UNDERSAMPLED FACE RECOGNITION WITH ONE-PASS DICTIONARY LEARNING

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

Undersampled face recognition deals with the problem in which, for each subject to be recognized, only one or few images are available in the gallery (training) set. Thus, it is very difficult to handle large intra-class variations for face images. In this paper, we propose a one-pass dictionary learning algorithm to derive an auxiliary dictionary from external data, which consists of image variants of the subjects not of interest (not to be recognized). The proposed algorithm not only allows us to efficiently model intra-class variations such as illumination and expression changes, it also exhibits excellent abilities in recognizing corrupted images due to occlusion. In our experiments, we will show that our method would perform favorably against existing sparse representation or dictionary learning based approaches. Moreover, our computation time is remarkably less than that of recent dictionary learning based face recognition methods. Therefore, the effectiveness and efficiency of our proposed algorithm can be successfully verified.

***Index Terms*—** Face recognition, sparse representation, dictionary learning

1. INTRODUCTION

Due to low intrusiveness and high uniqueness, face recognition has been among the most popular biometric approaches for identity recognition. Practically, face recognition is still a challenging task, since face images often exhibit pose, illumination, and expression variations, or even encounter corruptions due to occlusion or disguise. A traditional way to tackle the task of face recognition is to collect a sufficient amount of training data, which are expected to cover the aforementioned variations. Unfortunately, in real-world scenarios like those for surveillance or forensic purposes, the subject of interest might only have very few face images available for training (to be matched). This leads us to the challenging task of *undersampled* face recognition. If only one image can be observed in advanced, then a even more difficult problem of *single sample* face recognition needs to be addressed.

Approaches to undersampled or single sample face recognition can be typically categorized into two groups: local matching based methods [1, 2, 3] and sparse representation based methods [4, 5, 6, 7]. The former type of approaches

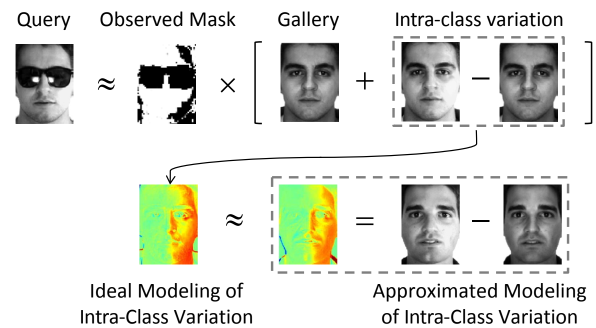

Fig. 1. Our proposed framework for undersampled face recognition. Note that we utilize external data from the subjects not of interest to model intra-class variations.

extract discriminative features from patches of face images, and integrate the classification results from each patch for determining the final output/decision. Several textural features such as local binary pattern (LBP) [1], Gabor filters [2], and local Gabor XOR pattern (LGXP) [3] have been utilized for feature extraction. A concern of local matching methods comes from the fact that local patches only carry limited information, especially for the case of single sample face recognition. When the difference between the variations of test images and those of the training ones is large, the recognition performance would degrade significantly.

Recently, sparse representation based classification (SRC) [8] has shown promising performance for robust face recognition [9, 10, 11]. It assumes that the test image belongs to the subspace spanned by the face images of the training dictionary. Because of this assumption, SRC requires a large amount of training images to construct the dictionary. To apply SRC for undersampled face recognition, approaches like [4, 5, 6, 7] applied an *auxiliary dictionary* for modeling intra-class variations. This auxiliary dictionary is constructed from external data with subjects *not* of interest. With the observed auxiliary dictionary, one can have one or very few images per person as the training dictionary.

Generally, there are two different ways to construct the auxiliary dictionary from external data. In [4, 5], the mean image of external data is subtracted from each of the external face images, and thus the resulting data matrix can be viewed as an auxiliary dictionary for modeling intra-class variations.

While promising results were reported in [4, 5], the direct use of external data as auxiliary dictionary might not always be preferable, since such data might contain noisy or redundant information. Dictionary learning techniques can be further applied to derive compact yet representative auxiliary dictionaries from external data [6, 7]. Nevertheless, in order to cover a sufficient amount of intra-class variations, the size of the auxiliary dictionary might still be large. As a result, its computation cost would be high.

