In Defense of Sparsity Based Face Recognition

Weihong Deng, Jiani Hu, Jun Guo
Beijing University of Posts and Telecommunications,
Beijing, 100876, China
{whdeng, jnhu, guojun}@bupt.edu.cn

Abstract

The success of sparse representation based classification (SRC) has largely boosted the research of sparsity based face recognition in recent years. A prevailing view is that the sparsity based face recognition performs well only when the training images have been carefully controlled and the number of samples per class is sufficiently large. This paper challenges the prevailing view by proposing a "prototype plus variation" representation model for sparsity based face recognition. Based on the new model, a Superposed SRC (SSRC), in which the dictionary is assembled by the class centroids and the sample-to-centroid differences, leads to a substantial improvement on SRC. The experiments results on AR, FERET and FRGC databases validate that, if the proposed prototype plus variation representation model is applied, sparse coding plays a crucial role in face recognition, and performs well even when the dictionary bases are collected under uncontrolled conditions and only a single sample per classes is available.

1. Introduction

The sparse representation-based classification (SRC) algorithm for face recognition was introduced by Wright *et al*. in a highly-cited papers [15]. The key idea of that paper is a judicious choice of dictionary: representing the test image as a sparse linear combination of the the training images themselves. Motivated by the conditional equivalence of the sparsity measured by ℓ_0 norm and ℓ_1 norm [4], the efficient ℓ_1-minimization technique was applied to find the sparsest coding vector. Finally, the test sample is classified by checking which class yields minimum representation error.

The success of SRC has largely boosted the research of sparsity based face recognition. Huang *et al*. introduced a transformation-invariant SRC for face recognition [5]. Zhou *et al*. combined SRC with markov random fields to recognize the disguise face with large contiguous occlusion [20]. Wagner *et al*. extended the SRC framework to simultaneously handle the mis-alignement, pose and illumination invariant recognition problem [12]. Based on the sparsity assumption, Zhang *et al*. applied the sparse coding to jointly address blind image restoration and blurred face recognition [18]. Yang *et al*. introduced a discriminative dictionary learning method to improve the accuracy and efficiency of face recognition [17].

Despite its simplicity and effectiveness, SRC has often been criticized for being excessively sensitivity on the quality and quantity of training samples, as stated in the review paper [14] by Wright *et al*.

> The sparse representation based face recognition assumes that the training images have been carefully controlled and that the number of samples per class is sufficiently large. Outside these operating conditions, it should not be expected to perform well. [14]

It is the purpose of this paper to challenge the common view that the sparsity based face recognition is inadequate with the uncontrolled training samples. The inferior performance of SRC can properly be traced to the training samples based dictionary that do not distinguish the class-specific prototype and the intra-class variation. It is shown in this paper that a simple variant of SRC, which represents the test sample as a sparse linear combination of the class centroid and the differences to the class centroid, leads to an enormous improvement under the uncontrolled training conditions. The added complexity of the algorithm is trivial. Our experimental results on the AR, FERET, and FRGC databases validate that, if the proposed prototype plus variation representation model is applied, sparse coding plays a crucial role in face recognition, and performs well even when the dictionary bases are collected under uncontrolled conditions and only a single sample per classes is available.

2. The Debates on SRC

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 sub-matrix $A_i \in \mathbb{R}^{d\times n_i}$ 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 [4]:

$$\hat{x}_1 = \arg\min_x \|Ax - y\|_2^2 + \lambda\|x\|_1 \tag{2}$$

where λ is a trade-off parameter between sparsity and reconstruction. Ideally, the nonzero entries in the estimate $\hat{x}_1$ will all be associated with the column of A from a single class.

2.1. Is the ℓ_1-norm sparsity useful?

A recent paper of Shi *et al.* [11] suggested that the ℓ_1 norm regularization is not useful for face recognition, and the computational ℓ_1-minimization problem can be simplified to the well-established least-square approximation problem

$$\min_x \|Ax - y\|_2^2 \tag{3}$$

where the objective is the sum of the squares of residuals. The solution to this problem is given by the so-called normal equations

$$A^T Ax = A^T y \tag{4}$$

If the columns of A are independent, the least-square approximation problem has the unique solution

$$x_2 = (A^T A)^{-1} A^T y \tag{5}$$

In [11], the experimental results on EYB and AR databases showed that a simple ℓ_2 approach by (5) is significantly more accurate than SRC, and thus concluded that the ℓ_1-norm regularization did not deliver the robust or performance desired. However, Wright *et al.* [15] clarified that the comparison is unfair for SRC: the ℓ_2 approach was based on the matrix A of 19800-dimensional measurements (i.e. the original image), but SRC relied on a reduced matrix A of 300-dimensional measurements (derived by random projections) for the ℓ_1-minimization. When the two methods are compared on a fair footing with the same number of observation dimensions, the usefulness of ℓ_1-minimization become apparent.

