LEARNING AUXILIARY DICTIONARIES FOR UNDERSAMPLED FACE RECOGNITION

Chia-Po Wei and *Yu-Chiang Frank Wang*

Research Center for Information Technology Innovation, Academia Sinica, Taipei, Taiwan
{cpwei, ycwang}@citi.sinica.edu.tw

ABSTRACT

In this paper, we address the problem of robust face recognition using *undersampled* data. Given only *one* or *few* face images per class, our proposed method not only handles test images with large intra-class variations such as illumination and expression, it is also able to recognize the corrupted ones due to occlusion or disguise. In our work, we advocate the learning of auxiliary dictionaries from the subjects *not* of interest. With the proposed optimization algorithm which jointly solves the tasks of auxiliary dictionary learning and sparse-representation based face recognition, our approach is able to model the above intra-class variations and corruptions for improved recognition. Our experiments on two face image datasets confirm the effectiveness and robustness of our approach, which is shown to outperform state-of-the-art sparse representation based methods.

Index Terms— Dictionary learning, sparse representation, face recognition

1. INTRODUCTION

For robust face recognition, one not only needs to recognize face images with possible illumination and expression variations, these images might also suffer from corruptions due to occlusion or disguise. Although a typical solution is to collect training data in advance for covering the above intra-class variations, it is generally not clear whether the collected training data will be sufficient for providing generalization for the designed classifiers. Moreover, in practical scenarios, only one or very few face images of the subject of interest might be captured during the data acquisition (i.e., training) stage. As a result, one would encounter the challenging task of *undersampled* face recognition.

Proposed by Wright *et al.* [1], sparse representation based classification (SRC) has emerged as a state-of-the-art face recognition approach due to its promising performance. SRC advances sparse coding and treats the training face images of different subjects as an *over-complete* dictionary. As reviewed in Section 2, it assumes that test images of these subjects will lie in the subspace spanned by the corresponding dictionary atoms (i.e., training data), so that recognition can be performed accordingly. To improve the performance, Wagner *et al.* [2] extended SRC and designed a projection-based

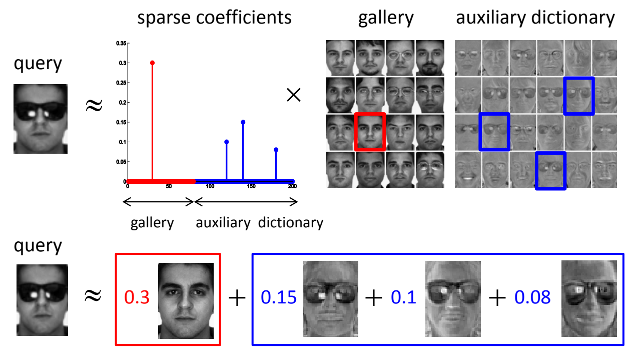


Fig. 1. Illustration of our proposed method, which utilizes external face images for modeling intra-class variations or occlusion. Note that the gallery set contains only one or few face images of the subjects of interest, while the auxiliary dictionary is to be learned from external data for observing possible image variants.

framework for covering possible face illumination variations. Nevertheless, SRC-based approaches require the collection of a *large* amount of training image data as the over-complete dictionary for recognition.

In order to address undersampled face recognition problems, Zhu *et al.* [3] proposed a multi-scale patch based collaborative representation (PCRC) approach. PCRC performs recognition based on patch-wise reconstruction error, but it is not designed to handle face images with occlusion. To introduce additional capabilities in modeling intra-class variations, recent works like AGL [4] and ESRC [5] utilized images collected from an external dataset which contains subjects *not* of interest. In particular, ESRC was proposed to handle image variants including the corrupted ones (e.g., occluded faces). However, images collected from external data might contain noisy, redundant, or undesirable information which would degrade the capability in covering intra-class variations. Without proper selection and integration of face images from both the gallery and external data, it is not guaranteed that the use of external data is preferable.

We note that dictionary learning has been shown as an effective technique for both data representation and classification [6]. Many efforts have been devoted to the learning of a proper dictionary for particular applications like image denoising [7, 8], image inpainting [9, 8], and image classification [10, 11, 12, 13, 14]. However, most existing dictionary learning approaches for face recognition like [10, 11, 12, 13, 14] require the training images to be clean (i.e., without occlusions). As pointed out in [15], the presence of corrupted face images in the training set would significantly degrade the recognition performance.

