Synergistic Generic Learning for Face Recognition From a Contaminated Single Sample per Person

Meng Pang, Yiu-Ming Cheung, *Fellow, IEEE*, Binghui Wang, *Student Member, IEEE*, and Jian Lou

***Abstract*— Single sample per person face recognition (SSPP FR), i.e., identifying a person (i.e., data subject) with a single face image only for training, has several attractive potential applications, but it is still a challenging problem. Existing generic learning methods usually leverage prototype plus variation (P+V) model for SSPP FR provided that face samples in the biometric enrolment database are variation-free and thus can be treated as the prototypes of data subjects. However, this condition is not satisfied when these samples are contaminated by nuisance facial variations in the wild, such as varied expressions, poor lightings, and disguises (e.g., wearing scarf). We call this new and practical problem SSPP FR with a *c*ontaminated biometric *e*nrolment database (SSPP-ce FR). Subsequently, a challenging issue will be raised on estimating proper prototypes from the contaminated enrolment samples in SSPP-ce FR. Moreover, the generated variation dictionary also needs to be enhanced because it is simply based on the subtraction of average face from the samples of the same data subject in the generic set, thus containing individual characteristics that can hardly be shared by other data subjects. To address these two issues, we propose a novel synergistic generic learning (SGL) method to study the SSPP-ce FR problem. Compared with the existing generic learning methods, SGL develops a new "learned P + learned V" model to identify new query samples. Specifically, it learns better prototypes for the contaminated samples in the biometric enrolment database by preserving their more discriminative subject-specific portions and learns a representative variation dictionary by extracting the less discriminative intra-subject variants from an auxiliary generic set. The experiments on various benchmark face datasets demonstrate the effectiveness of the proposed SGL method.**

***Index Terms*— Single sample per person, generic learning, prototype learning, variation dictionary learning.**

I. Introduction

FACE recognition (FR) with single sample per person (SSPP) enrolled in the biometric enrolment database[1] for training has received increased attention in information security, pattern recognition and computer vision because of its potential applications in criminal identification, access control, law enforcement, video surveillance, just to name a few [1]–[6]. However, in such scenarios, a flurry of conventional discriminative subspace learning methods such as linear discriminant analysis (LDA) [7] and other Fisher-based subspace learning methods [8]–[10] cannot be directly applied since the intra-class information is not available to be utilized. Moreover, many existing sparse representation and dictionary learning methods [11]–[14], e.g., Fisher discrimination dictionary learning (FDDL) [12], also suffer from heavy performance degeneration, since these methods always require intra-class information to learn the discriminative structured dictionary or need multiple training samples to reasonably represent query samples.

To address the SSPP FR problem, many attempts have been made in the literature, which can be roughly classified into two categories [15]: *patch-based* methods and *generic learning* methods. For patch-based methods [16]–[20], each face sample in the biometric enrolment database is first partitioned into several local patches. Then, these patches will be treated as independent samples for feature extraction and recognition. However, local feature extraction and discriminative learning from partitioned patches can be sensitive to image variations [21]. In contrast, generic learning methods [22]–[27] are based on the pivotal assumption that a query sample of a person (i.e., data subject) equals its **p**rototype **plus** the intra-subject **v**ariation (i.e., P+V model) [24]. Such methods treat each sample in the biometric enrolment database as the prototype of each data subject. Then, they introduce an auxiliary generic set with data subjects *not of interest* to generate an intra-subject variation dictionary, e.g., by subtracting the average face from the samples of each data subject in the generic set, to encode the difference between each query sample and the prototype. Consequently, generic learning methods usually outperform patch-based methods. In this paper, we therefore only focus on generic learning methods.

Despite promising performance achieved by the state-of-the-art generic learning methods, these methods assume that each sample in the biometric enrolment database should be a standard unoccluded face with neutral expression and under uniform lighting (like an ID photo). However, in real-world scenarios, some samples in the biometric enrolment database are likely to be collected in less constrained environments. For example, for criminal identification, the suspects can be illegal immigrants, smugglers, or people without residence registration, which is common in the border areas of a country. In such

Manuscript received August 21, 2018; revised December 6, 2018, January 31, 2019, March 25, 2019, and April 16, 2019; accepted May 20, 2019. Date of publication May 30, 2019; date of current version September 11, 2019. This work was supported in part by the National Natural Science Foundation of China under Grant 61672444 and Grant 61272366, in part by the SZSTI under Grant JCYJ20160531194006833, and in part by the Faculty Research Grant of Hong Kong Baptist University under Grant FRG2/17-18/082. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Christoph Busch. *(Corresponding author: Yiu-Ming Cheung.)*

M. Pang, Y.-M. Cheung, and J. Lou are with the Department of Computer Science, Hong Kong Baptist University, Hong Kong (e-mail: mengpang@comp.hkbu.edu.hk; ymc@comp.hkbu.edu.hk; jianlou@comp.hkbu.edu.hk).

B. Wang is with the Department of Electrical and Computer Engineering, Iowa State University, Ames, IA 50011 USA (e-mail: binghuiw@iastate.edu).

Digital Object Identifier 10.1109/TIFS.2019.2919950

[1]More standardized biometric vocabularies can refer to the website of https://www.christoph-busch.de/standards.html

scenarios, the enrolment samples (i.e., reference photos) of suspects are hardly acquired through standard photograph, but are likely to be provided by witnesses with unaligned mobile photos or by surveillance cameras with blurred videos. As a result, various nuisance facial variations, e.g., expressions, illuminations, shadows, poses, and disguises (e.g., wearing glasses or scarf) can exist in these enrolment samples, which will increase much more difficulty for practical SSPP FR. In this paper, we thus call this new and practical problem as SSPP FR with a *c*ontaminated biometric *e*nrolment database (SSPP-ce FR). Besides, the previous problem of SSPP FR with a *s*tandard biometric *e*nrolment database is called SSPP-se FR for convenience.

The new SSPP-ce FR problem will lead the existing generic learning methods to suffering from heavy performance degeneration (see Fig. 11). Specifically, a plausible reason can be twofold:

1) Existing generic learning methods assume that each sample in the biometric enrolment database is variation-free and thus can be treated as the appropriate prototype. However, this assumption could not hold in SSPP-ce FR due to the nuisance variations in these samples.
2) Existing generic learning methods usually subtract the average face from generic samples of the same data subject as the first step to generate intra-subject variations. However, the individual characteristics cannot always be removed from the variation dictionary, because the average face only represents a general feature for the whole samples of a data subject but not a specific feature for each sample.

As a result, the used P+V model in the existing generic learning methods can be unsuitable for this SSPP-ce FR problem, and will cause a query sample to be easily identified as a wrong enrolment data subject with the similar facial variation. In a nutshell, the SSPP-ce FR problem has yet to be studied in the literature.

To address the aforementioned two issues, we will propose a novel Synergistic Generic Learning (SGL) method to deal with the SSPP-ce FR problem. SGL is based on two crucial observations that each face image can be decomposed into: 1) a *less discriminative* intra-subject variant shared by different data subjects; and 2) a *more discriminative* subject-specific portion capturing individual uniqueness. Specifically, we first learn a variation dictionary by extracting the intra-subject variants from an auxiliary generic set. Then, we learn better prototypes for the contaminated samples in the biometric enrolment database by preserving their subject-specific portions. Finally, we leverage the learned variation dictionary and learned prototypes to perform SSPP FR.

To be more specific, *for variation dictionary learning*, we first decompose each sample in the auxiliary generic set into different components via representation bases learning and sparse coding. Considering that generic samples are often of high-dimensionality and thus the related optimization problem will be time-consuming, we then develop an equivalent low-rank factorization-based optimization problem and solve it efficiently. Subsequently, we leverage the Fisher information in the generic set to regroup the components with relatively low discriminative ability into a less discriminative part (LDP), to learn the intra-subject variation dictionary.

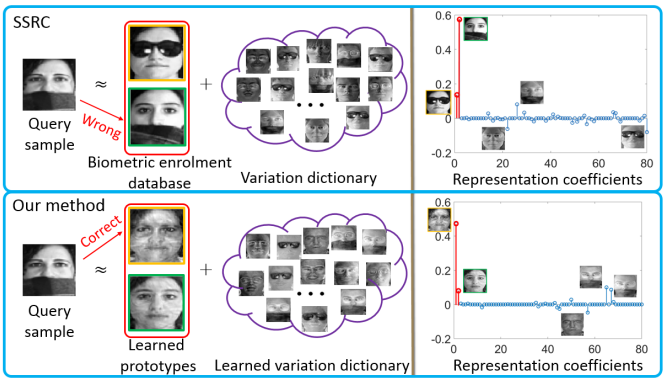

Fig. 1. A binary classification example of comparing SSRC with our SGL for SSPP-ce FR. According to the representation coefficients, a query sample wearing scarf is misclassified as the data subject wearing similar type of scarf for SSRC. In contrast, the query sample can be classified as the correct data subject in our SGL.

For prototype learning, we first detect the possible contaminated samples from the biometric enrolment database. Then, we leverage the learned representation bases to sparsely decompose them, and reuse Fisher information to regroup the components with relatively high discriminative ability into a more discriminative part (MDP), which separates the nuisance variations from the contaminated samples. In doing so, we can preserve the subject-specific portions of these contaminated samples and thus learn better prototypes. Furthermore, for the rest standard samples in the biometric enrolment database, we directly leverage them as the learned prototypes.

Compared with the state-of-the-art generic learning methods simply based on P+V model, SGL proposes a new "learned P + learned V" model that enables the prototype learning and variation dictionary learning to work collaboratively to identify new query samples. Consequently, our method is thus called synergistic generic learning. On the one hand, SGL pursues better prototypes for the contaminated samples in the biometric enrolment database, which narrows the gap between a query sample and the enrolment sample of the same data subject but with different types of variations, and meanwhile enlarges the gap between a query sample and the enrolment samples of different data subjects but with the similar type of variation. On the other hand, SGL efficiently learns a representative variation dictionary via extracting the LDP of each generic sample, which can decrease the individual characteristics and provide additive intra-subject variations to better reconstruct new query samples. A binary classification example of comparing our SGL with the popular superposed sparse representation based classification (SSRC) [24], [27] is illustrated in Fig. 1, which validates the superior performance of SGL over SSRC for SSPP-ce FR. We further verify the effectiveness of SGL on the AR, E-YaleB, CMU PIE, CAS-PEAL, FRGC v2.0 and LFW benchmark face datasets.

Moreover, it is worth noting that the convolutional neural networks (CNNs) based deep learning based methods [28]–[31], e.g., DeepID [28] and VGG-Face [29], have achieved great success in face identification with

TABLE I
THE LIST OF IMPORTANT ACRONYMS IN THE PROPOSED METHOD

Acronym	Meaning
SSPP-se FR	SSPP FR with a standard biometric enrolment database
SSPP-ce FR	SSPP FR with a contaminated biometric enrolment database
P+V	prototype plus variation
E+V	enrolment database plus variation
learned P + learned V	learned prototype plus learned variation
RBL	representation bases learning
FIFR	Fisher information-based feature regrouping
MDP	more discriminative part
LDP	less discriminative part

sufficient training samples and face verification. Although the CNNs cannot be directly applied in SSPP-se FR/SSPP-ce FR as it requires substantial training images to train the mass of parameters in the deep neural networks, some attempts have still been tried to employ the pre-trained neural networks to address the SSPP FR problem. For example, Parchami *et al*. [32] and Yang *et al*. [33] utilized the CNNs to extract the deep features of input images (or image patches), and collected external face datasets in the web to train the networks, which could benefit the generalization ability of the deep model. Motivated by these, in this work, we also consider to apply the pre-trained CNNs to generate high-semantic features for the enrolment and query samples, and explore the feasibility of combining our SGL method with the deep learning-based features to address the practical SSPP FR problem.

We highlight the contributions of our work as follows:

- To the best of our knowledge, this is the first work to study the challenging SSPP-ce FR problem, where the biometric enrolment database is contaminated by nuisance facial variations.
- We propose a novel SGL method to integrate variation dictionary learning and prototype learning into a unified framework. Moreover, SGL can be further combined with the deep learning-based features to enhance its performance for SSPP-ce FR.
- We present a new way to learn the variation dictionary and use low-rank factorization to solve it efficiently.

The rest of this paper is organized as follows. Section II reviews preliminaries and some related works. In Section III, we introduce the proposed SGL method in detail. In Section IV, we evaluate the performance of SGL and provide the corresponding experimental results. Finally, Section V gives the conclusion and future works.

II. PRELIMINARIES AND RELATED WORKS

A. Basic Notations

In this paper, matrices and vectors are represented with capital bold and lowercase bold symbols, respectively. $||.||_F$, $||.||_1$ and $||.||_2$ indicate the Frobenius, l_1 and l_2 norms, respectively, and $\mathbf{I}$ represents the identity matrix. Furthermore, considering that there are many acronyms in the proposed method, we thus summarize some important acronyms in Table I for clarity.

