This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2017.2695239, IEEE Access

A Survey of Dictionary Learning Algorithms for Face Recognition

Yong Xu, Senior *Member*, *IEEE*, Zhengming Li, Jian Yang, David Zhang, *Fellow*, *IEEE*

***Abstract*—During the past several years, as one of the most successful applications of sparse coding and dictionary learning, dictionary based face recognition has received significant attention. Although some surveys of sparse coding and dictionary learning have been reported, there is no specialized survey concerning dictionary learning algorithms for face recognition. This paper provides a survey of dictionary learning algorithms for face recognition. To provide a comprehensive overview, we not only categorize existing dictionary learning algorithms for face recognition but also present details of each category. Since the number of atoms has an important impact on classification performance, we also review the algorithms for selecting the number of atoms. Specifically, we select six typical dictionary learning algorithms with different numbers of atoms to perform experiments on face databases. In summary, this survey provides a broad view of dictionary learning algorithms for face recognition and advances study in this field. It is very useful for readers to understand the profiles of this subject and to grasp the theoretical rationales and potentials as well as their applicability to different cases of face recognition.**

Index Terms—dictionary learning; sparse coding; face recognition

I. INTRODUCTION

Face recognition is an important research topic in computer vision and pattern recognition. With inspiration from the sparsity mechanism of the human vision system and the success of sparse coding in image processing, the sparse representation based classification algorithm has received sufficient attention and achieved excellent performance in face recognition [1-2]. However, research has demonstrated that learning a desired dictionary from training data instead of using off-the-shelf bases(e.g., wavelets) could lead to state-of-the-art results in many practical applications, such as face recognition [3-4], de-noising [5-6], clustering [7-8], image super-resolution [9-10] image de-blurring [11-12] and image segmentation [13]. That is, the obtained dictionary plays an important role in the success of the sparse representation, which allows an input signal to be faithfully and discriminatively represented as a sparse linear combination of atoms. Therefore, many dictionary learning algorithms have been proposed for different applications. The characteristics of sparse coding and dictionary learning algorithms have been presented in the past years. Elad [14] offered a brief presentation of sparse and redundant representation modelling and outlined ten key future research directions for sparse coding. Rubinstein et al. [15] described the evolution process of how to obtain a dictionary by using mathematical and learned models. Tosic et al. [16] presented a broad overview of dictionary learning algorithms and showed their usage in various applications, such as audio-visual coding and stereo image approximation. Specifically, they discussed the discriminative power of sparse representations and outlined the benefits of dictionary learning in classification and face recognition applications. Cheng et al. [17] presented a survey of algorithms on sparse representation, learning and modelling with an emphasis on visual recognition, which addressed both theory and application aspects. Gangeha et al. [18] provided a review of supervised dictionary learning and sparse representation and divided the dictionaries into six categories based on the approach of using label information in learning the dictionary and/or sparse representation. Zhang et al. [19] presented a comprehensive overview of sparse representation, summarized various available sparse representation methods and discussed their motivations, mathematical representations and applications.

Although the above surveys provide a broad review of sparse coding and dictionary learning, there is no survey of dictionary learning algorithms for face recognition. For face recognition, because of varying poses, illuminations and facial expressions, a test sample usually cannot be well represented by original training samples. However, a dictionary is able to effectively model the pose, illumination and facial expression information including the corresponding variations, so a test sample can be better represented by atoms of the dictionary. There are a number of works concerning dictionary learning based face recognition over the past decade. Therefore, it is necessary to review the ideas, technical potential and performance of dictionary learning algorithms for face recognition. Moreover, this survey offers some in-depth insights into the studies of face recognition based on dictionary learning, including key points and some important details. In terms of the objectives of

This work was supported in part by Shenzhen Council for Scientific and Technological Innovation under Grant JCYJ2015033015522059, Foundation for Young Talents in Higher Education of Guangdong, grants No. 2015KQNCX08 and Open Fund Project of Fujian Provincial Key Laboratory of Information Processing and Intelligent Control (Minjiang University) (No. MJUKF201720).

Corresponding author：Y. Xu is with the Bio-Computing Research Center, Shenzhen Graduate School, Harbin Institute of Technology, Shenzhen 518055, China (e-mail: laterfall@hitsz.edu.cn).

Z. Li is with the Industrial Training Center, Guangdong Polytechnic Normal University, Guangzhou 510665, China. He is also with Fujian Provincial Key Laboratory of Information Processing and Intelligent Control (Minjiang University), Fuzhou 350121, China (e-mail: lizhengming2004@126.com).

J. Yang is with the School of Computer Science, Nanjing University of Science and Technology, Nanjing, Jiangsu 210094, China(e-mail: csjyang@mail.njust.edu.cn; tangjh1981@acm.org).

D. Zhang is with the Biometrics Research Centre, Department of Computing, Hong Kong Polytechnic University, Kowloon 999077, Hong Kong(e-mail: csdzhang@comp.polyu.edu.hk).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2017.2695239, IEEE Access

dictionary learning algorithms for face recognition, we can divide them into five categories, i.e., shared dictionary learning algorithms, class-specific dictionary learning algorithms, commonality and particularity dictionary learning algorithms, auxiliary dictionary learning algorithms and domain adaptive dictionary learning algorithms. A shared dictionary learning algorithm can capture the common characteristics of face images, and it usually cannot adequately preserve specific characteristics of face images of each class. A class-specific dictionary learning algorithm can capture the main characteristics of face images of each class, whereas it usually contains considerable redundant information. A commonality and particularity dictionary learning algorithm can not only preserve common characteristics of the face images but can also preserve specific characteristics of the face images of each class. An auxiliary dictionary learning algorithm uses images of external faces, i.e., outsiders, to represent possible variations of the face images. A domain adaptive dictionary learning algorithm applies domain adaption to face recognition, which can perform well in the case where the training and test samples do not have the same distribution. Additionally, we provide discussions of the atom selection methods, which play an important role in the process of dictionary learning.

The remainder of this paper is organized as follows. Section II introduces the shared dictionary learning algorithm. Section III presents the class-specific dictionary learning algorithm. Section IV gives an introduction to the commonality and particularity dictionary learning algorithm. The auxiliary dictionary learning algorithm is presented in Section V. The domain adaptive dictionary learning algorithm is presented in Section VI. The algorithm for selecting atoms is introduced in Section VII. The experimental results of seven dictionary learning and sparse coding algorithms are presented in Section VIII. Finally, the conclusions are presented in Section IX.

II. SHARED DICTIONARY LEARNING ALGORITHM

When the inter-class variation of the face images is small, a shared dictionary can adequately capture the main characteristics of the face images, such that the dictionary obtained using the training samples can represent a test sample. A shared dictionary learning algorithm only learns a dictionary by using training samples of all classes and expects the obtained dictionary to have discriminative ability for different classes. Then, a test sample can be represented by using the learned dictionary, and the representation coefficients are used for classification. The K-SVD algorithm is one of the most well-known shared dictionary learning algorithms [20]. Many variants of the original K-SVD algorithm have been proposed and applied in image de-noising and image reconstruction [21-23]. The K-SVD algorithm focuses on reconstruction. In general, it is used as a benchmark dictionary learning algorithm for face recognition.

A typical and significant shared dictionary learning algorithm was proposed by Jiang [24], which first assigned a label to each atom by using the K-SVD algorithm and then constructed a discriminative sparse coding error term by exploiting the labels of the atoms (LC-KSVD). Thus, it can improve the discriminative ability of the shared dictionary. The objective function of the LC-KSVD algorithm is as follows.

