Extended SRC: Undersampled Face Recognition via Intraclass Variant Dictionary

Weihong Deng, Jiani Hu, and Jun Guo

Abstract—Sparse Representation-Based Classification (SRC) is a face recognition breakthrough in recent years which has successfully addressed the recognition problem with sufficient training images of each gallery subject. In this paper, we extend SRC to applications where there are very few, or even a single, training images per subject. Assuming that the intraclass variations of one subject can be approximated by a sparse linear combination of those of other subjects, Extended Sparse Representation-Based Classifier (ESRC) applies an auxiliary intraclass variant dictionary to represent the possible variation between the training and testing images. The dictionary atoms typically represent intraclass sample differences computed from either the gallery faces themselves or the generic faces that are outside the gallery. Experimental results on the AR and FERET databases show that ESRC has better generalization ability than SRC for undersampled face recognition under variable expressions, illuminations, disguises, and ages. The superior results of ESRC suggest that if the dictionary is properly constructed, SRC algorithms can generalize well to the large-scale face recognition problem, even with a single training image per class.

Index Terms—Face recognition, sparse representation, undersampled problem, feature extraction.

1 Introduction

WITHIN the last two decades, face recognition systems were known to be critically dependent on discriminative feature extraction methods, such as Fisherfaces [1], [2], [3] and Laplacianfaces [4], [5], [6]. Of late, there has been a debate on the significance of feature extraction. Wright et al. have demonstrated that, once the test image can be approximated by a sparse linear combination of the training images, the choice of feature space is no longer critical [7]. This surprising claim is supported by the experimental results that Sparse Representation-Based Classification (SRC) with random projections-based features can outperform a number of conventional face recognition schemes, such as the nearest-neighbor classifier with Fisherfaces and Laplacianfaces-based features. It is commonly believed that SRC always requires a rich set of training images of each subject that can span the facial variation of that subject under testing conditions [8]. To fulfill this requirement, Wagner et al. [8] recently designed a system that acquires tens of images of each subject to cover all possible illumination changes. However, many important applications on law enforcement and homeland security can only offer a few, or even single, facial images per subject. This is often called the undersampled problem of face recognition, which has become one of the challenges in real-world applications.

In this paper, we propose an Extended Sparse Representation-Based Classifier (ESRC) for undersampled face recognition, which is effective even when there is only a single training image per subject. Taking advantage of the observation that the intraclass variability, caused by variable expressions, illuminations, and disguises, can be shared across different subjects, ESRC constructs an intraclass variant dictionary to represent the possible variation between the training and testing images. The recognition problem is cast as finding a sparse representation of the test image in terms of the training set as well as the intraclass variant bases, and the nonzero coefficients are expected to concentrate on the training samples with the same identity as the test sample and on the related intraclass variant bases. Fig. 1 illustrates this simple idea of ESRC to address the challenging face recognition problem despite disguise and side light.

Experimental results on the AR [9] and FERET [10] databases show that the usage of intraclass variant dictionary can largely improve the sparse representation-based face recognition accuracy. In most cases, ESRC can improve the accuracy of SRC by a margin as large as 5-40 percent. We empirically show that *adding the intraclass variant bases of 5-10 generic faces (outside the gallery) improves the recognition rate significantly*. On the feature representation of the proposed method, we make two valuable observations: 1) using local features, such as Gabor wavelet and Local Binary Pattern (LBP) instead of pixel feature can largely improve Sparse Representation-Based Performance; 2) for local features, dimension reduction such as random projections would lose useful information. In particular, Gabor feature-based classification using ESRC yields 99 percent accuracy on the AR data set, and LBP feature-based classification using ESRC achieves a 92.3 percent recognition rate on the most challenging FERET *dup2* probe set. These excellent results suggest that once the dictionary is properly constructed, SRC algorithms can generalize well to the large-scale face recognition problem, even when there is only a single training image per subject.

2 From SRC to ESRC

In this section, we first discuss the ability of SRC to address the densely sampled face recognition problem, and then present our intuition and algorithm that extend SRC to the undersampled problem.

2.1 Densely Sampled Problem: Dealing with Small Dense Noise

The densely sampled problem is defined as follows: Given sufficient training samples of the ith object class, any test sample from the same class will approximately lie in the linear span of the training samples associated with object i. Denote the training samples of all k classes as the matrix $A = [A_1, A_2, \ldots, A_k] \in \mathbb{R}^{d \times n}$, where the submatrix $A_i \in \mathbb{R}^{d \times n_i}$ stacks the training samples of class i. Then, the linear representation of a testing sample y can be rewritten as

$$y = Ax_0 + z, \tag{1}$$

where x_0 is a sparse vector whose entries are zeros except those associated with the ith class, and $z \in \mathbb{R}^d$ is a noise term with bounded energy $\|z\|_2 < \varepsilon$. The theory of compressed sensing reveals that if the solution of x_0 is sparse enough, it can be recovered efficiently by the following ℓ^1-minimization problem [11]:

$$\left(\ell_s^1\right) : \hat{x}_1 = \arg\min \|x\|_1, \ s.t. \ \|Ax - y\|_2 \leq \varepsilon. \tag{2}$$

Ideally, the nonzero entries in the estimate $\hat{x}_1$ will all be associated with the column of A from a single class.