Let $\mathbf{y} \in \Re^d$ be the query sample and $\mathbf{G} = [\mathbf{g}_1, \cdots, \mathbf{g}_n] \in \Re^{d\times n}$ be the SSPP set with n data subjects in the biometric enrolment database. The auxiliary generic data matrix is defined as $\mathbf{A} = [\mathbf{A}_1, \cdots, \mathbf{A}_m] = [\mathbf{a}_1, \cdots, \mathbf{a}_M] \in \Re^{d\times M}$ ($M = mT$), with m data subjects *not of interest* and each having T images with different types of variations (including one standard reference image), and $\mathbf{A}_i = [\mathbf{a}_{(i-1)\cdot T+1}, \cdots, \mathbf{a}_{i\cdot T}]$. In addition, we further define $\tilde{\mathbf{A}} = [\mathbf{A}^1, \cdots, \mathbf{A}^T]$ as the reordered generic data of $\mathbf{A}$ according to the type of variations, where $\mathbf{A}^j = [\mathbf{A}_1^j, \mathbf{A}_2^j, \cdots, \mathbf{A}_m^j] \in \Re^{d\times m}$ denotes the collection of the jth type of variation across all m data subjects, and $\mathbf{A}_i^j \in \Re^d$ is a vector representing the j-th image of $\mathbf{A}_i$. In particular, $\mathbf{A}^1$ is used to represent the reference subset in generic set that contain standard photos corresponding to m data subjects.

The variation dictionary generated by SSRC is denoted as $\mathbf{V}_s$, and the variation dictionary and the enrolment prototypes to be learned in the proposed SGL are denoted as $\mathbf{V}$ and $\mathbf{P}$, respectively.

B. Related Work

In this subsection, we review some classical and state-of-the-art patch-based methods and generic learning methods that address the previous SSPP-se FR problem, and further discuss their limitations when facing with the more challenging SSPP-ce FR problem.

Patch-based methods perform recognition based on the partitioned patches of each enrolment image data. Chen *et al*. [34] adopted linear discriminant analysis (LDA) to extract discriminative features of each patch. Zhu *et al*. [19] extended the collaborative representation based classification (CRC) [35] to its patch-based counterpart, i.e., PCRC, by integrating the CRC outputs of all partitioned patches. Lu *et al*. [16] proposed a discriminative multi-manifold analysis (DMMA) method provided that patches of each data subject lie in an individual manifold, thus converting FR to a manifold-to-manifold matching problem. Based on it, Yan *et al*. [36] proposed a multi-feature multi-manifold learning (M^3L) method by combining multiple local features of partitioned patches for recognition. Zhang *et al*. [37] extended DMMA and proposed a sparse discriminative multi-manifold embedding (SDMME) method by using sparse graph embedding to improve the discriminative abilities of patches in the manifold. Nevertheless, the major concern of these patch-based methods comes from the fact that patches extracted from the single sample of each data subject contain limited and highly correlated information, and discriminative learning from these partitioned patches are still sensitive to image variations. Particular

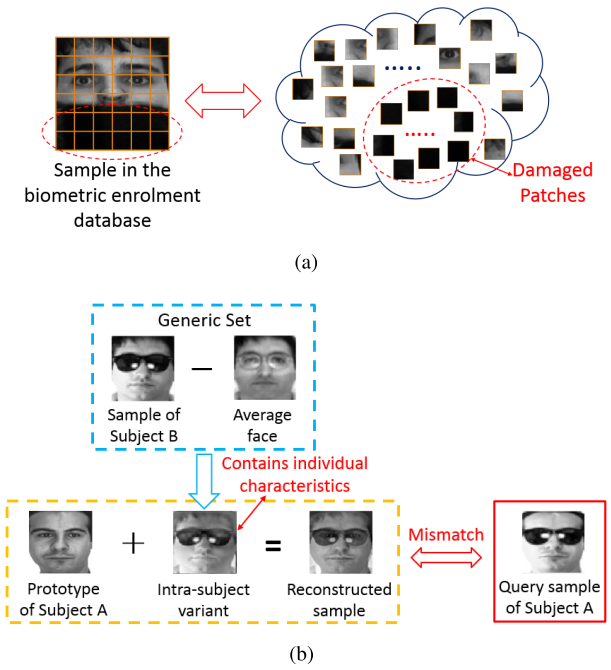

Fig. 2. (a) One sample wearing scarf in the contaminated biometric enrolment database (left) and its partitioned patches (right). In these patches, some of them are damaged and contain useless information for discriminative learning and feature extraction. (b) A failed reconstruction example of P+V model for SSPP FR. The generated intra-subject variant contains some individual characteristics from data subject B and thus making the reconstructed sample fail to estimate the query sample of data subject A.

in SSPP-ce FR, some patches in the contaminated samples may even be damaged and become meaningless for feature extraction and discriminative learning (see Fig. 2(a)).

Generic learning methods usually introduce an auxiliary generic set to provide new and useful information. Wang *et al.* [38] estimated the within-class scatter from the generic set by assuming that different sets of data subjects share similar intra-subject variations. Deng *et al.* [22] proposed an extended SRC (ESRC) framework by adding a generic variation set into the SSPP set for sparse coding. Moreover, Deng *et al.* [24] presented an SSRC-based P+V model provided that a query face equals its prototype plus the intra-subject variation. In their P+V model, the prototype is directly approximated by the sample in the biometric enrolment database since the enrolment sample is constrained to be standard and variation-free under their assumption. In other words, the P+V model is actually implemented as the *e*nrolment database plus variation (E+V) model. Formally, in the P+V/E+V model, a query sample $\mathbf{y}$ is considered as the combination of two different subsignals, i.e., enrolment samples dictionary $\mathbf{G}$ plus a variation dictionary $\mathbf{V}_s$ in the linear additive model:

$$\mathbf{y} = \mathbf{G}\boldsymbol{\beta} + \mathbf{V}_s\boldsymbol{\varphi} + \mathbf{e}, \tag{1}$$

where $\boldsymbol{\beta}$ indicates the sparse coefficient vector that selects a few of data subjects from the enrolment samples dictionary $\mathbf{G}$. The variation dictionary $\mathbf{V}_s$ is generated by simply subtracting the average face from generic samples of the same data subject, i.e., $\mathbf{V}_s = [\mathbf{A}_1 - \mathbf{c}_1\mathbf{1}'_T, \cdots, \mathbf{A}_m - \mathbf{c}_m\mathbf{1}'_T] \in \Re^{d\times M}$, where $\mathbf{c}_i = \frac{1}{T}\mathbf{A}_i\mathbf{1}_T \in \Re^{d\times 1}$ is the class centroid of the ith class. $\boldsymbol{\varphi}$ is another sparse coefficient vector that chooses a few types of variations from $\mathbf{V}_s$, and $\mathbf{e}$ is a small noise.

Then, the sparse coefficient vectors $\boldsymbol{\beta}$ and $\boldsymbol{\varphi}$ can be computed through solving the following l_1-minimization problem:

$$\begin{bmatrix}\boldsymbol{\beta}^* \\ \boldsymbol{\varphi}^*\end{bmatrix} = \arg\min_{\boldsymbol{\beta},\boldsymbol{\varphi}} \left\|\mathbf{y} - [\mathbf{G} \quad \mathbf{V}_s]\begin{bmatrix}\boldsymbol{\beta} \\ \boldsymbol{\varphi}\end{bmatrix}\right\|_2^2 + \lambda\left\|\begin{bmatrix}\boldsymbol{\beta} \\ \boldsymbol{\varphi}\end{bmatrix}\right\|_1, \tag{2}$$

where λ is a regularization parameter. Finally, similar to SRC, the test sample $\mathbf{y}$ will be classified into the data subject (i.e, class) with the smallest reconstruction residual.

Based on the P+V/E+V model, a variety of generic learning methods [21], [23], [25], [26] have been proposed to address the SSPP-se FR problem. For example, Yang *et al.* [21] proposed a sparse variation dictionary learning (SVDL) by additionally using the relationship between the enrolment database and generic set. Ji *et al.* [26] further extended SVDL by proposing a collaborative probabilistic labels (CPL), to utilize the contributions of different data subjects in the generic set. However, when handling the SSPP-ce FR problem, the linear additive P+V/E+V model in these methods will be impaired, as the contaminated enrolment samples can no longer be treated as appropriate prototypes. In such a case, the difference between a query sample and the contaminated enrolment one of the same data subject will be difficult to be measured, and thus the query sample is likely to be directly classified as the wrong data subject with the similar facial variation. Moreover, taking a step back, even the desired prototypes have been acquired, the generated variation dictionary in these methods still has limitation, since it always contains some unshareable individual characteristics (e.g., facial hair and contour) that can hardly be used to reconstruct query samples (see Fig. 2(b)). As a result, the existing generic learning methods will suffer from significant performance degradation. Hence, there is a need to develop a synergistic learning framework that simultaneously learns better prototypes for the biometric enrolment database as well as learns a more representative variation dictionary.

It is worth mentioning that the only related work concerning with SSPP-ce FR was in [39]. However, it needs to acquire the unknown query set in advance to estimate appropriate prototypes via a Gaussian mixture model (GMM), which may not be suitable for online learning. Considering that the setting in [39] is different from ours, we thus do not compare with it in the following experiments.

III. The Proposed Method

We first present the proposed SGL in two synergistic learning phases: *variation dictionary learning* and *prototype learning*. Then, we leverage the learned variation dictionary and learned prototypes to perform SSPP-ce FR. In the former phase, we extract the less discriminative part (LDP) of each generic sample, based on representation bases learning (RBL) and Fisher information-based feature regrouping (FIFR), to learn the variation dictionary. In the latter phase, we first detect the contaminated samples in the biometric enrolment database. Then, we use the learned representation

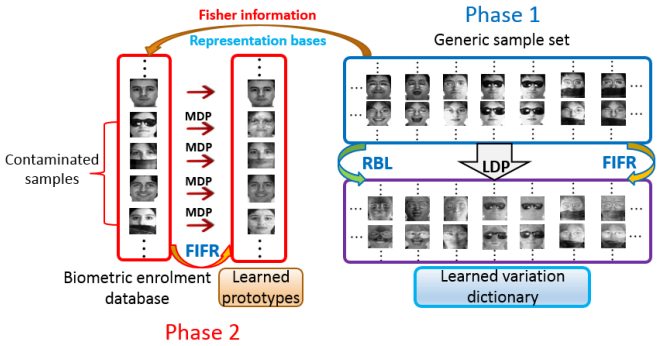

Fig. 3. The flowchart of the proposed SGL method.

bases to sparsely decompose these contaminated samples, and reuse Fisher information to extract their more discriminative part (MDP) for learning better prototypes. For the rest standard enrolment samples, we directly treat them as the learned prototypes. The flowchart of SGL is shown in Fig. 3.

A. Variation Dictionary Learning

In this phase, we learn a representative variation dictionary $\mathbf{V}$ by extracting the intra-subject variants from the generic set $\mathbf{A}$. The learning process involves representation bases learning and Fisher information-based feature regrouping.

1) Representation Bases Learning: In this step, we aim to decompose each generic sample into different representation components. Motivated by the success of dictionary learning and sparse coding in image processing [40]–[43], we learn the *enriched bases* (i.e., dictionary) $\mathbf{D} = [\mathbf{d}_1, \cdots, \mathbf{d}_K] \in \Re^{d\times K}$ to represent the generic set $\mathbf{A}$ under sparse constraints. Specifically, we minimize the following objective function:

$$\begin{aligned} &\min_{\mathbf{Z},\mathbf{D},\mathbf{E}} \|\mathbf{Z}\|_1 + \gamma \|\mathbf{E}\|_F^2 \\ &s.t.\ \ \mathbf{A} = \mathbf{D}\mathbf{Z} + \mathbf{E},\ \ \mathbf{d}_i'\mathbf{d}_i = 1, \forall i, \end{aligned} \tag{3}$$

where $\gamma > 0$ is a tradeoff parameter, $\mathbf{Z} = [\mathbf{z}_1, \cdots, \mathbf{z}_M] \in \Re^{K\times M}$ is the sparse coefficient matrix, and $\mathbf{E} \in \Re^{d\times M}$ is the residual matrix that encodes the generic sample noises.

The problem in Eq. (3) can be solved via Metaface Learning (MFL) [40] or other convex optimization toolboxes. However, due to the high dimensionality of each generic sample ($d \gg M \geq K$), it is often time-consuming to compute the solution for Eq. (3). To reduce the time cost, we thus develop a factorized problem with the low-rank factorization and provide a fast solution for Eq. (3) without performance loss, analogous to the work in [44]. Specifically, we first factorize the generic data $\mathbf{A}$ via the skinny singular value decomposition (SVD):

$$\mathbf{A} = \mathbf{U}_r\mathbf{\Sigma}_r\mathbf{H}_r', \tag{4}$$

where $\mathbf{\Sigma}_r = \text{diag}(\sigma_1, \cdots, \sigma_r)$ is a diagonal matrix with r positive singular values in a descending order, $\mathbf{U}_r \in \Re^{d\times r}$ and $\mathbf{H}_r \in \Re^{M\times r}$ are column-wise orthogonal matrices satisfying $\mathbf{U}_r'\mathbf{U}_r = \mathbf{H}_r'\mathbf{H}_r = \mathbf{I}$.

