Multiple Feature Fusion for Face Recognition

Shu Kong, Xikui Wang, Donghui Wang, Fei Wu

Abstract— **Recent studies show face recognition (FR) with additional features achieves better performance than that with single one. Different features can represent different characteristics of human faces, and utilizing different features effectively will have positive effect on FR. Meanwhile, the advances of sparse coding enable researchers to develop various recognition methods to cooperate with multiple features. However, even if these methods achieve very encouraging performances, there still exist some intrinsic problems. Firstly, these methods directly encode the multiple features over the original training set, by which way some redundant, noisy and trivial information are incorporated and the recognition performance can be compromised. Moreover, when the training data increase in number, the jointly-encoding process can be very time-consuming. Thirdly, these methods ignore some semantic relationships among the features, which can boost the FR performance. Thus, coarsely utilizing all the features not only adds extra computation burden, but also prevent further improvement. To address these issues, we propose to fuse the multiple features into a more preferable presentation, which is more compact and more discriminative for better FR performance. As well, we take advantage of the dictionary learning framework to derive an effective recognition scheme. We evaluate our model by comparing it with other state-of-the-art approaches, and the experimental results demonstrate the effectiveness of our approach.**

I. Introduction

Different human faces have different characteristics, and there are many algorithms designed to exploit them. Researchers have already realized studying multiple features jointly will have positive effect on face recognition (FR) performances [5], [26]. However, simply putting multiple features together will bring much redundant information which contributes little to the recognition tasks. Therefore pursuing effective and efficient methods is still an urgent problem.

Meanwhile, FR with supervised dictionary learning (DL) has raised a lot of attention in recent years [21], [8], [11]. With the label information and sparse representation over the learned dictionary, the classification-oriented dictionary is mainly derived in two ways [10]:

1) directly making the dictionary discriminative, such as learning a class-specific sub-dictionary for each class;
2) making the sparse coefficients discriminative to propagate the discrimination power to the dictionary.

Even through DL-based classification methods achieve very promising or even state-of-the-art performances on many public databases, they cannot take advantage of multiple features for further improvement.

In order to extend the capability of sparse coding framework to study multiple features jointly, researchers have proposed several methods [26], [27], [23]. Yuan and Yan propose a multi-task joint sparse representation based classification method (MTJSRC), which treats the recognition with multiple features as a multi-task problem, and each feature type is one task [26]. They assume that the coefficients share the same sparsity pattern among all the features. However, this assumption is too strict and is not held in practice. Therefore, Zhang *et al.* propose a joint dynamic sparse representation classification method (JDSRC) [27] to address this problem. They argue that the same sparsity pattern is shared among the coefficients at class-level, but not necessarily at atom-level. Yang *et al.* also address this problem by proposing a relaxed collaborative representation method (RCR), which assumes the sparse codes among different features should be similar in appearance [23].

All the above three methods elaborately consider the sparsity pattern among the coefficients of different features, and achieve very promising FR performances. However, these methods merely use the training data as an overall-dictionary, which can be very large when the number of training data increases. Also, simply taking all features into computation leads to a very time-consuming sparse coding procedure, and bring much more redundant information into dictionary. Furthermore, the different features are only connected through sparse coefficients but the internal relationship, which can semantically bridge different features to enhance FR performance, are not fully utilized. If we suppress the redundant and noisy information between different features and incorporate this relationship, we can further improve the performance on FR.

To this end, we propose a two-step method to learn a more compact and more discriminative dictionary:

1) We fuse the data into a more compact and more discriminative representation.
2) With this preferable data representation, we learn a core dictionary for better FR performance.

As shown in the experiment results, this two-step method achieves very decent performance with an easy implementation.

The rest of this paper is organized as follows. In Section II, we briefly introduce some tensor algebras used in our model and the related work. The proposed model is elaborated in Section III. Experiments are carried out in Section IV. Finally, we conclude our paper in Section V with discussions on future work.

This work is supported by 973 Program (No.2010CB327904) and Natural Science Foundations (No.61071218) of China.

The authors are with the Department of Computer Science and Technology, Zhejiang University, Hangzhou 310027, China {aimerykong, xkwang, wufei, dhwang}@zju.edu.cn

II. Preliminary

In this section, we first introduce some tensor algebras and notations that are used in our work. Then we discuss three closely related methods in FR with multiple features.

A. Tensor algebra with notations

We use the tensor algebra to formulate our multiple feature study problem. The computation and notation mainly follow [9], [18]. High-order tensors are denoted by boldface Euler script letters, *e.g.* $\mathcal{X}$. The mode-n flattening of tensor $\mathcal{X}$ is denoted by $\mathcal{X}_{(n)}$. The k-mode product of a K^{th}-order tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times \cdots \times I_{k-1} \times I_k \times I_{k+1} \times \cdots \times I_K}$ by matrix $\mathbf{U} \in \mathbb{R}^{J \times I_k}$ is expressed as $\mathcal{X} \times_k \mathbf{U} \in \mathbb{R}^{I_1 \times \cdots \times I_{k-1} \times J \times I_{k+1} \times \cdots \times I_K}$.