It should be noted that the robustness of ℓ_1-minimization is empirically justified only in the cases when the training images have been carefully controlled [15]. Once the training images contain contaminant, the low-dimensional linear models assumption of SRC would easily break down. As evidence, in Section 5 of [11] the training images in A are randomly selected from the AR database regardless of their nature, the simple ℓ_2 approach indeed outperform SRC. The experiments in [2] also found that SRC tended to recognize test images to the class with the same type of corruption. Low-rank matrix recovery technique was applied to recover the clean training images [2][6], but the performance improvement is limited.

2.2. Is the ℓ_1-norm sparsity crucial?

The discussion between Shi *et al.* [11] and Wright *et al.* [13] clarified that imposing the ℓ_1-norm sparsity is useful for face recognition, but did not confirm its necessity: Can the ℓ_1-regularization be replaced by other types of regularization? Recently, Zhang *et al.* [19] propose to replace the ℓ_1-norm regularization of SRC with the ℓ_2-norm regularization, which results in a (convex) quadratic optimization problem:

$$\hat{x}_2 = \arg\min_x \|Ax - y\|_2^2 + \lambda\|x\|_2^2 \tag{6}$$

where the parameter λ is chosen by the user to give the right trade-off between making the square error $\|Ax-y\|_2^2$ small, while keeping the ℓ_2-norm $\|x\|_2^2$ not too big. This regularized least-norm problem has the analytical solution

$$\hat{x}_2 = (A^T A + \lambda I)^{-1} A^T y \tag{7}$$

Since $A^T A + \lambda I \succ 0$ for any $\lambda > 0$, the regularized least-squares solutions requires no rank (or dimension) assumptions on matrix A. In an estimation setting, the regularization term penalizing large $\|x\|_2^2$ can be interpreted as our prior knowledge that $\|x\|_2^2$ is not too large.

The controlled experiments on EYB and AR databases [19] first reduced the dimensionality to $O(10^2)$ by PCA such that the linear equation $Ax = y$ become underdetermined. Under the underdetermined condition, ℓ_1-norm and ℓ_2-norm regularizations were fairly compared, and the results showed that the ℓ_2-regularized method, called collaborative representation based classification (CRC), had very competitive face recognition accuracy to the ℓ_2-regularized method (SRC), but with much lower complexity. Based on their results, the sparsity based face recognition seems to be useful, but not necessary.

As suggested in [19], if over-complete dictionaries are available for representing each class, the sparse solution by ℓ_1-norm regularization is arguably more discriminative than the dense solution of ℓ_2-norm. It is possible that the sample size per class used in [19] is still not enough to directly ensemble a over-complete dictionary for the face recognition problem. Wagner *et al.* [12] have proposed a system to collect sufficient samples for SRC, but it is still difficult for most real-world applications to acquire such number of samples. How to design an over-complete dictionary with limited sample size per class is an essential problem for sparsity based face recognition.

3. Prototype plus Variation Model and Algorithm

The previous studies in [11][13][19] have revealed the limitations of sparsity based recognition when the training images are corrupted and the number of samples per class is insufficient. This section introduces a prototype plus variation (P+V) model and a corresponding sparsity based classification algorithm to address these limitations of SRC.

3.1. Signal = Prototype + Variation

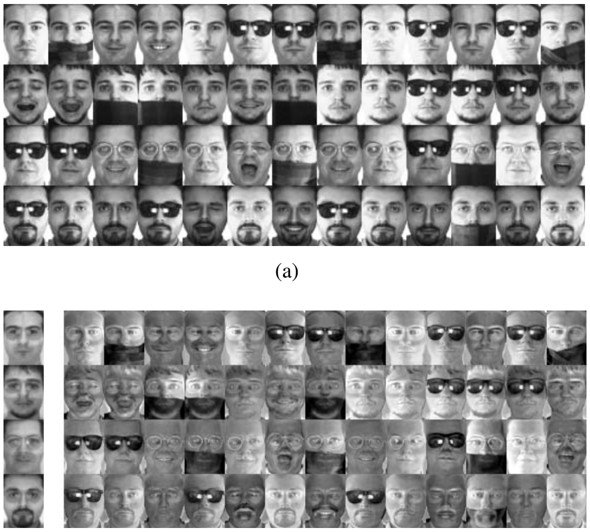
We assume that the observed signal is a superposition of two different sub-signals y_p, y_v and noise z (i.e. $y = y_p + y_v + z$). We further assume that y_p is sparsely generated using the model with a *prototype dictionary* (matrix) $P = [P_1, P_2, \ldots, P_k] \in \mathbb{R}^{d\times m}$, where the sub-matrix $P_i \in \mathbb{R}^{d\times m_i}$ stacks the m_i prototypical bases of class i. Similarly, y_v is sparsely generated using the model with a *variation dictionary* (matrix) $V \in \mathbb{R}^{d\times q}$ represents the universal intra-class variant bases, such as the unbalanced lighting changes, exaggerated expressions, or occlusions that cannot be modelled by the small dense noise z. Then, the linear representation of a testing sample y can be rewritten as