In this paper, we propose a SRC-based dictionary learning algorithm for robust face recognition using undersampled data. With the same setting as [5], we consider only *one* or *few* non-occluded training images are available for each subject of interest. Unlike [5] which directly utilizes face images with possible variants from an external dataset (e.g., subjects *not* of interest), we advocate the extraction of representative information from external data via dictionary learning, and this process is considered as *auxiliary dictionary learning*. It is worth noting that, while one can directly apply existing dictionary learning algorithms like K-SVD [16] to derive dictionaries for images from external datasets, the derived dictionary only exhibits representation ability and does not guarantee recognition performance for the subjects of interest. In our work, we jointly solve the tasks of auxiliary dictionary learning and SRC-based face recognition in a *unified* optimization framework (detailed in Section 3). This is the reason why our approach is able to improve the performance for robust face recognition under the scenario of undersampled training data.

Fig. 1 illustrates the idea of our proposed method. By learning an auxiliary dictionary from an external dataset together with SRC, the benefits of our approach are twofold. Firstly, we are able to address undersampled face recognition problem since only one or few training images of the subjects to be recognized are required for training. Thus, there is no need to collect a large training dataset for covering image variants for *all* subjects of interest (as SRC did). Secondly, the auxiliary dictionary learned from external data allows us to model intra-class variations including corruptions like occlusion. If a particular type of intra-class variations is of interest, our approach allows one to simply collect such images from subjects *not* of interest. By solving both auxiliary dictionary learning and face recognition in a unified framework, improved recognition performance can be expected.

2. RELATED WORK

2.1. SRC and Extended SRC

Sparse representation based classification (SRC) was recently proposed by Wright *et al.* [1] for face recognition. It considers each test image $\mathbf{y}$ as a sparse linear combination of a codebook $\mathbf{D} = [\mathbf{D}_1, \mathbf{D}_2, \cdots, \mathbf{D}_k]$, where $\mathbf{D}_i$ denotes the training images associated with class i. SRC calculates the sparse coefficient $\mathbf{x}$ of $\mathbf{y}$ by solving the following ℓ_1-minimization problem:

$$\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{D}\mathbf{x}\|_2^2 + \lambda \|\mathbf{x}\|_1. \tag{1}$$

Once (1) is solved, the test input $\mathbf{y}$ will be recognized as class j if it satisfies

$$j = \arg\min_i \|\mathbf{y} - \mathbf{D}\delta_i(\mathbf{x})\|_2, \tag{2}$$

where $\delta_i(\mathbf{x})$ is a vector whose only nonzero entries are the entries in $\mathbf{x}$ that are associated with class i. In other words, the test image $\mathbf{y}$ will be assigned to the class based on the class-wise minimum reconstruction error. The motivation behind SRC is that the test image $\mathbf{y}$ lies in the column subspace spanned by $\mathbf{D}_j$ if it belongs to the corresponding class. As a result, most non-zero elements of $\mathbf{x}$ will be mainly presented in the non-zero elements of $\delta_j(\mathbf{x})$ and thus result in the minimum reconstruction error.

A major limitation of SRC is its requirement of collecting a large amount of training data as the over-complete dictionary $\mathbf{D}$. To address this concern, Deng *et al.* [5] proposed Extended SRC (ESRC) by solving:

$$\min_{\mathbf{x}} \left\| \mathbf{y} - [\mathbf{D},\ \mathbf{A}] \begin{bmatrix} \mathbf{x^d} \\ \mathbf{x^a} \end{bmatrix} \right\|_2^2 + \lambda \|\mathbf{x}\|_1, \tag{3}$$

where $\mathbf{x} = [\mathbf{x^d};\ \mathbf{x^a}]$. It can be seen that ESRC introduces an *intra-class variant dictionary* $\mathbf{A}$ for SRC. Different from $\mathbf{D}$ in SRC, this intra-class variant dictionary $\mathbf{A}$ consists of image data collected from an external dataset (e.g., subjects not of interest), and the use of this dictionary is to model all possible variations of interests including occlusion. Similar to SRC, the classification criterion of ESRC needs to satisfy:

$$j = \arg\min_i \left\| \mathbf{y} - [\mathbf{D},\ \mathbf{A}] \begin{bmatrix} \delta_i(\mathbf{x^d}) \\ \mathbf{x^a} \end{bmatrix} \right\|_2. \tag{4}$$

Although observing intra-class variations using external data has shown to produce promising results by ESRC, the direct use of such data might degrade the performance due to the noise or redundant images presented. Another concern of ESRC is that, since the goal of collecting $\mathbf{A}$ is to cover all variations of concern, the size of $\mathbf{A}$ can become very large which makes the computation of (3) extremely expensive.