B. Related work

Various sparse representation based methods for FR are proposed in recent years. Wright *et al.* propose SRC [19] which achieves promising performance in FR. SRC uses the all training data as a predefined dictionary $\mathbf{D}$ to approximately represent the query image through sparse coding framework, where $\mathbf{D} = [\mathbf{x}_1, \ldots, \mathbf{x}_i, \ldots, \mathbf{x}_N]$ and $\mathbf{x}_i$ is the i^{th} training sample. The query image is assigned to the class according to the reconstruction error of each sub-dictionary. Even though SRC achieves quite good performance, one drawback of SRC is that it can only deal with single feature. To overcome this major drawback, researchers have developed several methods to extend SRC for multiple features FR.

One intuitive way to bring various features into computation is to use K different dictionaries $\mathbf{D}_k$'s, each one for each feature. Dictionary $\mathbf{D}_k$ consists of the k^{th} feature of all the training samples batched together directly. We name this extension as separate SRC (S-SRC) [27], as demonstrated in upper panel of Fig. 1, which constructs each dictionary for each feature independently, and summarizes the reconstruction errors of all features in each class for classification. Another way is to concatenate each different feature into a huge vector and calculate the reconstruction error of each class for classification. We name this extension as holistic SRC (H-SRC) [19]. Although S-SRC uses all features for classification, it fails to incorporate the correlation between different features.

Focusing on this, Yuan and Yan treat all the K features as K tasks, and solve the multi-task problem for multi-feature classification (MTJSRC) [26]. Their method assumes the sparse coefficients share the same sparsity pattern at atom-level, as demonstrated by Fig. 1 (a). Zhang *et al.* propose a joint dynamic sparse coding method (JDSRC) [27] to force the same class-level sparsity, but allows different atom-level sparsity within groups, as illustrated by Fig. 1 (b). Another method (RCR) proposed by Yang *et al.* [23] considers the coefficients corresponding to multiple features to be similar measured by ℓ_2-norm distance, as shown by Fig. 1 (c), rather than atom-level identity in [26] and group-level identity in [27]. Note that, in [23], the sparse coefficients corresponding to all the K features are not only forced to have the similar sparsity pattern in appearance, but also pushed to have similar non-zeros values. To summarize, given a query image $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_i, \ldots, \mathbf{x}_K] \in \mathbb{R}^{p \times K}$, in which $\mathbf{x}_k$ denotes the k^{th} feature[1], the three methods solve the following objective function to calculate the reconstruction error, but differing in the constraint $\Phi(\cdot)$ on the coefficients for each sparsity pattern:

$$\{\mathbf{a}_1, \ldots, \mathbf{a}_K\} = \underset{\mathbf{a}_1, \ldots, \mathbf{a}_K}{\operatorname{argmin}} \sum_{k=1}^{K} \|\mathbf{x}_k - \mathbf{D}_k \mathbf{a}_k\|_2^2 + \lambda \Phi(\mathbf{a}_1, \ldots, \mathbf{a}_K), \tag{1}$$

where $\mathbf{D}_k$ is the dictionary whose columns are the k^{th} feature vectors of training image. The final classification scheme of the three methods follows that of SRC, *i.e.* identifying the query image to the class which sub-dictionaries of all the K features produce the smallest reconstruction error in total.

III. Dictionary learning with multiple feature fusion

As demonstrated in Section II, all the three methods put too much effort in imposing constraint on coefficients and ignore the semantic relationship among different features. There are some drawbacks among these methods:

1) As the training sample number grows and feature number increases, the FR process will become more time-consuming.
2) Directly using the training sample features as dictionary atoms will bring much redundant information. These information will bring negative effect to the FR performance.
3) From classification perspective, different features of the same object will have some semantic relationships. Neglecting such connection will hinder further improvement for the FR performances.

Concerning above issues, we propose a method to learn a more discriminative representation for face images, which fuses the features for better FR. In this section, we elaborate the proposed method in detail.

A. General framework

Firstly, to overcome the first two problem, we propose to learn K dictionaries for all features, instead of using the training set as a predefined dictionary. This technique has been used in [22] to deal with single feature FR. It is worth mentioning that the order of dictionary atoms is critical to some methods (MTJSRC [26] and RCR [23]), since the classification relies on the sparsity pattern which correlates with the atom order. Therefore, only the method proposed in [27] can be extended in this way, since it considers group-level sparsity rather than atom-level sparsity. However, owing to different classification scheme, our method does not suffer from this problem. We postpone to discuss it in Subsection III-D.

Suppose we have already learned K dictionaries ($\mathbf{D}_k \in \mathbb{R}^{p \times d}$ for the k^{th} feature). We arrange them to a tensorial

[1]In this paper, we assume all features are with equal length, so that we can arrange them in order directly.