$$\begin{aligned}&\min_{\mathbf{D},\mathbf{X},\mathbf{W},\mathbf{A}} \left\|\mathbf{Y}-\mathbf{D}\mathbf{X}\right\|_2^2 + \alpha\left\|\mathbf{H}-\mathbf{W}\mathbf{X}\right\|_2^2 + \beta\left\|\mathbf{Q}-\mathbf{A}\mathbf{X}\right\|_2^2 \\ &\text{subject to } \forall \text{i}, \ \left\|x_i\right\|_0 \le T_0\end{aligned} \tag{1}$$

where $\mathbf{Y}=[y_1,\cdots,y_N]\in\Re^{n\times N}$ are the training samples, and n and N are the dimension and number of them, respectively. The label matrix of training samples $\mathbf{Y}$ is defined as $\mathbf{H}=\left[h_1,\cdots,h_N\right]\in\Re^{C\times N}$ ($h_i=\left[0,\cdots,1,\cdots,0\right]^T\in\Re^C$). Only the j-th entry of h_i is non-zero if training sample y_i is in the j-th class, and C is the class number of the training samples. $\mathbf{D}=[d_1,\cdots,d_K]\in\Re^{n\times K}$ is the learned dictionary, and K is the number of atoms. $\mathbf{X}=[x_1,\cdots,x_N]\in\Re^{K\times N}$ is the coding coefficient matrix. $\mathbf{W}$ is the classifier parameter, and $\left\|\mathbf{H}-\mathbf{W}\mathbf{X}\right\|_2^2$ is the classification error term. $\mathbf{Q}$ is the discriminative sparse code of training sample $\mathbf{Y}$, and it can be defined as $\mathbf{Q}=\left[q_1,\cdots,q_N\right]\in\Re^{K\times N}$. $q_i=\left[q_i^1,\cdots,q_i^K\right]^T=\left[0,\cdots,1,1,\cdots,0\right]^T\in\Re^K$ is the discriminative sparse code corresponding to training sample y_i. The non-zero values of q_i occur at those indices where training sample y_i and atom d_k share the same label. For example, for $\mathbf{D}=\left[d_1,\cdots,d_7\right]$ and $\mathbf{Y}=\left[y_1,\cdots,y_7\right]$, if y_1, y_2, d_1 and d_2 are from the first category, y_3, y_4, d_3 and d_4 are from the second category, and y_5, y_6, y_7, d_5, d_6 and d_7 are from the third category, then $\mathbf{Q}$ should be defined as

$$\mathbf{Q}=\begin{bmatrix}1&1&0&0&0&0&0\\1&1&0&0&0&0&0\\0&0&1&1&0&0&0\\0&0&1&1&0&0&0\\0&0&0&0&1&1&1\\0&0&0&0&1&1&1\end{bmatrix}.$$

Each column of $\mathbf{Q}$ corresponds to a discriminative sparse code for a training sample. $\mathbf{A}$ is the linear transformation matrix, and $\left\|\mathbf{Q}-\mathbf{A}\mathbf{X}\right\|_2^2$ is the discriminative sparse-code error term. α and β are the regularization parameters. T_0 is the sparsity constraint factor that limits the number of non-zero elements.

The LC-KSVD algorithm classification method is as follows.

(1) Test sample y_t can be represented by shared dictionary $\mathbf{D}$ as

$$\begin{aligned}&\arg\min_x \left\|y_t-\mathbf{D}x\right\|_2^2 \\ &s.t. \left\|x\right\|_0 \le T_0\end{aligned} \tag{2}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2017.2695239, IEEE Access

This problem can be solved by using the Orthogonal Matching Pursuit (OMP) algorithm [25]. We denote the obtained optimal representative coefficient by x^*.

(2) The label of test sample y_t can be obtained by using the following equation

$$j = \arg\max\left(\mathbf{W}x^*\right) \tag{3}$$

Classification parameter $\mathbf{W}$ can be calculated by using coding coefficients matrix $\mathbf{X}$ and label matrix $\mathbf{H}$ of the training samples as follows:

$$\mathbf{W} = \mathbf{H}\mathbf{X}^T\left(\mathbf{X}\mathbf{X}^T + \mathbf{I}\right)^{-1} \tag{4}$$

where $\mathbf{I}$ is an identity matrix.

The shared dictionary learning algorithms can also use the SVM methods to classify the test samples. Regardless, the coding coefficient matrix plays an important role in improving the classification performance of the shared dictionary learning algorithm. To improve the discriminative ability of the shared dictionary, many constrained models are proposed. We divide these constrained models into two categories, the locality constrained model and the label constrained model.

A. Locality Constrained Model

The locality information of data plays an important role in sparse coding and dictionary learning. In fact, locality is more essential than sparsity since locality leads to sparsity but not necessary vice versa [26]. Therefore, an increasing number of researchers focus on the locality preservation strategy when designing dictionary learning algorithms for face recognition. Their main goal is to preserve the structure information of training samples and to expect similar training samples to have more similar coding coefficients than other training samples, which is helpful for improving the discriminative ability of the dictionary. A typical algorithm was proposed by Zheng [27], which used the training samples to construct a Laplacian matrix for preserving the locality characteristics. Additionally, Gao [28] constructed a hypergraph Laplacian matrix to preserve local information of the training samples for improving the discriminative ability of the learned dictionary. However, face images of the same person vary with facial poses and expressions as well as illumination, so it is difficult to obtain a robust Laplacian matrix to accurately reflect the locality information of the training samples. Thus, it may degrade the discriminative ability of the learned dictionary. Jiang [29] modelled the problem of discriminative dictionary learning as a graph topology selection problem, which was solved by maximizing a monotonically increasing and submodular objective function. Haghiri [30] presented a discriminative dictionary learning algorithm that preserved the local structure of the training samples. Because the face images usually contain noise, this algorithm might not be robust. Liu [31] constructed a locality constrained dictionary learning algorithm by using the training samples and atoms to preserve the locality information. This can reduce the influence of noise to some extent. Yang [32] proposed a visual feature coding method by using the dictionary structure. Recently, the locality information of atoms has also been used to improve the discriminative ability of the shared dictionary [33].

B. Label Constrained Model

The label constrained model belongs to supervised learning. For face recognition, we usually encounter the problem of insufficiently available labelled samples. Therefore, the label information is very important for face recognition based on dictionary learning. Many dictionary learning algorithms based on the label constraint model have been proposed. For example, a typical algorithm was proposed by Zhang [34], which constructed a classification error term by using the labels of training samples with the goal of improving the discriminative ability of the dictionary. Shrivastava [35] proposed a discriminative dictionary learning algorithm by using partially labelled data. Pham [36] proposed a joint representation and classification framework that achieved the dual goals of finding the most discriminative sparse over-complete encoding and the optimal classifier parameters. Lin [37] proposed an incoherent dictionary learning algorithm by explicitly incorporating a correlation penalty into the dictionary learning model. Guo [38] used the labels of training samples to construct a pair-wise sparse code error term and then combined it with the classification error term to learn a discriminative dictionary for face verification and recognition problems. However, since face images usually contain noise, the coding coefficients would be contaminated and the discriminative ability of the learned dictionary may be degraded. In addition, using the label information of the original training samples to construct the discriminative term could not exploit the discriminative information hidden in the training samples. This also degrades the discriminative ability of the shared dictionary. Recently, the labels of atoms also have been used to improve the discriminative ability of the shared dictionary. Jiang [24] assigned a label to each atom by using the K-SVD algorithm and then constructed a discriminative sparse code error term by using the labels of atoms. Li [33] constructed a label that embedded within the atoms to improve the discriminative ability of the shared dictionary.

Moreover, the kernel method [39-40], non-negative constrained method [41], and the Bayesian method [42] are also used to improve the discriminative ability of the shared dictionary.

The shared dictionary learning algorithm can learn a dictionary for all classes, and the number of atoms is relatively small. However, the differences of the different classes may not be well conveyed in the shared dictionary. Moreover, the noise of face images can also reduce the robust representation ability of the shared dictionary.

III. CLASS-SPECIFIC DICTIONARY LEARNING ALGORITHM

Because face images of the same person vary with facial poses and expressions as well as illumination, the intra-class variation of face images is usually large and even greater than the inter-class variance of face images. Therefore, the class-specific dictionary learning algorithm is usually designed to capture the main characteristics of face images of each class. A class-specific dictionary learning algorithm first learns a dictionary for each class by using face images of the class, and then classifies the test face images by judging which class leads to the minimum reconstruction error. It exploits the reconstruction term to improve the discriminative ability of the learned dictionary. A typical algorithm was proposed by Yang [43], which learned a dictionary based on the Fisher discrimination criterion (FDDL). Specifically, a structured dictionary, whose atoms correspond to the class labels, is proposed, with which not only the representation residual can be used to distinguish different classes, but additionally the representation coefficients have smaller within-class scatter and larger between-class scatter. The objective function of the FDDL algorithm is formulated as follows.

$$\min_{\mathbf{D},\mathbf{X}} \sum_{i=1}^{C} \left(\begin{array}{l} \left\| \mathbf{Y}_i - \mathbf{D}\mathbf{X}_i \right\|_F^2 + \left\| \mathbf{Y}_i - \mathbf{D}_i \mathbf{X}_i^i \right\|_F^2 \\ + \sum_{\substack{j=1 \\ j \neq i}}^{C} \left\| \mathbf{D}_j \mathbf{X}_i^j \right\| \end{array} \right) + \alpha \left\| \mathbf{X} \right\|_1 + \beta \left(tr(S_W(\mathbf{X}) - S_B(\mathbf{X})) + \gamma \left\| \mathbf{X} \right\|_F^2 \right) \quad (5)$$

Where $\mathbf{Y} = [\mathbf{Y}_1, \mathbf{Y}_2 \cdots, \mathbf{Y}_C]$ is the training sample, $\mathbf{Y}_\mathrm{i} (i = 1, 2, \cdots C)$ is the i-th class training sample, $\mathbf{D} = [\mathbf{D}_1, \mathbf{D}_2, \cdots, \mathbf{D}_C]$ is the learned dictionary, and $\mathbf{D}_i (i = 1, \cdots C)$ the i-th sub-dictionary corresponding to the i-th class. $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_C]$ is the coding coefficient matrix, where $\mathbf{X}_i$ is the representation coefficient of $\mathbf{Y}_i$ over $\mathbf{D}$, and $\mathbf{X}_i^j$ is the representation coefficient of $\mathbf{Y}_i$ over $\mathbf{D}_j$. $S_W(\mathbf{X})$ is the within-class scatter of $\mathbf{X}$ and $S_B(\mathbf{X})$ is the between-class scatter of $\mathbf{X}$. $tr(\bullet)$ represents the trace of the matrix, and α, β and γ are the regularization parameters.