2.2. Dictionary Learning for Sparse Coding

Recent research progresses indicate that the learning of data or application-driven dictionaries might be preferable than using predefined ones [6]. In general, dictionary learning algorithms can be designed in an *unsupervised* or *supervised* way. Unsupervised dictionary learning aims at data representation [17, 16], and thus the corresponding reconstruction process can be applied to image synthesis tasks like image denoising.

For image classification, supervised dictionary learning aims at introducing data discrimination using some classification criteria. For example, Mairal *et al.* added an *softmax* cost function term [18] into the sparse representation formulation for recognition. Yang *et al.* imposed the Fisher discrimination criterion [14] on the objective function so that the learned dictionaries would favor data classification. Some recent works [19, 20, 11, 12] further integrated classifier design into the sparse representation framework, so that both classifier and dictionary will be jointly learned for improved recognition.

Unfortunately, the above dictionary learning approaches will not generalize well if the amount of training data is limited. In this paper, we propose to *learn* an auxiliary dictionary for observing intra-class variations using external data, while only one or few training images per class are required for the subjects of interest. Later in the experiments, we will confirm that our approach not only outperforms state-of-the-art SRC based methods, it would require fewer external data than the ESRC does while achieving better recognition performance.

3. AUXILIARY DICTIONARY LEARNING

3.1. Problem Formulation

We propose a *supervised* learning algorithm for solving undersampled face recognition problems. Inspired by [4, 5], we utilize images collected from external data, and we advocate the learning of an *auxiliary* dictionary from such data for modeling intra-class variations including possible occlusions. By jointly solving this learning algorithm with SRC in a unified framework, the recognition performance can be significantly improved. Table 1 highlights and compares the properties of recent SRC-based face recognition methods, which shows the advantages of our proposed method.

We now detail our proposed algorithm. To learn an auxiliary dictionary for modeling intra-class image variants, we collect images of p subjects from an external dataset. We separate these external images into a probe set $\mathbf{Y}^e$ and a gallery set $\mathbf{D}^e$ (note that the superscript e indicates external data). The probe set $\mathbf{Y}^e = [\mathbf{y}_1^e, \mathbf{y}_2^e, \ldots, \mathbf{y}_N^e] \in \mathbb{R}^{d\times N}$ consists of N d-dimensional images with different intra-class variations of interest (and to be modeled). The gallery $\mathbf{D}^e$ contains only one or few face images per subject for the setting of undersampled recognition. Suppose $\mathbf{D}^e \in \mathbb{R}^{d\times p}$ (i.e., one image per subject), our proposed algorithm solves the following problem during training:

$$\min_{\mathbf{A},\mathbf{X}} \sum_{i=1}^{N} \Big(\|\mathbf{y}_i^e - [\mathbf{D}^e, \mathbf{A}]\mathbf{x}_i\|_2^2 + \lambda \|\mathbf{x}_i\|_1 + \eta \left\| \mathbf{y}_i^e - \mathbf{D}^e \delta_{i_k}(\mathbf{x}_i^{\mathbf{d}}) - \mathbf{A}\mathbf{x}_i^{\mathbf{a}} \right\|_2^2 \Big), \tag{5}$$

where $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_N]$, and the auxiliary dictionary $\mathbf{A}$ is of size $d \times m$ (m specifies the number of dictionary atoms to be observed). The vector $\mathbf{x}_i = [\mathbf{x}_i^{\mathbf{d}};\ \mathbf{x}_i^{\mathbf{a}}]$ is the sparse coefficient for $\mathbf{y}_i^e$, in which $\mathbf{x}_i^{\mathbf{d}} \in \mathbb{R}^{p\times 1}$ indicates the coefficient associated with the gallery $\mathbf{D}^e$, and $\mathbf{x}_i^{\mathbf{a}} \in \mathbb{R}^{m\times 1}$ is the one with the learned auxiliary dictionary $\mathbf{A}$. The function $\delta_{i_k}(\mathbf{x}_i^{\mathbf{d}})$ produces the vector $\mathbf{x}_i^{\mathbf{d}}$ with only nonzero entries associated with class i_k (i_k denotes the label of $\mathbf{y}_i^e$ in the external data set). Parameters λ and η control the sparsity and the class-wise reconstruction error, respectively.