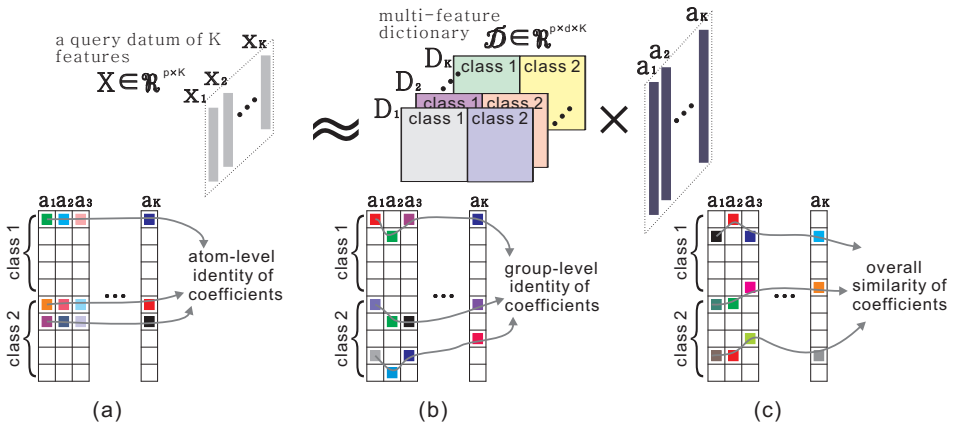

Fig. 1. The upper panel shows the K features of a query datum $\mathbf{X}$ are approximated by K dictionaries with K sparse coefficients. Three existing methods impose different constraints on the coefficients among the K coefficients: atom-level sparsity [26] shown in (a), group-level sparsity [27] in (b), and overall similarity [23] in (c).

representation $\mathcal{D} \in \mathbb{R}^{p\times d\times K}$ as illustrated by the upper panel of Fig. 1. Our goal is to utilize the relationship among these dictionaries for better FR performance and to lower the computational burden. In our work, we assume there exists a core dictionary $\mathcal{B} \in \mathbb{R}^{p\times d\times M}$ ($M < K$ or $M \ll K$) which can get rid of redundant information among different features. This dictionary $\mathcal{B}$ can be linearly transformed into $\mathcal{D}$. In other words, there is a transform matrix $\mathbf{W} \in \mathbb{R}^{K\times M}$, such that $\mathcal{D} = \mathcal{B} \times_3 \mathbf{W}$, which means we transform $\mathcal{B}$ into $\mathcal{D}$ along the third mode through the transformation matrix $\mathbf{W}$, as illustrated by Fig. 2. Therefore, with the core dictionary $\mathcal{B}$ and the transformation/fusion matrix $\mathbf{W}$, we rewrite Eq. 1 to derive the new objective function as below:

$$\begin{aligned}\{\mathbf{a}_1,\ldots,\mathbf{a}_K\} = \underset{\mathbf{a}_1,\ldots,\mathbf{a}_K}{\operatorname{argmin}} \sum_{k=1}^{K} \|\mathbf{x}_k - \mathbf{D}^k\mathbf{a}_k\|_F^2 + \lambda\Phi(\mathbf{a}_1,\ldots,\mathbf{a}_K),\\ \text{s.t. } \mathcal{D} = \mathcal{B}\times_3 \mathbf{W},\\ \mathbf{D}^k \text{is the } k^{th} \text{ slice of } \mathcal{D} \text{ along the third mode.}\end{aligned} \tag{2}$$

Here the Lagrange constraint $\Phi(\cdot)$ is imposed on the coefficients corresponding to the K multi-feature dictionaries, such as ℓ_1-norm penalty [19], group sparsity [27] and overall-similarity term [23].

However, if we base our FR method on Eq. 2, we still need to compute all the features of each query image, *i.e.* this core dictionary does not bring any computation benefit to our new model.

In order to explore the feature correlation explicitly, we take an alternative solution. Instead of transforming K-feature dictionary into M-dimensional core dictionary, we employ the fusion matrix $\mathbf{W}$ directly on query image $\mathbf{X}$, which gives us a compact representation $\mathbf{Y} \in \mathbb{R}^{p\times M}$ subject to $\mathbf{Y} = \mathbf{X}\mathbf{W}$.

Therefore, we have an alternative objective function as below:

$$\begin{aligned}\{\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_M\} = \underset{\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_M}{\operatorname{argmin}} \sum_{m=1}^{M} \|\mathbf{y}_m - \mathbf{B}_m\boldsymbol{\alpha}_m\|_F^2\\ + \lambda\Phi(\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_M),\end{aligned} \tag{3}$$

where $\mathbf{y}_m$ is the vector represent the m^{th} feature of fused datum $\mathbf{y}$, and $\mathbf{B}_m \in \mathbb{R}^{p\times d}$ is the m^{th} sub-dictionary of the core dictionary $\mathcal{B} \in \mathbb{R}^{p\times d\times M}$. As $M < K$ or $M \ll K$, solving the alternative objective function Eq. 3 is more computationally efficient than solving Eq. 2, and the core dictionary $\mathcal{B}$ can be fully exploited. Now, the questions are how to obtain such a desired fusion matrix $\mathbf{W}$ and how to learn such a core dictionary $\mathcal{B}$.

In this paper, we solve this problem through a two-step method, *i.e.* first learning the fusion matrix $\mathbf{W}$ and then learning the core dictionary $\mathcal{B}$.