The classification method is as follows:

First, test sample y_t is sparsely represented by sub-dictionary $\mathbf{D}_i$, as follows:

$$x_i = \min \left\| y_t - \mathbf{D}_i x_i \right\|_2^2 \ s.t. \left\| x_i \right\|_0 \leq T_0 \quad (6)$$

Then, test sample y_t is classified using

$$\mathrm{identity}(y_t) = \min \left\| y_t - \mathbf{D}_i x_i \right\|_2 \quad (7)$$

where $\mathrm{identity}(y_t)$ is the obtained label for y_t.

Shrivastava [44] proposed a non-linear discriminative dictionary learning algorithm by using the kernel trick. Chen [45] proposed a dictionary learning algorithm for video face recognition. Ma [46] integrated rank minimization into sparse representation for dictionary learning. Cai [47] proposed a support vector guided dictionary learning (SVGDL) algorithm by formulating the discrimination term as the weighted summation of the squared distances between all pairs of coding vectors. Zheng [48] proposed a discriminative dictionary learning algorithm, the Fisher discrimination K-SVD algorithm. Wang [49] proposed a supervised class-specific dictionary learning algorithm by incorporating the similarity constraint and dictionary incoherence terms. It not only captured the correlations between similar samples by sharing dictionaries but also encouraged dictionaries associated with different classes to be independent by enforcing the dictionary incoherence term.

The class-specific dictionary learning algorithm learns a sub-dictionary for face images of each class, and captures particular characteristic of face images of each class. A problem in this type of algorithm is that a class in the face recognition task may have only a few training samples so the information used to obtain a sub-dictionary is limited and the uncertainty of the atoms may increase.

IV. COMMONALITY AND PARTICULARITY DICTIONARY LEARNING ALGORITHM

For complex face recognition tasks, the intra-class variation of face images is usually large and even greater than the inter-class variation of face images. Therefore, the commonality and particularity dictionary learning algorithm is proposed to cope with the inter-class variance by using the commonality dictionary to address the intra-class variance by using the particularity dictionary. A commonality and particularity dictionary learning algorithm learns a particularity dictionary for each category that captures the most discriminative features of this category, and simultaneously learns a commonality dictionary whose atoms are shared by all the categories and only contribute to the representation of the data rather than discrimination. A typical algorithm is proposed by Wang [50], which designed a class-specific dictionary (called particularity) for each category to capture the most discriminative features of the category, and simultaneously learned a shared pattern pool (called commonality), whose atoms were shared by all the categories and only contributed to representation of the data rather than discrimination (DLSPC). The objective function of the DLSPC algorithm is

$$\min_{\mathbf{D},\mathbf{X}} \sum_{i=1}^{C} \left\{ \begin{array}{l} \left\| \mathbf{Y}_i - \mathbf{D}\mathbf{X}_i \right\|_F^2 + \left\| \mathbf{Y}_i - \mathbf{D}\mathbf{Q}_i \mathbf{Q}_i^T \mathbf{X}_i \right\|_F^2 \\ + \left\| \widetilde{\mathbf{Q}}_i^T \mathbf{X}_i \right\|_F^2 + \lambda \phi(\mathbf{X}_i) \end{array} \right\} + \eta \sum_{i=1}^{C+1} \sum_{\substack{j=1 \\ j \neq i}} \Omega(\mathbf{D}_i, \mathbf{D}_j) \quad (8)$$

where $\mathbf{Y}_i$ is the i-thclass training sample, $\mathbf{D} = [\mathbf{D}_1, \mathbf{D}_2, \cdots, \mathbf{D}_C, \mathbf{D}_{C+1}]$ is the learned dictionary, $\mathbf{D}_i (i = 1, 2, \cdots, C)$ represents the particularity of the i-th class, and $\mathbf{D}_{C+1}$ is the commonality dictionary. $\mathbf{X}_i$ refers to the coding coefficients of $\mathbf{Y}_i$ over $\mathbf{D}$. $\mathbf{Q}_i$ is a selection operator and is defined as $\mathbf{Q}_i = \left[q_1^i, \cdots, q_j^i, \cdots, q_{K_i}^i \right]$, in which the j-th column of $\mathbf{Q}_i$ is of the form $q_j^i = \left[\underbrace{0, \cdots, 0}_{\sum_{m=1}^{s-1} K_m}, \underbrace{\overbrace{0, \cdots, 0}^{j-1}, 1, 0, \cdots 0}_{K_s}, \underbrace{0, \cdots, 0}_{\sum_{m=s+1}^{i+1} K_m} \right]^T$. $\phi(x_i)$ is defined as the l_1-norm penalty. $\widetilde{\mathbf{Q}}_i = \left[\mathbf{Q}_i, \mathbf{Q}_{C+1} \right]$,

$\Omega\left(\mathbf{D}_i,\mathbf{D}_j\right)=\left\|\mathbf{D}_i^T\mathbf{D}_j\right\|_F^2$. For the DLSPC algorithm, there are two types of classification methods. λ and η are the parameters. The first type uses the global coding classifier as follows.

(1) Test sample y_t is represented by using the commonality and particularity dictionaries and the formula is

$$\mathbf{X}=\min_X\left\|y_t-[\mathbf{D},\mathbf{D}_1,\cdots,\mathbf{D}_C]\mathbf{X}\right\|_2+\left\|\mathbf{X}\right\|_1 \quad (9)$$

(2) Test sample y_t is assigned to the class with the minimum reconstruction error by using

$$\text{identity}\left(y_t\right)=\min_i\left\|y_t-\left[\mathbf{D},\mathbf{D}_i\right]\mathbf{X}_i\right\|_2^2 \quad (10)$$

The second type is the local coding classifier, which is implemented below.

(1) Test sample y_t is represented by using the particularity dictionary and the formula is

$$\mathbf{X}=\min_{\mathbf{X}}\left\|y_t-[\mathbf{D}_1,\cdots,\mathbf{D}_C]\mathbf{X}\right\|_2+\left\|\mathbf{X}\right\|_1 \quad (11)$$

(2) Test sample y_t is assigned to the class with the minimum reconstruction error by using

$$\text{identity}\left(y_t\right)=\min_i\left\|y_t-\mathbf{D}_i\mathbf{X}_i\right\|_2^2 \quad (12)$$

Sun [51] presented a dictionary learning model to improve sparse representation for image classification, with the goal of learning a class-specific dictionary for each class and a common dictionary shared by all classes. The model is composed of discriminative fidelity, a weighted group sparse constraint, and a class-specific dictionary incoherence term. Because every class must have sufficient representative training samples and the training data must be uncorrupted, Yang [52] proposed an analysis-synthesis commonality and particularity dictionary learning algorithm for face recognition.

The commonality and particularity dictionary learning algorithm not only preserves common characteristics of all face images but also preserves specific characteristics of face images of each class. Therefore, the commonality and particularity dictionary learning algorithm is very suitable for face recognition. However, to design a commonality dictionary and a particularity dictionary with the proper number of atoms is very important and has a severe effect on the performance of face recognition.

