label regression and reconstructed label information as two supervised items to improve the robustness and adaptability of the model on authentication tasks. 3) We design a numerical scheme with augmented the Lagrangian multiplier (ALM) method [21] and the alternative direction method of multipliers (ADMM) [22] to guarantee the convergence of the objective function. 4) We evaluate our approach on three public datasets containing three types of noise and interference.

II. Related Work

In this section, we review two aspects of related methods: low-rank representation and feature subspace learning.

A. Low rank representation

Given a set of data examples $X \in R^{d \times n}$ (d and n denote the numbers of dimensions and examples, respectively). Each example can be represented by a d-dimensional basis vector and an error vector. We then stack all vectors as columns to construct a dictionary A and an error matrix E. The object of LRR is to find the lowest-rank representation of data examples X with a model:

$$\min_{Z} \|Z\|_* + \lambda \|E\|_{2,1}, \text{ s.t.} X = AZ + E \tag{2.1}$$

where $\|Z\|_*$ denotes the nuclear norm of the representation coefficient matrix Z, $\|E\|_{2,1}$ denotes the $l_{2,1}$ norm of E that can be computed by $\|E\|_{2,1} = \sum_{j=1}^{n} \left(\sum_{i=1}^{d} \left([E]_{ij} \right)^2 \right)^{1/2}$, and λ is a tradeoff parameter.

B. Feature subspace learning

Feature subspace learning seeks a projection subspace that can project high-dimensional data to a low-dimensional space. In the process of learning the projection space, significant information is retained for minimum intra-class distance and maximum inter-class distance. Because of the application of this critical technique to improve the performance of face authentication models, feature subspace learning models have come to be widely used. Commonly used feature subspace learning models are divided into three categories: unsupervised methods, supervised methods, and semi-supervised methods.

Among the methods of supervised feature subspace learning, linear discriminant analysis (LDA) is a representative method. To further improve performance, many models based on LDA have been proposed, such as locality-sensitive discriminant analysis (LSDA) [23], local Fisher discriminant analysis (LFDA) [24], probabilistic LDA (PLDA) [25], and sparse discriminant analysis (SDA) [26]. Moreover, other state-of-the-art subspace methods are worth learning [27, 28].

III. Our Proposed Approach

In this section, our proposed robust feature model is described. We first formulate the objective function and then design a numerical iterative scheme to solve it efficiently.

A. Formulate the objective function

Motivated by the LRR, we initially design a feature subspace learning model based on the sample correlation constraint of low-rank representation. The LRR model can explore the structural similarity information of the training sample and eliminate a portion of noise in the sample data. The coefficient of low-rank representation is then used as measure similarity after sample denoising. Compared to the conventional models without consideration on the noise from data, the performance of the model with low-rank representation was better, and increased the robustness of the feature subspace. The objective function is expressed as:

$$\begin{aligned} &\sum_{ij} S_{ij} Z_{ij} \left\| P^T X_i - P^T X_j \right\|_F^2 + \|Z\|_* + \lambda_1 \|E\|_{2,1} \\ &s.t.\ X = XZ + E,\ P^T P = \mathrm{I}_p \end{aligned} \tag{3.1}$$

where $X = [X_1, X_2, ..., X_m]$ is the training set, P denotes the projection matrix, E denotes the error matrix, Z_{ij} denotes the representation coefficient used to measure the similarity between samples of the training set in the feature subspace. Once the training samples were projected into the feature subspace, the similarity between the two samples increased, Z_{ij} decreased, and vice versa. S_{ij} and Z_{ij} are elements of matrix S and Z, and S is a constant matrix containing only +1 and -1, which ensures that Z_{ij} is positive when two samples belong to the same class, and negative otherwise. We used the data matrix itself as dictionary A, such that $X = XZ + E$. To reduce redundancy in and compactness of the subspace, an orthogonal constraint on P was incorporated into our objective function. To increase robustness, we constrained the error matrix with the $l_{2,1}$ norm, which can better explore the relevance of the data and remove sample-specific corruptions.