In (5), the first term indicates data representation, the second term introduces the sparsity constraint, while the last term $\left\| \mathbf{y}_i^e - \mathbf{D}^e \delta_{i_k}(\mathbf{x}_i^{\mathbf{d}}) - \mathbf{A}\mathbf{x}_i^{\mathbf{a}} \right\|_2^2$ is the class-wise reconstruction error. In this last term, $\mathbf{D}^e \delta_{i_k}(\mathbf{x}_i^{\mathbf{d}})$ is associated with the SRC classification rule as shown in (2), and $\mathbf{A}\mathbf{x}_i^{\mathbf{a}}$ takes the observed auxiliary dictionary into consideration for handling possible intra-class variations or corruptions. This explains how we effectively integrate both auxiliary dictionary learning and the SRC objective function into a unified framework.

Table 1. Properties of different sparse representation based face recognition approaches. Note that both ESRC and our approach utilize external face images for recognition.

	Corrupted Training Data	Undersampled Gallery Set	Dictionary Learning
SRC [1]	$\times$	$\times$	$\times$
LR [15]	$\surd$	$\times$	$\times$
ESRC [5]	$\surd$	$\surd$	$\times$
Ours	$\surd$	$\surd$	$\surd$

3.2. Optimization

We now provide optimization details for solving (5) during training. Since (5) is nonlinear with respect to variables $\mathbf{X}$ and $\mathbf{A}$, we alternate between the stages of sparse coding and dictionary update for obtaining their optimal solutions.

3.2.1. Sparse Coding for Updating $\mathbf{X}$

In the sparse coding stage, we fix $\mathbf{A}$ and rewrite the minimization problem of (5) as follows:

$$\min_{\mathbf{X}} \left\| \begin{bmatrix} \mathbf{y}_i^e \\ \gamma \mathbf{y}_i^e \end{bmatrix} - \begin{bmatrix} \mathbf{D}^e & \mathbf{A} \\ \gamma \delta_{i_k}(\mathbf{D}^e) & \gamma \mathbf{A} \end{bmatrix} \begin{bmatrix} \mathbf{x}_i^{\mathbf{d}} \\ \mathbf{x}_i^{\mathbf{a}} \end{bmatrix} \right\|_2^2 + \lambda \|\mathbf{x}_i\|_1, \tag{6}$$

where $\gamma = \eta^{1/2}$ and $\delta_{i_k}(\mathbf{D}^e) \in \mathbb{R}^{d\times p}$ whose nonzero columns are those associated with class i_k. Since (6) has the same formulation as (1), one can apply existing techniques such as Homotopy, Iterative Shrinkage-Thresholding, or Augmented Lagrange Multiplier for solving (6). In our work, we adopt the Homotopy method due to its effectiveness and efficiency as suggested in [21].

3.2.2. Dictionary Update for $\mathbf{A}$

To learn the auxiliary dictionary $\mathbf{A}$, we first perform k-means clustering on external data to obtain m cluster centers for initialization (recall in Section 3.1 that m is the number of dictionary atoms to be observed). When updating $\mathbf{A}$, we fix $\mathbf{X}$ in (5) and minimize the following function:

$$\min_{\mathbf{A}} \sum_{i=1}^{N} \left(\left\| \begin{bmatrix} \mathbf{y}_i^e - \mathbf{D}^e \mathbf{x}_i^{\mathbf{d}} \\ \gamma \mathbf{y}_i^e - \gamma \mathbf{D}^e \delta_{i_k}(\mathbf{x}_i^{\mathbf{d}}) \end{bmatrix} - \begin{bmatrix} \mathbf{A}\mathbf{x}_i^{\mathbf{a}} \\ \gamma \mathbf{A}\mathbf{x}_i^{\mathbf{a}} \end{bmatrix} \right\|_2^2 \right), \tag{7}$$

in which $\gamma = \eta^{1/2}$. The minimization problem of (7) can be further expressed as

$$\min_{\mathbf{A}} \left\| \hat{\mathbf{Y}} - \mathbf{Z}\mathbf{A}\mathbf{X}_a \right\|_F^2, \tag{8}$$

where $\mathbf{X}_a = [\mathbf{x}_1^{\mathbf{a}}, \mathbf{x}_2^{\mathbf{a}}, \ldots, \mathbf{x}_N^{\mathbf{a}}]$, $\hat{\mathbf{Y}} = [\hat{\mathbf{y}}_1, \hat{\mathbf{y}}_2, \ldots, \hat{\mathbf{y}}_N]$, and

$$\hat{\mathbf{y}}_i = \begin{bmatrix} \mathbf{y}_i^e - \mathbf{D}^e \mathbf{x}_i^{\mathbf{d}} \\ \gamma \mathbf{y}_i^e - \gamma \mathbf{D}^e \delta_{i_k}(\mathbf{x}_i^{\mathbf{d}}) \end{bmatrix}, \mathbf{Z} = \begin{bmatrix} \mathbf{I} \\ \gamma \mathbf{I} \end{bmatrix}.$$