V. The Auxiliary Dictionary Learning Algorithm

Face recognition is a typical small sample size problem and insufficiently available samples have severe negative effects on dictionary learning algorithms for face recognition. To address this problem, the auxiliary dictionary learning algorithm has been proposed to improve the classification performance of the case where each class has only one training sample. A typical auxiliary dictionary learning algorithm for face recognition is proposed by Wei [53], which learns a robust auxiliary dictionary from a generic training set. The objective function of the algorithm is

$$\begin{aligned}\min_{D,X}\ &\sum_{i=1}^{N} f\left(y_i^e-\left[\mathbf{G}_e,\mathbf{D}_e\right]\begin{bmatrix}x_g^i\\x_d^i\end{bmatrix}\right)+\alpha\left\|x^i\right\|_1+\\&\beta f\left(y_i^e-\mathbf{G}_e\delta_{it}\left(x_g^i\right)-\mathbf{D}x_d^i\right)\end{aligned} \quad (13)$$

where $\mathbf{E}=\left[\mathbf{Y}_e,\mathbf{G}_e\right]$ are the auxiliary training samples, $\mathbf{Y}_e=\left[y_e^1,\cdots,y_e^N\right]$ is the probe set, and $\mathbf{G}_e$ is the gallery set. $\mathbf{D}_e$ is the auxiliary dictionary. $x^i=\left[x_g^i;x_d^i\right]$ is the sparse coefficient of y_e^i and $\mathbf{X}=\left[x^1,\cdots,x^N\right]$ is the sparse coefficient matrix for $\mathbf{Y}_e$. x_g^i and x_d^i indicate the coefficients associated with gallery G_e and auxiliary dictionary $\mathbf{D}_e$, respectively. Function $\delta_{it}\left(x_d^i\right)$ outputs a vector whose only nonzero entries are the entries in x_g^i that are associated with the it-th class (it denotes the label of y_e^i in the auxiliary training samples). Function $f\left(.\right)$ is defined as

$$\begin{aligned}&f\left(e_i\right)=-\frac{1}{2\mu}\begin{pmatrix}\ln\left(1+\exp\left(-\mu e_i^2+\mu\delta\right)\right)\\-\ln\left(1+\exp\mu\delta\right)\end{pmatrix}\\&f\left(e\right)=\sum_{i=1}^{d}f\left(e_i\right)\end{aligned} \quad (14)$$

where e_i is the i-th entry of $e=y-\left[\mathbf{Y},\mathbf{D}_e\right]x$.

The classification method is as follows.

(1) Normalize test sample y_t and the columns of $\mathbf{Y}$ to have unit l_2-norm.

(2)Initialize the weight matrix by setting $\mathbf{W}=\mathbf{I}$.

(3)Calculate the optimal solution of representation coefficient x and weight matrix W by using

$$x\leftarrow\min_x\left\|\mathbf{W}\left(y-\left[\mathbf{Y},\mathbf{D}_e\right]\begin{bmatrix}x_y\\x_d\end{bmatrix}\right)\right\|_2^2+\lambda\left\|x_1\right\| \quad (15)$$

$$e\leftarrow y-\left[\mathbf{Y},\mathbf{D}_e\right]x \quad (16)$$

$$\mathbf{W}\leftarrow diag\left(w\left(e_1\right),\cdots,w\left(e_d\right)\right)^{1/2} \quad (17)$$

$$w\left(e_k\right)=\frac{\exp\left(-\mu e_k^2+\mu\delta\right)}{1+\exp\left(-\mu e_k^2+\mu\delta\right)} \quad (18)$$

(4)Classify test sample y_t via the weighted reconstruction errors as follows in equation (19).

$$\text{identity}\left(y_t\right)=\min_{l\in\left\{1,2,\cdots,L\right\}}\left\|\mathbf{W}^*\left(y_t-\left[\mathbf{Y},\mathbf{D}_e\right]\begin{bmatrix}\delta\left(x_y^*\right)\\x_d^*\end{bmatrix}\right)\right\| \quad (19)$$

Supposing that the intra-class variations of one subject can be approximated by a sparse linear combination of other subjects, the extended SRC algorithm [54] applied to an auxiliary intra-class variant dictionary to model the possible variation between the training and testing images. The auxiliary intra-class variant dictionary is constructed by using either the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2017.2695239, IEEE Access

gallery faces themselves or generic faces that are outside the gallery. To compensate for the missing illumination information provided by multiple training images, Zhuang [55] used additional illumination examples of face images from one or more additional classes to construct an illumination dictionary. Then, they used the sparse illumination transfer (SIT) technique to transfer the pose and illumination information from the alignment stage to the recognition stage. Moreover, Deng [56] proposed a superposed SRC (SSRC) algorithm, in which the dictionary was assembled by the class centroids and sample-to-centroid differences, which led to a substantial improvement in the SRC algorithm. Gao [57] proposed an intra-class variance dictionary by using the gallery set.

When a suitable dictionary is learned by using the auxiliary dictionary learning algorithm, the classification performance will be improved. However, the method for selecting suitable auxiliary training samples for learning a dictionary is a key point of this algorithm.

VI. Domain Adaptive Dictionary Learning Algorithm

For face images, training samples and testing samples may come from different domains. In this case, if we still use conventional dictionary learning algorithms to learn dictionaries from the training samples, the performance of face recognition will degrade. A domain adaptive dictionary learning algorithm can adequately resolve this problem. It first learns a dictionary that can transfer the features of the source domain to the target domain. It then utilizes a source domain with sufficient labelled data to learn a classifier for a target domain which is usually collected from a different distribution. A typical domain adaptive dictionary learning algorithm is proposed by Zhu [58], which expanded the intra-class diversity of original training samples by virtue of collaboration with the source data. The objective function is defined as

$$\begin{aligned}&\min_{\mathbf{D}_t,\mathbf{D}_s,\mathbf{X}_t,\mathbf{A},\mathbf{W}} \left\|\mathbf{Y}_t-\mathbf{D}_t\mathbf{X}_t\right\|_2^2+\left\|\mathbf{Y}_s\mathbf{A}^T-\mathbf{D}_s\mathbf{X}_t\right\|_2^2\\&+\alpha\left\|\mathbf{Q}-\mathbf{B}\mathbf{X}_t\right\|_2^2+\beta\left\|\mathbf{H}-\mathbf{W}\mathbf{X}_t\right\|_2^2\\&\text{subject to }\forall\text{i},\ \left\|\left(x_t\right)_i\right\|_0\le T_0\end{aligned} \tag{20}$$

where $\mathbf{Y}_t=\left[y_t^1,\cdots,y_t^L\right]\in\Re^{n\times L}$ is the training sample of the target domain, and L and n are the number of training samples and dimensions, respectively. $\mathbf{Y}_s=\left[y_s^1,\cdots,y_s^M\right]$ is the training sample of the source domain, and M is the number of training samples. $\mathbf{D}_t$ is the learned target domain dictionary and $\mathbf{D}_s$ is the learned source domain dictionary. $\mathbf{X}_t$ is the coefficient matrix of the target domain and $\left(x_t\right)_i$ is the i-th column of coefficient matrix $\mathbf{X}_t$. $\mathbf{A}$ is a transformation matrix, and it can transform the source domain data to match the target domain data. Thus, $\mathbf{A}$ can be defined as

$$\mathbf{A}=\begin{pmatrix}\mathbf{A}_1&&&\\&\mathbf{A}_2&&\\&&\ddots&\\&&&\mathbf{A}_C\end{pmatrix},$$
$$A_C\left(i,j\right)=\begin{cases}1 & \text{if } \Lambda_C\left(i,j\right)=Max\left(\Lambda_C\left(:,j\right)\right)\\0 & \text{otherwise}\end{cases} \tag{21}$$

It is assumed that $\mathbf{Y}_t^c$ and $\mathbf{Y}_s^c$ are samples of the C-th category from the target and source domains, respectively. Λ_C is the Gaussian distance between each pair of samples $\mathbf{Y}_t^c$ and $\mathbf{Y}_s^c$. $\mathbf{Q}$ is the input signal of $\mathbf{Y}_t$, and it can be defined as $\mathbf{Q}=\left[q_1,q_2,\cdots,q_L\right]\in\Re^{L\times L}$, $q_i=\left[0,\cdots,1,1,\cdots 0\right]^T\in\Re^{L\times 1}$, and non-zero entries of q_i appear at those indices where y_t^i and x_t^k share the same class label. $\mathbf{H}$ is the label matrix of $\mathbf{Y}_t$. $\mathbf{W}$ is the coefficient of the linear classifier. $\mathbf{B}$ is the linear transformation matrix that maps the original sparse codes to the target discriminative sparse codes. The classification method is the same as the LC-KSVD algorithm.

Qiu [59] presented a function learning framework for the task of transforming a dictionary learned from a visual domain to another domain and maintaining a domain-invariant sparse representation of a signal. Huang [60] proposed a coupled dictionary and feature space learning algorithm for cross-domain image synthesis and recognition, which not only obtained a shared feature space for associating cross-domain image data for recognition purposes but also jointly updated the dictionaries in each image domain for improving the representation ability. Shekhar [61] proposed a generalized domain-adaptive dictionary learning algorithm by optimally representing both source and target domains with a shared dictionary. They jointly learned projections of data in the two domains, and the latent dictionary can succinctly represent both domains in the projected low-dimensional space. Ni [62] proposed interpolating subspaces via dictionary learning to link the source and target domains. These subspaces were able to capture the intrinsic domain shift and formed a shared feature representation for cross domain recognition. To compensate for the transformation of faces due to changes in viewpoint, illumination, and resolution, Qiu [63] proposed compositional dictionaries for domain adaptive face recognition.