To enhance the discriminability and robustness of the feature subspace and maintain the good performance of the model, especially under noise or interference, we incorporated a reconstructed class label information constraint into our framework. The constraint used the reconstructed label information to further enlarge the disparity between the intra-class scatter matrix and the inter-class scatter matrix. To make our approach more effective by improving its adaptability to classification tasks, we used label regression, which has the advantage of gathering samples of the same class in a cluster center with discriminative information for a large marginal space. To this end, the objective function for our proposed framework was rewritten as follows:

$$\begin{aligned}&\min_{P,Z,E}\frac{1}{2}\left\|Y-P^TX\right\|_F^2+\sum_{ij}S_{ij}Z_{ij}\left\|P^TX_i-P^TX_j\right\|_2^2+\|Z\|_*+\lambda_1\|E\|_{2,1}\\&\quad+\lambda_2\left(\left\|P^TXZ(\mathrm{I}-B_u)\right\|_F^2-\left\|P^TXZ(B_u-B_v)\right\|_F^2+\eta\left\|P^TXZ\right\|_F^2\right)\\&s.t.\ X=XZ+E,\ P^TP=\mathrm{I}_p\end{aligned}\tag{3.2}$$

where $\|\bullet\|_F^2$ denotes the Frobenius norm, $Y=[Y_1,Y_2,...,Y_m]$ is the label matrix, in which Y_i=[-1,-1,...,N-1,...,-1]$^T \in R^N$ denotes the i-th column of Y and N is the total number of classes. If the i-th sample belongs to the c-th class, its c-th element is $N-1$, whereas the others are -1. B_u and B_v are two constant coefficient matrices, $B_u(i,j)=(1/n_k)$ if X_i and X_j belong to the same class, where n_k is the number of samples in the k-th class. $U_e(i,j)=0$ and $U_f(i,j)=(1/n)$; on the contrary, $U_e(i,j)=0$, $U_f(i,j)=(1/n)$, where n is total number of samples. λ_1 , λ_2 , and η are positive scalars to balance the three terms and ensure the convergence of the objective function.

B. Solution scheme

In this section, we developed a numerical scheme to solve the objective function. However, the minimization problem in (3.2), in which all variables are jointed, is not convex. To solve the problem, inexact ALM and ADMM are adopted to obtain the optimal solution. Meanwhile, we add the auxiliary variables H and W to relax the minimization. Then, objection function can be rewritten as

$$\begin{aligned}&\min_{P,Z,E}\frac{1}{2}\left\|Y-P^TX\right\|_F^2+\sum_{ij}S_{ij}W_{ij}\left\|P^TX_i-P^TX_j\right\|_2^2+\|H\|_*+\lambda_1\|E\|_{2,1}\\&\quad+\lambda_2\left(\left\|P^TXZ(\mathrm{I}-B_u)\right\|_F^2-\left\|P^TXZ(B_u-B_v)\right\|_F^2+\eta\left\|P^TXZ\right\|_F^2\right)\\&s.t.\ X=XZ+E,P^TP=I_p,Z=W,Z=H\end{aligned}\tag{3.3}$$

Furthermore, by the ALM, the problem (3.3) can be converted to the following form

$$\begin{aligned}&\min_{P,Z,E,H,W}\frac{1}{2}\left\|Y-P^TX\right\|_F^2+\sum_{ij}S_{ij}W_{ij}\left\|P^TX_i-P^TX_j\right\|_2^2+\|H\|_*+\lambda_1\|E\|_{2,1}\\&\quad+\lambda_2\left(\left\|P^TXZ(\mathrm{I}-B_u)\right\|_F^2-\left\|P^TXZ(B_u-B_v)\right\|_F^2+\eta\left\|P^TXZ\right\|_F^2\right)\\&\quad+\mathrm{Tr}\left(Y_1(X-XZ-E)\right)+\mathrm{Tr}\left(Y_2(Z-H)\right)+\mathrm{Tr}\left(Y_3(Z-W)\right)\\&\quad+\frac{\mu}{2}\left(\|X-XZ-E\|_F^2+\|Z-H\|_F^2+\|Z-W\|_F^2\right)\\&s.t.\ P^TP=I_p\end{aligned}\tag{3.4}$$

where Y_1, Y_2 and Y_3 are the Lagrangian multipliers, $\mathrm{Tr}(\bullet)$ denotes the trace operation.

To solve problem (3.4), we alternately update the variables by iteratively solving each variable and fixing others. In this way, in the k-th iteration, by dropping the irrelevant terms of P, we have

$$\begin{aligned}&\min_{P}\frac{1}{2}\left\|Y-P^TX\right\|_F^2+\sum_{ij}S_{ij}W_{ij}\left\|P^TX_i-P^TX_j\right\|_2^2\\&\quad+\lambda_2\left(\left\|P^TXZ(\mathrm{I}-B_u)\right\|_F^2-\left\|P^TXZ(B_u-B_v)\right\|_F^2+\eta\left\|P^TXZ\right\|_F^2\right)\\&s.t.\ P^TP=I_p\end{aligned}\tag{3.5}$$