Based on the compressed sensing theory, SRC has been successfully applied on densely sampled face recognition: In light of the theory of illumination cones [12], a small number of illuminations can linearly represent a wide range of illuminations, and the noise term z is reasonable to counteract the sensory noise and the non-Lambertian effect on the facial surface. SRC has achieved nearly perfect accuracy on the Extended Yale B database by sampling 32 differently illuminated images per subject [7]. Face

• *The authors are with the Pattern Recognition and Intelligent System Laboratory, School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, PO Box 186, Beijing 100876, China. E-mail: {whdeng, jnhu, guojun}@bupt.edu.cn.*

Manuscript received 17 Aug. 2011; revised 3 Jan. 2012; accepted 10 Jan. 2012; published online 16 Jan. 2012.
Recommended for acceptance by M. Tistarelli.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number TPAMI-2011-08-0561.
Digital Object Identifier no. 10.1109/TPAMI.2012.30.

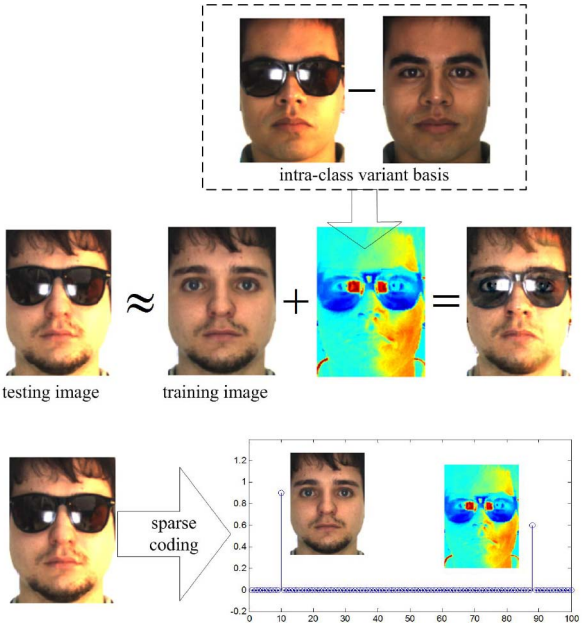


Fig. 1. The basic idea of ESRC. Top: The intraclass variant basis of a person can be shared by other people. For instance, a facial image with disguise and side light can be approximated by a natural (training) image of this person plus a variant basis image of another person. Bottom: The nonzero coefficients of the sparse representation are expected to concentrate on the training samples with the same identity as the test sample and on the related intraclass variant bases. Note that the intraclass variant bases can be acquired from any subject of the training set. By using the ℓ^1-minimization technique to find out the sparse representation, robust face recognition can be performed with a few, even a single, training images per person.

recognition across expression is also a densely sampled problem if the testing expressions, such as smile, anger, and scream, appear in the training images of the same identity. It has been evidenced by the excellent accuracy of SRC on the expressional faces of the AR database [7]. The disguise problem is solved by adding a complete set of single-pixel-based bases to the dictionary of SRC [7]. Recently, the uncontrolled illuminations, pose variations, and face alignment have been handled simultaneously by a deformable sparse recovery and classification algorithm [13].

2.2 Undersampled Problem: Dealing with Large Deviation

The undersampled problem is defined as follows: Given insufficient training samples of the ith object class, any test sample of the same class will largely deviate from the linear span of the training samples associated with object i. In the representation model (1), due to the lack of samples in A_i (we assume the test sample y belongs to class i), the noise term z becomes a large representation error. Since the assumptions of SRC are violated, the sparse representation $\hat{x}_1$ computed from the ℓ^1-minimization (2) is no longer useful for recognition. As evidence, a recent evaluation of SRC on the large-scale FERET database, where there is a single training image per subject, reported a recognition accuracy of only 20.3 percent on the *dup2* set [13]. In general, SRC is not designed for undersampled face recognition.

For instance, if the gallery set consists of a single natural image (such as a passport photograph) for each subject, the test images in real-world applications may contain complex variations of expressions, illuminations, and disguises. Since the test image deviates largely from the linear span of the correct gallery image, the nonzero coefficients of $\hat{x}_1$ in (2) would not concentrate on the correct gallery image. It is the significant difference between test image and gallery image that makes the sparse representation become uninformative. Fig. 2 illustrates some possible intraclass differences, displayed in downsampled image form, from four subjects of the AR database. One can see from the figure that the intraclass difference of the four

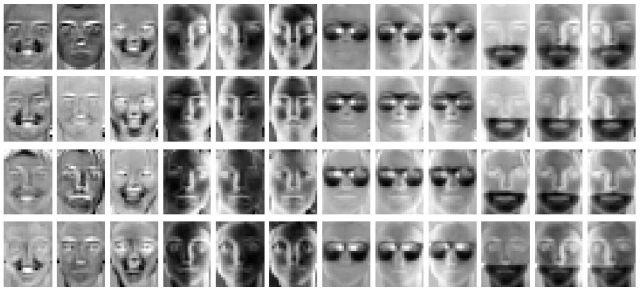


Fig. 2. Example of intraclass difference images of four subjects. Each row represents the bases of one subject which are simply generated by subtracting the (downsampled) natural image from the (downsampled) images with the variations of expressions, illuminations, and disguises.

subjects is similar since the shapes of human faces are highly correlated. Given a data set with a sufficiently large number of subjects, one can readily find these similarly shaped faces. Intuitively, we make the following assumption:

Assumption 1. *The intraclass variation of any gallery face can be approximated by a sparse linear combination of the intraclass differences from sufficient number of generic faces.*

Based on Assumption 1, the large deviation from the test image to the correct gallery images may be linearly approximated by the intraclass differences of generic subjects. If such an approximation could help compensate for the representation error z in model (1), it is possible that the ℓ^1-minimization (2) would become desirable again. It is this intuition that inspires us to extend SRC to undersampled face recognition by representing the intraclass variability by a set of universal bases for all subjects.