When a suitable dictionary is learned by using a domain adaptive dictionary algorithm, the classification performance of face recognition will be improved. However, adequately transferring the characteristics of the source domain to the target domain is very important for this type of algorithm.

VII. Numbers of THE Atoms

To better represent face images, it seems that a learned dictionary should contain as many atoms as possible to cover all the variations of face images for each subject. In general, a larger dictionary may provide a greater variety of illuminations, poses and occlusion of face images. However, a larger dictionary is not always better, as the dictionary might contain

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2017.2695239, IEEE Access

some similar elements or some elements that are seldom used for representation. To achieve excellent face recognition performance, a learned dictionary should have low reconstructive error, as well as compact representation and satisfactory discriminative ability. Specifically, the compact representation expects that the learned dictionary consists of incoherent atoms, and encourages similar signals, which are more likely from the same class, to be consistently described by a similar set of atoms with similar coefficients [65]. Therefore, many methods have been proposed for selecting a suitable number of atoms for different applications.

Mazhar [64] proposed an Enhanced K-SVD algorithm (EK-SVD), which combined the competitive agglomeration algorithm and the matching pursuit algorithm to develop a dictionary with an optimal size for a given dataset, without compromising its approximation accuracy. Qiu [65] used information theory to select atoms from an initial dictionary for image classification. Winn [66] used the pair-wise merging of visual words from an initially large dictionary to obtain an optimally compact visual dictionary. Krause [67] developed an efficient learning framework to construct signal dictionaries for sparse representation by selecting the dictionary columns from multiple candidate bases. Yaghoobi [68] presented an exemplar-based approach for the linear model (called the dictionary). Wang [69] proposed a semi-supervised robust dictionary learning algorithm, which designed a data adaptive dictionary by imposing structured sparsity on the data representation coefficients to automatically select prominent dictionary basis vectors, such that the optimal dictionary size was learned from input data in a principled way and no heuristic pre-specification was required. Lu [70] proposed a scale adaptive dictionary learning framework, which jointly estimated suitable scales and corresponding atoms in an adaptive way, without the need for prior information. They designed an atom counting function and developed a reliable numerical scheme to solve the challenging optimization problem.

Since the number of atoms varies widely, ranging from hundreds to hundreds of thousands, the comprehensive classification performance with different numbers of atoms has not been presented in previous literature. Therefore, in the next section we provide the experimental results of six dictionary learning algorithms with different numbers of atoms on five face databases. The experimental results can provide some in-depth insights to the performance of dictionary learning algorithms for face recognition and are helpful for researchers to use and design dictionary learning algorithms.

VIII. Experimental Results

In this section, we provide the experimental results of the K-SVD [20], D-KSVD [34], LC-KSVD [24], FDDL [43], SVGDL [47] and DLSPC [50] algorithms with different numbers of atoms on the Labeled Faces in the Wild (LFW) [71], the Georgia Tech (GT) [72], the Extended Yale B [73], the AR [74] and the CMU PIE [75] face databases. Moreover, to better show the representation ability of the learned dictionary, we also compare these dictionary learning algorithms with the SRC [1] algorithm.

A. Experiment Setting

In this subsection, we provide the implementation details of seven comparison algorithms. For the SRC algorithm, the implementation procedure is presented in [8] and the representation coefficients of test samples are obtained by using the DALM-fast method. For the D-KSVD, LC-KSVD, FDDL, SVGDL and DLSPC algorithms, we use the source codes provided by the authors. For the K-SVD algorithm, the K-SVD box is used to learn a dictionary, and it uses the same classification method as the D-KSVD and LC-KSVD algorithms. The codes used for the SRC and K-SVD algorithms can be downloaded at:http://www.yongxu.org/default.html. Since the LC-KSVD2 algorithm always achieves higher average recognition rates than the LC-KSVD1 algorithm, we use the LC-KSVD2 algorithm as the LC-KSVD algorithm in this paper. For the DLSPC algorithm, when the global coding classifier is used, we denote it as the DLSPC-G algorithm. When the local coding classifier is used, we denote it as the DLSPC-L algorithm.

B. Experimental Results on the LFW Database

The LFW database contains more than 13,000 images of faces collected from the web, and all of them are labelled with the name of the person pictured. The main goal is to study the problem of unconstrained face recognition. In the database, 1,680 of the people have two or more distinct photos. Following [76], we use a cropped version (LFW crop) of the LFW dataset, which retains only the centre portion of each image (i.e., the face) and almost all of the background is omitted. The LFW crop database was created due to concern about the misuse of the original LFW dataset, where the face matching accuracy can be unrealistically boosted through the use of the background portions of the images (i.e., exploitation of possible correlations between faces and backgrounds). For each LFW image, the area inside a fixed bounding box was extracted. The bounding box was at the same location for all images, with the upper-left and lower-right corners being (83, 92) and (166, 175), respectively. The extracted area was then scaled to a size of 64×64 pixels. The selection of the bounding box location was based on the positions of 40 randomly selected LFW faces. As the location and size of faces in the LFW database were determined through the use of an automatic face locator (detector), the cropped faces in the LFW crop database exhibit real-life conditions, including misalignment, scale variations, and in-plane as well as out-of-plane rotations.

In this experiment, we select and use a subset of the LFW crop database consisting of 1,215 images of 86 people. In this subset each person has approximately 11 to 20 images. Each image is resized to a 32×32 pixel image. Sample images from the LFW crop database are shown in Fig.1.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2017.2695239, IEEE Access

We randomly select ten images of each person as the training samples and reserve the remaining images as the test samples. For the SRC algorithm, the number of training samples of each class is varied from 2 to 10. For the K-SVD, D-KSVD and LC-KSVD algorithms, the number of atoms is varied from 172 to 860 with a step of 86. For the FDDL and SVGDL algorithms, the number of atoms of each class sub-dictionary is varied from 2 to 10. For the DLSPC algorithm, the learned dictionary contains two parts, one is the shared atoms and the other is the specific class atoms. The number of atoms of the shared dictionary is 86, and the number of atoms of each class is varied from 1 to 9. The seven comparison algorithms are each executed ten times and the average recognition rates are reported in Fig.2.

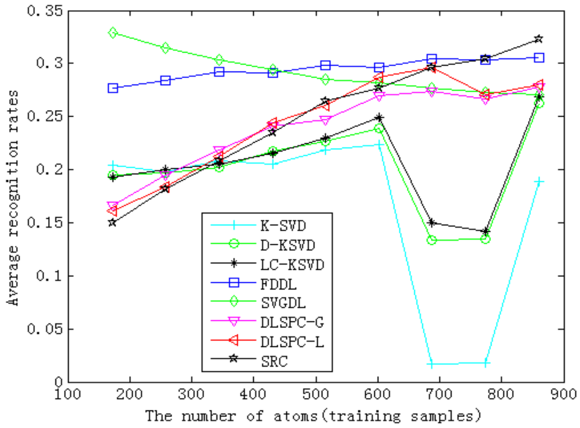

Fig.2. The average recognition rates of seven algorithms with different numbers of atoms

C. *Experimental Results on the GT Face Database*

The Georgia Tech (GT) face database was built at Georgia Institute of Technology, and contained images of 50 people taken in two or three sessions. The pictures show frontal and/or tilted faces with different facial expressions, lighting conditions and scales. Everyone in the database is represented by 15 colour JPEG images with a cluttered background and a resolution of 640×480 pixels. The average size of the faces in these images is 150×150 pixels. Each image was manually labelled to determine the position of the face in the image. We use the face image with the background removed. Each image is 30 by 40 pixels. Sample images from the GT face database are shown in Fig.3.

Fig.3.Example images from the GT face database

We randomly select ten images of each person as training samples and reserve the remaining images as test samples. For the SRC algorithm, the number of training samples of each class is varied from 2 to 10. For the K-SVD, D-KSVD and LC-KSVD algorithms, the number of atoms is varied from 100 to 500 with a step of 50. For the FDDL and SVGDL algorithms, the number of atoms of each class sub-dictionary is varied from 2 to 10. For the DLSPC algorithm, the learned dictionary contains two parts, one is the shared atoms and the other is the specific class atoms. The number of atoms in the shared dictionary is 50, and the number of atoms in each class is varied from 1 to 9. The seven comparison algorithms are each executed ten times and the average recognition rates are reported in Fig.4.