Then, we develop a factorized problem for the lower-dimensional generic data, i.e., $\mathbf{\Sigma}_r\mathbf{H}_r' \in \Re^{r\times M}$, as follows:

$$\begin{aligned} &\min_{\mathbf{W},\widehat{\mathbf{D}}} \ ||\mathbf{W}||_1 + \gamma\, ||\mathbf{\Sigma}_r\mathbf{H}_r' - \widehat{\mathbf{D}}\mathbf{W}||_F^2 \\ &s.t.\ \widehat{\mathbf{d}}_i'\widehat{\mathbf{d}}_i = 1, \forall i, \end{aligned} \tag{5}$$

Algorithm 1 Solving the Problem in Eq. (5)

Input: Data matrix to be processed: $\widehat{\mathbf{A}} \in \Re^{r\times M}$
Output: $\widehat{\mathbf{D}} = [\widehat{\mathbf{d}}_1, \widehat{\mathbf{d}}_2, \cdots, \widehat{\mathbf{d}}_K] \in \Re^{r\times K}$, $\mathbf{W} \in \Re^{K\times M}$

1: Initialization: each column of $\widehat{\mathbf{D}}$ is initialized as a random unit vector under l_2-norm constraint
2: **while** not converge **do**
3: Fix $\widehat{\mathbf{D}}$, update $\mathbf{W} \leftarrow \arg\min_{\mathbf{W}}(||\widehat{\mathbf{A}} - \widehat{\mathbf{D}}\mathbf{W}||_F^2 + \widehat{\gamma}||\mathbf{W}||_1)$
4: Fix $\mathbf{W}$, update $\widehat{\mathbf{D}}$. Let $\mathbf{W} = [\mathbf{w}_1; \cdots; \mathbf{w}_K]$, $\mathbf{w}_j \in \Re^{1\times M}$
5: **for** $i = 1 : K$ **do**
6: Fix $\widehat{\mathbf{d}}_j (\forall j \neq i)$, and define $\mathbf{S} = \widehat{\mathbf{A}} - \Sigma_{j\neq i}\widehat{\mathbf{d}}_j\mathbf{w}_j$
7: Update $\widehat{\mathbf{d}}_i \leftarrow \arg\min_{\widehat{\mathbf{d}}_i} ||\mathbf{S} - \widehat{\mathbf{d}}_i\mathbf{w}_i||_F^2$
8: Normalize $\widehat{\mathbf{d}}_i \leftarrow \widehat{\mathbf{d}}_i / ||\widehat{\mathbf{d}}_i||_2$
9: **end for**
10: **end while**

where $\mathbf{W} \in \Re^{K\times M}$ and $\widehat{\mathbf{D}} = [\widehat{\mathbf{d}}_1, \cdots, \widehat{\mathbf{d}}_K] \in \Re^{r\times K}$ are the coefficient matrix and representation bases for $\mathbf{\Sigma}_r\mathbf{H}_r'$, respectively. Subsequently, to solve the solution of the problem in Eq. (3), we develop the following theorem:

Theorem 1: Suppose $\{\boldsymbol{W}^*, \widehat{\boldsymbol{D}}^*\}$ is the optimal solution of the problem in Eq. (5), then the optimal solution $\{\boldsymbol{Z}^*, \boldsymbol{D}^*, \boldsymbol{E}^*\}$ of the primal problem in Eq. (3) can be defined as

$$\begin{cases} \boldsymbol{Z}^* = \boldsymbol{W}^* \\ \boldsymbol{D}^* = \boldsymbol{U}_r\widehat{\boldsymbol{D}}^* \\ \boldsymbol{E}^* = \boldsymbol{A} - \boldsymbol{U}_r\widehat{\boldsymbol{D}}^*\boldsymbol{W}^*. \end{cases} \tag{6}$$

*The detailed proof of **Theorem 1** is given in Appendix.*

Compared with the primal problem in Eq. (3), the problem in Eq. (5) reduces to a two-variable optimization problem and the data matrix to be processed is changed from $\mathbf{A} \in \Re^{d\times M}$ to the lower-dimensional $\mathbf{\Sigma}_r\mathbf{H}_r' \in \Re^{r\times M}$ ($r \ll d$). Therefore, the problem in Eq. (5) can be solved more efficiently. For simplicity, let $\widehat{\mathbf{A}} = \mathbf{\Sigma}_r\mathbf{H}_r'$ and $\widehat{\gamma} = \frac{1}{\gamma}$, then the problem in Eq. (5) can be rewritten as follows:

$$\begin{aligned} &\min_{\mathbf{W},\widehat{\mathbf{D}}} \ ||\widehat{\mathbf{A}} - \widehat{\mathbf{D}}\mathbf{W}||_F^2 + \widehat{\gamma}||\mathbf{W}||_1 \\ &s.t.\ \widehat{\mathbf{d}}_i'\widehat{\mathbf{d}}_i = 1, \forall i. \end{aligned} \tag{7}$$

Eq. (7) is solved via an alternating optimization, which is presented in **Algorithm 1**. In this algorithm, line 3 is optimized using the basis pursuit denoising (BPDN)-homotopy algorithm [45], and line 7 is solved by setting the derivative of $||\mathbf{S} - \widehat{\mathbf{d}}_i\mathbf{w}_i||_F^2$ w.r.t. $\widehat{\mathbf{d}}_i$ to be zero.

The convergence property of **Algorithm 1** can be empirically guaranteed. To be persuasive, we study the convergence of **Algorithm 1** on the AR, E-YaleB, CMU PIE, CAS-PEAL, FRGC v2.0 and LFW datasets, by plotting the curves where the objective function of Eq. (7), i.e., $J(\widehat{\mathbf{D}}, \mathbf{W})$, versus the iteration number is illustrated. It can be seen from Fig. 4 (a)-(f) that **Algorithm 1** converges very fast after a few number of iterations over six tested datasets.

After obtaining $\widehat{\mathbf{D}}$ and $\mathbf{W}$, the representation bases $\mathbf{D}$ and the sparse coefficient matrix $\mathbf{Z}$ can be obtained by leveraging the relationships in Eq. (6). Let $\mathbf{a}_{i,p} = \mathbf{d}_p z_{i,p}$ be the representation on the pth column of $\mathbf{D}$, where $z_{i,p}$ is the pth element of $\mathbf{z}_i$.

Then, for each generic sample $\{\mathbf{a}_i\}_{i=1}^M$, we have $\mathbf{a}_i \approx \mathbf{D}\mathbf{z}_i = \sum_{p=1}^K \mathbf{d}_p z_{i,p} = \sum_{p=1}^K \mathbf{a}_{i,p}$.

2) Fisher Information-Based Feature Regrouping: In this step, we aim to extract the LDP of each generic sample $\mathbf{a}_i$ from its K representation components $\mathbf{a}_{i,p}$ by leveraging the discriminative ability. Specifically, we use Fisher information [46], [47] to quantify the discriminative ability of the pth component $\mathbf{a}_{i,p}$. It is defined as follows:

$$dis(p) = \sum_{c=1}^{m} (\boldsymbol{\mu}_p^c - \boldsymbol{\mu}_p)^2 / \sum_{c=1}^{m} \frac{1}{T} \sum_{\mathbf{a}_i \in \mathbf{A}^c} (\mathbf{a}_{i,p} - \boldsymbol{\mu}_p^c)^2, \quad (8)$$

where $\mathbf{A}^c$ denotes the image set of the cth data subject, $\boldsymbol{\mu}_p$ is the mean vector of $\mathbf{a}_{i,p}$, i.e., $\boldsymbol{\mu}_p = \frac{1}{M}\sum_{i=1}^M \mathbf{a}_{i,p}$, and $\boldsymbol{\mu}_p^c$ is the mean vector of $\mathbf{a}_{i,p}$ belonging to the cth data subject. It is worth noting that, $dis(p)$ also reflects the discriminative ability as well as the importance of the representation basis $\mathbf{d}_p$ because all $\{\mathbf{a}_{i,p}\}_{i=1}^M$ share a common basis $\mathbf{d}_p$ (refer to $\mathbf{a}_{i,p} = \mathbf{d}_p z_{i,p}$). Hence, the larger value $dis(p)$ is, the better discriminative ability $\mathbf{d}_p$ has.

Then, we superpose the components with relatively smaller $dis(p)$ to generate the LDP of each generic sample. To this end, we first reorder the K components of each $\mathbf{a}_i$ according to their $dis(p)$ in an *ascending* order. Next, we select the first ρK components, e.g., $\{\mathbf{a}_{i,1}, \mathbf{a}_{i,2}, \cdots, \mathbf{a}_{i,\rho K}\}$, and regroup them into the $LDP(\mathbf{a}_i)$. Specifically, for each generic sample $\mathbf{a}_i$, the regroup formula can be rewritten as:

$$\begin{aligned} LDP(\mathbf{a}_i) &= \mathbf{a}_{i,1} + \mathbf{a}_{i,2} + \cdots + \mathbf{a}_{i,\rho K} \\ &= \mathbf{d}_1 z_{i,1} + \mathbf{d}_2 z_{i,2} + \cdots + \mathbf{d}_{\rho K} z_{i,\rho K} \end{aligned} \quad (9)$$

where $\rho < 1$ is a scalar to determine the fraction of K components to be selected to generate the LDP. In this feature regrouping, we allocate equal weight to each representation component $\mathbf{a}_{i,j}$ to retain the natural weight defined by $z_{i,j}$, which can reflect the degree of correlations between $\mathbf{a}_{i,j}$ (or $\mathbf{d}_j$) and the generic sample $\mathbf{a}_i$. Finally, we concatenate $LDP(\mathbf{a}_i)$ of all generic samples to construct the learned variation dictionary $\mathbf{V}$ as follows:

$$\mathbf{V} = [LDP(\mathbf{a}_1), \cdots, LDP(\mathbf{a}_M)]. \quad (10)$$

In Fig. 5, we illustrate the top-5 learned representation bases w.r.t. the lowest $dis(p)$ and the corresponding coefficients, of a generic sample wearing sunglasses on the AR dataset. Furthermore, in this figure, we also present the extracted LDP and the remaining part when ρ is set to be 0.1, 0.3 and 0.5, respectively. From Fig. 5, we have the following observations:

- First, the representation bases with lower $dis(p)$ indeed reflect the nuisance variation parts (e.g., expressions and disguises) possessing lower discriminative abilities of different data subjects.
- Second, the coefficient of each representation base can indicate the degree of correlation between the generic sample and the base.
- Third, by superposing several representation components w.r.t. lower $dis(p)$, the extracted LDP can well characterize the variation of sunglasses in the generic sample, while the remaining part preserves the clean portion of the generic sample.
- Forth, different values of ρ can generate richer variations to enrich the learned variation dictionary.

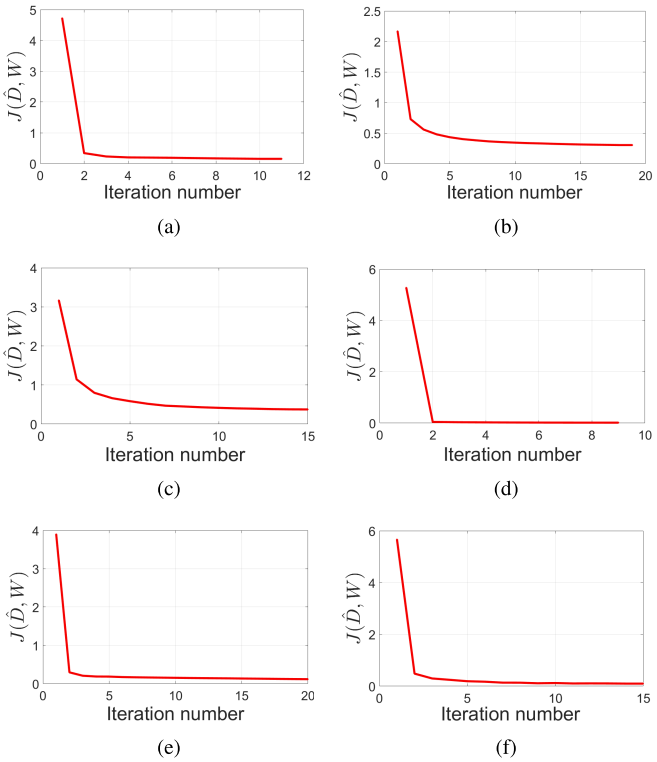

Fig. 4. Convergence study of **Algorithm 1** on the (a) AR, (b) E-YaleB, (c) CMU PIE, (d) CAS-PEAL, (e) FRGC v2.0, and (f) LFW datasets, respectively. The maximum iteration number is set as 50, and the objective function gap that terminates the iteration is set as 1e-3.