2.3 Extended Sparse Representation Classifier

Let a basis matrix D_I represent the universal intraclass variant bases. When the gallery data matrix A does not contain sufficient samples, the model (1) can be modified to account for large variation between the training and test images by writing

$$y = Ax_0 + D_I\beta_0 + z, \tag{3}$$

where the intraclass variant matrix D_I usually represents unbalanced lighting changes, exaggerated expressions, or occlusions that cannot be modeled by the small dense noise z. If there are redundant and overcomplete facial variant bases in D_I, the combination coefficients in β_0 are naturally sparse. Hence, the sparse representation x_0 and β_0 can be recovered simultaneously by ℓ^1-minimization.

Since, we assume that the intraclass variations of different subjects are sharable, the variant bases could be acquired either from the gallery samples themselves (if there are multiple samples per subject) or from the subjects outside the gallery. In general, the bases of dictionary D_I can be generated in various ways as long as they can reflect the intraclass difference.[1] For instance, given a data set with multiple images per subject, the m_i samples of subject i, stacked as vectors, form a matrix $D_i \in \mathbb{R}^{d\times m_i}$, $i = 1, \ldots, l$, $\sum_{i=1}^{l} m_i = m$. If there is a sample that is labeled as "natural" for each subject, the variant bases can be obtained by subtracting the natural image from other images of the same class:

1. In fact, the generic samples themselves can be used as the bases of the intraclass variant dictionary because any intraclass difference is a linear combination of them. The purpose of using class differences as bases is to remove the specific identity information and extract common expressions, illuminations, and occlusions (see Fig. 2 for examples) so that ℓ^1-minimization may yield more stable and informative results. As evidence, experiments show that the sample-differences yield superior performance to the raw samples for constructing intraclass variant dictionary (see Fig. 7 for details).

$$D_I^{(1)} = [D_1^- - a_1^* e_1, \ldots, D_l^- - a_l^* e_l] \in \mathbb{R}^{d\times(m-l)}, \quad (4)$$

where $e_i = [1, \ldots, 1] \in \mathbb{R}^{1\times(m_i-1)}$, a_i^* is the natural samples in class i, and D_i^- is the reduced data matrix of class i removing the natural sample. If the "natural" sample is not available, the intraclass variant bases could be calculated as follows:

$$D_I^{(2)} = [D_1 - c_1 e_1, \ldots, D_l - c_l e_l] \in \mathbb{R}^{d\times m}, \quad (5)$$

where $e_i = [1, \ldots, 1] \in \mathbb{R}^{1\times m_i}$, c_i is the class centroid of class i. In addition, the pairwise difference between samples can also be utilized to form intraclass variant bases. Let columns of matrix $P_i \in \mathbb{R}^{d\times[m_i(m_i-1)/2]}$ be the pairwise difference vectors between the samples of class i; an overcomplete variant dictionary could be constructed as follows:

$$D_I^{(3)} = [P_1, \ldots, P_l] \in \mathbb{R}^{d\times\sum_i[m_i(m_i-1)/2]}. \quad (6)$$

Based on the model (3), we propose an Extended Sparse Representation-Based Classification which casts the recognition problem as finding a sparse representation of the test image in term of the training set as well as the intraclass variant bases. The nonzero coefficients are expected to concentrate on the training samples with the same identity as the test sample and on the intraclass variant bases.

To illustrate how Algorithm 1 works, Fig. 3 shows the sparse coefficients of a test image with sunglasses using ESRC. In this example, we use the downsampled images of size 27×20 as features (see Section 3.2 for details). The first 80 coefficients, i.e., $\hat{x}_1$, correspond to 80 gallery images (with a single image per subject), and the remaining 120 coefficients, i.e., $\hat{\beta}_1$, correspond to the variant bases computed from 10 generic subjects. There are 10 variant bases, marked by red circles, which describe the differences between the generic faces with and without sunglasses. As expected, one can see from the figure that the test image with sunglasses is actually the sparse linear combination of the gallery image of the same identity (without sunglasses) and several intraclass variant bases related to sunglasses. Since the coefficients are sparse and the dominant coefficient is associated with subject six, the smallest residual corresponds to the correct subject. Besides the disguise case, similar working schemes have been observed on recognizing images under variable illuminations and expressions.

Algorithm 1. Extended Sparse Representation-Based Classification

1: **Input**: a matrix of training samples $A = [A_1, A_2, \ldots, A_k] \in \mathbb{R}^{d\times n}$ for k classes, a matrix of intraclass variant bases $D_I \in \mathbb{R}^{d\times p}$ (the dictionary size p depends on the data source and construction method), a test sample $y \in \mathbb{R}^d$, and an optimal error tolerance $\varepsilon > 0$.

2: Normalize the columns of A and D_I to have unit ℓ^2-norm.