$$y = P\alpha_0 + V\beta_0 + z \tag{8}$$

where we assume that the samples from the class i are formed by taking the same sparse linear combination α_0 with nonzero elements corresponding to P_i, but a different variation term β_0 that represents the style of this face: it describes systematic contribution to the image from uncontrolled viewing conditions. Note that β_0 differs for each view condition and so it tells us nothing about identity. If the number of classes k is reasonably large, the combination coefficients in α_0 is naturally sparse. If there are redundant and overcomplete facial variant bases in V, the combination coefficients in β_0 are naturally sparse. Hence, the sparse representation α_0 and β_0 can be recovered simultaneously by ℓ_1-minimization.

In general, P+V model has two advantages over the traditional sparse model in (3):

- P+V model improves the robustness against the contaminative training samples. By separating the image contaminations to the variation matrix that is shared by all classes, the class-specific prototypes would become clean and natural, and thus the classification would not be deteriorated by the corrupted training sample.
- P+V model requires less samples per class to construct an over-complete dictionary. As the variation matrix is shared by all classes, the dictionary size of the class i is expanded from m_i to $m_i + q$. Once q is sufficiently large, the overcomplete dictionary for each class can be readily constructed.



(a)

(b)

Figure 1. The illustrative examples of the prototype plus variation (P+V) model. (a) the randomly selected training images from AR database. (b) the first column contains the “prototypes” derived by averaging the images of the same subject, and the rest columns are the “sample-to-centroid” variation images.

3.2. A Superposed SRC Algorithm

To show the strength of the P+V model, we propose a very simple classification algorithm according to this model and demonstrate its effectiveness on face recognition under uncontrolled training conditions. Given a data set with multiple images per subject, the n_i samples of subject i, stacked as vectors, form a matrix $A_i \in \mathbb{R}^{d\times n_i}$,$i = 1, \ldots, k$, $\sum_{i=1}^{k} n_i = n$. The prototype matrix can be represented as follows

$$P = [c_1, \ldots, c_i, \ldots, c_k] \in \mathbb{R}^{d\times k} \tag{9}$$

where $c_i = \frac{1}{n_i} A_i e_i$ is the geometric centroid of class i, and $e_i = [1, \ldots, 1]^T \in \mathbb{R}^{n_i\times 1}$. As the prototypes are represented by class centroids, the variation matrix is naturally constructed by the sample based difference to the centroids as follows:

$$V = [A_1 - c_1 e_1^T, \ldots, A_k - c_k e_k^T] \in \mathbb{R}^{d\times n} \tag{10}$$

where c_i is the class centroid of class i. Fig. 1 illustrates an typical examples of the prototype and variation matrices. When number of samples per class is insufficient, and in particular when only a single sample per class is available, the intra-class variation matrix would become collapsed. To address this difficult, one can acquire the variant bases from the generic subjects outside the gallery, as the P+V model has assumed that the intra-class variations of different subjects are sharable.

Based on the P+V model in (8), we further propose an Superposed Sparse Representation-based Classification

(SSRC) which casts the recognition problem as finding a sparse representation of the test image in term of a superposition of the class centroids as and the intra-class differences. The nonzero coefficients are expected to concentrate on the centroid of the same class as the test sample and on the related intra-class differences.

Algorithm 1. Superposed Sparse Representation based Classification (SSRC)

1: **Input**: a matrix of training samples $A = [A_1, A_2, \ldots, A_k] \in \mathbb{R}^{d \times n}$ for k classes, and an regularization parameter $\lambda > 0$. Compute the prototype matrix P according to (9), and the variation matrix V according to (10). When the sample size per class is insufficient, the matrix V can be computed from a set of generic samples outside the gallery.