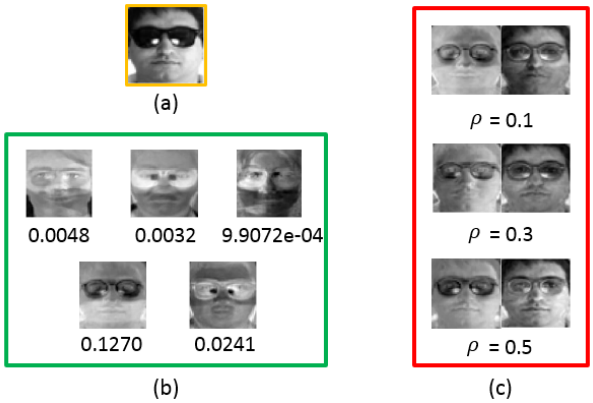

Fig. 5. (a) A selected generic sample wearing sunglasses on the AR dataset. (b) The top-5 learned representation bases w.r.t. the lowest $dis(p)$ and the corresponding coefficients of this generic sample. (c) The extracted LDP (left) and the remaining part (right) of this generic sample when $\rho = \{0.1, 0.3, 0.5\}$.

B. Prototype Learning

1) Contaminated Enrolment Sample Detection: In the biometric enrolment database, not all samples but a portion of them will be contaminated by nuisance variations. Therefore, there is no need to perform prototype learning on all enrolment samples. In this step, we detect the possible contaminated samples from the biometric enrolment database in advance.

As described in [21], a face sample could be represented as a combination of generic samples, and it is believed that a standard enrolment sample would be better represented by the standard generic samples while the contaminated enrolment ones would be better represented by the generic ones with the similar variations. Motivated by these, we thus leverage the representation residual to differentiate the contaminated and standard enrolment samples. Specifically, we let $\tilde{\mathbf{A}} = [\mathbf{A}^1, \mathbf{A}^c]$ denote the reordered generic data of $\mathbf{A}$ according to the type of variations, where $\mathbf{A}^1$ represents the reference subset containing m standard images, and $\mathbf{A}^c = [\mathbf{A}^2, \cdots, \mathbf{A}^T]$ denotes the contaminated subset with $T-1$ types of variations across all m data subjects. Next, we employ $\tilde{\mathbf{A}}$ to collaboratively represent each enrolment sample $\mathbf{g}_i$, i.e., $\tilde{\mathbf{A}}\mathbf{u}_i \approx \mathbf{g}_i$, and compute the representation coefficients by

$$\mathbf{u}_i = (\tilde{\mathbf{A}}\tilde{\mathbf{A}}' + \mu\mathbf{I})^{-1}\tilde{\mathbf{A}}\mathbf{g}_i, \tag{11}$$

where $\mu\mathbf{I}$ is added into $\tilde{\mathbf{A}}\tilde{\mathbf{A}}'$ to avoid singularity, and μ is empirically set as 0.01. Note that, the coefficient $\mathbf{u}_i$ can also be obtained by solving the following sparse optimization problem: $\min ||\mathbf{g}_i - \tilde{\mathbf{A}}\mathbf{u}_i||_2^2 + ||\mathbf{u}_i||_1$. Subsequently, we compute the representation residuals of $\mathbf{g}_i$ w.r.t. $\mathbf{A}^1$ and $\mathbf{A}^c$, respectively, as follows:

$$r_1(\mathbf{g}_i) = ||\mathbf{g}_i - \tilde{\mathbf{A}}\delta_1(\mathbf{u}_i)||_2^2, \tag{12}$$

$$r_c(\mathbf{g}_i) = ||\mathbf{g}_i - \tilde{\mathbf{A}}\delta_c(\mathbf{u}_i)||_2^2, \tag{13}$$

where $\delta_1(\mathbf{u}_i)$ and $\delta_c(\mathbf{u}_i)$ indicate the vectors whose nonzero entries are the entries associated with $\mathbf{A}^1$ and $\mathbf{A}^c$, respectively. Finally, we determine the category of $\mathbf{g}_i$ by using the smaller residual between $r_1(\mathbf{g}_i)$ and $r_c(\mathbf{g}_i)$, i.e., if $r_c(\mathbf{g}_i) \leq r_1(\mathbf{g}_i)$, $\mathbf{g}_i$ is classified into the contaminated category; otherwise, $\mathbf{g}_i$ will be treated as the standard sample. It is worth mentioning that, the adopted detection method is a simple collaborative representation-based binary classification scheme that is designed for the constrained and less constrained datasests (e.g., AR), and may not handle the unconstrained datasets well. In Fig.6, an illustration example of the binary classification on the AR dataset is presented, where we can see that the contaminated sample wearing sunglasses and the standard one in the biometric enrolment database can be well differentiated.

Suppose n_c samples are detected as the possible contaminated enrolment samples, while the rest n_s ($n_s = n - n_c$) samples are classified as the standard enrolment samples. Formally, we define the detected contaminated enrolment sample set and the standard enrolment sample set as $\mathbf{G}^c = [\mathbf{g}_1^c, \cdots, \mathbf{g}_{n_c}^c]$ and $\mathbf{G}^s = [\mathbf{g}_1^s, \cdots, \mathbf{g}_{n_s}^s]$, respectively, where $\mathbf{G}^c \cup \mathbf{G}^s = \mathbf{G}$.

2) Prototype Learning: After the detection of the contaminated enrolment samples, we then perform prototype learning on these samples by preserving their MDP so as to learn better prototypes. Moreover, for the rest standard enrolment samples, we directly leverage them as the learned prototypes.

In this step, the challenge is that we cannot directly learn discriminative representation bases for the contaminated enrolment samples, as the Fisher information is not available in the SSPP-based biometric enrolment database. Fortunately, the previous works [38], [48] have demonstrated that different

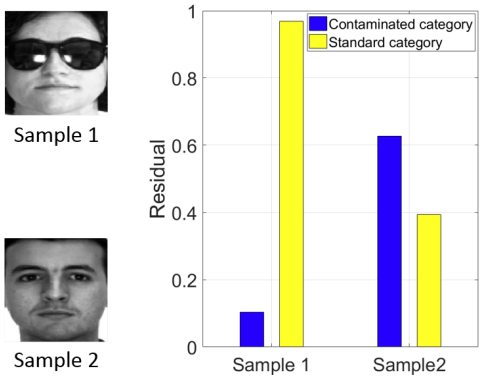

Fig. 6. A binary classification example for detecting the contaminated samples in the biometric enrolment database. According to the representation residuals, sample 1 (one contaminated sample wearing sunglasses) and sample 2 (one standard sample) can be well differentiated.

data subjects can share common representation bases and similar intra-subject variations. Motivated by this, we then transfer the Fisher information in the generic set to the biometric enrolment database via the learned representation bases $\mathbf{D}$, and thus enabling FIFR to be feasible to preserve the MDP from the contaminated enrolment samples.

Specifically, we first use the learned representation bases $\mathbf{D}$ to sparsely decompose each contaminated enrolment sample $\mathbf{g}_i^c$ into K representation components $\{\mathbf{g}_{i,p}^c\}_{p=1}^K$, with the sparse coefficient vector $\boldsymbol{\alpha}_i \in \Re^{K\times 1}$ by solving

$$\min_{\boldsymbol{\alpha}_i} ||\mathbf{g}_i^c - \mathbf{D}\boldsymbol{\alpha}_i||_2^2 + ||\boldsymbol{\alpha}_i||_1. \tag{14}$$

Then, $\mathbf{g}_i^c$ can be approximately represented as

$$\mathbf{g}_i^c \approx \mathbf{D}\boldsymbol{\alpha}_i = \sum\nolimits_{p=1}^{K} \mathbf{d}_p\alpha_{i,p} = \sum\nolimits_{p=1}^{K} \mathbf{g}_{i,p}^c, \tag{15}$$

where $\alpha_{i,p}$ is the pth element of $\boldsymbol{\alpha}_i$.

Subsequently, we reuse Fisher information, i.e., $dis(p)$, from the generic set, to perform feature regrouping based on $\{\mathbf{g}_{i,p}^c\}_{p=1}^K$ for each $\mathbf{g}_i^c$. As stated above, larger $dis(p)$ indicates a more discriminative ability of the representation basis $\mathbf{d}_p$. Thus, we reorder the values of $dis(p)$ in a *descending* order with the new indexes $\{ind(1), \cdots, ind(K)\}$, and rearrange the representation bases as $\{\mathbf{d}_{ind(1)}, \cdots, \mathbf{d}_{ind(K)}\}$. Consequently, the former bases have better discriminative abilities. Then, we superpose the first τK components $\mathbf{g}_{i,ind(p)}^c$ (i.e., $\mathbf{g}_{i,ind(p)}^c = \mathbf{d}_{ind(p)}\alpha_{i,ind(p)}$) to generate the MDP of each $\mathbf{g}_i^c$ as

$$MDP(\mathbf{g}_i^c) = \mathbf{g}_{i,ind(1)}^c + \mathbf{g}_{i,ind(2)}^c + \cdots + \mathbf{g}_{i,ind(\tau K)}^c, \tag{16}$$

where $\tau < 1$ is a scalar that determines the fraction of K components to be selected to generate the MDP. The rest components are considered as the nuisance facial variations or trivial structures. In this feature regrouping, we still allocate equal weights to these representation components, to make the generated MDP correctly present the target data subject but not other irrelevant data subjects.

Finally, we concatenate the extracted MDP from the contaminated enrolment sample set, i.e., $MDP(\mathbf{G}^c) = [MDP(\mathbf{g}_1^c), \cdots, MDP(\mathbf{g}_{n_c}^c)]$, with the standard enrolment

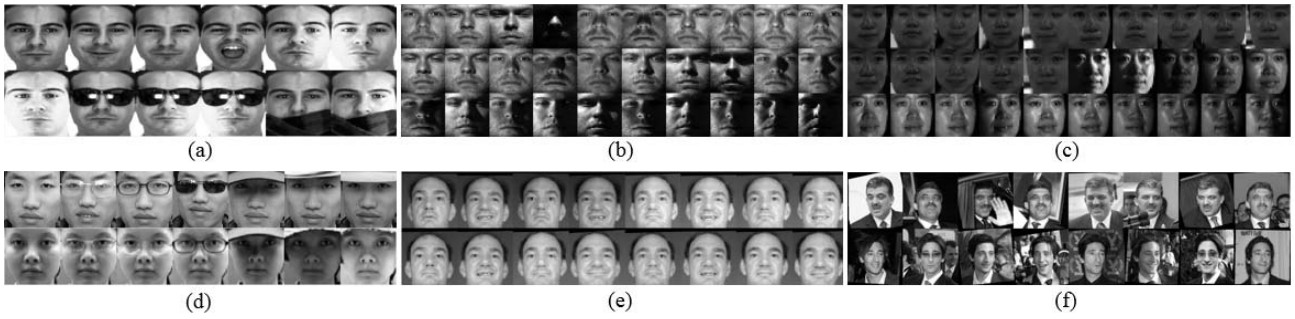

Fig. 7. Illustration of some selected samples from six benchmark datasets. (a) AR. (b) E-YaleB. (c) CMU PIE. (d) CAS-PEAL (e) FRGC v2.0. (f) LFW-a.

sample set, i.e., $\mathbf{G}_s$, to construct the learned prototypes for the biometric enrolment database, i.e., $\mathbf{P} = [MDP(\mathbf{G}^c), \mathbf{G}^s]$.

C. Face Recognition

In the recognition stage, given a query sample $\mathbf{y}$, we first solve the following P+V based optimization problem with the learned prototypes $\mathbf{P}$ and learned variation dictionary $\mathbf{V}$:

$$\begin{bmatrix}\boldsymbol{\beta}^* \\ \boldsymbol{\varphi}^*\end{bmatrix} = \arg\min_{\boldsymbol{\beta},\boldsymbol{\varphi}} \left\| \mathbf{y} - [\mathbf{P} \quad \mathbf{V}] \begin{bmatrix}\boldsymbol{\beta} \\ \boldsymbol{\varphi}\end{bmatrix} \right\|_2^2 + \lambda \left\| \begin{bmatrix}\boldsymbol{\beta} \\ \boldsymbol{\varphi}\end{bmatrix} \right\|_1, \quad (17)$$

where λ is a regularization parameter, $\boldsymbol{\beta} \in \Re^n$ and $\boldsymbol{\varphi} \in \Re^q$ denote the coefficient vectors of $\mathbf{P}$ and $\mathbf{V}$, respectively. In this paper, Eq. (17) is also optimized utilizing the BPDN-homotopy algorithm [45].