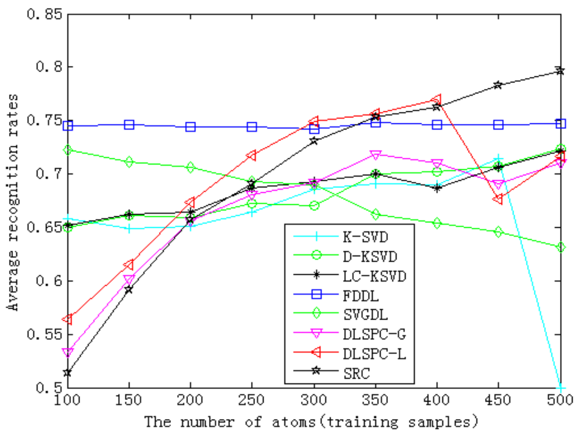

Fig.4. The average recognition rates of seven algorithms with different numbers of atoms

D. *Experimental Results on the Extended Yale B Face Database*

The Extended Yale B face database consists of 2,414 front-face images of 38 people which were taken under various illumination conditions and expressions. There are approximately 59 to 64 images for each person and each image was normalized to the size of 32×32 pixels. Examples of images from the Extended Yale B face database are shown in Fig.5.

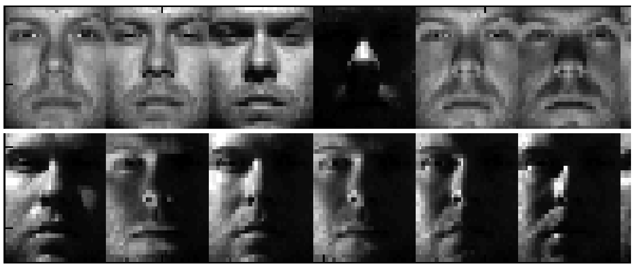

Fig.5. Example images from the Extended Yale B face database

We randomly select 32 images of each person as training samples and reserve the remaining images for testing. For the SRC algorithm, the number of training samples of each class is varied from 2 to 32. For the K-SVD, D-KSVD and LC-KSVD algorithms, the number of atoms is varied from 76 to 1216 with a step of 38. For the FDDL and SVGDL algorithms, the number of atoms is varied from to 2 to 32 in each class sub-dictionary. For the DLSPC algorithm, the learned dictionary contains two

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2017.2695239, IEEE Access

parts, one is the shared atoms, and the other is the specific class atoms. The number of atoms in the shared dictionary is 38, and the number of atoms of each class is varied from 1 to 31. The seven comparison algorithms are each executed ten times and the average recognition rates are reported in Fig.6.

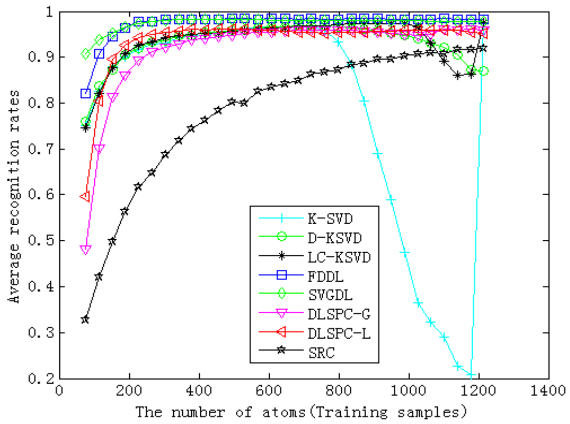

Fig.6. The average recognition rates of seven algorithms with different numbers of atoms

E. Experimental Results on the AR Face Database

The AR face database contains over 4,000 images of 126 people. There are 26 face images of each person taken during two sessions, and each image is taken under various lighting conditions. For each person, there are 12 images including those with the sunglasses and scarves. Following [77], a subset of the AR face database consisting of 3,120 images from 120 people is used in this experiment. The resolution of the AR images was 40×50 pixels. Images of one person from the AR face database are shown in Fig.7.

Fig.7. Example images from the AR face database

We randomly select ten images of each class as training samples and reserve the remaining images for testing. For the SRC algorithm, the number of training samples of each class is varied from 2 to 10. For the K-SVD, D-KSVD and LC-KSVD algorithms, the number of atomsis varied from 240 to 1200 with a step of 120. For the DLSPC algorithm, the learned dictionary contains two parts, one is the shared atoms and the other is the specific class atoms. The number of atoms in the shared dictionary is 120, and the number of atoms of each class is varied from 1 to 9. The seven comparison algorithms are each executed ten times and the average recognition rates are reported in Fig.8.

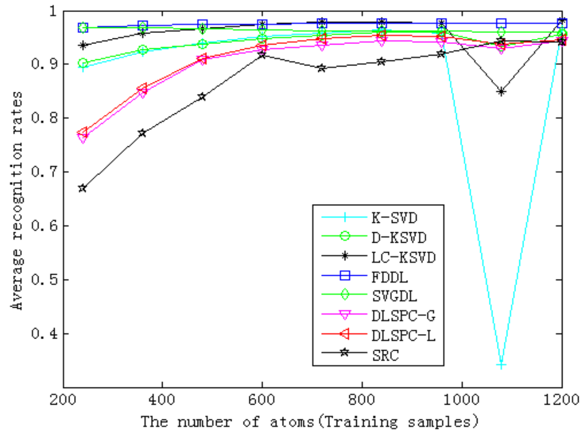

Fig.8. The average recognition rates of seven algorithms with different numbers of atoms

F. Experimental Results on the PIE Face Database

The PIE face database consists of 41,368 front-face images of 68 people, and the face images of each person are captured under 13 different poses, 43 different illumination conditions, and with 4 different facial expressions. Some sample images from the PIE face database are shown in Fig.9.

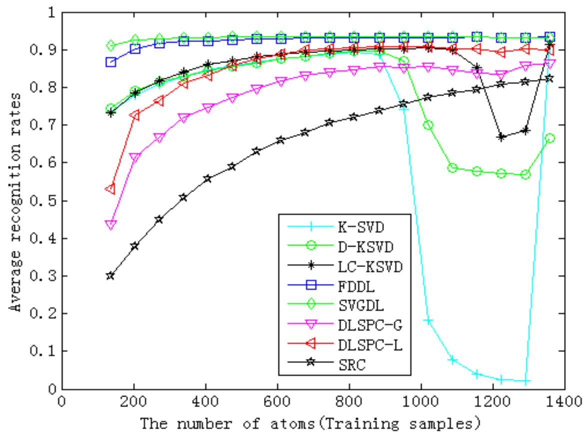

Fig.10. The average recognition rates of seven algorithms with different numbers of atoms

Following [78], we choose the five near frontal poses (C05, C07, C09, C27, and C29) of each subject and use all images under different illuminations and expressions. Thus, we obtain

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2017.2695239, IEEE Access

170 images for each individual. Every image is normalized to the size of 32×32 pixels. We randomly select twenty images of each person as training samples and reserve the remaining images as test samples. For the SRC algorithm, the number of training samples of each class is varied from 2 to 20. For the K-SVD, D-KSVD and LC-KSVD algorithms, the number of atoms is varied from 136 to 1360 with a step of 68. For the DLSPC algorithm, the learned dictionary contains two parts, one is the shared atoms and the other is the specific class atoms. The number of atoms in the shared dictionary is 68, and the number of atoms of each class is varied from 1 to 19. The seven comparison algorithms are each executed ten times and the average recognition rates are reported in Fig.10.

G. Analysis of Experimental Results

In the above sections, the experimental results on the five face databases were elaborated. We summarize the experimental results as follows.

1) Fig.2, Fig.4, Fig.6, Fig.8 and Fig.10 show that the average recognition rates of the FDDL and SVGDL algorithms are insensitive to the number of atoms. For the Extended Yale B, AR and PIE face databases, the average recognition rates of the DLSPC-G and DLSPC-L algorithms increase with the number of atoms. However, the average recognition rates of the DLSPC-G and DLSPC-L algorithms do not always increase with the number of atoms on the LFW and GT face databases. The average recognition rates of the K-SVD, D-KSVD and LC-KSVD algorithms on the five face databases do not always increase with the number of atoms. Since the K-SVD, D-KSVD and LC-KSVD algorithms are all shared dictionary learning algorithms, the FDDL and SVGDL algorithms are both specific class dictionary learning algorithms, and the DLSPC algorithm is the commonality and particularity dictionary learning algorithm, the experimental results demonstrate that the specific class dictionary learning algorithm is less sensitive to the variation in the number of atoms than the shared dictionary learning algorithm and the commonality and particularity dictionary learning algorithm.
2) When the number of atoms is equal to the number of training samples, the average recognition rates of the D-KSVD, LC-KSVD, FDDL, SVGDL, DLSPC-G and DLSPC-L algorithms on the five face databases are higher than the average recognition of the SRC algorithm in most cases. This is mainly because the pose, illumination and expression information in face images can be implicitly encoded into the learned dictionaries, such that the learned dictionaries can have more powerful representation ability than the original training samples.
3) When the number of training samples increases, the FDDL and SVGDL algorithms achieve higher average recognition rates than the D-KSVD, LC-KSVD, DLSPC-G and DLSPC-L algorithms in most cases. This demonstrates that the specific class dictionary learning algorithm can preserve main characteristics of face images better than the shared dictionary learning algorithm and the commonality and particularity dictionary learning algorithm.
4) For the SRC algorithm, when the number of the training samples increases, the average recognition rate also increases in most cases. This demonstrates that the number of training samples plays an important role in face recognition.