In this paper, we focus on undersampled face recognition with auxiliary dictionary learning, as illustrated in Figure 1. Similar to [4, 5, 6, 7], we utilize external data with subjects not of interest. However, unlike these prior approaches, our auxiliary dictionary would vary with the test input images. The benefit of this property is that the auxiliary dictionary does not need to cover all possible intra-class variations when recognizing each face image. We propose a *one-pass dictionary learning* algorithm for addressing the tasks of learning the auxiliary dictionary and recognizing the test input simultaneously. Our algorithm not only models intra-class variations accurately (which is dependent on the test input image), the size of our auxiliary dictionary can also be very compact for reduced computation purposes. Later in our experiments, we will verify that our proposed method performs favorably against state-of-the-art SRC-based face recognition approaches. In addition, we will also show that reduced computation time for dictionary learning can be achieved.

2. RELATED WORK

2.1. SRC and Extended SRC (ESRC)

Proposed by Wright *et al.* [8], sparse representation based classification (SRC) has been successfully applied for solving robust face recognition problems. SRC assumes that a test image $\mathbf{y} \in \mathbb{R}^d$ can be represented as a sparse linear combination of the columns of a training dictionary $\mathbf{X}$. More precisely, SRC solves the following L_1-minimization problem to obtain the sparse coefficient $\boldsymbol{\alpha}$ of $\mathbf{y}$:

$$\min_{\boldsymbol{\alpha}} \|\mathbf{y} - \mathbf{X}\boldsymbol{\alpha}\|_2^2 + \lambda \|\boldsymbol{\alpha}\|_1. \tag{1}$$

Once the solution of (1) is obtained, the test input $\mathbf{y}$ will be recognized as class k^* by:

$$k^* = \arg\min_k \|\mathbf{y} - \mathbf{X}\delta_k(\boldsymbol{\alpha})\|_2, \tag{2}$$

where $\delta_k(\boldsymbol{\alpha})$ is a vector whose only nonzero entries are the entries in $\boldsymbol{\alpha}$ that are associated with class k. That is, the test sample $\mathbf{y}$ is assigned to the class with the minimum class-wise reconstruction error.

In addition to requiring the training images to be registered/aligned, SRC is based on the collection a sufficiently large amount of such images as the dictionary $\mathbf{X}$. In other words, a direct use of SRC for the task of undersampled or single sample face recognition would lead to degraded performance. In view of this issue, Deng *et al.* [4] proposed Extended SRC (ESRC), which considers the following minimization problem:

$$\min_{\boldsymbol{\alpha}} \left\| \mathbf{y} - [\mathbf{X},\ \mathbf{D}_{\mathrm{ESRC}}] \begin{bmatrix} \boldsymbol{\alpha}_{\mathbf{x}} \\ \boldsymbol{\alpha}_{\mathbf{d}} \end{bmatrix} \right\|_2^2 + \lambda \|\boldsymbol{\alpha}\|_1, \tag{3}$$

where $\boldsymbol{\alpha} = [\boldsymbol{\alpha}_{\mathbf{x}};\ \boldsymbol{\alpha}_{\mathbf{d}}]$. For ESRC, the training dictionary $\mathbf{X}$ contains only one or few images per subject, while the intra-class variant dictionary $\mathbf{D}_{\mathrm{ESRC}}$ consists of image data from an external dataset with subjects not of interest. Different from SRC, ESRC applies the following criterion to classify the test sample $\mathbf{y}$:

$$k^* = \arg\min_k \left\| \mathbf{y} - [\mathbf{X},\ \mathbf{D}_{\mathrm{ESRC}}] \begin{bmatrix} \delta_k(\boldsymbol{\alpha}_{\mathbf{x}}) \\ \mathbf{v}_{\mathbf{d}} \end{bmatrix} \right\|_2. \tag{4}$$

We note that, the difference between (2) and (4) is that, instead of applying the operator $\delta_k(\cdot)$ to the entire coefficient vector $\boldsymbol{\alpha}$ in (2), ESRC only applies $\delta_k(\cdot)$ to $\boldsymbol{\alpha}_{\mathbf{x}}$. This is because that $\boldsymbol{\alpha}_{\mathbf{d}}$ is not associated with any label information (i.e., subject identity) of interest.