Then, $\mathbf{y}$ can be classified as the prototype (i.e., class) according to the smallest reconstruction residual $r_k(\mathbf{y})$ among all classes, i.e., $Identity(\mathbf{y}) = \arg\min_k r_k(\mathbf{y})$, and $r_k(\mathbf{y})$ is computed by

$$r_k(\mathbf{y}) = \left\| \mathbf{y} - [\mathbf{P} \quad \mathbf{V}] \begin{bmatrix}\delta_k(\boldsymbol{\beta}^*) \\ \boldsymbol{\varphi}^*\end{bmatrix} \right\|_2^2, \quad (18)$$

where $\delta_k(\boldsymbol{\beta}^*)$ is a vector whose nonzero entries are the entries in $\boldsymbol{\beta}^*$ that are associated with class k.

IV. Experimental Results

This section conducts six experiments to demonstrate the performance of the proposed SGL method[2]. Specifically, in Subsection IV-B, we evaluate the performance of the learned variation dictionary in SGL on the AR, Extended YaleB (E-YaleB), CMU PIE, CAS-PEAL datasets. In Subsection IV-C, we verify the effectiveness of the proposed detection strategy in SGL for detecting the contaminated and standard samples in the biometric enrolment database. In Subsection IV-D, we evaluate the performance of SGL for SSPP-ce FR on the above four benchmark datasets. In Subsection IV-E, we analyze the computational complexity of SGL. In Subsection IV-F, to mimic the real-world face retrieval scenarios, we evaluate the performance of SGL on three enlarged biometric enrolment databases for SSPP-ce FR. Lastly, in Subsection IV-G, motivated by the great success of deep learning [49], [50], we further evaluate the performance of SGL by combining it with the deep learning-based features on the Face Recognition Grand Challenge version 2.0 (FRGC v2.0) and the unconstrained labeled Faces in the Wild (LFW) datasets.

All experiments are carried out on a host (CPU: Dual 4-core Intel Xeon X5570 2.93GHz 8MB L3 Cache, Memory: 32GB).

A. Dataset Description

The AR dataset [51] contains over 4,000 face images of 126 data subjects from two sessions (i.e., S-I and S-II), and each session has 13 images per data subject, which involve different facial variations of expressions, illuminations and disguises.

The E-YaleB dataset [52] contains 2,414 images of 38 data subjects under various lighting conditions, which are divided into five subsets (i.e., 7, 12, 12, 14 and 19 images per data subject). Subset 1 is under normal lighting condition (lighting angle: 0°-12°), Subsets 2-3 characterize slight-to-moderate luminance variations (13°-25° and 26°-50°), and Subsets 4-5 characterize severe light variations (51°-77° and >77°).

The CMU PIE dataset [53] consists of 41,368 images of 68 data subjects, and each data subject has 43 different illuminations, 13 different poses, and 4 different expressions.

The CAS-PEAL dataset [54] contains 99,594 images of 1,040 data subjects (595 males and 445 females) with variations including expression, facing direction, accessory, lighting, age, etc. It is believed to be the largest public dataset with occluded face images available.

The FRGC v2.0 dataset [55] consists of 50,000 images of 4,003 data subjects with two different facial expressions, taken under different illumination conditions.

The LFW dataset [56] contains over 13,000 face images of 5,749 data subjects collected in unconstrained environments with large variations in expressions, poses, illuminations, etc.

Fig. 7 shows some samples of the images on the AR, E-YaleB, CMU PIE, CAS-PEAL, FRGC v2.0 and LFW datasets.

B. Evaluation of the Learned Variation Dictionary

In this subsection, our purpose is to validate the effectiveness of the learned variation dictionary in SGL. Hence, we choose standard unoccluded faces as the samples of the biometric enrolment database and thus prototype learning phase is unnecessary in this case. Accordingly, we denote SGL without prototype learning as *SGL w/o PL*, and evaluate its performance on SSPP-se FR.

[2]The code is provided at https://github.com/PangMeng92/SGL_v1.git

TABLE II
DATASET CONFIGURATION

Dataset	Dimension	Evaluated data subjects	Generic data subjects
AR	2304	50	50
E-YaleB	2304	20	18
CMU PIE	4096	40	28
CAS-PEAL	2304	50	100

On the AR dataset, we randomly select 50 data subjects from S-I for evaluation, and select another 50 data subjects as the generic set. The frontal face images with neutral expression and under normal illumination are used as the biometric enrolment samples, and the rest 12 images of each data subject are arranged to form 5 probe sets (expression, illumination, illumination+sunglasses, illumination+scarf and disguise). Moreover, we also use the face images from S-II as another 5 probe sets for evaluation. On the E-YaleB dataset, we choose the first 20 data subjects for evaluation, and use the rest 18 data subjects as the generic set. The first image of each data subject in Subset 1 is selected as the biometric enrolment sample, and Subsets 2-5 form 4 probe sets. On the CMU PIE dataset, we use the first 40 data subjects for evaluation, and use the rest 28 data subjects as the generic set. The neutral face images taken in the C27 subset (frontal pose) are used as the biometric enrolment samples, and the rest images with the poses C27, C05, C07, C09 and C29 as the probe samples. On the CAS-PEAL dataset, we use 150 data subjects from the Normal and the Accessory categories, thus each data subject has 1 neutral image, and 6 images wearing different glasses and hats. The first 50 data subjects are selected for evaluation, and the rest 100 data subjects are used as the generic set. The neutral images of the evaluated data subjects are used to build the biometric enrolment database, and the rest 6 images are used as 2 probe sets (glasses and hats). The configurations of the four tested datasets are listed in Table II.

We choose 5 representative methods for comparison, including the baseline SRC [11], and 4 latest generic learning methods, i.e., ESRC [22], SSRC [24], SVDL [21] and CPL [26]. Among the 6 comparing methods, we implement CPL by ourselves, and obtain the codes of other 5 methods from the original authors. It is noteworthy that the variation dictionaries in SVDL and CPL are both based on the generated one in SSRC. As to our SGL w/o PL, the Fisher information-based feature regrouping (FIFR) is the key step to extract the LDP from the generic set for learning the variation dictionary. If this step is removed, the learned variation dictionary will degenerate to the original generic set (we denote it as *Dict-Deg* for convenience). In the experiment, to explore the importance (or not) of the introduction of Fisher information in the variation dictionary learning, we also report the recognition results of Dict-Deg for reference. Moreover, we further report the performance of the incorporation of SGL w/o PL with the one-pass dictionary learning (OPDL) framework [57], i.e., SGL w/o PL+OPDL, to investigate the flexibility of the learned variation dictionary.

Regarding the parameters setting, the regularization parameters λ in SRC, ESRC, SSRC and Dict-Deg are searched from $\{0.001, 0.005, 0.01, 0.05, 0.1\}$ to achieve their best results over four tested datasets. For SVDL and CPL, the parameters are set according to the suggestions in [21] and [26], respectively. Specifically, the parameters λ_1, λ_2 and λ_3 of SVDL are set to be 0.001, 0.01 and 0.0001, respectively, and the parameters λ, δ_1, δ_2 of CPL are set to be 0.01, 0.3, 3, respectively. For our SGL w/o PL, there are three major parameters, i.e., $\widehat{\gamma}$ in Eq. (7), λ in Eq. (17) and ρ in Eq. (9). Through experiments, we observe that $\widehat{\gamma}$ has a little effect on the performance within the range from 0.001 to 0.1 and is fixed as 0.001. λ is empirically set within the range from 0.001 to 0.01 to achieve optimal stable results. We also fix it as 0.001. ρ is to calculate the LDP of generic samples and preferably set below 0.5 to acquire useful intra-subject variations. In this case, we set ρ as $\{0.1, 0.3, 0.5\}$ to enrich the learned variation dictionary. The other parameters in SGL w/o PL+OPDL are set according to the suggestions in [57].

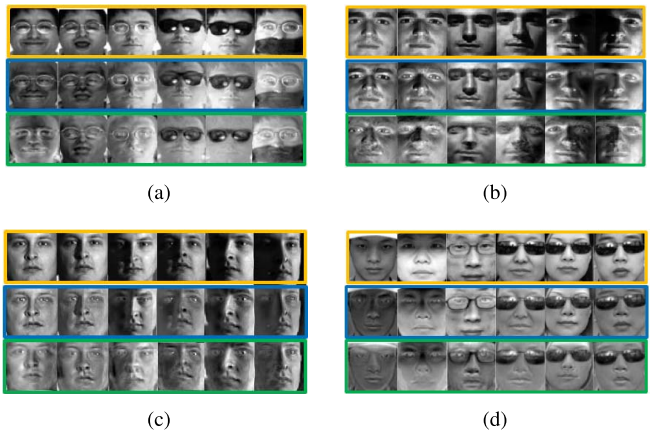

Fig. 8. The generic samples (the first line) and the variation dictionaries learned by SSRC (the second line) and by our SGL (the third line) on the (a) AR, (b) E-YaleB, (c) CMU PIE, and (d) CAS-PEAL datasets, respectively. The variation dictionaries learned by our SGL contain less individual characteristics compared with the ones generated by SSRC.

Fig. 8 (a)-(d) show the variation dictionaries learned by our method and by SSRC on the four tested datasets, respectively. It can be seen that the variation dictionaries learned by our method can decrease some sample-specific details (e.g., facial hair and contour) of the generic samples, and contain less individual characteristics compared with the variation dictionary generated by SSRC.

Table III lists the recognition results of different methods. First, our methods perform the best among the comparing generic learning methods. In particular, SGL w/o PL delivers 6.90%, 4.53%, 6.56% and 5.00% improvements on average over SSRC on the AR, E-YaleB, CMU PIE and CAS-PEAL, respectively. Besides, SGL w/o PL (with FIFR) consistently performs much better than the Dict-Deg over the four tested datasets, which verifies the importance of the introduction of Fisher information in variation dictionary learning. Moreover, we also observe that, by integrating the OPDL framework, SGL w/o PL has an extra enhanced performance in almost all cases we have tried. Overall, the promising performances of SGL w/o PL and SGL w/o PL+OPDL demonstrate the effectiveness and flexibility of the learned variation dictionary in SGL. Second, SVDL and CPL perform better than

TABLE III

RECOGNITION RATES (%) OF DIFFERENT METHODS ON THE AR, E-YALEB, CMU PIE AND CAS-PEAL BENCHMARK DATASETS FOR SSPP-SE FR

Probe set		Baseline	Generic learning methods					Our methods	
		SRC	Dict-Deg	ESRC	SSRC	SVDL	CPL	SGL w/o PL	SGL w/o PL +OPDL
AR	S-I: Expression	84.00	78.67	85.33	87.67	89.33	86.67	92.67	**95.33**
	S-I: Illumination	74.00	94.00	94.33	94.67	95.00	95.00	**99.33**	**99.33**
	S-I: Ill.+Sunglasses	41.33	64.00	83.33	86.67	86.67	87.00	91.33	**92.30**
	S-I: Ill.+Scarf	26.67	53.33	74.67	77.00	78.67	79.00	86.00	**88.67**
	S-I: Disguise	36.00	77.00	80.00	81.00	87.00	84.00	90.00	**94.00**
	S-II: Expression	62.00	70.67	71.33	73.67	74.00	74.00	78.00	**80.00**
	S-II: Illumination	42.67	76.00	78.00	81.00	84.67	83.33	88.67	**90.00**
	S-II: Ill.+Sunglasses	21.33	46.67	59.33	64.33	68.67	62.67	72.00	**75.33**
	S-II: Ill.+Scarf	16.00	44.67	56.67	60.33	65.30	62.00	70.33	**72.00**
	S-II: Disguise	26.00	61.00	64.00	68.00	70.00	65.00	75.00	**77.00**
	Average	43.00	66.60	74.70	77.43	79.93	77.87	84.33	**86.40**
E-YaleB	Subset 2 (13°-25°)	96.67	99.83	99.83	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**
	Subset 3 (26°-50°)	56.67	96.67	95.33	97.33	98.58	98.33	**100.00**	**100.00**
	Subset 4 (51°-77°)	15.00	56.79	61.43	69.29	71.07	70.71	78.93	**81.43**
	Subset 5 (>77°)	7.63	16.58	16.05	18.95	20.79	20.37	24.74	**27.32**
	Average	43.99	67.47	68.16	71.39	72.61	72.35	75.92	**77.19**
CMU PIE	C27 subset	44.44	69.72	73.89	75.94	75.00	76.67	80.00	**82.56**
	C05 subset	29.50	52.00	56.75	64.50	64.75	64.25	70.50	**71.75**
	C07 subset	32.00	43.50	48.25	55.00	55.75	55.50	60.50	**64.00**
	C09 subset	31.50	49.75	54.25	58.50	56.75	58.25	62.00	**64.75**
	C29 subset	15.50	39.00	34.00	39.75	40.50	43.75	**52.50**	51.75
	Average	30.59	50.79	53.43	58.54	58.55	59.68	65.10	**66.96**
CAS-PEAL	Accessory: Glasses	94.00	96.00	96.00	97.33	98.00	97.33	**98.67**	**98.67**
	Accessory: Hats	52.67	63.33	64.00	68.00	68.67	68.67	76.67	**77.73**
	Average	73.34	79.67	80	82.67	83.34	83.00	87.67	**88.20**

SSRC and ESRC. It is because that they additionally use the relationships between the generic and enrolment samples. Third, SRC performs the worst and is not competitive with generic learning methods in almost all cases, especially under illumination variation and disguises.