Let $F(\mathbf{A})$ denote the objective function of (8) and $\mathbf{a}_j$ be the jth column of $\mathbf{A}$. Since (8) is an unconstrained optimization problem and $F(\mathbf{A})$ is a quadratic function of $\mathbf{A}$, the optimal solution of (8) can be analytically obtained by setting the partial derivatives of $F(\mathbf{A})$ with respect to $\mathbf{a}_j$ equal to zero, i.e.,

$$\begin{aligned}\frac{\partial F}{\partial \mathbf{a}_j} &= \sum_{i=1}^{N} -2x_{ij}\left(\mathbf{Z}^T\hat{\mathbf{y}}_i - \mathbf{Z}^T\mathbf{Z}\mathbf{A}\mathbf{x}_i^{\mathbf{a}}\right)\\ &= \sum_{i=1}^{N} -2x_{ij}\left(\mathbf{Z}^T\hat{\mathbf{y}}_i - (1+\gamma^2)\sum_{\ell=1}^{m} x_{i\ell}\mathbf{a}_\ell\right) = 0\end{aligned}$$

for $j = 1, 2, \ldots, m$, where x_{ij} is the (i, j) entry of $\mathbf{X}_{\mathbf{a}}$. As a result, $\mathbf{A}$ can be derived by solving of a linear system. More precisely, by taking the transpose of the above m linear equations, we have

$$\mathbf{U}\begin{bmatrix}\mathbf{a}_1^T\\ \mathbf{a}_2^T\\ \vdots\\ \mathbf{a}_m^T\end{bmatrix} = \mathbf{V}, \text{ where } \mathbf{V} = \sum_{i=1}^{N}\begin{bmatrix}x_{i1}\hat{\mathbf{y}}_i^T\mathbf{Z}\\ x_{i2}\hat{\mathbf{y}}_i^T\mathbf{Z}\\ \vdots\\ x_{im}\hat{\mathbf{y}}_i^T\mathbf{Z}\end{bmatrix} \tag{9}$$

and

$$\mathbf{U} = (1+\gamma^2)\sum_{i=1}^{N}\mathbf{x}_i^{\mathbf{a}}(\mathbf{x}_i^{\mathbf{a}})^T. \tag{10}$$

In view of (9), the auxiliary dictionary $\mathbf{A}$ is obtained by solving the linear system $\mathbf{U}\mathbf{A}^T = \mathbf{V}$.

3.3. Testing

For the query $\mathbf{y}$ and undersampled training data $\mathbf{D}$ of the subjects *of* interest (different from $\mathbf{y}^e$ and $\mathbf{D}^e$ in the training stage), we apply (3) together with the auxiliary dictionary $\mathbf{A}$ learned from external data to calculate the coefficient $\mathbf{x}$. Once $\mathbf{x}$ is obtained for the query $\mathbf{y}$, recognition is achieved by (4).

4. EXPERIMENTAL RESULTS

4.1. Extended Yale B Database

In our experiments, we first consider the Extended Yale B database [22], which contains frontal-face images of 38 subjects with about 64 images for each person. The face images are taken under various illumination conditions [23]. All images are converted into grayscale and are cropped to 192×168 pixels prior to our experiments (see examples in Fig. 2). We select 32 from the 38 subjects to be recognized, and the remaining 6 subjects are considered as external data (i.e., subjects not of interest) for auxiliary dictionary learning.

For the 32 subjects of interest, we select 3 images from each of the 32 subjects as the gallery, and the remaining 61 images for testing. The three gallery images are A+000 E+00, A-085 E+20, and A+085 E+20 (A+085 refers to 85 degrees azimuth, and E+20 refers to 20 degrees elevation [22]). For the training stage of auxiliary dictionary learning using external data (i.e., the six subjects *not* of interest), we choose the same images at A+000 E+00, A-085 E+20, and A+085 E+20 as the gallery $\mathbf{D}^e$, and thus $\mathbf{D}^e$ contains a total of 6×3 images. The probe $\mathbf{Y}^e$ consists of the remaining 61 images of these 6 subjects. The number m of the atoms for the auxiliary dictionary $\mathbf{A}$ is set as $61 \times k$ $(1 \leq k \leq 6)$. We will vary k and evaluate the performance of our approach. The parameters λ and η in (5) are set to 0.001 and 1, respectively.