As noted in Section 1, ESRC directly applies external data as $\mathbf{D}_{\mathrm{ESRC}}$, which may cause $\mathbf{D}_{\mathrm{ESRC}}$ to be noisy or contain undesirable artifacts. Furthermore, ESRC relies on the intra-class variant dictionary $\mathbf{D}_{\mathrm{ESRC}}$ for modeling not only intra-class variations but also occlusion. This might not be preferable, since the type of occlusion typically cannot be known in advance during the collection of external data.

2.2. Robust Sparse Coding (RSC)

In (1), SRC measures the reconstruction error in terms of L_2-norm. It is known that L_2-norm amplifies entries with large magnitudes. Therefore, L_2-norm characterization might not be suitable for reconstructing or recognizing occluded images. Based on this observation, Yang *et al.* [12] proposed robust sparse coding (RSC), which iteratively solves the coefficient $\boldsymbol{\alpha}$ and a matrix $\mathbf{W}$:

$$\min_{\boldsymbol{\alpha}} \|\mathbf{W}(\mathbf{y} - \mathbf{X}\boldsymbol{\alpha})\|_2^2 + \lambda \|\boldsymbol{\alpha}\|_1 \tag{5}$$

with

$$\mathbf{W} = \mathrm{diag}(w(e_1), w(e_2), \ldots, w(e_d))^{1/2}, \tag{6}$$

$$w(e_k) = \frac{\exp(-\mu e_k^2 + \mu\delta)}{1 + \exp(-\mu e_k^2 + \mu\delta)}, \tag{7}$$

where e_k is the kth entry of $\mathbf{e} = \mathbf{y} - \mathbf{X}\boldsymbol{\alpha}$, and μ and δ are given parameters. Note that $\mathbf{W}$ is a diagonal matrix whose nonzero entries are the weights of entries of $\mathbf{e}$. It can be seen that, the idea of RSC is to assign small weights for entries of $\mathbf{e}$ with large magnitudes. As a result, the influence of poorly reconstructed pixels can be suppressed. In the same spirit of SRC, RSC classifies $\mathbf{y}$ according to

$$k^* = \arg\min_k \|\mathbf{W}(\mathbf{y} - \mathbf{X}\delta_k(\boldsymbol{\alpha}))\|_2. \tag{8}$$

Although RSC has improved SRC for the task of robust face recognition, it has the same limitation as SRC, i.e., the training dictionary $\mathbf{X}$ needs to contain a sufficient number of training images to represent subjects of interest. In other words, RSC cannot be directly applied to the problems of undersampled or single sample face recognition.

3. FACE RECOGNITION VIA ONE-PASS DICTIONARY LEARNING

3.1. Extending ESRC for Robust Face Recognition

We now present our proposed algorithm for undersampled (including single sample) face recognition. As illustrated in Figure 1, our method allows the query images to exhibit illumination and expression variations, or even corrupted due to occlusion. Extended from ESRC, we advocate the learning of an auxiliary dictionary for modeling intra-class variations. Different from ESRC which directly applies external data as the auxiliary dictionary for handling occlusion, we further tackle such problems by RSC, which views occluded image regions as poorly reconstructed pixels. This makes our approach suitable for real-world applications, since the information about occlusion might not be available during the collection of external data.

We first define the notations for the sake of clarity. Let $\mathbf{y} \in \mathbb{R}^d$ be the query image and $\mathbf{X} \in \mathbb{R}^{d\times N}$ be the gallery matrix consisting of a total of N training images. The auxiliary dictionary $\mathbf{D} \in \mathbb{R}^{d\times M}$ to be learned is derived from external data. It is worth repeating that, external data contains face images of the subjects *not* of interest (i.e., subjects not to be recognized). The algorithm for deriving $\mathbf{D}$ will later be detailed in Section 3.2.

For the recognition stage, we apply RSC and solve the following optimization problem:

$$\min_{\boldsymbol{\alpha}} \left\| \mathbf{W}\left(\mathbf{y} - [\mathbf{X},\ \mathbf{D}]\begin{bmatrix}\boldsymbol{\alpha}_{\mathbf{x}}\\ \boldsymbol{\alpha}_{\mathbf{d}}\end{bmatrix}\right)\right\|_2^2 + \lambda\|\boldsymbol{\alpha}\|_1. \tag{9}$$