B. Learning the fusion matrix $\mathbf{W}$

Suppose we have C classes of training data with K features, and there are N_c face images in the c^{th} class. The i^{th} image is denoted as $\mathbf{X}_i \in \mathbb{R}^{p\times K}$. Multiple features can be beneficial to FR, since they bring much valuable information, which will boost the recognition performance. However, more features also bring in more redundancy. To balance the two aspects, fusing multiple feature is a good choice. We hope the fused features are more discriminative for better recognition and more compact for efficient computation. Fisher criterion [2] is the method that increases discrepancy between classes and coherency within classes. Maximizing Fisher criterion is a desirable way to achieve this purpose. Thus, we derive the fusion matrix $\mathbf{W}$ as below:

$$\begin{aligned}\mathbf{W} &= \underset{\mathbf{W}}{\operatorname{argmax}} \frac{\sum_{c=1}^{C} N_c\|(\bar{\mathbf{X}}_c - \bar{\mathbf{X}})\mathbf{W}\|_F^2}{\sum_{c=1}^{C}\sum_{i\in\mathcal{I}_c}\|(\mathbf{X}_i - \bar{\mathbf{X}}_c)\mathbf{W}\|_F^2}\\ &= \underset{\mathbf{W}}{\operatorname{argmax}} \frac{\operatorname{tr}\left\{\mathbf{W}^T\left\{\sum_{c=1}^{C} N_c(\bar{\mathbf{X}}_c - \bar{\mathbf{X}})^T(\bar{\mathbf{X}}_c - \bar{\mathbf{X}})\right\}\mathbf{W}\right\}}{\operatorname{tr}\left\{\mathbf{W}^T\left\{\sum_{c=1}^{C}\sum_{i\in\mathcal{I}_c}(\mathbf{X}_i - \bar{\mathbf{X}}_c)^T(\mathbf{X}_i - \bar{\mathbf{X}}_c)\right\}\mathbf{W}\right\}},\end{aligned} \tag{4}$$

where $\mathcal{I}_c$ is the index set of images from class c, $\bar{\mathbf{X}}_c = \frac{1}{N_c}\sum_{i\in\mathcal{I}_c}\mathbf{X}_i$ is the mean of the c^{th} class, and similarly, $\bar{\mathbf{X}} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{X}_i$ is the global mean. Let $\mathbf{S}_b = \sum_{c=1}^{C} N_c(\bar{\mathbf{X}}_c - \bar{\mathbf{X}})^T(\bar{\mathbf{X}}_c - \bar{\mathbf{X}})$ and $\mathbf{S}_w = \sum_{c=1}^{C}\sum_{i\in\mathcal{I}_c}(\mathbf{X}_i - \bar{\mathbf{X}}_c)^T(\mathbf{X}_i - \bar{\mathbf{X}}_c)$. Thus, solving Eq. 4 to derive $\mathbf{W}$ is equivalent to calculating

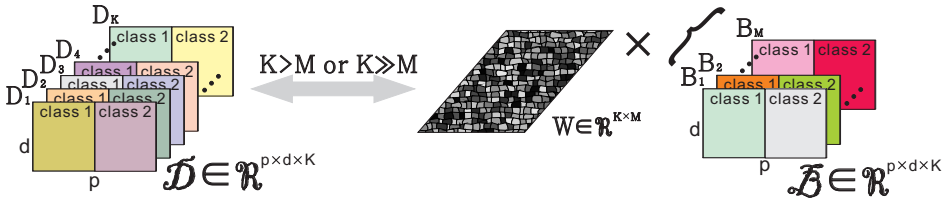

Fig. 2. The multi-feature dictionary $\mathcal{D}$ can be constructed from the core dictionary $\mathcal{B}$ and the transformation $\mathbf{W}$.

the generalized eigenvalue problem [6]: $\mathbf{S}_b\mathbf{w} = \lambda\mathbf{S}_w\mathbf{w}$, for $\lambda \neq 0$. In detail, we have $\mathbf{W} = [\mathbf{w}_1, \ldots, \mathbf{w}_m, \ldots, \mathbf{w}_M]$, where $\mathbf{w}_m$ is the eigenvector corresponding to the m^{th} largest eigenvalue of $\mathbf{S}_w^{-1}\mathbf{S}_b$. Note that Eq. 4 is not the same as the classical linear discriminant analysis (LDA) [6], but is a special case of two-dimensional LDA [24] or multilinear discriminant analysis [20] that only deals with the relationship of multi-feature information along the 2^{nd}-mode. Moreover, the tensorial application, which is brought in to resolve the multi-feature learning, can alleviate overfitting problem to some extent, especially when the training sample number is limited [13].

C. Learning the core dictionary $\mathcal{B}$

After obtaining the desired fusion matrix $\mathbf{W}$, we can fuse the K features into more compact and more discriminative M features through $\mathbf{Y}_i = \mathbf{X}_i\mathbf{W} \in \mathbb{R}^{p\times M}$ for $i = 1, \ldots, N$.

With the fused training data, we divide the dictionary $\mathcal{B} \in \mathbb{R}^{p\times d\times M}$ into MC sub-dictionaries $\mathbf{B}_m^{(c)}$ for c^{th} individual and m^{th} feature. There are many efficient and effective methods to learn such classification-oriented dictionaries [21], [8], [11]. In this paper, for simplicity and effectiveness, we choose K-SVD algorithm [1] to learn each sub-dictionary for each feature.

D. Classification scheme

As introduced in Section II, the three methods [26], [27], [23] impose different structured sparsity contraints $\Phi(\cdot)$ on the coefficients for multi-feature dictionaries. Especially, in [27], group sparsity constraint is proposed to explore the sparsity pattern at group-level, but allowing discriminative pattern at atom-level within each group. In our method, we propose a similar group-level sparsity constraint, but in a strict fashion. When considering class c, we only allow the atoms from class c to contribute to the representation of the query image.