3: Solve the ℓ^1-minimization problem

$$\begin{bmatrix}\hat{x}_1\\ \hat{\beta}_1\end{bmatrix} = \arg\min \left\| \begin{bmatrix} x\\ \beta\end{bmatrix}\right\|_1, \ s.t. \ \left\|[A, D_I]\begin{bmatrix} x\\ \beta\end{bmatrix} - y\right\|_2 \le \varepsilon \quad (7)$$

where $x, \hat{x} \in \mathbb{R}^n$, $\beta, \hat{\beta} \in \mathbb{R}^p$.

4: Compute the residuals

$$r_i(y) = \left\| y - [A, D_I]\begin{bmatrix}\delta_i(\hat{x}_1)\\ \hat{\beta}_1\end{bmatrix}\right\|_2, \quad (8)$$

for $i = 1, \ldots, k$, where $\delta_i(\hat{x}_1) \in \mathbb{R}^n$ is a new vector whose only nonzero entries are the entries in $\hat{x}_1$ those are associated with class i.

5: **Output**: $Identity(y) = \arg\min_i r_i(y)$.

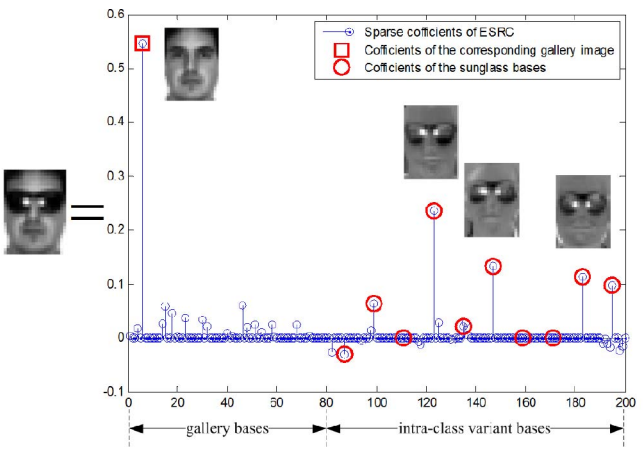


Fig. 3. Recognition with 27×20 downsampled images as features using ESRC. The test image belongs to subject six with sunglasses. The values of the sparse coefficients recovered from ESRC are plotted together with the four bases that correspond to the four largest sparse coefficients.

3 Experimental Results

In this section, we present experiments on publicly available databases for face recognition to demonstrate the efficacy of the proposed ESRC. For fair comparisons, both SRC and ESRC use the Homotopy[2] method [15], [11] to solve the ℓ^1-minimization problem with the error tolerance $\varepsilon = 0.05$ and identical parameters[3] so that the performance difference will be solely induced by the adoption of intraclass variant dictionary.

3.1 Recognition from Insufficient Training Samples

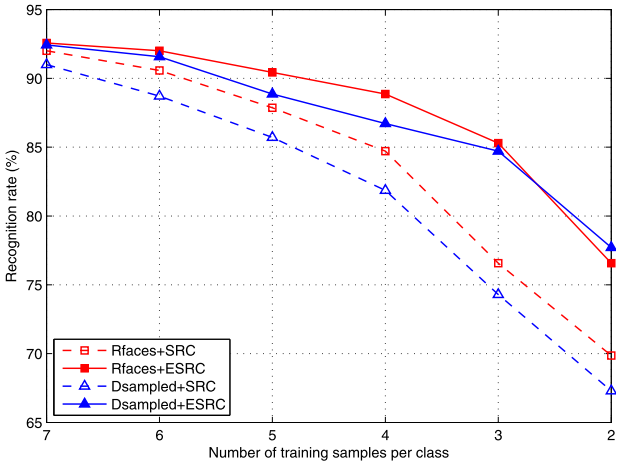
The first experiment is designed to test the hypothesis that given insufficient training images, ESRC can largely improve the generalization ability of SRC by exploiting the observation that intraclass variation can be shared across different subjects. Specifically, we use the AR database, which consists of over 4,000 frontal images for 126 individuals [9]. In the experiment, we chose a subset of the data set consisting of 50 male subjects and 50 female subjects, and the images are cropped with dimension 165×120. For each subject, 14 images with only illumination change and expression are selected: the seven images from Session 1 for training and the other seven from Session 2 for testing. Gabor feature is effective to improve the conventional face recognition algorithms [16], [17]; we therefore apply it to SRC/ESRC to test whether similar improvement can be achieved. The images are first resized to a resolution of 128×128, and then the 10,240D Gabor feature vector is extracted according to [3]. We selected a dimension of 540 for Pixel and Gabor-based randomfaces, and a resolution of 27×20 for downsampled images.

To test the undersampled effect, we reduce the number of training samples per class one by one from seven to two. The intraclass variant dictionary of ESRC is computed from the training samples according to (5). Fig. 4 shows the comparative performance of SRC and ESRC. As expected, SRC deteriorates rapidly as the number of training images decreases. In all 24 test cases (4 features $\times$ 6 sample sizes), ESRC performs better than SRC. In general, the superiority of ESRC becomes more and more significant as the sample size decreases.