Note that $\mathbf{W}$ in (9) is updated according to (6) and (7), in which e_k is the kth entry of the vector indicating the reconstruction error $\mathbf{e} = \mathbf{y} - [\mathbf{X},\ \mathbf{D}]\boldsymbol{\alpha}$. We have $\boldsymbol{\alpha} = [\boldsymbol{\alpha}_{\mathbf{x}};\ \boldsymbol{\alpha}_{\mathbf{d}}]$. The parameters μ and δ in (7) are determined based on the rules provided in [12]. Once both $\boldsymbol{\alpha}$ and $\mathbf{W}$ are obtained, we finally classify $\mathbf{y}$ by:

$$k^* = \arg\min_k \left\| \mathbf{W}\left(\mathbf{y} - [\mathbf{X},\ \mathbf{D}]\begin{bmatrix}\delta_k(\boldsymbol{\alpha}_{\mathbf{x}})\\ \mathbf{v}_{\mathbf{d}}\end{bmatrix}\right)\right\|_2, \tag{10}$$

where the operator δ_k is defined as the one in (2).

3.2. One-Pass Dictionary Learning (OPDL)

3.2.1. External Data with Same Image Variants

For ESRC or related undersampled recognition approaches, external data are collected with the goal that the images of these subjects not of interest would cover necessary intra-class variations. For simplicity, we first consider the scenario in which the external data and the gallery images have the *same* intra-class variations. For example, the number of image variants for illumination changes are known in advance, and thus one can collect external data which contains such image variants but just from the subjects not of interest. Later in Section 3.2.2, this assumption will be relaxed.

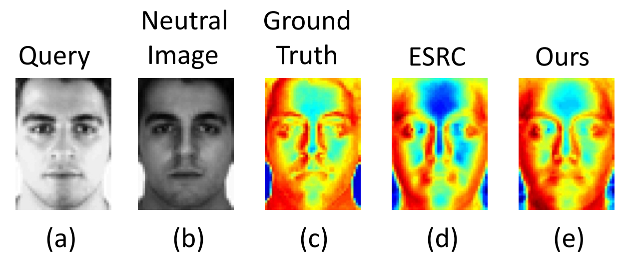

Fig. 2. Modeling intra-class variations using external data. Note that the query and neutral (gallery) images are shown in (a) and (b), respectively. The ground truth intra-class variation between (a) and (b) is depicted in (c), while the intra-class variations derived by ESRC and our method are shown in (d) and (e), respectively.

We now provide the definitions of the notations for the sake of clarity. Given an external data matrix $\mathbf{E} \in \mathbb{R}^{d\times(KV)}$ with K subjects not of interest, each with V different types of image variants, our goal is to learn an auxiliary dictionary $\mathbf{D}$ from $\mathbf{E}$ so that intra-class variations can be properly described. Assume that $\mathbf{E} = [\mathbf{E}_1, \mathbf{E}_2, \cdots, \mathbf{E}_K]$, where $\mathbf{E}_k \in \mathbb{R}^{d\times V}$ is the external data matrix of subject k with $k \in \{1, 2, \ldots, K\}$. More precisely, we have $\mathbf{E}_k = [\mathbf{E}_k^1, \mathbf{E}_k^2, \cdots, \mathbf{E}_k^V]$, in which the vector $\mathbf{E}_k^v$ is the vth image of subject k with a particular illumination or expression variation. In addition, we further define $\mathbf{E}^v \in \mathbb{R}^{d\times K}$ as a sub-matrix of $\mathbf{E}$, and $\mathbf{E}^v = [\mathbf{E}_1^v, \mathbf{E}_2^v, \cdots, \mathbf{E}_K^v]$ with $v \in \{1, 2, \ldots, V\}$. In other words, $\mathbf{E}^v$ indicates the collection of the vth type of image variants across all K subjects in $\mathbf{E}$.

Without the loss of generality, we assume that the gallery matrix $\mathbf{X}$ has a particular type of image variants $g \in \{1, 2, \ldots, V\}$. As for the query $\mathbf{y}$, its image variant is of type $q \in \{1, 2, \ldots, V\}$, which is not necessarily the same as g. Now, we decompose the query image $\mathbf{y}$ into $\mathbf{y}^g + \tilde{\mathbf{y}}$, where $\mathbf{y}^g$ indicates the portion which can be recovered by the gallery images $\mathbf{X}$ (whose intra-class variation is of type g only). On the other hand, $\tilde{\mathbf{y}}$ would contain the remaining difference in terms of intra-class variations, and thus we have $\tilde{\mathbf{y}} = \mathbf{y} - \mathbf{y}^g$. As a result, by utilizing the external data $\mathbf{E}$, $\tilde{\mathbf{y}}$ can be described by $\mathbf{E}^q - \mathbf{E}^g$. This results in the observed auxiliary dictionary as

$$\mathbf{D} = \mathbf{E}^q - \mathbf{E}^g, \tag{11}$$

in which both q and g are among $\{1, 2, \ldots, V\}$ but unknown.