We can use local sparse coding based method for classification [21], [11] under this constraint, which is to calculate the reconstruction error of each sub-dictionary for the query image. This becomes a least square problem which is much easier and more efficient than solving LASSO problem [17] or group-level sparsity problem [25].

The detailed classification procedure is sketched as below:

1) Given a query face image $\mathbf{X} \in \mathbb{R}^{p\times K}$ consisting of K features, apply the fusion matrix $\mathbf{W}$ to it and derive the fused datum $\mathbf{Y} = \mathbf{X}\mathbf{W} \in \mathbb{R}^{p\times M}$.
2) Calculate the reconstruction error by the corresponding sub-dictionary $\mathbf{B}_m^c$ of the c^{th} class and m^{th} feature, for $c = 1, \ldots C$ and $m = 1, \ldots, M$:
$$e_c^{(m)} = \min_{\boldsymbol{\alpha}} \|\mathbf{y}_m - \mathbf{B}_m^c\boldsymbol{\alpha}\|_2^2, \tag{5}$$
3) Summarize the reconstruction error of all the M features of each class: $e_c = \sum_{m=1}^{M} e_c^{(m)}$, and identify the query image to the class which produces the smallest reconstruction error on all the M features: $label(\mathbf{X}) = \mathrm{argmin}_c\, e_c$.

IV. Experiment

In this section, we evaluate our method by experiments on three databases: Extended Yale B [12], AR [14] and CMU-PIE [15]. To fairly demonstrate the effectiveness of our method, we compare it with some closely related approaches. These methods include holistic SRC (H-SRC) [19], separate SRC (S-SRC) [27], MTJSRC [26], JDSRC [27] and RCR [23]. H-SRC and S-SRC act as baseline methods, in which H-SRC concatenates all the features into a huge one, while S-SRC, as an intuitive extension of SRC, separately reconstructs multiple features and then summarizes the reconstruction error of each feature for classification.

We extract ten types of features in each image for multi-feature FR, one is the original gray-scale image, and the other nine are low-level visual features generated from the original image, as illustrated in Fig. 3. The features are[2]:

1) original gray-scale image.
2) the image after histogram equalization.
3) low-frequency Fourier feature [16] of original image with cut-off frequency 8.
4) low-frequency Fourier feature of histogram equalized image with cut-off frequency 8.
5) edge image of original image by Sobel operator [7] (threshold=t_{s1}) with 3×3 mean filter.
6) edge image of histogram equalized image by Sobel operator (threshold=t_{s2}) with 3×3 mean filter.
7) edge image of original image by Canny operator [4] (threshold=t_{c1} and $\sigma = \sigma_1$) with 3×3 mean filter.
8) edge image of histogram equalized image by Canny operator (threshold=t_{c2} and $\sigma = \sigma_2$) with 3×3 mean filter.

[2]Since images from different databases are with various resolutions and under different illumination condition, here we only give the notation of each parameter of the feature extraction algorithms. The parameter value for each database is listed in Table I

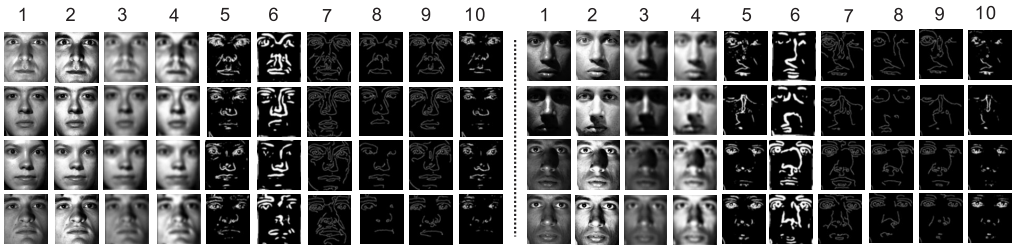

Fig. 3. The ten features of four persons. The left panel shows the features of four persons, on each row, and the right one displays the features of one same person under four different illumination conditions.

TABLE I
DIFFERENT FEATURE EXTRACTION PARAMETER VALUES FOR EACH DATABASE.

Parameter Value	Extended Yale B	AR	CMU-PIE
t_{s1}	15	12	3
t_{s2}	20	11	8
t_{c1}	0.3	0.1	24
σ_1	0.5	15	0.15
t_{c2}	0.2	0.1	1.0
σ_2	1.0	1.1	0.4
t_{c3}	0.5	0.1	0.25
σ_3	1.3	0.6	0.9
t_{s4}	32	25	15

9) edge image of original image by Canny operator (threshold=t_{c3} and $\sigma = \sigma_3$) with 3×3 mean filter.
10) edge image of original image by Sobel operator (threshold=t_{s4}) with 3×3 mean filter.

The feature 9 and 10 use same algorithm as feature 7 and 5, but we force the edge image only containing major edges of original image. Note that, MTJSRC only uses two types of features in [26], whereas it performs better on the ten features in this paper than that in [26]; JDSRC manually selects the face regions in [27] for multi-region FR, yet we run it on the global features to report the results; RCR also considers sophisticatedly segmenting the face images for multi-region FR, but we use RCR for multi-feature FR to report its performance.