Fig. 2. Example images of the Extended Yale B database.

We first compare our method with ESRC [5] with different sizes of the external data (i.e, different k). Fig. 3(a) shows the comparisons of these two approaches using LBP or pixel-based features (we apply the settings of LBP as [24] did). From this figure, it is clear that our method outperformed ESRC especially when k is smaller. We note that, it is expected that the difference between these two approaches would become smaller as k increased. This is because that ESRC did not perform any learning on external data, and the use of more external data would give comparable performance as learning-based approaches did (but would be more expensive in terms of both computation and storage costs).

We now compare the performance of different SRC-based approaches: Nearest Neighbor (NN), SRC [1], KSVD [16], and ESRC over different feature dimensions. While NN and SRC and NN do not utilize any external data, we fix the number of external subjects $k = 2$ for KSVD, ESRC, and our approach. For KSVD, we directly apply the KSVD algorithm to learn a dictionary from external data for deriving $\mathbf{A}$, and use (4) as the classification rule. We apply Randomfaces [1] to perform dimensionality reduction, and show the performance comparisons using LBP and pixel-based features in Figs. 3(b) and (c), respectively. From these two figures, we see that our method not only outperformed NN and SRC, we also consistently produced the improved performance over different dimensions when comparing to other SRC-based methods. Although KSVD also performed auxiliary dictionary learning, its performance did not differ much from that of ESRC (without learning). This is because that KSVD only aims at learning dictionaries for data representation. This also confirms that the integration of both auxiliary dictionary learning and the classification rule into a unified optimization framework (like our proposed one in (5)) is preferable.

4.2. AR Database

The AR database [25] contains over 4,000 frontal images of 126 individuals. The images are taken under different variations, including illumination, expression, and facial occlusion/disguise in two separate sessions. More specifically, there are thirteen images for each session, in which three im-

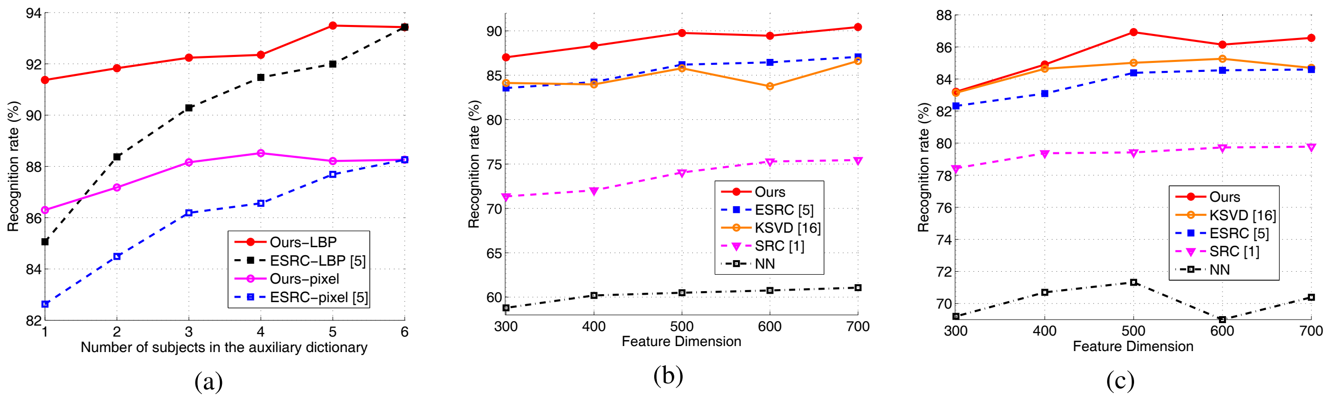


Fig. 3. Performance comparisons on the Extended Yale B database with (a) different number of subjects in the auxiliary dictionary, (b) different dimensions of LBP features, and (c) different dimensions of pixel-based features.

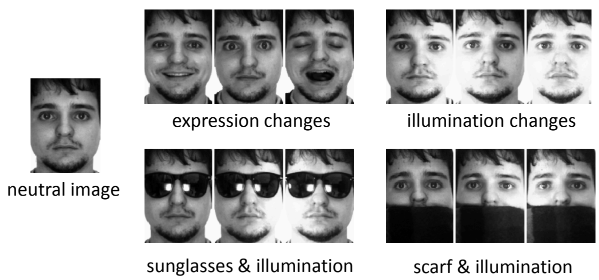


Fig. 4. Example images of the AR database. Note that only the neural image of each subject is included in the gallery set, while the rest are to be recognized.

ages with sunglasses, another three with scarfs, and the remaining seven are with illumination and expressions variations. In our experiments, all images are cropped to 165×120 pixels and converted to grayscale. We choose a subset of AR consisting of 50 men and 50 women (as [1] did). From the first session, we select 80 subjects of interest for training and testing, and the remaining 20 subjects are considered as external data for auxiliary dictionary learning.