To learn the auxiliary dictionary $\mathbf{D}$, it is obvious that we need to determine q and g, i.e., the types of image variants which the query $\mathbf{y}$ and gallery $\mathbf{X}$ belong to, respectively. Based on the use of external data $\mathbf{E}$ which contains V types of variations of interest, q and g can be determined by:

$$\begin{aligned} q &= \arg\min_{v} \operatorname{dist}(\mathbf{y}, \mathbf{m}^v), \\ g &= \arg\min_{v} \operatorname{dist}(\bar{\mathbf{x}}, \mathbf{m}^v), \end{aligned} \tag{12}$$

where $\mathbf{m}^v = \frac{1}{K}\sum_{k=1}^{K}\mathbf{E}_k^v$, $\bar{\mathbf{x}} = \frac{1}{N}\sum_{n=1}^{N}\mathbf{X}(:,n)$, and $\operatorname{dist}(\cdot,\cdot)$ calculates the Euclidean distance between the two inputs. The rationale behind (12) is that, compared to the different face images of the same subject, face images of the same type of variations across different subjects tend to be close to each other [13]. Once q and g are determined, the auxiliary dictionary in (11) can be derived. It can be seen that, one of the advantages of our method is its reduced training time compared to existing dictionary learning based approaches, which typically require iterative optimization in their learning processes.

It is worth noting that, in the standard ESRC [4], its auxiliary dictionary is constructed by

$$\mathbf{D}_{\text{ESRC}} = [\mathbf{E}^1 - \mathbf{E}^g, \mathbf{E}^2 - \mathbf{E}^g, \cdots, \mathbf{E}^V - \mathbf{E}^g], \tag{13}$$

where g is the type of image variants which $\mathbf{X}$ belongs to. In other words, ESRC utilizes all the variation differences observed from external data as the auxiliary dictionary. Different from ESRC, our method only learns the intra-class variation which is directly associated with the query image. This not only leads to better modeling capability, computation costs are also greatly reduced. Figure 2 shows an image-pair selected from the AR database [14], in which our approach better describes the query image than ESRC does due to improved intra-class variation (i.e., $\mathbf{D}\boldsymbol{\alpha}_{\mathbf{d}}$) observed.

3.2.2. External Data with Similar Image Variants

In Section 3.2.1, we have the assumption of both q and g being among $\{1, 2, \ldots, V\}$. That is, the external data $\mathbf{E}$ is expected to encompass all V types of image variants as those presented in the gallery set $\mathbf{X}$. We now further relax this assumption for more general and practical scenarios.

Let $\bar{\mathbf{E}}$ be defined as $\bar{\mathbf{E}} = [\mathbf{m}^1, \mathbf{m}^2, \cdots, \mathbf{m}^V] \in \mathbb{R}^{d\times V}$, where $\mathbf{m}^v = \frac{1}{K}\sum_{k=1}^{K}\mathbf{E}_k^v$. This matrix $\bar{\mathbf{E}}$ can be viewed as the collection of V types of image variants. Since the gallery set $\mathbf{X}$ and the query input $\mathbf{y}$ do not necessarily have the same type of variation $v \in \{1, 2, \ldots, V\}$, it would be more challenging to apply the techniques presented in Sections 3.1 and 3.2.1 for performing recognition.