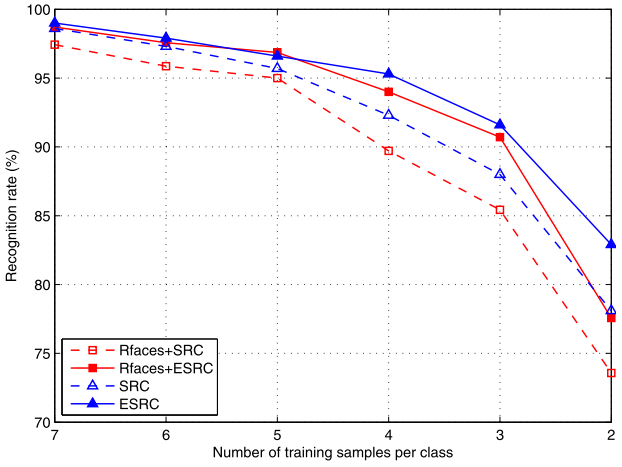
By globally comparing Figs. 4a with 4b, one can immediately find that Gabor feature-based classification is more accurate than

2. This optimization method had the best accuracy and fastest speed on the comparative study in [14], and its source code was downloaded at http://www.eecs.berkeley.edu/~yang/software/l1benchmark/.

3. To remove the bias caused by random number generator, we settle the initial state to 0 for all random data partitions and all the random projections of Randomfaces.



(a) Pixel feature based classification



(b) Gabor feature based classification

Fig. 4. The comparative recognition rates of the AR data set as the number of training images per class decreases.

pixel-based classification for both SRC and ESRC, which is consistent with the results found in the studies of the conventional algorithms [16], [17]. In particular, as shown in Fig. 4b, Gabor feature-based classification using ESRC achieves 99 percent recognition rate when all seven training images are available. This may be because Gabor feature is invariant to illumination change, image misalignment ,and expressional distortion, which makes the facial images of each class constrained to a linear subspace.

On the other hand, the accuracy of the best performed ESRC drops to about 83 percent using two training samples per class, which indicates that the intraclass variant dictionary from a few gallery images themselves brings about a limited effect. For satisfactory performance on undersampled face recognition, one could resort to the dictionary constructed from the generic subjects with sufficiently sampled images, which will be tested in the following experiment.

3.2 Recognition from One Single Training Sample

The second experiment is designed to test the robustness of ESRC against more severe intraclass variability using a single training sample per subject. In this experiment, we chose a random subset of the AR data set consisting of 80 subjects. For each subject, 13 images in Session 1 were selected: the single image with natural expression and illumination for training, the other 12 images with illumination change, expressions and facial disguises for testing. The images are cropped with dimension 165×120 and converted to gray scale. We selected a feature dimension of 540 for Pixel and

Fig. 5. The cropped images of one person of the AR data set. The single natural image is used for training, while the other 12 images with severe variation are used for testing.

Gabor-based randomfaces, and a resolution of 27×20 for downsampled images.

Fig. 5 shows the 13 images of one subject in this test. As there is only single natural training image, the tested classifiers need to simultaneously handle the variations in expression, illumination, and disguise of the testing images. To construct the intraclass variant dictionary for ESRC, another 20 subjects (not overlapping with the 80 testing subjects) are selected, also with 13 images per subject. The intraclass variant dictionary contains 240 bases (12 bases for each generic subject), which are computed by (4). Fig. 2 has displayed the bases (in downsampled image form) of four subjects. During the running of ESRC, we find that the number of nonzero coefficients, i.e., $\|\hat{\beta}_1\|_0$, ranges from about 20 to 60 out of the total 240 coefficients, which validates Assumption 1 about the sparse linear combination of intraclass differences.

Table 1 enumerates the recognition error rates for this experiment, and one can see from the table that the error reduction by switching SRC to ESRC is significant for all five types of features. Further, we define an Error Reduction Rate (ERR), denoted by a notion $\downarrow$, to characterize the proportion of the errors reduced by switching SRC to ESRC. For instance, since the downsampled image-based ESRC reduces the error rate from 43.44 to 11.98 percent, the ERR is $\downarrow$72.42 percent [(43.44-11.98)/43.44], suggesting that 72.42 percent recognition errors can be avoided by using ESRC instead of SRC. For the five types of tested features, the ERR is about $\downarrow$60-$\downarrow$78 percent, strongly proving the effectiveness of ESRC. In particular, only 5 percent error rate is achieved by Gabor feature-based classification using ESRC.

As illustrated in Fig. 5, these test images of this experiment contain four variabilities: expression, illumination, disguise, and disguise+illmination. To better understand the effects of ESRC, Table 2 separately enumerates the error rates of the four test variabilities. Across all five tested feature types, the error reduction rates for illumination and disguise ($\downarrow$63-$\downarrow$100 percent) are notably higher than those for expression ($\downarrow$28-$\downarrow$50 percent). The relatively low ERRs for expression indicate that the expression change is more sensitive to the specific facial shape of different subjects than the illumination and disguise, and it is more difficult to sparsely

TABLE 1
Comparative Error Rates of SRC and ESRC
on the AR Database Using a Single Training Sample per Person

Feature	Dim	Error Rate (%)		ERR
		SRC	ESRC	
Dsampled	27×20	43.44	11.98	$\downarrow$72.42%
Pixel-Rfaces	540	40.10	16.04	$\downarrow$60.00%
Pixel	19800	40.31	10.63	$\downarrow$73.65%
Gabor-Rfaces	540	28.96	11.46	$\downarrow$60.43%
Gabor	10240	22.71	**5.00**	$\downarrow$77.98%

TABLE 2
Comparative Error Rates of SRC and ESRC on the AR Database Using a Single Training Sample per Person