2: Derive the projection matrix $\Phi \in \mathbb{R}^{d \times p}$ by applying PCA on the training samples A, and project the prototype and variation matrices to the p-dimensional space.

$$P \leftarrow \Phi^T P,\ V \leftarrow \Phi^T V \qquad (11)$$

3: Normalize the columns of P and V to have unit ℓ_2-norm, and solve the ℓ_1-minimization problem

$$\begin{bmatrix} \hat{\alpha}_1 \\ \hat{\beta}_1 \end{bmatrix} = \arg\min \left\| [P, V] \begin{bmatrix} \alpha \\ \beta \end{bmatrix} - y \right\|_2^2 + \lambda \left\| \begin{bmatrix} \alpha \\ \beta \end{bmatrix} \right\|_1, \qquad (12)$$

where $\alpha, \hat{\alpha} \in \mathbb{R}^k$, $\beta, \hat{\beta} \in \mathbb{R}^n$.

4: Compute the residuals

$$r_i(y) = \left\| y - [P, V] \begin{bmatrix} \delta_i(\hat{\alpha}_1) \\ \hat{\beta}_1 \end{bmatrix} \right\|_2, \qquad (13)$$

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

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

3.3. Related Works and Discussions

There are several previous methods that aim to improve the robustness of SRC by appending additional bases to the conventional dictionary of training images. Wright *et al.* addressed the disguise problem by adding a complete set of single-pixel based bases to the dictionary of SRC [15]. Yang and Zhang [16] used the Gabor features for SRC with a learned Gabor occlusion dictionary to reduce the computational cost. Deng *et al.* introduced Extended SRC (ESRC) method to address the undersampled problem of SRC by representing the typical facial variations in an additional dictionary [3]. These methods are effective to improve the robustness against the corruption of the test images, but they are still sensitive to the corruption of the training images.

Table 1. Comparative recognition rates of SSRC and other recognition methods. The results of the first five rows are cited from [11] under identical experimental settings.

Algorithms	Dictionary Size	Accuracy
ℓ_2[11]	19800×1300	95.89±2.35%
Nearest Subspace	19800×1300	92.34±4.16%
Random OMP	300×1300	84.85±3.43%
Hash OMP	300×1300	86.92±3.44%
SRC	300×1300	93.12±2.94%
SRC	300×1300	92.82±0.95%
ESRC	300×2600	96.88±0.71%
SSRC	300×1400	**98.31±0.44**%
SRC	19800×1300	93.75±1.01%
ESRC	19800×2600	97.36±0.59%
SSRC	19800×1400	**98.58±0.40**%

The proposed P+V model and the corresponding SSRC algorithm, for the first time, design the dictionary by the decomposition of the training samples into the separated parts of prototypes and variations. Therefore, the P+V model based classification is expected to be robust against the corruption of both the training and test images. A recent work of Chen *et al.* [2] also aimed to address the training corruption problem, but they only filtered out the corruption by low-rank and sparse decomposition, without any concern of the typical intra-class variations in the dictionary setting.

4. Experimental Study

This section presents experiments on publicly available databases to demonstrate the efficacy of the proposed SSRC. For fair comparisons, SRC [15], ESRC [3], and SSRC use the Homotopy[1] method [8][4] to solve the ℓ_1-minimization problem with the regularization parameter $\lambda = 0.005$ and identical parameters, so that the performance difference will be solely induced by the different choice of dictionary.

4.1. Recognition with Uncontrolled Training Set

The AR database consists of over 3,000 frontal images of 126 individuals. There are 26 images of each individual, taken at two different occasions [7]. The faces in AR contain variations such as illumination change, expressions and facial disguises (i.e. sun glasses or scarf). We randomly selected 100 subjects (50 male and 50 female) for our experiments, and the images are cropped with dimension 165×120.

The first experiment is a reproduction of that in the sec-

[1] This optimization method had acceptable accuracy and fastest speed on the comparative study in [19], and its source code was downloaded at http://www.users.ece.gatech.edu/~sasif/homotopy/

tion 5 of [11]. Specifically, for each subject, the 26 images are randomly permuted and then the first half is taken for training and the rest for testing. In this way, we have 1300 training images and 1300 test images. For statistical stability, 10 different training and test set pairs are generated by randomly permuting, and averaged accuracy and standard deviation are reported.

We first evaluate SRC in the both 300 dimensional eigenspace and the 19800 dimensional image space. SRC obtains a better recognition rate of 93.75% on the full image dimension, which is compared to a 95.89% recognition rate obtained with basic ℓ_2 approach [11]. As suggested by Wright *et al.* [13], SRC performs worse because the randomly selected training set contains corruption images occlusion that would break the sparsity assumption.