To address this problem, we first calculate $\bar{\mathbf{x}}$ as the mean of the column vectors of the gallery images $\mathbf{X}$. The coefficient $\boldsymbol{\beta} \in \mathbb{R}^{V\times 1}$ for estimating the image variant of $\bar{\mathbf{x}}$ can be determined by RSC, i.e.,

$$\boldsymbol{\beta} = \arg\min_{\boldsymbol{\beta}} \left\|\mathbf{W}(\bar{\mathbf{x}} - \bar{\mathbf{E}}\boldsymbol{\beta})\right\|_2^2 + \lambda\left\|\boldsymbol{\beta}\right\|_1, \tag{14}$$

in which $\mathbf{W}$ is updated via (6) and (7). Similarly, we calculate $\boldsymbol{\gamma} \in \mathbb{R}^{V\times 1}$ for estimating the image variant of $\mathbf{y}$:

$$\boldsymbol{\gamma} = \arg\min_{\boldsymbol{\gamma}} \left\|\mathbf{W}(\mathbf{y} - \bar{\mathbf{E}}\boldsymbol{\gamma})\right\|_2^2 + \lambda\left\|\boldsymbol{\gamma}\right\|_1. \tag{15}$$

It can be seen that, $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ describe the image variants of the gallery and query images using external data, respectively. More specifically, the image variants presented in the gallery/query images are modeled as linear combinations of those observed in $\bar{\mathbf{E}}$.

Once $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are obtained, the auxiliary dictionary can be constructed as:

$$\mathbf{D} = \sum_{v=1}^{V}\boldsymbol{\gamma}(v)\mathbf{E}^v - \boldsymbol{\beta}(v)\mathbf{E}^v = \sum_{v=1}^{V}(\boldsymbol{\gamma}(v) - \boldsymbol{\beta}(v))\mathbf{E}^v. \tag{16}$$

It can be seen that, $\mathbf{D}$ models the difference between the estimated image variant of $\bar{\mathbf{x}}$ and that of $\mathbf{y}$. Recall that, in Section 3.2.1, the gallery $\mathbf{X}$ is assumed to have the same image variants as $\mathbf{E}^g$ does, while the variants to be observed in the query $\mathbf{y}$ is assumed to be equivalent to that of $\mathbf{E}^q$. As a result, we would observe $\boldsymbol{\beta}(v) \approx 1$ when $v = g$, and $\boldsymbol{\beta}(v) \approx 0$ otherwise; similarly, we have $\boldsymbol{\gamma}(v) \approx 1$ when $v = q$, otherwise $\boldsymbol{\gamma}(v) \approx 0$. In other words, if both gallery an external data exhibit the same types of image variants, (16) can be simplified and turns into (11). Therefore, (16) can be considered as more general formulation for learning the auxiliary dictionary.

4. EXPERIMENTS

4.1. Face Recognition Using External Data with Same Image Variants

We consider the AR database [14], which contains over 4,000 face images of 126 individuals in frontal pose. The images are captured under different facial expressions, illumination conditions, and facial occlusion/disguise in 2 separate sessions. Each subject in one session has 13 images (3 images are with sunglasses, another 3 are with scarves, and the remaining 7 images are with illumination and expressions variations).

In our experiments, we consider a subset of AR with 50 men and 50 women. All images are cropped to 165×120 pixels and converted to grayscale. From these 100 subjects, we randomly choose 80 subjects of interest (40 men and 40 women) for training and testing, and the remaining 20 subjects are viewed as external data for learning the auxiliary dictionary. As a result, both images to be recognized and those in the external data exhibit the same types of intra-class variations. For the setting of undersampled face recognition, we select only the *neutral* image of each of the 80 subjects in Session 1 as the gallery, and the remaining images in Sessions 1 and 2 are for testing (see Figure 3 for examples). The external dataset $\mathbf{E}$ consists of 20 subjects per session, and each subject has 13 images (i.e., 13 types of variations). In other words, we have $K = 20 \times 2 = 40$ and $V = 13$.

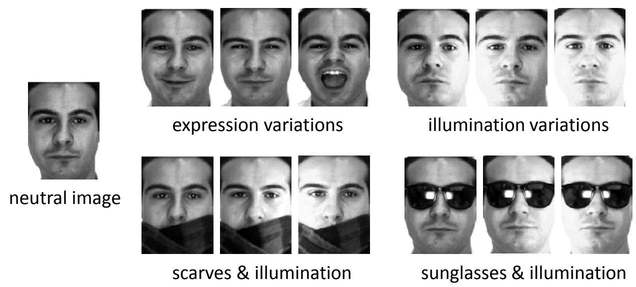

Fig. 3. Example images of the AR database.