	Feature	Dsampled Image	Pixel-Rfaces	Pixel	Gabor-Rfaces	Gabor
Variablity	Dim	27×20	540	19800	540	10240
Expression	SRC	14.2	17.1	17.9	15.0	11.7
	ESRC	7.5 (↓47%)	15.4 (↓10%)	10.8 (↓40%)	10.8 (↓28%)	**5.8** (↓50%)
Illumination	SRC	19.2	16.3	14.6	3.8	0.8
	ESRC	1.3 (↓93%)	2.1 (↓87%)	0.8 (↓95%)	1.3 (↓66%)	**0.0** (↓100%)
Disguise	SRC	56.3	46.9	48.1	36.3	28.1
	ESRC	14.4 (↓74%)	16.9 (↓64%)	10.6 (↓78%)	7.5 (↓79%)	**5.6** (↓80%)
Disguise+ Illumination	SRC	77.2	71.9	72.5	54.7	44.7
	ESRC	22.2 (↓71%)	26.6 (↓63%)	17.8 (↓75%)	21.6 (↓61%)	**7.8** (↓83%)

represent the expressional variations. In general, our ESRC methods provide a novel and unified solution on the four variabilities.

Even with this excellent performance, an interesting question remains: How many generic subjects are needed to construct the intraclass variant dictionary? Fig. 6 shows a plot of error rate versus the number of generic persons in the dictionary. For all kinds of features, the intraclass variant bases of a small number of subjects are sufficient to largely improve error rate. In particular, the intraclass variant bases of a single generic subject reduce the error rate from 43.44 to 23.85 percent, using downsampled image-based ESRC. This finding suggests that, once the intraclass variant bases are properly designed according to the testing condition, a dictionary of 5-10 subjects is enough to dramatically boost face recognition performance.

The final test of this experiment evaluates several options for computing the bases of the intraclass dictionary: difference to natural image ($D_I^{(1)}$), difference to the class centroid ($D_I^{(2)}$), pairwise difference ($D_I^{(3)}$), and original generic samples themselves. Previous tests show that 10 generic subjects are enough to improve performance significantly; we therefore use the 130 images of these generic subjects to compute the dictionaries. Fig. 7 shows that the algorithms with the three sample difference-based dictionaries perform almost equally, which is notably better than the one with raw generic data. The similar performance of the first three dictionaries indicates that ℓ^1-minimization is effective to recover the sparse combination regardless of the dictionary size. Certainly, because $D_I^{(1)}$ and $D_I^{(2)}$ contain much fewer bases than $D_I^{(3)}$, one should use the former two for the computational efficiency of ESRC.

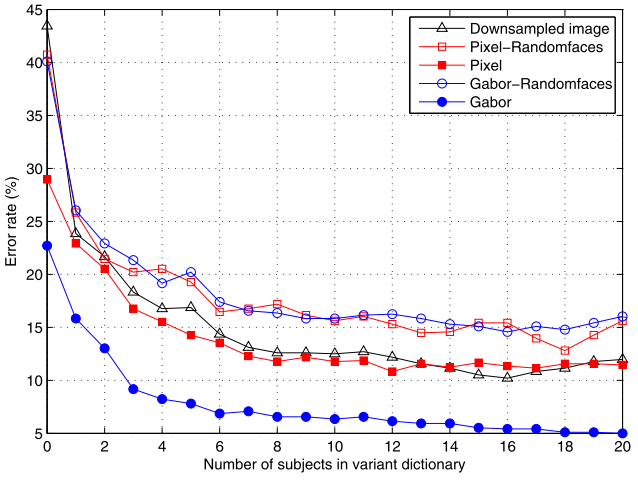


Fig. 6. The recognition error rate of ESRC as a function of the number of subjects in the intraclass variant dictionary.

3.3 Large-Scale Recognition Despite Complex Variation

The last experiment is designed to test the robustness of ESRC against complex facial variation in the real-world applications. The experiment follows the standard data partitions of the FERET database.

- *Generic training set* contains 1,002 images of 429 people, which are listed in the FERET standard training CD.
- *Gallery training set* contains 1,196 images of 1,196 people.
- *fb probe set* contains 1,195 images taken with an alternative facial expression.
- *fc probe set* contains 194 images taken under different lighting conditions.
- *dup1 probe set* contains 722 images taken in a different time.
- *dup2 probe set* contains 234 images taken at least a year later, which is a subset of the *dup1* set.

Note that the intraclass variability of the FERET database is more difficult to represent than those of the AR database since those of the AR data set are taken in a single laboratory circumstance, but the FERET database is acquired in multiple sessions over several

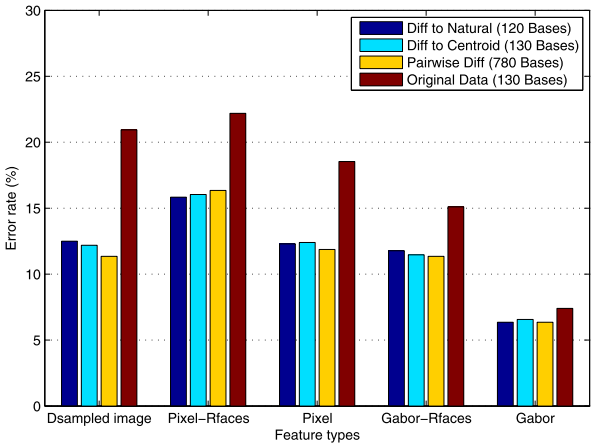


Fig. 7. The recognition error rate of ESRC using differently constructed intraclass variant dictionaries of 10 generic subjects.