C. Detection of the Contaminated Enrolment Samples

In this subsection, we verify the effectiveness of the proposed detection strategy for detecting the contaminated and standard samples in the biometric enrolment database. This detection process can be viewed as a binary classification, and the contaminated and standard categories are denoted as the "positive" and "negative" classes, respectively. We construct four contaminated biometric enrolment databases by randomly selecting face images of all data subjects from AR, E-YaleB, CMU PIE and CAS-PEAL datasets, respectively, and perform the binary classification experiments on the four datasets.

Two metrics, i.e., accuracy (ACC) and true positive rate (TPR), will be employed to measure the binary classification results. Please see [58], [59] for the detailed definitions of the two metrics. Each classification experiment is repeated 10 times, and the average ACC, TPR and the corresponding standard errors on the above four datasets are reported in Table IV. We can observe that the proposed detection strategy achieves high classification accuracies, which indicates that the contaminated and standard samples in the biometric enrolment database are well separated over the four datasets. In addition, the TPR metric is more of our concern, as it reflects the detection rates of the contaminated enrolment samples. It is clear that the average TPR exceeds 90% in all the cases, which suggests that the majority of the contaminated enrolment samples are successfully detected using our strategy. For ease of observation, three examples of the contaminated enrolment samples misclassified into the standard category on the AR, CMU PIE and CAS-PEAL datasets are shown in Fig. 9. It can be observed that, the three misclassified samples are with mild expressions and are actually close to the standard photos, which can also serve as reasonable prototypes and will not heavily impair the performance of SGL.

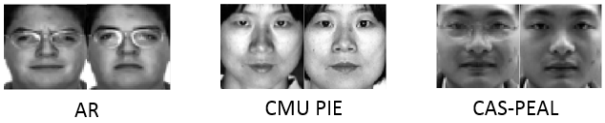

Fig. 9. The examples of the misclassified contaminated enrolment samples (left) and the standard photos for reference (right) on the AR, CMU PIE and CAS-PEAL datasets.

TABLE IV

PERFORMANCE OF THE PROPOSED STRATEGY FOR DETECTING THE CONTAMINATED SAMPLES IN THE BIOMETRIC ENROLMENT DATABASE

Dataset	ACC	TPR
AR	94.40%±1.26%	95.18%±2.00%
E-YaleB	99.00%±2.11%	96.67%±4.03%
CMU PIE	94.58%±2.46%	99.98%±0.29%
CAS-PEAL	88.83%±4.97%	95.02%±5.51%

D. Evaluation of SGL on SSPP-ce FR

In this subsection, we evaluate the performance of SGL on the AR, E-YaleB, CMU PIE and CAS-PEAL datasets, where the biometric enrolment database is contaminated by different nuisance facial variations.

For each tested dataset, we build 5 biometric enrolment databases with the contamination ratios of 10%, 30%, 50%, 70% and 90%, respectively, and all contaminated samples

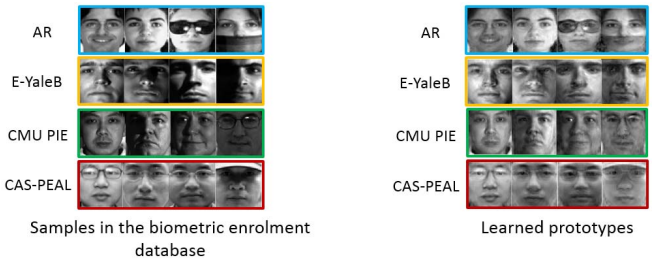

Fig. 10. Some contaminated samples in the biometric enrolment database and the corresponding learned prototypes by our SGL method on the AR, E-YaleB, CMU PIE and CAS-PEAL datasets.

are randomly corrupted by different variations. For example, on the AR dataset, each contaminated enrolment sample is randomly selected from the expression, illumination, illumination+sunglasses, illumination+scarf and disguises subsets, while the rest samples of each data subject are used for testing. The partitions of the evaluated and generic data subjects on the four datasets follow the settings in Table II. We repeat each experiment 5 times and report the average results.

We choose 8 methods for comparison, including the baseline SRC and CRC, 2 state-of-the-art patch-based methods, i.e., DMMA [16] and SDMME [37], and 5 state-of-the-art generic learning methods, i.e., ESRC, SSRC, SVDL, robust auxiliary dictionary learning (RADL) [23] and CPL. We implement DMMA and SDMME by ourselves, and obtain the source code of RADL from the authors. By default, the non-overlapping partitioned patch size for DMMA and SDMME is set as 8×8 pixels. The values of other parameters k_1, k_2, k, and σ in DMMA are empirically tuned to be 30, 2, 2, and 100, respectively. For SDMME, the l_1-ls toolbox is used to solve its l_1-minimization problem in accordance with [37], and the balance factor λ is tuned to be 0.001. For RADL, the parameters λ and η are set as 0.0001 and 1, respectively, as suggested in [23]. In addition, the parameters in SRC, CRC, ESRC, SSRC, SVDL, CPL, and ρ in Eq. (9), $\widehat{\gamma}$ in Eq. (7), λ in Eq. (17) of SGL are kept the same as that in Subsection IV-B. Besides, τ in Eq. (16) of SGL is empirically set within the range of 0.8 to 0.95 to preserve the discriminative MDP of contaminated enrolment samples.

Some contaminated samples in the biometric enrolment database and the corresponding learned prototypes by SGL on the AR, E-YaleB, CMU PIE and CAS-PEAL datasets are illustrated in Fig. 10. We observe that SGL can successfully separate the linear variations such as bad lightings and shadows from the contaminated enrolment samples, while cannot handle the nonlinear variations of expressions and poses well. A plausible reason is that the prototype learning in SGL is dependent on the linear-based FIFR and thus is difficult to deal with the nonlinear variations. In general, compared with the original contaminated enrolment samples, the learned prototypes in SGL is closer to the neutral image of the target data subject, which can narrow the gap between a query sample and the same data subject in the biometric enrolment database but with different types of variations, and enlarge the gap between a query sample and different enrolment data subjects but with the similar type of variation.

Fig. 11 shows the performance of all the involved methods, and we have the following observations:

- First, as the contamination ratio increases, all the methods will suffer from performance decline to different extents.
- Second, SGL performs better than the other comparing methods. For example, on the AR dataset, as the contamination ratio increases from 10% to 30%, 50%, 70% and 90%, SGL has a gain over SSRC, from 4.17% to 6.20%, 9.17%, 9.93% and 10.10%, respectively. The effectiveness of SGL can be attributed to its "learned P + learned V" model, which makes the learned prototypes and learned representative variation dictionary work collaboratively, thus both of them could contribute to the final recognition performance.
- Third, the generic learning SVDL, CPL and RADL perform poorly in most cases, especially when the contamination ratio is high. This is because that the dictionaries in these methods are generated provided that samples in the biometric enrolment database are standard and thus are not suitable to handle the SSPP-ce FR problem.
- Forth, the patch-based SDMME and DMMA perform the worst, again demonstrating that patch-based methods are inferior to generic learning methods for SSPP-ce FR.

E. Computational Complexity Analysis

In the training stage, we mainly analyze the computational complexity of the variation dictionary learning, because the size of generic set is always far larger than that of enrolment database ($M \gg n$) and the processing of the generic set can be the major cost w.r.t. both time and memory complexities.

The variation dictionary learning phase includes two steps: representation bases learning (RBL) and Fisher information-based feature regrouping (FIFR). In RBL, we use the low-rank factorization, i.e., $\widehat{\mathbf{A}} \in \Re^{r \times M}$, as the input for **Algorithm 1** to acquire the intermediate solution $\widehat{\mathbf{D}}$, then in turn compute the bases $\mathbf{D}$ via the relationships in Eq. (6). Supposing t and q denote the number of iterations for **Algorithm 1** and the l_1-minimization problem in line 3 of **Algorithm 1**, respectively, then the time complexities for computing $\mathbf{W}$ and $\widehat{\mathbf{D}}$ are $O(tqr^2M + tqKM)$ [60] and $O(trKM)$, respectively. Besides, the computation of $\mathbf{D}$ from $\widehat{\mathbf{D}}$ via Eq. (6) requires $O(drM)$. In FIFR, the time complexity for constructing the variation dictionary is $O(rKM)$. Overall, the time complexity of variation dictionary learning is summarized as $O(tqr^2M + tqKM + drM)$ (usually $tq > r$). Moreover, the memory cost in this phase is $O(dM)$.

Note that, in RBL, the used strategy for computing $\mathbf{D}$ with the low-rank factorization is more beneficial to the time cost, compared with the strategy directly leveraging the original data $\mathbf{A}$ (we call *Strategy-Ori* for convenience) to compute $\mathbf{D}$. To verify this point, we take CMU PIE dataset for testing, and record the average running time of Strategy-Ori and ours as 110.2404s and 26.3671s, respectively, which clearly verifies that our strategy is more time-saving than the original feature based Strategy-Ori.

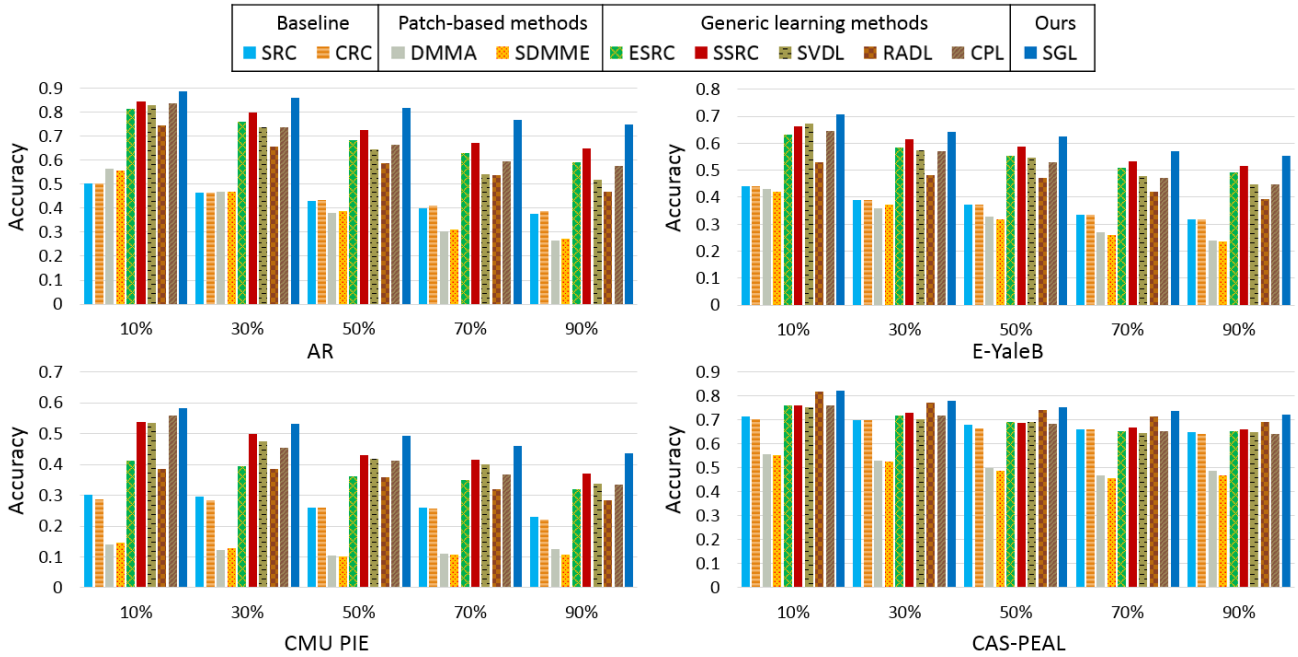

Fig. 11. Performance of different methods for SSPP-ce FR on the AR, E-YaleB, CMU PIE and CAS-PEAL datasets, where the contaminated samples occupy 10%, 30%, 50%, 70% and 90% of the total samples in the biometric enrolment database, respectively.

Furthermore, the recognition time in testing stage is more of our concern to measure the reality of SGL, since the training stage is always an offline process in real-world face retrieval scenarios. In the testing stage, SGL leverages the "learned P + learned V" model based sparse optimization for recognition and costs only 0.0719s on average, which is time-efficient and far less than the acceptable 0.5s. In a summary, the computational time of SGL will not limit its applications from the practical viewpoint.