We compare our method with several state-of-the-art sparse representation or dictionary learning based approaches (using pixel-based or Gabor features): SRC [8], RSC [12], AGL [15], SVDL [7], ESRC [4], ADL [6]. In addition, we consider a baseline method of ESRC+RSC, which utilizes (10) as the classification algorithm with $\mathbf{D}$ directly derived from ESRC instead of via our proposed one-pass dictionary learning. The pixel-based feature vector is obtained by downsampling the cropped image to 55×40 pixels. The Gabor feature vector of length 2,304 is derived from evaluating the Gabor kernel at three scales and four orientations, see [16] for more detailed information. For this and all subsequent experiments, the parameter λ in (9) is set as 10^{-4}. For dictionary learning based approaches of SVDL, ADL, ESRC, and ours, the dictionary size is fixed as 26 for fair comparisons (as suggested in [6]). In other words, once our $\mathbf{D}$ is determined by (11), we randomly select and keep 26 out of $K = 40$ atoms as the final dictionary to use.

Table 1 compares the recognition rates of different methods. We note that the gallery matrix $\mathbf{X}$ is obtained from Session 1, while the query image $\mathbf{y}$ is chosen from Sessions 1 or 2. As expected, recognizing images from Session 2 is more challenging. From Table 1, we see that our method outperformed other SRC-based approaches across different features and sessions. SVDL is observed to produce a lower recognition rate than ESRC. This is because that the auxiliary dictionary of SVDL is not designed to handle occlusion, while ESRC views occlusion as intra-class variations (directly from external data). It is worth repeating that our method achieved higher recognition rates than ESRC+RSC. While both ESRC+RSC and our method adopt the same classification algorithm presented in Section 3.1, the only difference lies in the way for auxiliary dictionary learning. That is, ESRC+RSC applies ESRC directly, while ours utilizes the proposed one-pass dictionary learning algorithm. This verifies the effectiveness of our proposed algorithm for modeling intra-class variations.

4.2. Face Recognition Using External Data with Similar Image Variants

For the experiments in which the external data and gallery images do not have exactly the same image variants, we consider the use of the AR database as the gallery images, and a different dataset of Extended Yale B [17, 18] as external data. In other words, we have the subjects of and not of interest selected from distinct datasets. We note that, the Extended Yale B database contains 38 subjects, and each of the subject has about 64 frontal face images. Different from the AR database which contains images with illumination, expression, and occlusion changes, Extended Yale B only consists of images with illumination variations (and at different lighting angles as those in AR).

Table 1. Recognition rates of the AR database using external data selected from a disjoint subset of the same dataset. Note that * indicates methods without using external data.

Methods	Pixel-based		Gabor	
	Session 1	Session 2	Session 1	Session 2
SRC* [8]	57.50	43.17	75.31	57.50
RSC* [12]	73.33	54.62	92.08	72.98
ESRC[4]	81.67	66.15	87.92	71.54
ESRC+RSC	86.67	70.19	95.21	79.04
AGL [15]	73.54	51.35	80.31	60.58
ADL [6]	86.67	72.69	92.92	80.58
SVDL [7]	69.58	56.44	86.98	66.44
Ours	**91.56**	**78.17**	**96.56**	**85.19**

We compare our method with SRC-based approaches that use auxiliary dictionary for modeling intra-class variations: SVDL [7], ADL [6], and ESRC [4]. Table 2 lists the recognition results of different approaches. From Table 2, we see that the recognition rates of ADL and ESRC degraded remarkably when the external data was changed from AR to Extended Yale B. This is because that ADL and ESRC rely on the auxiliary dictionary for handling occlusion. When external data (e.g., Extended Yale B) does not contain such image variants, ADL and ESRC would have difficulties in recognizing occluded query images. Our method does not suffer from this problem, since our algorithm does not treat occlusion as intra-class variations, and deals with such image variants via RSC instead (i.e., via updating $\mathbf{W}$). We also note that, since learning the auxiliary dictionary using Extended Yale B will not be expected to cover all image variants, the recognition rates were generally lower than those using a disjoint subset of AR as external data (i.e., results presented in Section 4.1). Nevertheless, from the above experiments, our method achieved improved performance and performed favorably against state-of-the-art SRC or dictionary learning based approaches for undersampled face recognition problems.

4.3. Remarks on Computational Time

Finally, we provide remarks on computation time for the approaches which require dictionary learning. Even this learning stage can be performed offline, it is desirable to be able to solve this problem efficiently. This is because that, in practi-

Table 2. Recognition of the AR database with external data selected from (a disjoint subset of) AR or Extended Yale B. Note that query images are selected from Sessions 1 or 2 of AR, while the gallery images are from Session 1 only.