However, one should not deny the the usefulness of the sparsity based recognition according to the above results, as we find that *the discrimination power of sparse representation relies heavily on the suitable choice of dictionary*. Specifically, we fairly compare SRC, ESRC, and SSRC in both the 300 dimensional eigenspace and the 19800 dimensional image space. The comparative results are reported in Table 1. By simply re-designing the dictionary by the P+V model, the SSRC dramatically boost the sparsity based recognition accuracy to over 98%. The ESRC method, which appends an intra-class dictionary to the training samples, also increases the accuracy to about 97%, but using a much larger dictionary of 2600 bases. Clearly, sparsity based classification can outperform the ℓ_2 approach even using drastically lower dimensional features.

The second experiment is a reproduction of that in [2] which specifically evaluates the robustness of the sparsity based face recognition by considering the following three scenarios of corrupted training images as follows:

- *Sunglasses*: Seven neutral images plus one randomly chosen image with sunglasses at session 1 are selected for training, and the remaining neutral images (all from session 2) plus the rest of the images with sunglasses (two taken at session 1 and three at session 2) for testing. In total, there are 8 training images and 12 test images per person.

- *Scarf*: Seven neutral images plus one randomly chosen image with scarf at session 1 are selected for training, and the remaining neutral images (all from session 2) plus the rest of the images with sunglasses (two taken at session 1 and three at session 2) for testing. In total, there are 8 training images and 12 test images per person.

- *Sunglasses+Scarf*: Seven neutral images and two corrupted images (one with sunglasses and the other with scarf) at session 1 are selected for training. In total, there are 9 training images and 17 test images (seven neutral images at session 2 plus the remaining ten occluded images) are available for this case.

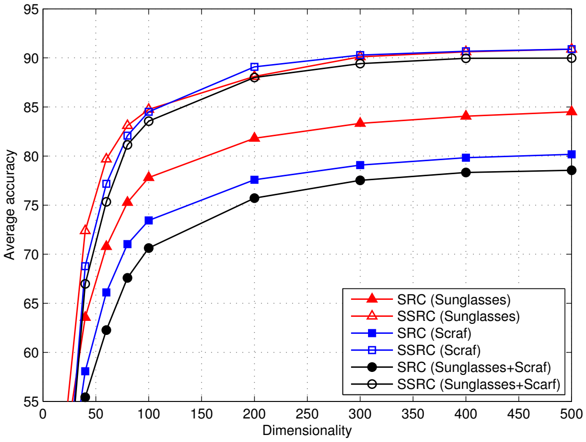


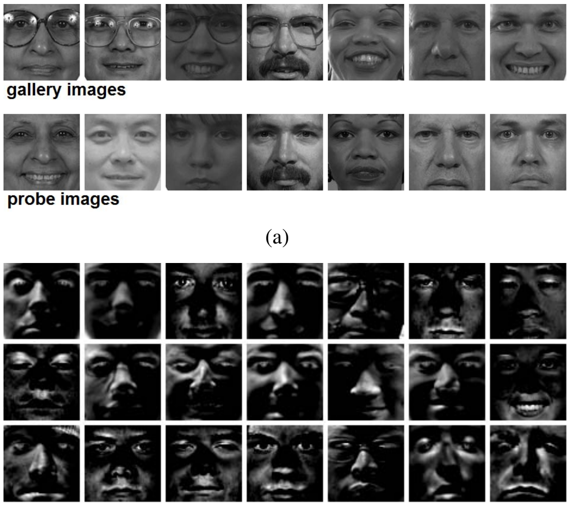
Figure 2. The comparative recognition rates between SRC and SSRC on the AR data set with different kinds of corrupted training images.

We vary the dimension of the eigenspace from 20 to 500, and compare the recognition performance of between SRC and SSRC. Each scenario is repeated three times, and the averaged performance is reported. Fig. 2 shows the comparative recognition rates between SRC and SSRC on the AR data set with different kinds of corrupted training images, and one can see from the figure that SSRC outperforms SRC by a margin about 6% to 12%, depending on the percentage of occlusion. Specifically, SRC performs better on the sunglasses scenario (about 84% accuracy with 20% occlusion) than the scarf scenario (about 80% accuracy with 40% occlusion), followed by the sunglasses+scarf scenario (about 78% accuracy). The performance of SRC deteriorates when the percentage of occlusion involved in the training images increases, and this is an observation consistent with the common criticism on SRC with uncontrolled training images [14]. In contrast, The accuracy of SSRC reaches about 90% in all the three scenarios. Besides the boosted accuracy, SSRC displays the stability against various kinds of corruption in the training images.

Table 2 summarizes the performance comparisons among different approaches under three different scenarios. The average accuracies of the first five methods are cited from [2], of which the best-performed method, denoted as LR+SI+SRC, applied low-rank matrix recovery with structural incoherence to filter out the corruption of the training images. LR+SI+SRC method achieves very competitive performance when the dimension is as low as 100, since it use only the low-rank components of the training images for recognition. However, as the dimension increasing, LR+SI+SRC cannot capture more information for recognition, and thus becomes significantly worse than SSRC when

Table 2. Comparative recognition rates of SSRC and other recognition methods. The results of the first five rows are cited from [2] under identical experimental settings.