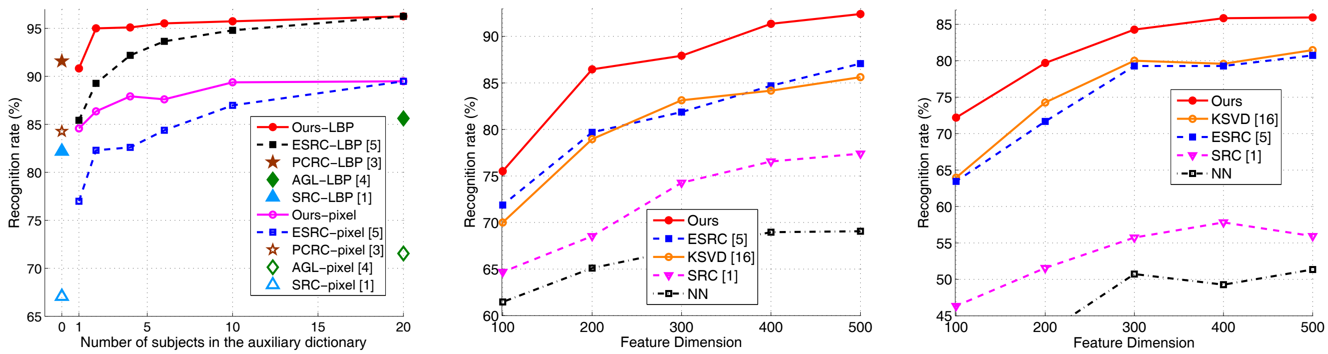
For the scenario of undersampled face recognition, we select only the *neutral* image of each of the 80 subjects as the gallery, and the rest for testing (see Fig. 4 for example). To learn the auxiliary dictionary $\mathbf{A}$ using external data, we also choose neutral images as the gallery $\mathbf{D}^e$ (and thus $\mathbf{D}^e$ contains a total of 20 images). The probe $\mathbf{Y}^e$ consists of the remaining images from the above 20 external subjects. The number of atoms m of the auxiliary dictionary $\mathbf{A}$ is $(13-1) \times k$ $(1 \leq k \leq 20)$. It is worth noting that the setting for the AR database is more challenging than that of the Extended Yale B database. We not only need to deal with image variants of illumination, expression, and occlusion, we only require one neutral image for each person as the gallery for recognition.

By varying the size k of the external data, we consider several SRC-based approaches (using LBP or pixel-based features) for comparisons: SRC, ESRC, AGL [4], and PCRC [3] (both AGL and PCRC are designed to handle only one training image per subject). We show the performance comparisons in Fig. 5(a). Note that we have $k = 0$ for SRC and PCRC, since they did not consider any external data. From this figure, we see that our method outperformed others when utilizing external data for auxiliary dictionary learning. If no external data is available, methods of ESRC and ours would turn into SRC. It is worth noting that although AGL considered external data for designing its classifiers, its recognition rate was much lower than ours. This is due to the fact that AGL did not consider the possible corruption of training data. Similar to our observations for the Extended Yale B dataset, our method required a smaller size of external data than ESRC did for producing comparable performance.

Finally, we compare the performance of NN, SRC, KSVD, ESRC, and ours over a range of dimensionality (with $k = 2$). We did not consider PCRC or AGL for this comparison, since PCRC is a patch-based method, and AGL is extended from LDA. In other words, the feature dimensions of both PCRC and AGL are fixed (e.g., only $80-1$ for AGL). We plot the performance comparisons using LBP and pixel-based features in Figs. 5(b) and (c), respectively. From both figures, it can be seen that our method outperformed others over all feature dimensions considered. It is worth repeating that the intra-class variations for AR are more challenging than those of Extended Yale B. From the above experiments, the effectiveness of our proposed algorithm is successfully verified. We confirm that a joint optimization framework which solves auxiliary dictionary learning and SRC would be preferable for undersampled face recognition problems.