Methods	Session 1		Session 2	
	AR	Ext. Yale	AR	Ext. Yale
ESRC [4]	87.92	77.81	71.54	59.42
ADL [6]	92.92	80.83	80.58	62.50
SVDL [7]	86.98	85.31	66.54	66.44
Ours	**96.56**	**93.65**	**85.19**	**79.81**

Table 3. Computation time (in seconds) for auxiliary dictionary learning using Extended Yale B. Ours* refers to the use of a disjoint subset of AR for OPDL (i.e., Section 4.1).

Method	ADL [6]	SVDL [7]	Ours	Ours*
Time (secs)	4192	282	0.64	0.016

cal scenarios, one might encounter new types of image variants for recognition and thus need to update the dictionary accordingly. Table 3 lists the computation time of selected dictionary learning based methods. From this table, we see that our proposed OPDL algorithm required the least computation time. Compared to other dictionary learning based approaches, the computation advantage of our method comes from the simpler formulation for dictionary learning and easier optimization (i.e., one-pass learning vs. iterative optimization). Therefore, our proposed method for undersampled face recognition is favorable in terms of the required computation time. The runtime estimates in Table 3 were obtained on an Intel Quad Core PC with 2.33 GHz processors and 4G RAM.

5. CONCLUSION

We presented a one-pass dictionary learning algorithm for undersampled face recognition. Unlike traditional approaches which require the collection of a sufficient amount of training data for dictionary learning, we propose to learn the auxiliary dictionary efficiently from external data, aiming at modeling possible image variants for improved image representation and recognition. Our formulation further incorporates the technique of robust sparse coding. This allows us to deal with occluded query inputs, even if there is no prior knowledge on the type of occlusion. Our experiments using two face datasets confirmed the effectiveness of our proposed method, which is also preferable over existing dictionary learning based approaches in terms of computation time.

Acknowledgement This work is supported in part by the Ministry of Science and Technology of Taiwan via MOST103-2221-E-001-021-MY2 and NSC102-2221-E-001-005-MY2.

6. REFERENCES

[1] T. Ahonen, A. Hadid, and M. Pietikainen, “Face description with local binary patterns: Application to face recognition,” *IEEE PAMI*, 2006.

[2] J. Zou, Q. Ji, and G. Nagy, “A comparative study of local matching approach for face recognition,” *IEEE TIP*, 2007.

[3] S. Xie, S. Shan, X. Chen, and J. Chen, “Fusing local patterns of gabor magnitude and phase for face recognition,” *IEEE TIP*, 2010.

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

[5] L. Zhuang et al., “Single-sample face recognition with image corruption and misalignment via sparse illumination transfer,” in *Proc. IEEE CVPR*, 2013.

[6] C.-P. Wei and Y.-C. F. Wang, “Learning auxiliary dictionaries for undersampled face recognition,” in *Proc. IEEE ICME*, 2013.

[7] M. Yang, L. V. Gool, and L. Zhang, “Sparse variation dictionary learning for face recognition with a single training sample per person,” in *Proc. ICCV*, 2013.

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

[9] J. Yang, L. Zhang, Y. Xu, and J.-Y. Yang, “Beyond sparsity: The role of L_1-optimizer in pattern classification,” *Pattern Recognition*, 2012.

[10] Z. Wang, J. Yang, N. Nasrabadi, and T. Huang, “A max-margin perspective on sparse representation-based classification,” in *Proc. ICCV*, 2013.

[11] X. Chen and P. J. Ramadge, “Collaborative representation, sparsity or nonlinearity: What is key to dictionary based classification?” in *Proc. IEEE ICASSP*, 2014.

[12] M. Yang, L. Zhang, J. Yang, and D. Zhang, “Robust sparse coding for face recognition,” in *Proc. IEEE CVPR*, 2011.

[13] D.-A. Huang and Y.-C. F. Wang, “With one look: Robust face recognition using single sample per person,” in *ACM Multimedia*, 2013.

[14] A. M. Martinez and R. Benavente, “The AR face database,” *CVC Technical Report*, 1998.

[15] 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.

[16] M. Yang and L. Zhang, “Gabor feature based sparse representation for face recognition with Gabor occlusion dictionary,” in *Proc. ECCV*, 2010.

[17] 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.

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