Fig. 8. The cropped images of one person from the FERET database.

TABLE 3
Comparative Recognition Rates of SRC and ESRC on the FERET Database Using the FERET'96 Testing Protocol

	Feature	Dsampled Image	Pixel-Rfaces	Pixel	Gabor-Rfaces	Gabor	LBP-Rfaces	LBP
Probe set	Dim	24×24	540	16384	540	10240	540	15104
fb	SRC	86.4	82.4	85.3	89.5	92.8	91.5	96.7
	ESRC	94.8(+8.4)	91.5(+9.1)	92.8(+7.5)	94.1(+4.6)	**97.3**(+4.5)	95.2(+3.7)	**97.3**(+0.6)
fc	SRC	69.6	75.8	76.3	96.4	97.4	72.7	93.3
	ESRC	67.5(–2.1)	78.9(+3.1)	79.4(+3.1)	96.9(+0.5)	**99.0**(+1.6)	71.1(–1.6)	95.4(+2.1)
dup1	SRC	62.7	60.9	63.7	63.0	72.7	75.2	87.7
	ESRC	75.6(+12.9)	73.1(+12.2)	77.0(+13.3)	73.5(+10.5)	85.0(+12.3)	81.0(+5.8)	**93.8**(+6.1)
dup2	SRC	52.6	53.0	55.6	70.1	76.5	69.7	83.8
	ESRC	62.4(+9.8)	59.8(+6.8)	66.2(+10.6)	72.6(+2.5)	85.9(+9.4)	71.4(+1.7)	**92.3**(+8.5)

years. The image is first normalized by an affine transformation that sets the centered intereye line horizontal and 70 pixel apart, and then cropped to the size of 128×128 with the centers of the eyes[4] located at (29, 34) and (99, 34) to extract the pure face region. No further preprocessing procedure is carried out in our experiments, and Fig. 8 shows some cropped images which are used in our experiments.

For comprehensive results, we compare ESRC with SRC by the classification of eight types of features, as listed in Table 3. For detailed procedures of Gabor and LBP feature extraction please refer to [17]. The intraclass dictionary of ESRC is computed by (5) using the generic training set. Out of the 28 test cases (7 features $\times$ 4 probe sets), ESRC raises the recognition rates in 26 cases. The best recognition rate in each probe set is achieved by ESRC. Specifically, Gabor feature-based ESRC, with 99 percent accuracy on the *fc* set, is the best to handle the illumination changes, while LBP feature-based ESRC, with 93.8 percent accuracy on the *dup1* set and 92.3 percent accuracy on *dup2* set, is expert in addressing aging effects. The LBP feature is also robust to the registration error [18], which is not explicitly tested in our experiment. Across all tested features, the boost on recognition accuracy is significant on the *dup1* and *dup2* image sets, which are acquired in uncontrolled settings that are close to real-world conditions, indicating that the intraclass variability of face is sharable even in a complex situation. Compared to the performance enhancement on the AR data set (Table 2), the enhancement by ESRC is not significant on the *fc* set, which may be because the generic training set of FERET does not contain sufficient illumination variations. Another interesting finding is that, for both SRC and ESRC, using original high-dimensional local feature vector instead of a random projections (Randomfaces)-based one indeed increases the accuracy by over 10 percent on *dup1* and *dup2* set.

Finally, we discuss some computational issues in this large-scale experiment. We implement the ESRC algorithm using 64-bit Matlab platform on a PC with Dual Core 2.93 GHz Pentium CPU and 4 GB memory. In this experiment, a simple heuristic is applied to accelerate the residual-based classification: only compute the residuals for the 10 classes associated with the largest entries in $\hat{x}_1$. This heuristic approach speeds up the classification by a factor of 120 without any loss of accuracy.[5] With this acceleration, ESRC (including both ℓ^1-minimization and classification) takes only 1.2 seconds (on average) per test image using the 540D LBP-Rfaces feature, and 13.1 seconds (on average) per test image using the 15,104D LBP feature.

4. We have found that there were slight errors on the eye coordinates of the standard FERET distribution and remarked the eye coordinates of all the FERET images accurately. To reproduce our experimental results, the updated eye-coordinate file is available upon request. Note that our (manual) accurate face alignment makes our SRC results on FERET significantly different from Wagner et al.'s [13], where the system aligns the faces in a fully automated way.

5. For all probe images, we have validated that the smallest residual of the 1,196 classes is from one of the 10 classes with largest coefficients in $\hat{x}_1$. Therefore, this heuristic approach significantly accelerates the computation, but does not affect the accuracy of ESRC.

4 Conclusion

The experiments suggest a number of conclusions:

1. When the training images of each class are insufficient to linearly represent the testing variability, ESRC raises the recognition rates of SRC by using the intraclass variant dictionary. The superiority of ESRC appears to be more significant as the number of training images decreases.
2. In the limit with a single training image per subject, ESRC still works effectively and generalizes well to large-scale databases using the intraclass variant dictionary constructed from generic subjects that are not in the gallery set.
3. If the generic images sufficiently cover the testing conditions, adding the intraclass variant bases of 5-10 generic classes to the intraclass variant dictionary improves recognition rate significantly.
4. For both SRC and ESRC, local features such as Gabor-based features and LBP features yield much better recognition rates than pixel-based features, which suggests that their invariant properties make the samples of each class more constrained to a linear subspace.
5. When the training images of each class are insufficient, dimension reduction of local features, such as random projections, would lose useful information for both SRC and ESRC.