To simplify (3.5), we convert it into the following form:

$$\begin{aligned}&\min_{P}\frac{1}{2}\left\|Y-P^TX\right\|_F^2+Tr\left(P^TXLX^TP\right)\\&\quad+\lambda_2\left(\left\|P^TXZ(\mathrm{I}-B_u)\right\|_F^2-\left\|P^TXZ(B_u-B_v)\right\|_F^2+\eta\left\|P^TXZ\right\|_F^2\right)\\&s.t.\ P^TP=I_p\end{aligned}\tag{3.6}$$

where $L=D-SW$ denotes the graph Laplacian matrix, and D is a diagonal matrix with $D_{ii}=\frac{\sum(SW)_{*i}+\sum(SW)_{i*}}{2}$. Equation (3.6) can be regard as $\mathcal{L}_P$ -norm based minimization problem. Due to the orthogonal constraint, minimization cannot be considered as a simple quadratic problem, which could be solved by the method proposed in [29] with the derivative of (3.6) as follows.

$$\begin{aligned}\frac{\partial\mathcal{L}_P}{\partial P}&=XX^TP-XY^T+2XLX^TP+2\lambda_2XZ(\mathrm{I}-B_u)(\mathrm{I}-B_u)^TZ^TX^TP\\&\quad+2\lambda_2XZ(B_u-B_u)(B_u-B_u)^TZ^TX^TP+2\lambda_2\eta XZ^TZ^TXP\end{aligned}\tag{3.7}$$

Then, ignoring the terms independent with respect to H in (3.4), the objection function can be rewritten as

$$\min_{G}\frac{1}{\mu}\|H\|_*+\frac{1}{2}\left\|H-\left(Z^k+\frac{Y_2^k}{\mu}\right)\right\|_F^2\tag{3.8}$$

By using the singular value shrinkage operator [30], equation (3.8) can be effectively solved.

Similarly, by fixing other variables, we obtain the problem about W. For optimizing the solution, the compact form is formulated as follows.

$$\min_{R}\left\|W-\left(Z^k+\frac{Y_3^k}{\mu}\right)\right\|_F^2+SW\otimes D^{k+1}\tag{3.9}$$

where D is a matrix with $D_{ij}=\left\|P^{T(k+1)}X_i-P^{T(k+1)}X_j\right\|_2^2$. Due to SW and D are independent, W can be solved with the following element-wise form

$$\min_{R}\left\|W-\left(Z^k+\frac{Y_3^k}{\mu}\right)\right\|_F^2+\left\|SW\otimes D^{k+1}\right\|_1\tag{3.10}$$

The problem (3.10) can been seen as the weighted l_1-norm minimization problem, which can be solved by the method in [31].

After solving the auxiliary variables, the objective function with respect to Z becomes:

$$\begin{aligned}\min_{Z} \lambda_2 \left(\left\| P^{T(k+1)} XZ (\mathrm{I} - B_u) \right\|_F^2 - \left\| P^{T(k+1)} XZ (B_u - B_v) \right\|_F^2 + \eta \left\| P^{T(k+1)} XZ \right\|_F^2 \right) \\ + \frac{\mu}{2} \left(\left\| X - XZ - E^k - \frac{Y_1^k}{\mu} \right\|_F^2 + \left\| Z - H^{k+1} - \frac{Y_2^k}{\mu} \right\|_F^2 + \left\| Z - W^{k+1} - \frac{Y_3^k}{\mu} \right\|_F^2 \right)\end{aligned} \tag{3.11}$$

By setting the derivative with respect to Z to be zero, we obtain

$$\begin{aligned}&\left(\left(X^T P^{k+1} P^{T(k+1)} X \right)^{-1} \left(2I + X^T X + 2\lambda_2 \eta X^T P^{k+1} P^{T(k+1)} X \right) \right) Z^{k+1} \\ &+ Z^{k+1} \left(2\lambda_2 \left((I - B_u)(I - B_u)^T - (B_u - B_v)(B_u - B_v)^T \right) \right) \\ &= \left(X^T P^{k+1} P^{T(k+1)} X \right)^{-1} \left(H^{k+1} + W^{k+1} - X^T E^k + X^T X - \left(-X^T Y_1^k + Y_2^k + Y_3^k \right) / \mu \right)\end{aligned} \tag{3.12}$$

Equation (3.12) is a standard Sylvester equation that can be effectively solved using existing tools in [32].