5. CONCLUSION

We presented an auxiliary dictionary learning algorithm for solving undersampled face recognition problems. Our proposed method is able to model intra-class variations (e.g., illumination and expression changes) and corruptions (e.g., occlusion or disguise) using a small amount of external face data. Using the observed auxiliary dictionary, our approach can be applied to the scenarios in which only one or few face images of each subject of interest are available (i.e., as the



(a) (b) (c)

Fig. 5. Performance comparisons on the AR database with (a) different number of subjects in the auxiliary dictionary, (b) different dimensions of LBP features, and (c) different dimensions of pixel-based features.

gallery set). Experimental results confirmed the effectiveness of our method, which was shown to outperform state-of-the-art sparse representation based approaches with or without using external face data.

Acknowledgement This work is supported in part by National Science Council of Taiwan via NSC100-2221-E-001-018-MY2.

6. REFERENCES

[1] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE PAMI*, 2009.

[2] A. Wagner, J. Wright, A. Ganesh, Z. Zhou, H. Mobahi, and Y. Ma, "Towards a practical face recognition system: Robust alignment and illumination by sparse representation," *IEEE PAMI*, 2012.

[3] P. Zhu, L. Zhang, Q. Hu, and S. C.K. Shiu, "Multi-scale patch based collaborative representation for face recognition with margin distribution optimization," in *Proc. ECCV*, 2012.

[4] Y. Su, S. Shan, X. Chen, and W. Gao, "Adaptive generic learning for face recognition from a single sample per person," in *Proc. IEEE CVPR*, 2010.

[5] W. Deng, J. Hu, and J. Guo, "Extended SRC: Undersampled face recognition via intraclass variant dictionary," *IEEE PAMI*, 2012.

[6] I. Tošić and P. Frossard, "Dictionary learning: What is the right representation for my signal?," *IEEE Signal Process. Mag.*, 2011.

[7] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE TIP*, 2006.

[8] J. Mairal, M. Elad, and G. Sapiro, "Sparse representation for color image restoration," *IEEE TIP*, 2008.

[9] O. G. Guleryuz, "Nonlinear approximation based image recovery using adaptive sparse reconstructions and iterated denoising - part I: Theory," *IEEE TIP*, 2006.

[10] I. Ramirez, P. Sprechmann, and G. Sapiro, "Classification and clustering via dictionary learning with structured incoherence and shared features," in *Proc. IEEE CVPR*, 2010.

[11] Q. Zhang and B. Li, "Discriminative K-SVD for dictionary learning in face recognition," in *Proc. IEEE CVPR*, 2010.

[12] Z. Jiang, Z. Lin, and L. S. Davis, "Learning a discriminative dictionary for sparse coding via label consistent K-SVD," in *Proc. IEEE CVPR*, 2011.

[13] J. Yang, J. Wang, and T. Huang, "Learning the sparse representation for image classification," in *Proc. IEEE ICME*, 2011.

[14] M. Yang, L. Zhang, X. Feng, and D. Zhang, "Fisher discrimination dictionary learning for sparse representation," in *Proc. ICCV*, 2011.

[15] C.-F. Chen, C.-P. Wei, and Y.-C. F. Wang, "Low-rank matrix recovery with structural incoherence for robust face recognition," in *Proc. IEEE CVPR*, 2012.

[16] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trnas. Signal Process.*, 2006.

[17] K. Engan, S. O. Aase, and J. H. Husoy, "Method of optimal directions for frame design," in *Proc. IEEE ICASSP*, 1999.

[18] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, "Discriminative learned dictionaries for local image analysis," in *Proc. IEEE CVPR*, 2008.

[19] D.-S. Pham and S. Venkatesh, "Joint learning and dictionary construction for pattern recognition," in *Proc. IEEE CVPR*, 2008.

[20] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, "Supervised dictionary learning," in *NIPS*. 2009.

[21] A. Y. Yang, S. S. Sastry, A. Ganesh, and Y. Ma, "Fast ℓ_1-minimization algorithms and an application in robust face recognition: A review," in *Proc. IEEE ICIP*, 2010.

[22] A. S. Georghiades, P. N. Belhumeur, and D. J. Kriegman, "From few to many: Illumination cone models for face recognition under variable lighting and pose," *IEEE PAMI*, 2001.

[23] K.-C. Lee, J. Ho, and D. J. Kriegman, "Acquiring linear subspaces for face recognition under variable lighting," *IEEE PAMI*, 2005.

[24] T. Ahonen et al., "Face description with local binary patterns: Application to face recognition," *IEEE PAMI*, 2006.

[25] A. M. Martinez and R. Benavente, "The AR face database," *CVC Technical Report*, 1998.