Although we have shown that directly adding sample differences as dictionary bases can improve SRC significantly, a well-learned dictionary matrix may lead to higher performance with a smaller number of bases [19]. We are studying ways to learn universal intraclass variant dictionaries for unconstrained face recognition. Furthermore, rejecting imposters is more challenging than identifying the correct gallery subjects in face-recognition practice [20], and we are working toward the undersampled open-set recognition via sparse representation.

Acknowledgments

The authors would like to thank the anonymous reviewers for their thoughtful and constructive remarks that are helpful to improve the quality of this paper. This work was supported by the National Natural Science Foundation of China (NSFC) under Grant Nos. 61002051 and 61005025, Major Project of National Science and Technology under Grant No. 2011ZX03002-005-01,

and the Fundamental Research Funds for the Central Universities under Grant Nos. 2011RC0102, 2009RC0106, and 2011RC0115.

REFERENCES

[1] P.N. Belhumeour, J.P. Hespanha, and D.J. Kriegman, "Eigenfaces versus Fisherfaces: Recognition Using Class Specific Linear Projection," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 19, no. 7, pp. 711-720, July 1997.

[2] D.L. Swets and J.J. Weng, "Using Discriminant Eigenfeatures for Image Retrieval," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 18, no. 8, pp. 831-836, Aug. 1996.

[3] C. Liu and H. Wechsler, "Gabor Feature Based Classification Using the Enhanced Fisher Linear Discriminant Model for Face Recognition," *IEEE Trans. Image Processing,* vol. 11, no. 4, pp. 467-476, Apr. 2002.

[4] X. He, S. Yan, Y. Hu, P. Niyogi, and H.-J. Zhang, "Face Recognition Using Laplacianfaces," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 27, no. 3, pp. 328-340, Mar. 2005.

[5] X. He and P. Niyogi, "Locality Preserving Projections," *Proc. Conf. Advances in Neural Information Processing System,* 2003.

[6] W. Deng, J. Hu, and J. Guo, "Comments on 'Globally Maximizing, Locally Minimizing: Unsupervised Discriminant Projection with Applications to Face and Palm Biometrics'," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 30, no. 8, pp. 1503-1504, Aug. 2008.

[7] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma, "Robust Face Recognition via Sparse Representation," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 31, no. 2, pp. 210-227, Feb. 2009.

[8] A. Wagner, J. Wright, A. Ganesh, Z. Zhou, and Y. Ma, "Towards a Practical Face Recognition System: Robust Registration and Illumination by Sparse Representation," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* pp. 597-604, 2009.

[9] A.M. Martinez and R. Benavente, "The AR Face Database," CVC Technical Report #24, June 1998.

[10] P.J. Phillips, H. Moon, P. Rizvi, and P. Rauss, "The Feret Evaluation Method for Face Recognition Algorithms," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 22, no. 10, pp. 1090-1104, Oct. 2000.

[11] D. Donoho and Y. Tsaig, "Fast Solution of ℓ_1-Norm Minimization Problems When the Solution May Be Sparse," *IEEE Trans. Information Theory,* vol. 54, no. 11, pp. 4789-4812, Nov. 2008.

[12] A. Georghiades, P. Belhumeur, and D. Kriegman, "From Few to Many: Illumination Cone Models for Face Recognition under Variable Lighting and Pose," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 23, no. 6, pp. 643-660, June 2001.

[13] A. Wagner, J. Wright, A. Ganesh, Z. Zhou, H. Mobahi, and Y. Ma, "Toward a Practical Face Recognition System: Robust Alignment and Illumination by Sparse Representation," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 34, no. 2, pp. 372-386, Feb. 2012.

[14] A. Yang, A. Ganesh, S. Sastry, and Y. Ma, "Fast l1-Minimization Algorithms and an Application in Robust Face Recognition: A Review," Technical Report UCB/EECS-2010-13, Univ. of California, 2010.

[15] M. Osborne, B. Presnell, and B. Turlach, "A New Approach to Variable Selection in Least Squares Problems," *IMA J. Numerical Analysis,* vol. 20, no. 3, pp. 389-403, 2000.

[16] W. Deng, J. Hu, J. Guo, W. Cai, and D. Feng, "Emulating Biological Strategies for Uncontrolled Face Recognition," *Pattern Recognition,* vol. 43, no. 6, pp. 2210-2223, 2010.

[17] W. Deng, J. Hu, J. Guo, W. Cai, and D. Feng, "Robust, Accurate and Efficient Face Recognition from a Single Training Image: A Uniform Pursuit Approach," *Pattern Recognition,* vol. 43, no. 5, pp. 1748-1762, 2010.

[18] C. Chan and J. Kittler, "Sparse Representation of (Multiscale) Histograms for Face Recognition Robust to Registration and Illumination Problems," *Proc. IEEE 17th Int'l Conf. Image Processing,* pp. 2441-2444, 2010.

[19] M. Elad and M. Aharon, "Image Denoising via Sparse and Redundant Representations over Learned Dictionaries," *IEEE Trans. Image Processing,* vol. 15, no. 12, pp. 3736-3745, Dec. 2006.

[20] W. Deng, J. Guo, and J. Hu, "Comment on '100 Percent Accuracy in Automatic Face Recognition'," *Science,* vol. 321, no. 5891, p. 912, 2008.