Next, ignoring the irrelevant terms of E, we have

$$\min_{E} \frac{\lambda_1}{\mu} \|E\|_{2,1} + \frac{1}{2} \left\| E - \left(X - XZ^{k+1} + \frac{Y_1^k}{\mu} \right) \right\|_F^2 \tag{3.13}$$

The solution to minimize the problem is presented in [33], let $\Theta = X - XZ^{k+1} + \frac{Y_1^k}{\mu}$, the i-th column of E^{k+1} is

$$E_i^{k+1} = \begin{cases} \dfrac{\|\Theta_i\|_2 - \lambda_1}{\|\Theta_i\|_2}, & \text{if } \lambda_1 < \|\Theta_i\|_2 \\ 0, & \text{otherwise} \end{cases} \tag{3.14}$$

Meanwhile, the Lagrange multiplier and parameters μ are updated in each iteration. Algorithm 1 describes the overall description of the solution scheme.

Algorithm 1 Scheme for discriminative feature subspace learning

Input: training data X, label matrix Y, λ_1, λ_2, η, $Z = H = W = 0$, $E = 0$, $Y_1 = Y_2 = Y_3 = 0$, $\mu = 0.6$, $\mu_{\max} = 10^{10}$, $\rho = 1.1$

Output: ***P***

While not convergence ***do***

1. Update P^{k+1} using (3.7)
2. Update H^{k+1} using (3.8);
3. Update W^{k+1} using (3.10);
4. Update Z^{k+1} using (3.11);
5. Update E^{k+1} using (3.13);
6. Update the Lagrangian multipliers and parameter:

$$Y_1^{k+1} = Y_1^k + \mu \left(X - XZ^{k+1} - E^{k+1} \right)$$

$$Y_2^{k+1} = Y_2^k + \mu \left(Z^{k+1} - H^{k+1} \right)$$

$$Y_3^{k+1} = Y_3^k + \mu \left(Z^{k+1} - W^{k+1} \right)$$

$\mu = \min(\mu_{\max}, \rho\mu)$;

end while

IV. EXPERIMENTAL RESULTS AND DISCUSSION

In this section, we used three public face datasets to evaluate the proposed feature model. We added three kinds of noises and interference to the dataset: Gaussian white noise, streak interference, and facial occlusion.

The detailed description of the datasets is as follows.

Extended YaleB dataset The Extended YaleB dataset contains 2414 images of 38 persons, and there are about 64 images under different lighting conditions of each person. Some sample images are shown in Fig 1(a). The sizes of each image is 32×32 pixels. Half of the images of each person are used for training, the rest for testing.

AR dataset The AR dataset has 3120 images of 120 individuals. Each individual have 26 face images with sunglasses or scarfs, different expressions and illumination conditions. Some samples images are shown in Fig 1(b). In our experiments, each image is cropped and then resized to 55×40. We randomly select half of the images of each individual as the training set, and the remaining for testing.

ORL dataset The ORL dataset contains 400 images of 40 subjects. The images were shot under different illumination and with various expression. Some samples images are shown in Fig 1(c). The images of the each subject used in our experiment are normalized to 32×32. Five images of each subject are used for training, and the rest for testing.

In our experiments, we compared the proposed approach with several prevalent methods for feature extraction, such as the PCA, LDA, LSDA, latent LRR, DLRDSR, and SFE-ALR. To balance the different features of the models, the same SRC classifiers were used for all test datasets. For SRC, atoms in the dictionary were used as training samples, and we assessed the results of recognition by using the minimum class-specific regression error [34]. All experiments for each dataset were run five times, and the average recognition results with standard deviations are shown in tables. Some images with different interferences on the three datasets are shown in Fig. 2.

A. Gaussian white noise

Face authentication on smart devices at night is a challenging problem because thermal noise is generated on the screen during imaging. Theoretically, thermal noise can be considered to be Gaussian white noise. We thus added this with different variances to the three datasets.

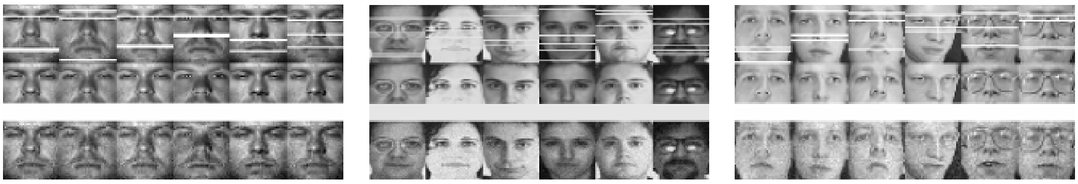
It Tables 1–3 show that the recognition rates of all methods with different degrees of Gaussian white noise were not considerably different from those of the clean datasets. This means that Gaussian white noise did not cause a significant drop in the recognition rate, and cannot be used to determine whether the model is robust.