Methods	Dimension=500			Dimension=100		
	Sunglasses	Scarf	Sunglasses +Scarf	Sunglasses	Scarf	Sunglasses +Scarf
Fisherfaces	–	–	–	72.50	57.67	61.80
NN	66.47	56.53	57.55	65.06	54.56	55.41
LLC+SRC [1]	84.47	76.61	79.03	79.14	70.08	72.04
SRC	84.22	76.25	78.00	79.92	71.70	71.59
LR+SI+SRC [2]	85.42	84.36	81.62	**85.27**	81.67	**81.37**
SRC	84.50±0.58	80.17±0.46	78.55±0.69	77.81±0.34	73.44±0.42	70.63±1.62
ESRC	89.33±0.65	87.31±0.71	85.65±0.66	82.28±0.65	78.92±0.68	77.29±0.69
SSRC	**90.89**±0.24	**90.89**±0.59	**89.98**±0.39	84.75±0.17	**84.50**±0.58	79.61±0.59



(a)

(b)

Figure 3. (a) The cropped images of some gallery images and corresponding probe images in the FERET database. (b) Example images of the differences to the class centroid computed from the FRGC version 2 database.

the dimension is equal to 500.

4.2. Recognition with Uncontrolled and Overcomplete Dictionary

This experiment is designed to test the robustness of SSRC against complex facial variation in the real-world applications. The experiment follows the standard data partitions of the FERET database [10] and :

- *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.

The images are first normalized by a similarity transformation that sets the centered inter-eye line horizontal and 70 pixel apart, and then cropped to the size of 128×128 with the centers of the eyes 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. 3(a) shows some cropped images which are used in our experiments. Note that the images of FERET database has complex intraclass variability, since they are acquired in multiple sessions during several years.

As there is only a single sample per gallery class, we construct the intra-class variation matrix from the standard training image set of the FRGC Version 2 database [9], which contains 12,766 frontal images of 222 people taken in the uncontrolled conditions. Fig. 3(b) shows some intra-class differences computed by (10) from this image set. Note that the collection of the FRGC database is totally independent from the FERET database. Hence, in this experiment, the variation matrix is required to universally represent the complex facial variations under uncontrolled conditions.

For comprehensive results, we also extract the Gabor feature and the LBP feature for classification besides the pixel intensity. For each feature, we test the recognition performance in the reduced PCA dimension of 125, 250, and 1000 respectively. In total, there are 36 test cases (4 probes×3 features×3 dimensions) and Table 3 lists the comparative performance between SRC and SSRC in all cases. Further, we define a Error Reduction Rate (ERR), denoted by a notion ↓, to characterize the proportion of the

Table 3. Comparative recognition rates of SRC and SSRC on FERET Database. The notation ↓ indicates the percentage of the recognition errors that are reduced by switching from SRC to SSRC.

Features	Methods	Dimension=1000				Dimension=250				Dimension=125			
		fb	fc	dup1	dup2	fb	fc	dup1	dup2	fb	fc	dup1	dup2
Intensity	SRC	85.2	76.3	63.9	57.3	84.3	75.8	62.5	53.0	80.7	64.9	57.8	46.6
	SSRC	87.9	91.8	68.6	67.5	88.3	87.1	65.2	61.5	85.6	80.9	61.6	56.4
		↓18%	↓65%	↓13%	↓24%	↓25%	↓47%	↓7%	↓18%	↓25%	↓46%	↓9%	↓18%
Gabor	SRC	93.0	97.4	73.0	78.6	88.6	94.3	63.6	70.5	83.5	90.2	53.5	61.5
	SSRC	96.7	99.5	80.7	85.5	93.2	96.4	68.6	76.9	89.0	94.3	57.6	67.9
		↓53%	↓81%	↓29%	↓32%	↓40%	↓37%	↓14%	↓22%	↓33%	↓42%	↓9%	↓17%
LBP	SRC	96.9	93.8	87.7	85.0	95.1	86.1	83.4	77.4	91.5	68.0	76.5	68.8
	SSRC	98.0	99.5	90.6	90.2	96.7	93.8	85.7	80.8	94.6	83.5	79.5	74.8
		↓35%	↓92%	↓24%	↓35%	↓33%	↓55%	↓14%	↓15%	↓36%	↓48%	↓13%	↓19%

errors reduced by switching SRC to SSRC. For instance, since the 1000 dimensional LBP-PCA feature based SSRC improves the accuracy from 85.0% to 90.2% on the fc probe set, the ERR is ↓35% (=100×(15.0-9.8)/15.0), suggesting that 35% recognition errors can be avoided by using SSRC instead of SRC.