A. Face Recognition

In this subsection, we compare our method with H-SRC, S-SRC, MTJSRC, JDSRC and RCR in face recognition application on three face databases: Extended Yale B, AR and CMU-PIE. The different settings of each database are described as below:

- **Extended Yale B [12]**: This database contains 2,414 frontal face images of 38 persons, about 64 frontal images for each individual under different poses and illumination conditions. All images are manually aligned, cropped, and resize to 168×192.
- **AR [14]**: This database contains 3,120 images of 120 individuals, and 26 images for each individual. These images are captured into two separate sessions, each session contains 13 images with different facial expressions, illuminations and occlusions. In this setting, we combine these two sessions together.
- **CMU-PIE [15]**: The CMU-PIE dataset contains $41,368$ images of 68 people. Each person under 13 different poses, 43 different illumination conditions, and with 4 different expressions. Here we select a subset provided by [3] which contains 5 frontal poses (C05, C07, C09, C27, C29). So, there are 170 images for each individuals.

For each database, we randomly select half images of each person for training and the rest for testing. Thus, we have about 32 in Extend Yale B, 13 in AR and 85 in CMU-PIE training images for each individual. We learn a d-atom dictionary $\mathbf{B}_m^c$ for each fused dimensionality m of every class c, where $d = 10$ for Extend Yale B, $d = 5$ for AR and $d = 20$ for CMU-PIE.

We randomly select K features from the ten features, where K ranges from 4 to 10. For our method, we fuse K features into three. For H-SRC, we concatenate all features into a single huge vector, and run FR in the SRC fashion. For S-SRC, we calculate the reconstruction error of each feature individually, and summarize them up for FR. For each K, we run each method 10 times on each database. We list the mean accuracies with standard deviations of each method given $K = 4$ and $K = 10$ on each database in Table II. The curve figures of all results on each database are shown in Fig. 4, Fig 5 and Fig 6 for Extended Yale B, AR and CMU-PIE, respectively.

As showed by the results, H-SRC achieves good performance, while S-SRC derives no better accuracy. This is because H-SRC and S-SRC blindly use multiple feature without utilizing the relationship among them. When different features share many common patterns with each other, *e.g.* the same occlusions on different person in AR, this will lead to misclassification. MTJSRC, JDSRC and RCR clearly improve the results over H-SRC and S-SRC, owing to their reasonable structured constraints on the sparse coefficients, which bridge the multiple features to enhance recognition performance. However, with the proposed fusion method, ours achieves the best recognition rate among all the methods. This illustrates the effectiveness of the proposed method in fusing all these low-level features for better FR. As we take more features into the learning phrase, the accuracies increase correspondingly, since this combines more information into our dictionary.

TABLE II

RECOGNITION ACCURACIES VARYING FUSING FEATURES NUMBER K. WE FUSED K FEATURES INTO 3 FEATURES FOR OUR METHOD.

Accuracy	Extended Yale B		AR		CMU-PIE	
	$K=4$	$K=10$	$K=4$	$K=10$	$K=4$	$K=10$
H-SRC	97.09 ± 0.30	97.40 ± 0.27	92.99 ± 0.31	93.43 ± 0.13	94.39 ± 0.36	94.44 ± 0.14
S-SRC	94.13 ± 0.31	94.85 ± 0.34	90.54 ± 0.24	90.76 ± 0.42	92.04 ± 0.13	92.61 ± 0.46
MTJSRC	98.24 ± 0.36	98.56 ± 0.21	94.55 ± 0.32	95.59 ± 0.44	95.61 ± 0.17	96.92 ± 0.14
JDSRC	98.39 ± 0.24	99.03 ± 0.27	95.18 ± 0.14	95.66 ± 0.30	96.46 ± 0.21	97.51 ± 0.20
RCR	98.04 ± 0.47	98.47 ± 0.11	94.33 ± 0.32	95.42 ± 0.48	95.83 ± 0.41	97.24 ± 0.10
Ours	98.54 ± 0.22	98.83 ± 0.17	95.30 ± 0.44	96.22 ± 0.48	97.13 ± 0.32	97.43 ± 0.33

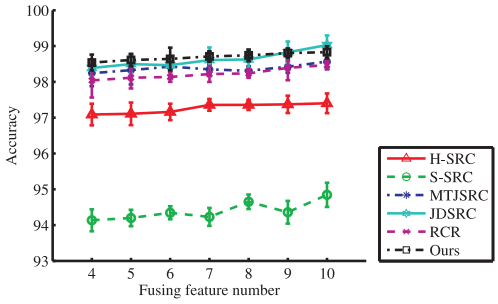

Fig. 4. Recognition accuracy curves by varying fusing feature number on Extended Yale B.

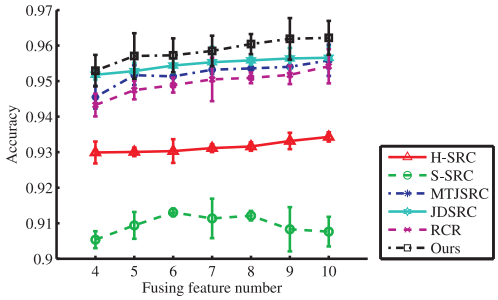

Fig. 5. Recognition accuracy curves by varying fusing feature number feature on AR.