IX. Conclusion

In this survey, we provide a current review of the existing dictionary learning algorithms for face recognition, which covers five types of major and very different algorithms, i.e., the shared dictionary learning algorithm, the class-specific dictionary learning algorithm, the commonality and particularity dictionary learning algorithm, the auxiliary dictionary learning algorithm and the domain adaptive dictionary learning algorithm. Additionally, we offer experimental results of different dictionary learning and sparse coding algorithms with different numbers of atoms in face databases. Experimental results show that the specific class dictionary learning algorithms are less sensitive to the variety of the number of atoms than the shared dictionary learning algorithms and the commonality and particularity dictionary learning algorithms. This review offers important ideas and cues for designing dictionary learning algorithms for face recognition.

References

[1] 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.

[2] M. Yang, L. Zhang, "Gabor feature based sparse representation for face recognition with gabor occlusion dictionary," in *Proc. Eur. Conf. Comput. Vis.*, Sep. 2010, pp. 448-461.

[3] K.-K. Huang, D.-Q. Dai, C.-X. Ren, and Z.-R. Lai, "Discriminative kernel collaborative representation with locality constrained dictionary for face recognition," *IEEE Trans. Neural Netw. Learn. Syst.*, 2016, Doi: 10.1109/TNNLS.2016.2522431.

[4] X.-Y. Jing, F. Wu, X. Zhu, X. Dong, F. Ma, and Z. Li, "Multi-spectral low-rank structured dictionary learning for face recognition," *Pattern Recognit.*, vol. 59, pp. 14-25, Nov. 2016.

[5] R. Giryes, M. Elad, "Sparsity-based poisson denoising with dictionary learning," *IEEE Trans. Image Process.*, vol. 23, no.12, pp. 5057-5069, Dec. 2014 .

[6] Y. Fu, A. Lam, I. Sato, and Y. Sato, "Adaptive spatial-spectral dictionary learning for hyperspectral image denoising," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 343-351.

[7] Y. C. Chen, C. S. Sastry, V. M. Patel, P. J. Phillips, and R. Chellappa, "In-plane rotation and scale invariant clustering using dictionaries," *IEEE Trans. Image Process.*, vol. 22, no. 6, pp. 2166-2180, Jun. 2013.

[8] L. Jing, M. K. Ng, T. Zeng, "Dictionary learning-based subspace structure identification in spectral clustering," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 24, no. 8, pp. 1188-1199, Aug. 2013.

[9] J. Yang, Z. Wang, Z. Lin, S. Cohen, and T. Huang, "Coupled dictionary training for image super-resolution," *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3467-3478, Aug. 2012.

[10] J. Yang, J. Wright, T. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE Trans. Image Process.*, vol. 19, no. 11, pp. 2861-873, Nov. 2010.

[11] S. Xiang, G. Meng, Y. Wang, C. Pan, and C. Zhang, "Image deblurring with coupled dictionary learning," *Int. J. Comput. Vis.*, vol. 114, no. 2, pp. 248-271, Sep. 2015.

[12] L. Ma, L. Moisan, J. Yu, and T. Zeng, "A dictionary learning approach for poisson image deblurring,"*IEEE Trans. Med. Imaging.*, vol. 32, no. 7, pp. 1277-1289, Jul. 2013.

[13] K. Cao, E. Liu, and A. K. Jain, "Segmentation and enhancement of latent fingerprints: a coarse to fine ridge structure dictionary," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 9, pp. 1847-1859, Sep. 2014 .

[14] M. Elad, "Sparse and redundant representation modeling—What next?," *IEEE Signal Process. Lett.*, vol.19, no. 12, pp. 922-928, Dec. 2012.

[15] R. Rubinstein, A. M. Bruckstein, and M. Elad, "Dictionaries for sparse representation modeling," *Proc. IEEE*, vol. 98, no. 6, pp. 1045-1057, Jun. 2010.

[16] I. Tosic, P. Frossard, “Dictionary learning,” *IEEE Signal Process. Mag.*, vol. 28, no. 2, pp. 27-38, Mar. 2011.

[17] H. Cheng, Z. Liu, L. Yang, and X. Chen, “Sparse representation and learning in visual recognition: theory and applications,” *Signal Process.*, vol. 93, no. 6, pp. 1408-1425, Jun. 2013.

[18] M. J. Gangeha, A. K. Farahat, A. Ghodsid, and M. S. Kamel, “Supervised dictionary learning and sparse representation-a review”, arXiv:1502.05928, Feb. 2015.

[19] Z. Zhang, Y. Xu, J. Yang, X. Li, and D. Zhang, “A survey of sparse representation: algorithms and applications,” *IEEE Access*, vol.3, pp. 490-530, May. 2015.

[20] M. Aharon, M. Elad, and A. M. Bruckstein, “K-SVD: an algorithm for designing of over-complete dictionaries for sparse representation,” *IEEE Trans. Signal Proces.*, vol.54, no. 11, pp. 4311-4322, Nov. 2016.

[21] K. Schnass, “On the identifiability of overcomplete dictionaries via the minimisation principle underlying K-SVD,” *Appl. Comput. Harmon. A.*, vol. 37, no. 3, pp. 464-491, Nov. 2014.

[22] R. Rubinstein, T. Peleg, and M. Elad, “Analysis K-SVD: a dictionary-learning algorithm for the analysis sparse model,” *IEEE Trans. Signal Proces.*, vol. 61, no. 3, pp. 661-677, Feb. 2013.

[23] E. M. Eksioglu, O. Bayir, “K-SVD meets transform learning: transform K-SVD,” *IEEE Signal Process. Lett.*, vol. 21, no.3, pp. 347-351, Mar. 2014.

[24] Z. Jiang, Z. Lin, and L. S. Davis,” Learning a discriminative dictionary for sparse coding via label consistent K-SVD,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2011, pp. 1697-1704.

[25] J. A. Tropp, A. C. Gilbert, “Signal recovery from random measurements via orthogonal matching pursuit,” *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655-4666, Dec. 2007.

[26] K. Yu, T. Zhang, and Y. Gong, “Nonlinear learning using local coordinate coding,” in *Proc. Adv. Neural Inf. Process. Syst.,* Dec. 2009, pp. 2223-2231.

[27] M. Zheng, J. Bu, C. Chen, C. Wang, L. Zhang, G. Qiu, and D. Cai, “Graph regularized sparse coding for image representation,” *IEEE Trans. Image Process.*, vol. 20, no.5, pp. 1327-1336, May. 2011.

[28] S. Gao, I. Tsang, and L. Chia, “Laplacian, Sparse coding, hypergraph laplacian sparse coding, and applications,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 92-104, Jan. 2013.

[29] Z. Jiang, G. Zhang, and L. S. Davis, “Submodular dictionary learning for sparse coding,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 3418-3425.

[30] S. Haghiri, H. R. Rabiee, A. Soltani-Farani, S. A. Hosseini, and M. Shadloo, “Locality preserving discriminative dictionary learning,” in *Proc. IEEE Int. Conf. Image Process.*, Oct. 2014, pp. 5242-5246.

[31]W. Liu, Z. Yu, L. Lu, Y. Wen, H. Li, and Y. Zou, “KCRC-LCD: Discriminative kernel collaborative representation with locality constrained dictionary for visual categorization,” *Pattern Recognit.*, vol. 48, no. 10, pp. 3076-3092, Oct. 2015.

[32] Y.-B. Yang, Q.-H. Zhu, X.-J. Mao, and L.-Y. Pan, “Visual feature coding for image classification integrating dictionary structure,” *Pattern Recognit.*, vol. 48, no. 10, pp. 3067-3075, Oct. 2015.

[33] Z. Li, Z. Lai, Y. Xu, J. Yang, and D. Zhang, “A locality-constrained and label embedding dictionary learning algorithm for image classification,” *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 2, pp. 278-293, Feb. 2017.

[34] Q. Zhang, B. Li, “Discriminative K-SVD for dictionary learning in face recognition,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 2691-2698.