Although the variation matrix is constructed from the FRGC database, SSRC improve the recognition rates on the FERET database in all the 36 test cases, indicating that the intra-class variability of face is sharable even when the generic data are collected from different conditions and camera set-ups. In addition, in term of the ERR, performance enhancement by replacing SRC with SSRC is notable on in all test cases. These results suggest that the P+V model is feasible for various feature representations, and thus it can be integrated with more informative features to address uncontrolled face recognition problem. For instance, LBP feature based SSRC achieves over 90% accuracy on all the four probe sets.

It should be mentioned that similar experimental results has been reported on ESRC method [3], but its intra-class variant dictionary are constructed from the generic training set of FERET database. There may be some implicit correlation, or even overlap, between the generic training set and the test sets of the FERET database. Therefore, the results of Deng *et al.* may not be feasible on the real-world applications. In contrast, *our experiment, for the first time, justifies the effectiveness of the sparsity based face recognition when the dictionary bases are collected from the uncontrolled conditions that are independent from the test condition.*

4.3. ℓ_1-norm versus ℓ_2-norm regularization with Over-complete Dictionary

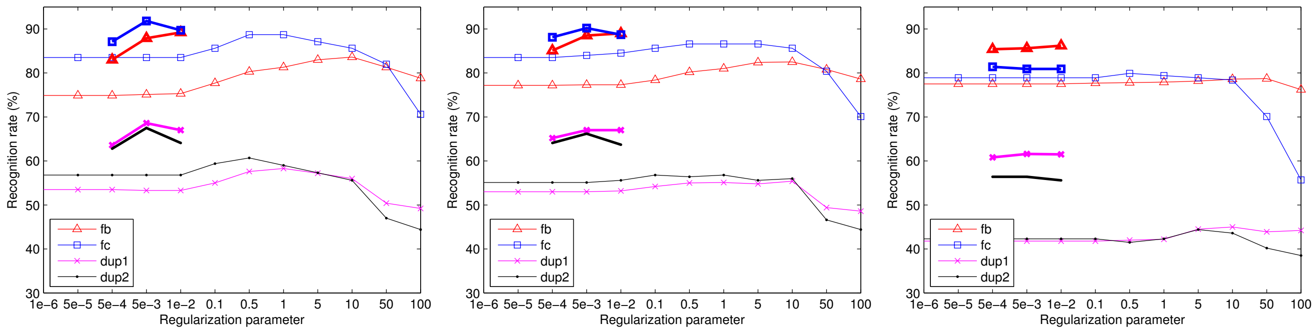
Based on the results on the FERET database, we further investigate the role of sparsity in face recognition with an uncontrolled and over-complete dictionary. In particular we evaluate whether the ℓ_1-norm regularization of SSRC can be replaced by the ℓ_2-norm that is much more computationally efficient. For this purpose, we replace the ℓ_1 norm regularization in (14) with the the ℓ_2 norm as follows.

$$\begin{bmatrix} \hat{\alpha}_2 \\ \hat{\beta}_2 \end{bmatrix} = \arg\min \left\| [P, V] \begin{bmatrix} \alpha \\ \beta \end{bmatrix} - y \right\|_2^2 + \lambda \left\| \begin{bmatrix} \alpha \\ \beta \end{bmatrix} \right\|_2^2, \tag{14}$$

Then test the performance of the ℓ_2-regularized minimization by increasing the parameter λ from 0.000001 to 100.

The comparative results on the varying dimensional PCA space are shown in Fig. 4. When the value of λ is relatively large in the range of $[0.1,\ 10]$, ℓ_2-norm regularization obtain its optimal performance. However, the optimal performance of ℓ_2-norm regularization is significantly lower than that of SSRC (ℓ_1-norm regularization) tested with limited number of $\lambda = \{0.0005, 0.005, 0.01\}$. The superiority of SSRC seems more apparent on the dup1 and dup2 set. Additionally, the Homotopy used in our experiments is far from the optimal solver of ℓ_1-minimization, the performance of SSRC might be further improved by more accurate solvers. This implies that ℓ_1-norm indeed play a crucial role in face recognition given an uncontrolled and over-complete dictionary.