B. FR under different fused dimension

In this section, we inspect the effect of fused dimension number, *i.e.* fused feature number M on FR performance. We implement our experiment on the three different face recognition databases: Extended Yale B, AR and CMU-PIE. For each database, we randomly select half images per each individual as training samples, and the other half for testing. All the ten features are fused into M compact new features, where M ranges from 1 to 6. The experimental result is listed in Table III and the corresponding figure is plotted in Fig. 7.

From this figure, we can see that, when M is small, the recognition performance is relatively lower. Since the dictionary becomes too compact, the intrinsic relationship of different features is not fully exploited, which leads to a relatively lower recognition performance. As M increases, the intrinsic relationship between each feature is well reserved and utilized by fusion matrix $\mathbf{W}$, which makes the recognition accuracy keep growing and gradually becomes stable.

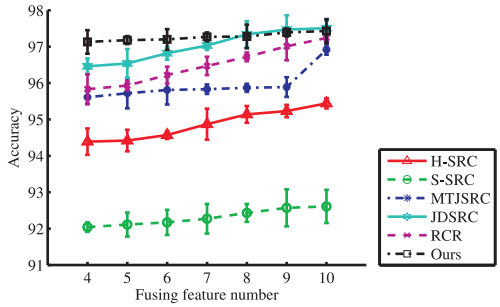

Fig. 6. Recognition accuracy curves by varying fusing feature number on CMU-PIE.

TABLE III

RECOGNITION PERFORMANCE ON VARIOUS NUMBER OF FUSED DIMENSION M. IN THIS SETTING, WE TAKE ALL THE 10 FEATURES FUSED INTO M FEATURES.

M	Extended Yale B	AR	CMU-PIE
1	96.75 ± 0.32	95.19 ± 0.34	95.23 ± 0.50
2	98.45 ± 0.38	95.94 ± 0.14	95.82 ± 0.42
3	98.72 ± 0.21	96.31 ± 0.24	96.18 ± 0.74
4	98.86 ± 0.11	96.25 ± 0.23	96.55 ± 0.60
5	98.98 ± 0.16	96.39 ± 0.11	96.40 ± 0.59
6	98.83 ± 0.19	96.39 ± 0.13	96.59 ± 0.55

C. FR under different training number

In this section, we carry out experiments on Extended Yale B, AR and CMU-PIE to demonstrate how different training number affect FR performance. In this setting, we take $K = 10$ features, and for our method, we fused them into $M = 3$ features. We use 3 different training sample number settings, numbered $1, 2$ and 3, for the above three databases. For Extend Yale B, we use $10, 20$ and 30 training samples per individual respectively. For AR, we use $6, 9$ and 13 training sample per individual respectively. For CMU-PIE, we use $60, 90$ and 110 training sample per individual respectively. The recognition result is listed in Table IV. The recognition performance grows as the training number increases, since there is more discriminative information combined in dictionary. Moreover, when the training sample number is small, the performance of our method is much better than the others, which verifies the effectiveness of multilinear learning to alleviate small sample size problem, as illustrated in [13].

TABLE IV

RECOGNITION ACCURACY ON VARIOUS TRAINING SAMPLE NUMBER. WE USE ALL $K = 10$ FEATURES IN THIS SETTING FOR EACH METHOD. FOR OURS, WE FUSED THE TEN FEATURES INTO $M = 3$ NEW FEATURES.

	Database	H-SRC	S-SRC	MTJSRC	JDSRC	RCR	Ours
Setting-1	Extend Yale B	89.02 ± 0.42	84.48 ± 0.21	88.08 ± 0.45	88.52 ± 0.24	87.84 ± 0.37	$\mathbf{90.83 \pm 0.72}$
	AR	86.24 ± 0.37	84.14 ± 0.18	87.79 ± 0.62	87.89 ± 0.42	87.61 ± 0.11	$\mathbf{88.41 \pm 0.58}$
	CMU-PIE	91.64 ± 0.45	89.37 ± 0.26	92.93 ± 0.35	93.69 ± 0.14	93.22 ± 0.78	$\mathbf{95.79 \pm 0.56}$
Setting-2	Extend Yale B	96.06 ± 0.47	93.09 ± 0.15	97.22 ± 0.51	97.29 ± 0.23	96.94 ± 0.39	$\mathbf{97.82 \pm 0.23}$
	AR	90.82 ± 0.23	88.63 ± 0.14	92.03 ± 0.40	92.66 ± 0.34	92.20 ± 0.24	$\mathbf{92.96 \pm 0.35}$
	CMU-PIE	94.83 ± 0.40	92.80 ± 0.77	96.27 ± 0.22	96.51 ± 0.49	96.14 ± 0.40	$\mathbf{97.08 \pm 0.82}$
Setting-3	Extend Yale B	96.39 ± 0.26	94.47 ± 0.58	97.95 ± 0.24	98.70 ± 0.49	98.21 ± 0.23	$\mathbf{98.88 \pm 0.48}$
	AR	94.08 ± 0.69	91.93 ± 0.55	95.40 ± 0.12	96.13 ± 0.34	95.38 ± 0.13	$\mathbf{96.30 \pm 0.22}$
	CMU-PIE	95.72 ± 0.43	93.10 ± 0.14	96.94 ± 0.18	96.75 ± 0.56	96.68 ± 0.94	$\mathbf{97.48 \pm 0.17}$

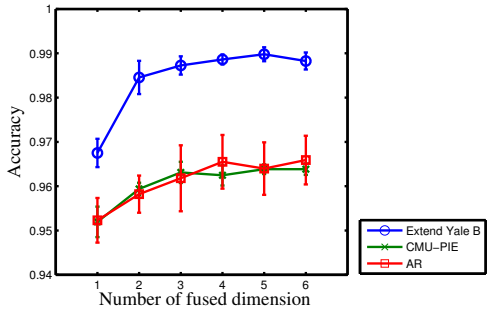

Fig. 7. Curves of Recognition Accuracy varying fused dimension M.