[35] A. Shrivastava, V. M. Patel, and R. Chellappa, “Non-linear dictionary learning with partially labeled data,” *Pattern Recognit.*, vol. 48, no. 11, pp. 3283-3292, Nov. 2015.

[36] D. Pham, S.Venkatesh, “Joint learning and dictionary construction for pattern recognition,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2008, pp. 1-8.

[37] T. Lin, S. Liu, and H. Zha, “Incoherent dictionary learning for sparse representation,” in *Proc. IEEE Int. Conf. Image Process.*, Oct. 2012, pp. 1237-1240.

[38] H. Guo, Z. Jiang, and S. D. Larry, “Discriminative dictionary learning with pair-wise constraints,” in *Proc. Asian Conf. Comput. Vis.*, Nov. 2012, pp. 328-342.

[39] M. J. Gangeh, A. Ghodsi, and M. S. Kamel, “Kernelized supervised dictionary learning,” *IEEE Trans. Signal Process.*, vol. 61, no. 19, pp. 4753-4767, Oct. 2013.

[40] M. Harandi, C. Sanderson, C. Shen, and B. Lovell, “Dictionary learning and sparse coding on grassmann manifolds: an extrinsic solution,” in *Proc. IEEE Int. Conf. Comput. Vis.*, pp. 3120-3127, Dec. 2013.

[41] Q. Pan, D. Kong, C. Ding, and B. Luo, “Robust non-negative dictionary learning,” in *Proc the Association for the Advancement of Artificial Intelligence*, Jul. 2014, pp. 2027-2033.

[42] N. Akhtar, F. Shafait, and A. Mian, “Discriminative Bayesian Dictionary Learning for Classification,” *IEEE Trans. Pattern Anal. Mach. Intell.*,vol. 38, no. 12, pp. 2374-2388, Dec. 2016.

[43] 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.

[44] A. Shrivastava, H.V. Nguyen, V. M. Patel, and R. Chellappa, “Design of non-linear discriminative dictionaries for image classification,” in *Proc. Asian Conf. Comput. Vis.*, pp. 660-674, Nov. 2012.

[45] Y. Chen, V. M. Patel, P. Jonathon Phillips, and R. Chellappa, “Dictionary-based face recognition from video,” in *Proc. Eur. Conf. Comput. Vis.*, Oct. 2012, pp. 766-779.

[46] L. Ma, C. Wang, B. Xiao, and W. Zhou, “Sparse representation for face recognition based on discriminative low-rank dictionary learning,” *in Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 2586-2593, Jun. 2012 .

[47] S. Cai, W. Zuo, L. Zhang, X. Feng, and P. Wang, “Support vector guided dictionary learning,” in *Proc. Eur. Conf. Comput. Vis.*, Sep. 2014, pp. 624-639.

[48] H. Zheng, D. Tao, “Discriminative dictionary learning via Fisher discrimination K-SVD algorithm,” *Neurocomputing*, vol. 62, pp. 9-15, Aug. 2015.

[49] H. Wang, C. Yuan, W. Hu, and C. Sun, “Supervised class-specific dictionary learning for sparse modeling in action recognition,” *Pattern Recognit.*, vol. 45, no. 11, pp. 3902-3911, Nov. 2012.

[50] D. Wang, S. Kong, “A classification oriented dictionary learning model: Explicitly learning the particularity and commonality across categories,” *Pattern Recognit.*, vol. 47, no. 2, pp. 885-898, Feb. 2014.

[51] Y. Sun, Q. Liu, J. Tang, and D. Tao, “Learning discriminative dictionary for group sparse representation,” *IEEE Trans. Image Process.*, vol. 23, no. 9, pp. 3816-3828, Sep. 2014.

[52] M. Yang, W. Liu, W. Luo, and L. Shen, “Analysis-synthesis dictionary learning for universality-particularity representation based classification,” in *Proc the Association for the Advancement of Artificial Intelligence*, Feb. 2016, pp. 2251-2257.

[53] C.-P. Wei, Y.-C. Frank Wang, “Undersampled face recognition via robust auxiliary dictionary learning,” *IEEE Trans. Image Process.*, vol. 24, no. 6, pp. 1722-1734, Jun. 2015.

[54] W. Deng, J. Hu, and J. Guo, “Extended SRC: under sampled face recognition via intraclass variant dictionary,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34. No. 9, pp. 1864-1870, Sep. 2012.

[55] L. Zhuang, A.Y. Yang, Z. Zhou, S. S. Sastry, and Y. Ma, “Single-sample face recognition with image corruption and misalignment via sparse illumination transfer,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 3546-3553.

[56] W. Deng, J. Hu, and J. Guo, “In defense of sparsity based face recognition”, in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 399-406, Jun. 2013.

[57] S. Gao, K. Jia, L. Zhuang, and Y. Ma, “Neither global nor local: regularized patch-based representation for single sample per person face recognition,” *Int. J. Comput.Vis.*,vol. 111. No. 3, pp. 365-383, Feb. 2015.

[58] F. Zhu, L. Shao, “Weakly-supervised cross-domain dictionary learning for visual recognition,” *Int. J. Comput.Vis.*, vol. 109, no. 1, pp. 42-59, Aug. 2014 .

[59] Q. Qiu, M. P. Vishal, T. Pavan, and C. Rama, “Domain adaptive dictionary learning,” in *Proc. Eur. Conf. Comput. Vis.*, Oct. 2012, pp. 631-645.

[60] D. Huang, Y. Wang, “Coupled dictionary and feature space learning with applications to cross-domain image synthesis and recognition,” in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 2496-2503 .

[61] S. Shekhar, V. M. Patel, H. V. Nguyen, and R. Chellappa, “Generalized domain-adaptive dictionaries,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 361-368.

[62] J. Ni, Q. Qiu, and R. Chellappa, “Subspace interpolation via dictionary learning for unsupervised domain adaptation,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 692-699 .

[63] Q. Qiu, R. Chellappa, “Compositional dictionaries for domain adaptive face recognition,” *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 5152-5165, Dec. 2015.

[64] R. Mazhar, P. D. Gader, “EK-SVD: optimized dictionary design for sparse representations,” in *Proc. Int. Conf. Pattern Recognit.*,Dec. 2008, pp. 1-4.

[65] Q. Qiu, V. M. Patel, and R. Chellappa, “Information-theoretic dictionary learning for image classification,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 11, pp. 2173-2184, Nov. 2014.

[66] J. Winn, A. Criminisi, T. Minka, “Object categorization by learned universal visual dictionary,” in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2005, pp. 1800-1807.

[67] A. Krause, V. Cevher, "Submodular dictionary selection for sparse representation," In *Proc.Int. Conf. Mach. Learn.*, pp. 567-574, Aug. 2010.

[68] M. Yaghoobi, L. Daudet, and M. E. Davies, "Dictionary subselection using an over-complete joint sparsity model," *IEEE Trans. Signal Process.*, vol. 62, no. 17, pp. 4547-4556, Sep. 2014.

[69] H. Wang, F. Nie, W. Cai, and H. Huang, "Semi-supervised robust dictionary learning via efficient Ll-norms minimization," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 1145-1152.

[70] C. Lu, J. Shi, and J. Jia, "Scale adaptive dictionary learning," *IEEE Trans. Image Process.*, vol. 23, no. 2, pp. 837-847, Feb. 2014.

[71] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: a database for studying face recognition in unconstrained environments," University of Massachusetts, Amherst, Technical Report 07-49, Oct. 2007.

[72] N. Goel, G. Bebis, and A. Nefian, "Face recognition experiments with random projection," in *Proc. SPIE 5779, Biometric Technology for Human Identification II,* Mar. 2005, pp. 426-437.

[73] 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.

[74] A. M. Martinez, R. Benavente, The AR face database, CVC Technical Report #24, Jun. 1998.

[75] T. Sim, S. Baker, and M. Bsat, "The CMU pose, illumination, and expression database," *IEEE Trans. Pattern Anal. Mach. Intell.,* vol. 25, no. 12, pp. 1615-1618, Dec. 2003.

[76] C. Sanderson, B.C. Lovell, "Multi-region probabilistic histograms for robust and scalable identity inference," in *Proc. Third International Conference on Advances in Biometrics,* Jun. 2009, pp. 199-208.

[77] J. Yang, D. Zhang, J.-Y. Yang, and B. Niu, "Globally maximizing, locally minimizing: unsupervised discriminant projection with applications to face and palm biometrics," *IEEE Trans. Pattern Anal. Mach. Intell.,* vol. 29, no. 4, pp. 650-664, Apr. 2007.

[78] D. Cai, X. He, J. Han, and H. Zhang, "Orthogonal Laplacianfaces for face recognition," *IEEE Trans. Image Process.*, vol. 15, no. 11, pp. 3608-3614, Nov. 2006.