It should be mentioned that our observation on ℓ_1-norm sparsity is different from that by Zhang *et al.* [19]. Indeed, both observations are valid, but under different dictionary settings. Zhang *et al.* directly ensemble the controlled training samples themselves to construct an under-complete dictionary, and thus both ℓ_1-norm and ℓ_2 norm regularization can provides reasonable results. The dictionary of SSRC contains an over-complete set of intra-class variation bases, and most of which are irrelevant to the test sample. The dense combination of the irrelevant bases would mislead the classification, and thus the ℓ_1-minimization technique is more desirable than ℓ_2 to select a small number of relevant bases from an over-complete set of bases.



(a) dimension=1000 (b) dimension=250 (c) dimension=125

Figure 4. The recognition rates of SSRC with ℓ_1-regularization (plotted by thick symbols) and ℓ_2-regularization (plotted by the thin symbols) as a function of the value of λ.

5. Conclusions

It has been shown in this paper that a simple separation between the prototype and variation components leads to an enormous improvement on sparsity based face recognition under uncontrolled training conditions. The proposed SSR-C algorithm performs best in several experiments on which SRC was previously criticized to perform poorly. The added complexity of the algorithm is trivial. In particular when only a single sample per class is available, ℓ_1-norm regularization based sparse coding the algorithms accurately find out the intra-class variation bases from an over-complete dictionary that is constructed from uncontrolled generic images outside the gallery. Our preliminary results suggest that the proposed prototype plus variation model provides a widely applicable framework to address uncontrolled face recognition problem.

6. Acknowledgements

This work was partially sponsored by NSFC (National Natural Science Foundation of China) under Grant No. 61005025, No. 61002051, and No. 61273217. This work was also supported by ”the Fundamental Research Funds for the Central Universities”.

References

[1] Y. Chao, Y. Yeh, Y. Chen, Y. Lee, and Y. Wang. Locality-constrained group sparse representation for robust face recognition. In *ICIP*, 2011.

[2] C. Chen, C. Wei, and Y. Wang. Low-rank matrix recovery with structural incoherence for robust face recognition. In *CVPR*, 2012.

[3] W. Deng, J. Hu, and J. Guo. Extended src: Undersampled face recognition via intraclass variant dictionary. *IEEE TPAMI*, 34(9):1864–1870, 2012.

[4] D. Donoho and Y. Tsaig. Fast solution of ℓ_1-norm minimization problems when the solution may be sparse. *Information Theory, IEEE Transactions on*, 54(11):4789–4812, 2008.

[5] J. Huang, X. Huang, and D. Metaxas. Simultaneous image transformation and sparse representation recovery. In *CVPR*, 2008.

[6] L. Ma, C. Wang, B. Xiao, and W. Zhou. Sparse representation for face recognition based on discriminative low-rank dictionary learning. In *CVPR*, 2012.

[7] A. M. Martinez and R. Benavente. The ar face database. *CVC Technical Report #24*, June 1998.

[8] M. Osborne, B. Presnell, and B. Turlach. A new approach to variable selection in least squares problems. *IMA journal of numerical analysis*, 20(3):389, 2000.

[9] P. J. Phillips, P. J. Flynn, T. Scruggs, K. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek. Overview of the face recognition grand challenge. *CVPR*, 2005.

[10] P. J. Phillips, H. Moon, P. Rizvi, and P. Rauss. The feret evaluation method for face recognition algorithms. *IEEE TPAMI*, 22:0162–8828, 2000.

[11] Q. Shi, A. Eriksson, A. van den Hengel, and C. Shen. Is face recognition really a compressive sensing problem? In *CVPR*, 2011.

[12] A. Wagner, J. Wright, A. Ganesh, Z. Zhou, and Y. Ma. Towards a practical face recognition system: robust registration and illumination by sparse representation. In *CVPR*, 2009.

[13] J. Wright, A. Ganesh, A. Yang, Z. Zhou, and Y. Ma. Sparsity and robustness in face recognition. *arXiv preprint arXiv:1111.1014*, 2011.

[14] J. Wright, Y. Ma, J. Mairal, G. Sapiro, T. Huang, and S. Yan. Sparse representation for computer vision and pattern recognition. *Proceedings of the IEEE*, 98(6):1031–1044, 2010.

[15] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma. Robust Face Recognition via Sparse Representation. *IEEE TPAMI*, 31(2):210–227, 2009.

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

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

[18] H. Zhang, J. Yang, Y. Zhang, N. Nasrabadi, and T. Huang. Close the loop: Joint blind image restoration and recognition with sparse representation prior. In *ICCV*, 2011.

[19] L. Zhang, M. Yang, and X. Feng. Sparse representation or collaborative representation: Which helps face recognition? In *ICCV*, 2011.

[20] Z. Zhou, A. Wagner, H. Mobahi, J. Wright, and Y. Ma. Face recognition with contiguous occlusion using markov random fields. In *ICCV*, 2009.