V. CONCLUSION

Recently, some methods are developed to deal with multiple types of features through sparse coding techniques. These methods directly use the original training set as multiple dictionaries, and impose some sparsity constraints on the coefficients for FR. We assume the multiple features have some intrinsic relationships, which bridges all these features for better FR performance. With this assumption, we propose a novel method to generate a more compact and more discriminative dictionary for classification, and to fuse the multiple features into a more preferable representation. Through experimental validation, we show our method outperforms other state-of-the-art methods on multi-feature face recognition task.

Despite the decent performance, there is still large room to improve our proposed method. One limitation of our method is that we assume the multiple features have the same length/dimension, and the features used in this paper are global ones. Undoubtedly, local features of different dimensions should be taken into consideration for better classification performance.

REFERENCES

[1] M. Aharon, M. Elad, and A. M. Bruckstein. K-svd: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Trans. Signal Processing*, 54(11):4311–4322, 2006.

[2] C. Bishop et al. *Pattern recognition and machine learning*, volume 4. springer New York, 2006.

[3] D. Cai, X. He, and J. Han. Spectral regression for efficient regularized subspace learning. In *Proc. Int. Conf. Computer Vision (ICCV'07)*, 2007.

[4] J. Canny. a computational approach to edge detection. *IEEE Trans. Pattern Anal. Mach. Intelligence*, 8(6):679–698, 1986.

[5] L. Cao, J. Luo, F. Liang, and T. S. Huang. Heterogeneous feature machines for visual recognition. In *ICCV*, 2009.

[6] K. Fukunaga. *Introduction to Statistical Pattern Classification*. Academic Press, San Diego, California, USA, 1990.

[7] R. C. Gonzalez and R. E. Woods. *Digital Image Processing*. Prentice Hall, Upper Saddle River, New Jersey, USA, third edition, 2008.

[8] Z. Jiang, Z. Lin, and L. S. Davis. Learning a discriminative dictionary for sparse coding via label consistent k-svd. In *CVPR*, 2011.

[9] T. G. Kolda and B. W. Bader. Tensor decompositions and applications. *SIAM Review*, 51(3):455–500, September 2009.

[10] S. Kong and D. Wang. A brief summary of dictionary learning based approach for classification. *CoRR abs/1205.6544*, 2012.

[11] S. Kong and D. Wang. A dictionary learning approach for classification: separating the particularity and the commonality. In *ECCV*, 2012.

[12] K. Lee, J. Ho, and D. Kriegman. Acquiring linear subspaces for face recognition under variable lighting. *IEEE Trans. Pattern Anal. Mach. Intelligence*, 27(5):684–698, 2005.

[13] H. Lu, K. Plataniotis, and A. Venetsanopoulos. A survey of multilinear subspace learning for tensor data. *Pattern Recognition*, 2011.

[14] A. Martinez. The ar face database. *CVC Technical Report*, 24, 1998.

[15] T. Sim, S. Baker, and M. Bsat. The cmu pose, illumination, and expression (pie) database. In *Automatic Face and Gesture Recognition, 2002. Proceedings. Fifth IEEE International Conference on*, pages 46–51. Ieee, 2002.

[16] Y. Su, S. Shan, X. Chen, and W. Gao. hierarchical ensemble of global and local classifiers for face recognition. In *ICCV*, 2007.

[17] R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B 58*, 15:267–288, 1996.

[18] D. Wang and S. Kong. Feature selection from high-order tensorial data via sparse decomposition. *Pattern Recognition Letters*, 2012.

[19] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma. Robust face recognition via sparse representation. *IEEE Trans. Pattern Anal. Mach. Intelligence*, 2009.

[20] S. Yan, D. Xu, Q. Yang, L. Zhang, and X. Tang. Multilinear discriminant analysis for face recognition. *IEEE Trans. Image Processing*, 2007.

[21] M. Yang, L. Zhang, X. Feng, and D. Zhang. Fisher discrimination dictionary learning for sparse representation. In *ICCV*, 2011.

[22] M. Yang, L. Zhang, J. Yang, and D. Zhang. metaface learning for sparse representation based face recognition. In *ICIP*, 2010.

[23] M. Yang, L. Zhang, D. Zhang, and S. Wang. relaxed collaborative representation for pattern classification. In *CVPR*, 2012.

[24] J. Ye, R. Janardan, and Q. Li. Two-dimensional linear discriminant analysis. In *NIPS*, 2004.

[25] M. Yuan and Y. Lin. Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(1):49–67, 2005.

[26] X.-T. Yuan and S. Yan. visual classification with multi-task joint sparse representation. In *CVPR*, 2010.

[27] H. Zhang, N. M. Nasrabadi, Y. Zhang, and T. S. Huang. Multi-observation visual recognition via joint dynamic sparse representation. In *ICCV*, 2011.