F. Evaluation on Enlarged Biometric Enrolment Database

In practice, among the thousands of data subjects in the biometric enrolment database, the target data subjects of interests (e.g., criminals) usually occupy only a small portion, while most of them are irrelevant. In this subsection, we aim to mimic such face retrieval scenarios and to further explore whether the performance of SGL will be severely impaired when facing with the enlarged biometric enrolment database. To this end, we leverage the contaminated biometric enrolment database with 50% contamination on the AR dataset (50 data subjects of interests) for evaluation, then expand it to three larger biometric enrolment databases with 1000, 1500, and 2000 data subjects, respectively, by collecting standard or near standard samples from other three benchmark face datasets, i.e., CAS-PEAL, FRGC v2.0 and FERET [61], with the similar shooting environment. The rest samples of each data subject on the AR dataset are used for recognition. We repeat each experiment 5 times by randomly constructing the contaminated biometric enrolment database, and the average accuracies of SGL when the enlarged database sizes equal 1000, 1500, and 2000, respectively, are shown in Table V. It can be observed that, the performances of SGL with the enlarged biometric enrolment databases in the three cases drop slightly compared with the result of 81.97% (see Fig. 11) directly using the original contaminated enrolment database, which indicates that SGL is robust against the irrelevant data subjects on the enlarged biometric enrolment database.

TABLE V
PERFORMANCE OF SGL ON THE AR DATASET WITH 50% CONTAMINATION, WHEN THE SIZE OF BIOMETRIC ENROLMENT DATABASE IS ENLARGED TO 1000, 1500, AND 2000, RESPECTIVELY

Enlarged database size	Accuracy (%)
1000	79.23
1500	78.67
2000	77.90

G. Evaluation on Deep Learning-Based Features

In this subsection, we evaluate the performance of our SGL with two latest deep learning-based features, i.e., Openface feature [62] and LightenedCNN feature (using Model A) [63]. In the first part, we perform SSPP-ce FR experiments on the FRGC v2.0 and LFW datasets, and compare SGL with 6 representative methods, i.e., SRC, CRC, ESRC, SSRC, SVDL and CPL, using raw pixels, Openface feature, and LightenedCNN feature, respectively. On the FRGC v2.0 dataset, a subset of 5000 images of 250 data subjects are used. The first 200 data subjects are selected for evaluation while the rest 50 data subjects for generic learning. On the LFW dataset, we use a subset of 158 data subjects with no less than 10 image per data subject from LFW-a to form the evaluation and generic sets. The first 80 data subjects are selected for evaluation and the rest 78 data subjects for generic learning. For each data subject, we randomly select one image as the enrolment sample, and another one from the rest images for

TABLE VI

ACCURACIES (%) AND STANDARD ERRORS (%) OF DIFFERENT METHODS ON THE FRGC V2.0 DATASET, WITH RAW PIXELS, DEEP LEARNING-BASED OPENFACE FEATURE (DIMENSION = 128) AND LIGHTENEDCNN FEATURE (DIMENSION = 256)

Methods	Raw pixels	OpenFace feature	LightenedCNN feature
SRC	59.80±4.32	87.40±4.47	90.90±2.56
CRC	62.30±5.79	87.60±1.75	90.90±3.03
ESRC	67.40±4.23	89.60±3.96	92.10±3.45
SSRC	69.80±4.78	91.70±2.41	93.90±2.70
SVDL	67.80±4.74	92.10±2.43	93.20±3.09
CPL	70.90±6.36	91.60±2.46	92.40±3.10
SGL	**76.40±4.38**	**95.20±1.68**	**96.40±1.39**

TABLE VII

ACCURACIES (%) AND STANDARD ERRORS (%) OF DIFFERENT METHODS ON THE LFW DATASET, WITH RAW PIXELS, DEEP LEARNING-BASED OPENFACE FEATURE (DIMENSION = 128) AND LIGHTENEDCNN FEATURE (DIMENSION = 256)

Methods	Raw pixels	OpenFace feature	LightenedCNN feature
SRC	13.75±2.25	53.00±4.01	88.25±3.60
CRC	14.50±2.09	54.40±4.26	87.75±3.35
ESRC	17.20±2.23	54.75±3.47	89.00±3.89
SSRC	20.75±3.24	56.00±4.18	91.25±2.34
SVDL	19.63±2.35	56.25±3.04	90.50±2.44
CPL	20.33±2.21	56.65±2.50	89.45±1.99
SGL	**24.50±3.97**	**60.21±3.39**	**94.20±2.18**

TABLE VIII

ACCURACIES (%) OF SGL + LIGHTENEDCNN AND THE COMPARING DEEP LEARNING-BASED METHODS ON THE LFW DATASET

Methods	Accuracy (%)
DeepID	70.70
VGG-Face	84.67
JCR-ACF	86.00
SGL+LightenedCNN	93.11

testing. Note that, on the LFW dataset, the face images are all collected in a totally unconstrained environment, which means that the LFW dataset does not contain a neutral prototype (i.e., a standard sample) for each data subject. So, the proposed detection strategy will not be adopted for this unconstrained dataset. We repeat each experiment 5 times, and report the average recognition accuracies (± standard errors) of different methods on the FRGC v2.0 and LFW datasets in Table VI and Table VII, respectively.

In the second part, we further compare SGL using LightenedCNN feature, i.e., SGL+LightenedCNN, with three deep learning based methods, i.e., DeepID [28], VGG-Face [29], and the state-of-the-art **J**oint and **C**ollaborative **R**epresentation with local **A**daptive **C**onvolution **F**eature (JCR-ACF) [33], on the LFW dataset. The experimental configuration follows the protocol in JCR-ACF, and the results of DeepID and JCR-ACF are reported in accordance with [33]. For VGG-Face, the outputs from the 34th layer are used as the input features for the nearest neighbor classifier with cosine distance. The results of SGL+LightenedCNN and the other deep learning based methods are listed in Table VIII.

From Table VI-VII, we can observe that our SGL still outperforms the other 6 comparing methods with all the three types of features. For example, on the FRGC v2.0 dataset, SGL improves 5.6%, 3.1%, and 2.5% w.r.t. the average accuracies compared with the second best method using the raw pixel, Openface feature, and LightenedCNN feature, respectively. In addition, it is clear that the Openface and LightenedCNN features indeed greatly enhance the performance of SGL compared with raw pixels, which again demonstrates the discriminating power of the deep learning-based features. Furthermore, from Table VIII, our SGL+LightenedCNN has shown to achieve the highest accuracy among the comparing deep learning based methods, and even delivers 7.11% improvement over the state-of-the-art JCR-ACF method. In general, the inspiring results in Table VI-VIII provide a feasible direction to address the practical SSPP-ce FR problem by combining our SGL with the deep learning-based features.

V. CONCLUSION AND FUTURE WORKS

This paper has proposed a novel synergistic generic learning (SGL) method, which is the first attempt to study the new and challenging SSPP-ce FR problem. SGL has presented a synergistic *prototype learning* plus *variation dictionary learning* framework to better use the P+V model, compared with the state-of-the-art generic learning methods that only focus on generating variation dictionary. Experiments on various benchmark face datasets have demonstrated the effectiveness of the proposed SGL method.

Even though SGL achieves promising performance for SSPP-ce FR, there still exist three challenges we have not addressed. First, SGL learns better prototypes for the contaminated samples in the biometric enrolment database by separating the nuisance variations and preserving the more discriminative subject-specific portions. Since this learning process is based on the linear-based feature regrouping, some nonlinear facial variations (e.g., expressions and poses) may not be successfully removed under the circumstances. Second, still for prototype learning, when some crucial regions (e.g., eyes) in the enrolment sample are corrupted, it will be quite difficult to recover the original ones precisely, as these regions of this data subject are probably not included in the auxiliary generic set for learning the corresponding representation bases. Third, for variation dictionary learning, it is observed that, although the learned variation dictionary in SGL can decrease some individual characteristics, it seems not such smooth and contains noises in some cases, which may be unfavorable for the reconstruction of query samples. We will leave the study of the three mentioned issues as the future research work.

APPENDIX
THE PROOF OF THEOREM 1

Proof: First, according to the relations in Eq. (6), the feasibility of $\{\mathbf{Z}^*, \mathbf{D}^*, \mathbf{E}^*\}$ can be easily verified by

$$\mathbf{D}^*\mathbf{Z}^* + \mathbf{E}^* = \mathbf{D}^*\mathbf{W}^* + (\mathbf{A} - \mathbf{D}^*\mathbf{W}^*) = \mathbf{A}.$$

Besides, we can also verify that

$$(\mathbf{d}_i^*)'\mathbf{d}_i^* = (\mathbf{U}_r\widehat{\mathbf{d}}_i^*)'\mathbf{U}_r\widehat{\mathbf{d}}_i^* = (\widehat{\mathbf{d}}_i^*)'\mathbf{U}_r'\mathbf{U}_r\widehat{\mathbf{d}}_i^* = (\widehat{\mathbf{d}}_i^*)'\widehat{\mathbf{d}}_i^* = 1.$$

Next, to prove that $\{\mathbf{Z}^*, \mathbf{D}^*, \mathbf{E}^*\}$ is optimal to the problem in Eq. (3), we need to show that any feasible solutions $\{\mathbf{Z}, \mathbf{D}, \mathbf{E}\}$

satisfy the inequality:

$$\|\mathbf{Z}\|_1 + \gamma \|\mathbf{E}\|_F^2 \geq \|\mathbf{Z}^*\|_1 + \gamma \|\mathbf{E}^*\|_F^2.$$

Supposing $\mathbf{D}$ can be factorized as $\mathbf{D} = \mathbf{U}_r \mathbf{\Phi}$, we have the following inequalities:

$$\begin{aligned}\|\mathbf{Z}\|_1 + \gamma \|\mathbf{E}\|_F^2 &= \|\mathbf{Z}\|_1 + \gamma \|\mathbf{\Sigma}_r \mathbf{H}_r' - \mathbf{\Phi}\mathbf{Z}\|_F^2 \\ &\geq \|\mathbf{Z}\|_1 + \gamma \|\mathbf{\Sigma}_r \mathbf{H}_r' - \widehat{\mathbf{D}}^* \mathbf{Z}\|_F^2 \\ &\geq \|\mathbf{W}^*\|_1 + \gamma \|\mathbf{\Sigma}_r \mathbf{H}_r' - \widehat{\mathbf{D}}^* \mathbf{W}^*\|_F^2 \\ &= \|\mathbf{Z}^*\|_1 + \gamma \|\mathbf{E}^*\|_F^2.\end{aligned}$$

The second and the third inequalities hold because $\widehat{\mathbf{D}}^*$ and $\mathbf{W}^*$ are the optimal solutions of the problem in Eq. (5). The last equality holds because $\|\mathbf{Z}^*\|_1 = \|\mathbf{W}^*\|_1$ and $\|\mathbf{E}^*\|_F^2 = \|\mathbf{U}_r(\mathbf{\Sigma}_r \mathbf{H}_r' - \widehat{\mathbf{D}}^* \mathbf{W}^*)\|_F^2 = \|\mathbf{\Sigma}_r \mathbf{H}_r' - \widehat{\mathbf{D}}^* \mathbf{W}^*\|_F^2$. □

REFERENCES

[1] X. Tan, S. Chen, Z.-H. Zhou, and F. Zhang, "Face recognition from a single image per person: A survey," *Pattern Recognit.*, vol. 39, no. 9, pp. 1725–1745, 2006.

[2] L. Best-Rowden, H. Han, C. Otto, B. F. Klare, and A. K. Jain, "Unconstrained face recognition: Identifying a person of interest from a media collection," *IEEE Trans. Inf. Forensics Security*, vol. 9, no. 12, pp. 2144–2157, Dec. 2014.

[3] S. Gao, Y. Zhang, K. Jia, J. Lu, and Y. Zhang, "Single sample face recognition via learning deep supervised autoencoders," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 10, pp. 2108–2118, Oct. 2015.

[4] Z. He, S. Yi, Y.-M. Cheung, X. You, and Y. Y. Tang, "Robust object tracking via key patch sparse representation," *IEEE Trans. Cybern.*, vol. 47, no. 2, pp. 354–364, Feb. 2017.

[5] F. Mokhayeri, E. Granger, and G. A. Bilodeau, "Domain-specific face synthesis for video face recognition from a single sample per person," *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 3, pp. 757–772, Mar. 2019.

[6] X. Lan, M. Ye, R. Shao, B. Zhong, P. C. Yuen, and H. Zhou, "Learning modality-consistency feature templates: A robust RGB-infrared tracking system," *IEEE Trans. Ind. Electron.*, to be published.

[7] P. N. Belhumeur, J. P. Hespanha, and D. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 711–720, Jul. 1997.