B. Streak interference

When we perform face authentication on a smart device, there is occasional partial streak interference. Transverse streak interference can be regarded as objects occluding the front of the face. An example of vertical streak interference is rainfall. To simulate this situation, we added random

(a) Extended YaleB (b) AR (c) ORL

Fig.1 Original sample images



(a) Extended YaleB (b) AR (c) ORL

Fig. 2. Sample images with different interferences

Table 1 Recognition rates(%) of comparison methods on Extend YaleB datasets

Methods	PCA	LDA	LSDA	Latent LRR	DLRDSR	SFE-ALR	Our
Original	80.88	82.19	86.71	93.97	92.36	92.92	96.04
variance 10	80.37	81.62	84.26	93.53	91.64	92.49	95.97
variance 50	80.63	80.96	83.87	93.18	91.73	92.64	95.92
variance 100	80.50	81.18	83.11	93.74	91.35	92.14	95.68

Table 2 Recognition rates(%) of comparison methods on AR datasets

Methods	PCA	LDA	LSDA	Latent LRR	DLRDSR	SFE-ALR	Our
Original	82.41	93.28	82.35	94.98	90.85	95.26	96.29
variance 10	80.24	93.43	82.17	94.05	90.39	95.24	96.02
variance 50	80.29	93.95	82.54	93.90	89.93	95.71	95.90
variance 100	79.33	93.71	82.06	93.52	90.12	94.95	95.43

Table 3 Recognition rates(%) of comparison methods on ORL datasets

Methods	PCA	LDA	LSDA	Latent LRR	DLRDSR	SFE-ALR	Our
Original	88.17	91.33	87.16	93.34	92.43	93.67	93.33
variance 10	88.05	91.52	87.39	93.28	91.79	92.56	93.60
variance 50	87.48	90.76	87.72	91.50	92.35	93.52	93.38
variance 100	88.33	90.17	86.64	92.17	91.60	93.18	93.13

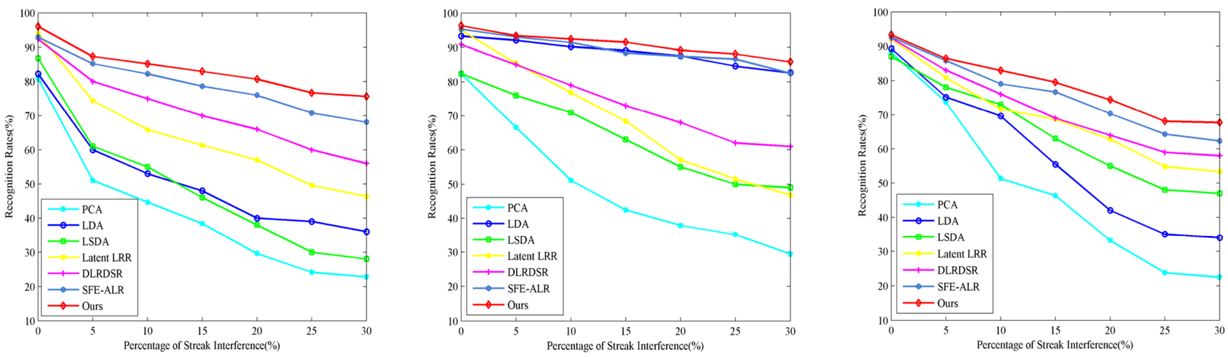
streak interference at different levels. The curve of the recognition rate on each dataset under different proportions of streak interference is drawn in Fig. 3.

Fig. 3 shows that our method was better than other comparison methods in case of streak interference. Our approach maximized inter-class distance and minimized intra-class distance. The experimental results indicate that our model has better robustness and discrimination than that of comparison methods.

C. *Partial facial occlusion*

Sometimes, only half the face is aligned with the camera when we perform face authentication. To verify that models can effectively recognize faces under these conditions, we processed the dataset by removing partial faces. We then experimented on the processed dataset.