[8] Y. Zhou and S. Sun, "Manifold partition discriminant analysis," *IEEE Trans. Cybern.*, vol. 47, no. 4, pp. 830–840, Apr. 2017.

[9] Y. Zhou and Y.-M. Cheung, "Probabilistic rank-one discriminant analysis via collective and individual variation modeling," *IEEE Trans. Cybern.*, to be published.

[10] M. Pang, Y.-M. Cheung, R. Liu, J. Lou, and C. Lin, "Toward efficient image representation: Sparse concept discriminant matrix factorization," *IEEE Trans. Circuits Syst. Video Technol.*, to be published.

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

[12] M. Yang, L. Zhang, X. Feng, and D. Zhang, "Sparse representation based Fisher discrimination dictionary learning for image classification," *Int. J. Comput. Vis.*, vol. 109, no. 3, pp. 209–232, Sep. 2014.

[13] M. Pang, B. Wang, Y.-M. Cheung, and C. Lin, "Discriminant manifold learning via sparse coding for robust feature extraction," *IEEE Access*, vol. 5, pp. 13978–13991, 2017.

[14] P. Zhou, C. Zhang, and Z. Lin, "Bilevel model-based discriminative dictionary learning for recognition," *IEEE Trans. Image Process.*, vol. 26, no. 3, pp. 1173–1187, Mar. 2017.

[15] T. Pei, L. Zhang, B. Wang, F. Li, and Z. Zhang, "Decision pyramid classifier for face recognition under complex variations using single sample per person," *Pattern Recognit.*, vol. 64, pp. 305–313, Apr. 2017.

[16] J. Lu, Y.-P. Tan, and G. Wang, "Discriminative multimanifold analysis for face recognition from a single training sample per person," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 39–51, Jan. 2013.

[17] F. Liu, J. Tang, Y. Song, L. Zhang, and Z. Tang, "Local structure-based sparse representation for face recognition," *ACM Trans. Intell. Syst. Technol.*, vol. 7, no. 1, pp. 2:1–2:20, 2015.

[18] F. Liu, J. Tang, Y. Song, Y. Bi, and S. Yang, "Local structure based multi-phase collaborative representation for face recognition with single sample per person," *Inf. Sci.*, vols. 346–347, pp. 198–215, Jun. 2016.

[19] P. Zhu, L. Zhang, Q. Hu, and S. C. K. Shiu, "Multi-scale patch based collaborative representation for face recognition with margin distribution optimization," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Oct. 2012, pp. 822–835.

[20] M. Pang, Y.-M. Cheung, B. Wang, and R. Liu, "Robust heterogeneous discriminative analysis for face recognition with single sample per person," *Pattern Recognit.*, vol. 89, pp. 91–107, May 2019.

[21] M. Yang, L. Van Gool, and L. Zhang, "Sparse variation dictionary learning for face recognition with a single training sample per person," in *Proc. Int. Conf. Comput. Vis. (ICCV)*, Dec. 2013, pp. 689–696.

[22] W. Deng, J. Hu, and J. Guo, "Extended SRC: Undersampled face recognition via intraclass variant dictionary," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 9, pp. 1864–1870, Sep. 2012.

[23] C.-P. Wei and Y.-C. F. Wang, "Undersampled face recognition via robust auxiliary dictionary learning," *IEEE Trans. Image Process.*, vol. 24, no. 6, pp. 1722–1734, Jun. 2015.

[24] W. Deng, J. Hu, and J. Guo, "In defense of sparsity based face recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2013, pp. 399–406.

[25] L. Zhuang, T.-H. Chan, A. Y. Yang, S. S. Sastry, and Y. Ma, "Sparse illumination learning and transfer for single-sample face recognition with image corruption and misalignment," *Int. J. Comput. Vis.*, vol. 114, no. 2, pp. 272–287, 2014.

[26] H.-K. Ji, Q.-S. Sun, Z.-X. Ji, Y.-H. Yuan, and G.-Q. Zhang, "Collaborative probabilistic labels for face recognition from single sample per person," *Pattern Recognit.*, vol. 62, pp. 125–134, Feb. 2017.

[27] W. Deng, J. Hu, and J. Guo, "Face recognition via collaborative representation: Its discriminant nature and superposed representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 10, pp. 2513–2521, Oct. 2017.

[28] Y. Sun, X. Wang, and X. Tang, "Deep learning face representation from predicting 10,000 classes," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2014, pp. 1891–1898.

[29] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," in *Proc. Brit. Mach. Vis. Conf.*, 2015, vol. 1. no. 3, p. 6.

[30] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "DeepFace: Closing the gap to human-level performance in face verification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2014, pp. 1701–1708.

[31] J.-C. Chen, V. M. Patel, and R. Chellappa, "Unconstrained face verification using deep CNN features," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2016, pp. 1–9.

[32] M. Parchami, S. Bashbaghi, and E. Granger, "CNNs with cross-correlation matching for face recognition in video surveillance using a single training sample per person," in *Proc. IEEE Int. Conf. Adv. Video Signal Based Surveill. (AVSS)*, Aug./Sep. 2017, pp. 1–6.

[33] M. Yang, X. Wang, G. Zeng, and L. Shen, "Joint and collaborative representation with local adaptive convolution feature for face recognition with single sample per person," *Pattern Recognit.*, vol. 66, pp. 117–128, Jun. 2017.

[34] S. Chen, J. Liu, and Z.-H. Zhou, "Making FLDA applicable to face recognition with one sample per person," *Pattern Recognit.*, vol. 37, no. 7, pp. 1553–1555, 2004.

[35] L. Zhang, M. Yang, and X. Feng, "Sparse representation or collaborative representation: Which helps face recognition?" in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Nov. 2011, pp. 471–478.

[36] H. Yan, J. Lu, X. Zhou, and Y. Shang, "Multi-feature multi-manifold learning for single-sample face recognition," *Neurocomputing*, vol. 143, pp. 134–143, Nov. 2014.

[37] P. Zhang, X. You, W. Ou, C. L. P. Chen, and Y.-M. Cheung, "Sparse discriminative multi-manifold embedding for one-sample face identification," *Pattern Recognit.*, vol. 52, pp. 249–259, Apr. 2016.

[38] J. Wang, K. N. Plataniotis, J. Lu, and A. N. Venetsanopoulos, "On solving the face recognition problem with one training sample per subject," *Pattern Recognit.*, vol. 39, no. 9, pp. 1746–1762, 2006.

[39] Y. Gao, J. Ma, and A. L. Yuille, "Semi-supervised sparse representation based classification for face recognition with insufficient labeled samples," *IEEE Trans. Image Process.*, vol. 26, no. 5, pp. 2545–2560, May 2017.

[40] M. Yang, L. Zhang, J. Yang, and D. Zhang, "Metaface learning for sparse representation based face recognition," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2010, pp. 1601–1604.

[41] L. Shao, R. Yan, X. Li, and Y. Liu, "From heuristic optimization to dictionary learning: A review and comprehensive comparison of image denoising algorithms," *IEEE Trans. Cybern.*, vol. 44, no. 7, pp. 1001–1013, Jul. 2014.

[42] V. M. Patel, T. Wu, S. Biswas, P. J. Phillips, and R. Chellappa, "Dictionary-based face recognition under variable lighting and pose," *IEEE Trans. Inf. Forensics Security*, vol. 7, no. 3, pp. 954–965, Jun. 2012.
[43] X. Lan, S. Zhang, P. C. Yuen, and R. Chellappa, "Learning common and feature-specific patterns: A novel multiple-sparse-representation-based tracker," *IEEE Trans. Image Process.*, vol. 27, no. 4, pp. 2022–2037, Apr. 2018.
[44] S. Xiao, W. Li, D. Xu, and D. Tao, "FaLRR: A fast low rank representation solver," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 4612–4620.
[45] D. L. Donoho and Y. Tsaig, "Fast solution of ℓ_1-norm minimization problems when the solution may be sparse," *IEEE Trans. Inf. Theory*, vol. 54, no. 11, pp. 4789–4812, Nov. 2008.
[46] M. Welling, "Fisher linear discriminant analysis," Dept. Comput. Sci., Univ. Toronto, Toronto, ON, Canada, Tech. Rep. 2005-03-01, 2005, vol. 3, no. 1.
[47] L. Zhang, P. Zhu, Q. Hu, and D. Zhang, "A linear subspace learning approach via sparse coding," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Nov. 2011, pp. 755–761.
[48] Y. Su, S. Shan, X. Chen, and W. Gao, "Adaptive generic learning for face recognition from a single sample per person," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2010, pp. 2699–2706.
[49] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015.
[50] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Netw.*, vol. 61, pp. 85–117, Jan. 2015.
[51] A. Martinez and R. Benavente, "The AR face database," Comput. Vis. Center, Barcelona, Spain, Tech. Rep. 24, Jun. 1998.
[52] A. S. Georghiades, P. N. Belhumeur, and D. Kriegman, "From few to many: Illumination cone models for face recognition under variable lighting and pose," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 6, pp. 643–660, Jun. 2001.
[53] T. Sim, S. Baker, and M. Bsat, "The CMU pose, illumination, and expression (PIE) database," in *Proc. 5th IEEE Int. Conf. Autom. Face Gesture Recognit. (FG)*, May 2002, pp. 53–58.
[54] W. Gao *et al.*, "The CAS-PEAL large-scale Chinese face database and baseline evaluations," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 38, no. 1, pp. 149–161, Jan. 2008.
[55] P. J. Phillips *et al.*, "Overview of the face recognition grand challenge," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 1, Jun. 2005, pp. 947–954.
[56] L. Wolf, T. Hassner, and Y. Taigman, "Effective unconstrained face recognition by combining multiple descriptors and learned background statistics," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 10, pp. 1978–1990, Oct. 2011.
[57] C.-P. Wei and Y.-C. F. Wang, "Undersampled face recognition with one-pass dictionary learning," in *Proc. Int. Conf. Multimedia Expo (ICME)*, Jun./Jul. 2015, pp. 1–6.
[58] A. Maratea, A. Petrosino, and M. Manzo, "Adjusted F-measure and kernel scaling for imbalanced data learning," *Inf. Sci.*, vol. 257, pp. 331–341, Feb. 2014.
[59] W. Wang, X. Wang, D. Feng, J. Liu, Z. Han, and X. Zhang, "Exploring permission-induced risk in Android applications for malicious application detection," *IEEE Trans. Inf. Forensics Security*, vol. 9, no. 11, pp. 1869–1882, Nov. 2014.
[60] A. Y. Yang, A. Ganesh, Z. Zhou, S. S. Sastry, and Y. Ma, "Fast ℓ_1-minimization algorithms for robust face recognition," 2010, *arXiv:1007.3753*. [Online]. Available: https://arxiv.org/abs/1007.3753
[61] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, "The FERET evaluation methodology for face-recognition algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 10, pp. 1090–1104, Oct. 2000.
[62] B. Amos, B. Ludwiczuk, and M. Satyanarayanan, "OpenFace: A general-purpose face recognition library with mobile applications," CMU School Comput. Sci., Carnegie Mellon Univ., Pittsburgh, PA, USA, Tech. Rep. CMU-CS-16-18, 2016.
[63] X. Wu, R. He, Z. Sun, and T. Tan, "A light CNN for deep face representation with noisy labels," *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 11, pp. 2884–2896, Nov. 2018.

Meng Pang received the B.Sc. degree in embedded engineering and the M.Sc. degree in software engineering from the Dalian University of Technology, Dalian, China, in 2013 and 2016, respectively. He is currently pursuing the Ph.D. degree with the Department of Computer Science, Hong Kong Baptist University, Hong Kong. His research interests include image processing, pattern recognition, and data mining.

Yiu-Ming Cheung (SM'06–F'18) received the Ph.D. degree from the Department of Computer Science and Engineering, The Chinese University of Hong Kong, Hong Kong. He is currently a Full Professor with the Department of Computer Science, Hong Kong Baptist University, Hong Kong. His research interests include machine learning, pattern recognition, visual computing, and optimization. He is a fellow of IET, BCS, and RSA, and a Distinguished Fellow of IETI. He serves as an Associate Editor for the IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, the IEEE TRANSACTIONS ON CYBERNETICS, and *Pattern Recognition*, to name a few.

Binghui Wang (S'16) received the B.Sc. degree in network engineering and the M.Sc. degree in software engineering from the Dalian University of Technology, Dalian, China, in 2012 and 2015, respectively. He is currently pursuing the Ph.D. degree in electrical and computer engineering with Iowa State University, Ames, IA, USA. His research interests include machine learning, big data mining, data-driven security and privacy, and adversarial machine learning.

Jian Lou received the Ph.D. degree from the Department of Computer Science, Hong Kong Baptist University. His research interests include statistical learning, numerical optimization and method, and privacy preserving for machine learning.