The experimental results in Tables 4–6 indicate that when the dataset contained partial facial occlusion, the recognition rate decreased slightly, but our method had a



(a) Extended YaleB (b) AR (C) ORL

Fig. 3. Recognition rate versus different proportions of streak interference on datasets

Table 4 Recognition rates(%) of comparison methods on Extended YaleB datasets

Methods	PCA	LDA	LSDA	Latent LRR	DLRDSR	SFE-ALR	Our
Original	80.88	82.19	86.71	93.97	92.36	92.92	96.04
20% occlusion	78.46	80.72	84.31	90.23	91.13	91.16	95.24
30% occlusion	77.26	79.24	82.87	88.96	89.64	90.97	94.23
40% occlusi0n	76.78	77.95	81.58	87.05	87.55	89.25	93.88

Table 5 Recognition rates(%) of comparison methods on AR datasets

Methods	PCA	LDA	LSDA	Latent LRR	DLRDSR	SFE-ALR	Our
Original	82.41	93.28	82.35	94.98	90.85	95.26	96.29
20% occlusion	78.76	93.14	80.83	87.62	88.72	95.14	96.14
30% occlusion	76.71	91.43	79.15	85.48	87.09	94.38	95.62
40% occlusi0n	76.48	89.86	77.69	84.37	86.58	94.29	95.43

Table 6 Recognition rates(%) of comparison methods on ORL datasets

Methods	PCA	LDA	LSDA	Latent LRR	DLRDSR	SFE-ALR	Our
Original	88.17	91.33	87.16	93.34	92.43	93.67	93.33
20% occlusion	88.08	90.14	86.64	91.17	91.24	89.93	92.51
30% occlusion	87.67	90.31	85.75	87.33	90.53	88.17	91.70
40% occlusi0n	84.41	87.83	83.57	81.56	89.37	82.85	91.51

better recognition rate than the other methods. It is clear from the results that although all models extracted features from partially occluded faces, our algorithm optimized the feature extraction process, and its performance was stable.

D. *Recognition rates versus different training numbers*

We discuss the impact of different training numbers on recognition rates. Experiments are implemented on the original images and its occupied version with 5% streak interference. We evaluate our approach on Extended YaleB and AR datasets, and show the recognition results in Fig. 4.

E. *Discussion on parameters and convergence*

There are five model parameters in our algorithm. Next, we will discuss the impact of parameter values on performance. Among these parameters, μ and ρ are set depend on [21] to ensure that the objective function converges. For the other parameters λ_1, λ_2 and η, we discuss the effect on the recognition results with their variational values by choosing ORL as the test dataset. The recognition rate curves are shown in Fig. 5. We can see that recognition rate is not sensitive to parameter values changing, and model performance is very stable over a wide range of parameter values on both of the original data and streak interference data.

Meanwhile, to prove that our objective function can converge quickly, we choose ORL dataset to evaluate our model and plot the convergence curve of the objective function values versus the iterative step. The convergence curve is show in Fig. 6. We can see that our model have great convergence performance.

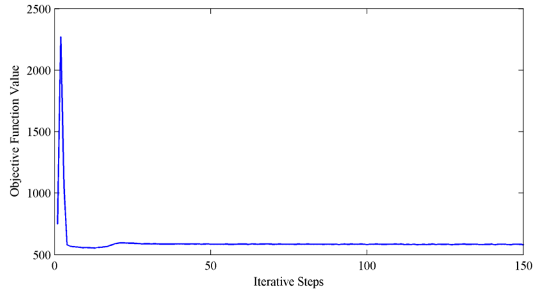
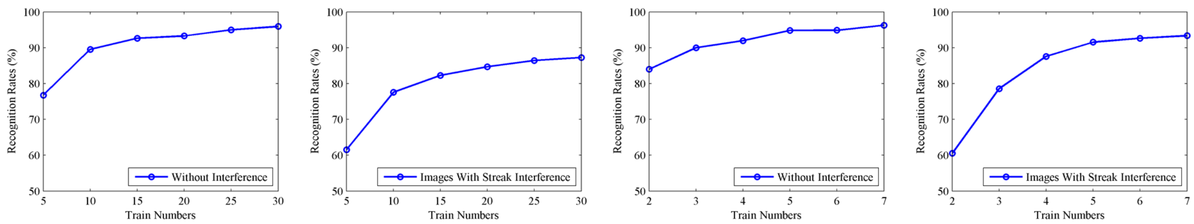
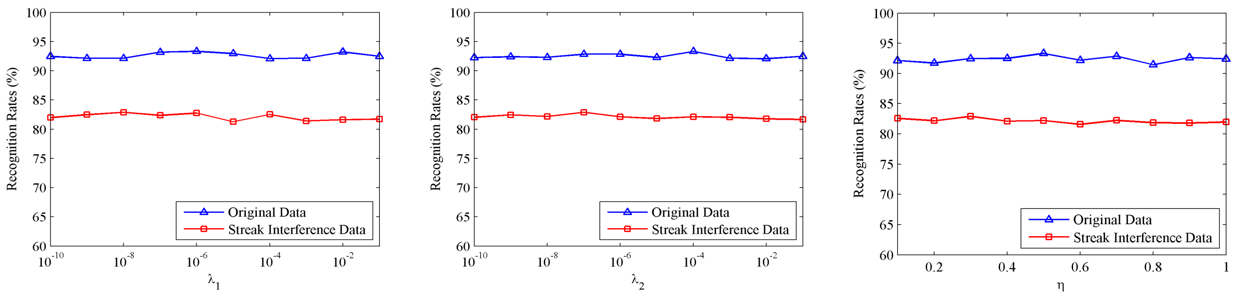


Fig. 6. Recognition rate versus iterative steps on ORL datasets



(a) Extended YaleB (b) Extended YaleB (c) AR (d) AR

Fig. 4. Recognition rate versus different train numbers on Extended YaleB and AR datasets



(a) λ_1 (b) λ_2 (c) η

Fig. 5. Recognition rate versus different variational parameters

F. *Complexity analysis*

In Algorithm 1, Steps 1 to 4 will consume the most time. The computational cost of solving P and H is $O\left(n^3\right)$ due to the singular value decomposition. The computational cost of solving W is $O\left(n^2\right)$, which can be seen as a weighted l_1-norm minimization problem. For Z, the solution to a standard Sylvester equation in (3.11) costs approximately $O\left(n^3\right)$. Hence, the overall computation complexity of the model is $O\left(tn^3\right)$, where t is the number of iterations.

V. CONCLUSION

In this paper, robust feature modeling is proposed for face authentication on smart devices. The proposed approach learns a robust feature subspace with low-rank constraints and label supervision. An iterative scheme with ALM and ADMM was designed to guarantee the convergence of the objective function. The proposed approach was tested on three public face datasets, where we simulated possible interference with face authentication. The experimental results demonstrated the superiority of our approach, especially under different kinds of interference. In future work, we intend to simplify the feature model while maintaining robustness, and to design our model as a semi-supervised or an unsupervised one.

REFERENCES

[1] R. Murmuria, A. Stavrou, D. Barbará, and D. Fleck, "Continuous authentication on mobile devices using power consumption, touch gestures and physical movement of users," in International Workshop on Recent Advances in Intrusion Detection, 2015, pp. 405-424.

[2] M. Muaaz and R. J. I. T. o. M. C. Mayrhofer, "Smartphone-based gait recognition: from authentication to imitation," IEEE Transactions on Mobile Computing, vol. 16, no. 11, pp. 3209-3221, 2017.

[3] S. Li, A. Ashok, Y. Zhang, C. Xu, J. Lindqvist, and M. Gruteser, "Whose move is it anyway? Authenticating smart wearable devices using unique head movement patterns," in Pervasive Computing and Communications (PerCom), 2016 IEEE International Conference on, 2016, pp. 1-9.

[4] P. Tome, R. Vera-Rodriguez, J. Fierrez, and J. J. F. s. i. Ortega-Garcia, "Facial soft biometric features for forensic face recognition," Forensic science international, vol. 257, pp. 271-284, 2015.

[5] T. Nakamura, V. Goverdovsky, D. P. J. I. T. o. I. F. Mandic, and Security, "In-ear EEG biometrics for feasible and readily collectable real-world person authentication," IEEE Transactions on Information Forensics and Security, vol. 13, no. 3, pp. 648-661, 2018.

[6] D. Addy and P. Bala, "Physical access control based on biometrics and GSM," in Advances in Computing, Communications and Informatics (ICACCI), 2016 International Conference on, 2016, pp. 1995-2001.

[7] A. K. Jain, and K. J. I. C. Nandakumar, "Biometric Authentication: System Security and User Privacy," IEEE Computer, vol. 45, no. 11, pp. 87-92, 2012.

[8] A. Li, Z. Wu, H. Lu, D. Chen, and G. J. I. A. Sun, "Collaborative self-regression method with nonlinear feature based on multi-task learning for image classification," IEEE Access, vol. 6, pp. 43513-43525, 2018.

[9] D. He, and D. J. I. S. J. Wang, "Robust biometrics-based authentication scheme for multiserver environment," IEEE Systems Journal, vol. 9, no. 3, pp. 816-823, 2015.

[10] J. Ortega-Garcia, J. Bigun, D. Reynolds, and J. J. I. s. p. m. Gonzalez-Rodriguez, "Authentication gets personal with biometrics," IEEE signal processing magazine, vol. 21, no. 2, pp. 50-62, 2004.

[11] M. Turk and A. J. J. o. c. n. Pentland, "Eigenfaces for recognition," Journal of cognitive neuroscience, vol. 3, no. 1, pp. 71-86, 1991.

[12] A. Pentland, B. Moghaddam, and T. Starner, "View-based and modular eigenspaces for face recognition," 1994.

[13] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," Yale University New Haven United States, 1997.

[14] B. Duc, S. Fischer, and J. J. I. T. o. I. P. Bigun, "Face authentication with Gabor information on deformable graphs," IEEE Transactions on Image Processing, vol. 8, no. 4, pp. 504-516, 1999.

[15] K. Jonsson, J. Kittler, Y. Li, J. J. I. Matas, and V. Computing, "Support vector machines for face authentication," Image and Vision Computing, vol. 20, no. 5-6, pp. 369-375, 2002.

[16] G. Liu and S. Yan, "Latent low-rank representation for subspace segmentation and feature extraction," in Computer Vision (ICCV), 2011 IEEE International Conference on, 2011, pp. 1615-1622.

[17] X. Fang et al., "Approximate low-rank projection learning for feature extraction," IEEE Transactions on Neural Networks and Learning Systems, no. 99, pp. 1-14, 2018.

[18] L. Ma, C. Wang, B. Xiao, and W. Zhou, "Sparse representation for face recognition based on discriminative low-rank dictionary learning," in Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, 2012, pp. 2586-2593.

[19] F. S. Samaria, "Face recognition using hidden Markov models," University of Cambridge Cambridge, UK, 1994.

[20] S. Lucey and T. Chen, "A GMM parts based face representation for improved verification through relevance adaptation," in Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on, 2004, vol. 2, pp. II-II.

[21] Z. Lin, M. Chen, and Y. J. a. p. a. Ma, "The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices," arXiv preprint arXiv:1009.5055, 2010.

[22] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. J. F. Eckstein, and T. i. M. learning, "Distributed optimization and statistical learning via the alternating direction method of multipliers," Foundations and Trends® in Machine learning, vol. 3, no. 1, pp. 1-122, 2011.

[23] D. Cai, X. He, K. Zhou, J. Han, and H. Bao, "Locality Sensitive Discriminant Analysis," in IJCAI, 2007, vol. 2007, pp. 1713-1726.

[24] M. J. J. o. m. l. r. Sugiyama, "Dimensionality reduction of multimodal labeled data by local fisher discriminant analysis," Journal of machine learning research, vol. 8, no. May, pp. 1027-1061, 2007.

[25] S. Ioffe, "Probabilistic linear discriminant analysis," in European Conference on Computer Vision, 2006, pp. 531-542.

[26] L. Clemmensen, T. Hastie, D. Witten, and B. J. T. Ersbøll, "Sparse discriminant analysis," Technometrics, vol. 53, no. 4, pp. 406-413, 2011.

[27] C. Peng, Z. Kang, H. Li, and Q. Cheng, "Subspace clustering using log-determinant rank approximation," in Proceedings of the 21th ACM SIGKDD international conference on Knowledge Discovery and Data Mining, 2015, pp. 925-934.

[28] Z. Kang, C. Peng, and Q. J. N. Cheng, "Kernel-driven similarity learning," Neurocomputing, vol. 267, pp. 210-219, 2017.

[29] Z. Wen and W. J. M. P. Yin, "A feasible method for optimization with orthogonality constraints," Mathematical Programming, vol. 142, no. 1-2, pp. 397-434, 2013.

[30] E. J. Candès, X. Li, Y. Ma, and J. J. J. o. t. A. Wright, "Robust principal component analysis?," Journal of the ACM (JACM), vol. 58, no. 3, p. 11, 2011.

[31] R. H. Bartels and G. W. J. C. o. t. A. Stewart, "Solution of the matrix equation AX+ XB= C [F4]," Communications of the ACM, vol. 15, no. 9, pp. 820-826, 1972.

[32] J. Yang and Y. J. S. j. o. s. c. Zhang, "Alternating direction algorithms for \ell 1-problems in compressive sensing," SIAM journal on scientific computing, vol. 33, no. 1, pp. 250-278, 2011.

[33] J. Yang, W. Yin, Y. Zhang, and Y. J. S. J. o. I. S. Wang, "A fast algorithm for edge-preserving variational multichannel image restoration," SIAM Journal on Imaging Sciences, vol. 2, no. 2, pp. 569-592, 2009.

[34] A. Li, D. Chen, Z. Wu, G. Sun, and K. J. P. o. Lin, "Self-supervised sparse coding scheme for image classification based on low rank representation," PloS one, vol. 13, no. 6, p. e0199141, 2018.