Local Structure based Sparse Representation for Face Recognition

FAN LIU, College of Computer and Information, HoHai University
JINHUI TANG*, Nanjing University of Science and Technology
YAN SONG, Nanjing University of Science and Technology
LIYAN ZHANG, Nanjing University of Aeronautics and Astronautics
ZHENMIN TANG, Nanjing University of Science and Technology

This paper presents a simple yet effective face recognition method, called local structure based sparse representation classification (LS_SRC). Motivated by the "divide-conquer" strategy, we firstly divide the face into local blocks, and classify each local block, and then integrate all the classification results to make the final decision. To classify each local block, we further divide each block into several overlapped local patches and assume that these local patches lie in a linear subspace. This subspace assumption reflects the local structure relationship of the overlapped patches and makes sparse representation based classification (SRC) feasible even when encountering the single sample per person (SSPP) problem. To lighten the computing burden of LS_SRC, we further propose the local structure based collaborative representation classification (LS_CRC). Moreover, the performance of LS_SRC and LS_CRC can be further improved by using the confusion matrix of the classifier. Experimental results on four public face databases show that our methods not only generalize well to SSPP problem but also have strong robustness to occlusion, little pose variation and the variations of expression, illumination, time.

Categories and Subject Descriptors: I.5.4 [Pattern Recognition]: Applications
General Terms: Design, Algorithms, Performance
Additional Key Words and Phrases: Face recognition, single sample per person problem, sparse representation, collaborative representation, local structure, confusion matrix, Bayesian inference

1. INTRODUCTION
Face recognition is an active research topic that has attracted significant attention in the domain of computer vision and pattern recognition for many years [Zhao et al. 2003], due to its scientific challenges and potential applications. In the prior literatures, plenty of research efforts have been dedicated into face recognition problem and tremendous progress has been achieved [Zhang et al. 1997; Tenenbaum et al. 2000; Belkin and Niyogi 2003; Wright et al. 2009; Zhang et al. 2011; An et al. 2013; Kafai et al. 2014].

In the last two decades, face recognition systems are considered to be critically dependent on discriminative feature extraction, about which many approaches have been proposed, such as Eigenfaces [Zhang et al. 1997], Fishfaces [Belhumeur et al. 1997], [Tenenbaum et al. 2000], LLE [Roweis and Saul 2000], LPP [He et al. 2005] and Laplacian Eigenmap [Belkin and Niyogi 2003]. Of late, the significance of feature extraction...
has been debated. Wright et al. indicated 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 [Wright et al. 2009]. Thus, they proposed a sparse representation based classification algorithm (SRC). Later, Zhang et al. [Zhang et al. 2011] proposed a collaborative representation based classification method with regularized least squares (CRC_RLS) to further reduce the complexity of SRC. Another representative method [Li et al. 2015] for data representation was developed by learning subspace with guide of high-level semantic information. However, SRC and CRC both rely on a rich set of training samples of each subject. Therefore, they suffer serious performance drop when encountering the single sample per person (SSPP) problem.

In this paper, we propose a simple yet effective method, called local structure based sparse representation classification (LS_SRC)\(^1\) to solve SSPP problem. Motivated by the “divide-conquer” [Stout 1987] strategy, we firstly divide each face into many local blocks, and make the classification decision for each of them, and then aggregate all the classification results to generate the final decision. To classify each local block, we further divide each block into overlapped patches and assume that these patches lie in a linear subspace due to the fact that these patches are strongly similar. Intuitively, if the training sample and test sample describe the same person, as to each pair of corresponding local blocks, the subspaces that the local patches lie in should be very close. Therefore, the central patch of the test block can be approximately represented by a linear combination of the patches in the corresponding block from the same class. Based on this idea, SRC is feasible even when encountering SSPP problem. Then we combine the classification outputs of all the local blocks by majority voting. To lighten the computing burden of LS_SRC, we also present the local structure based collaborative representation classification (LS_CRC) which has competitive FR accuracy but with significantly lower complexity than LS_SRC. Although LS_SRC and LS_CRC both have achieved good performance, their class probabilities are probably wrong due to the error of the classifier itself. In [Liu and Nakagawa 2000], Liu et al. proposed to use the confusion matrix to describe the error of a distance-based classifier with multiple output classes. Inspired by this idea, we use the confusion matrix to describe the error of our proposed vote-based classifier and further improve the performance of LS_SRC and LS_CRC. From the respect of probability, the confusion matrix can also be regarded as the prior information of the classifier. By Bayesian inference, we can infer more accurate class probabilities which are not observable.

To evaluate the proposed methods, we perform a series of experiments on four public datasets including Extended Yale B, PIE, AR and LFW face databases. Experimental results demonstrate that the proposed methods not only outperform SRC and CRC, but also have strong robustness to occlusion, little pose variation and the variations of expression, illumination and time. Moreover, we also compare our proposed methods with several state-of-art methods dealing with SSPP problem. Experimental results show that our proposed methods also generalize well on SSPP problem.

This paper is an extension to what we have presented earlier in our conference paper [Liu et al. 2014]. We highlight the newly incorporated work as follows:

---

1\(^{\text{Part of this material has been presented in IEEE ICIP 2014 [Liu et al. 2014]}}\)
We have conducted more experiments to verify the effectiveness of the proposed methods. Previously, we only provided the experimental results for SSPP problem. In this paper, we not only conduct more experiments to demonstrate the effectiveness of our methods but also analyze the influence of the parameters.

We have analyzed the advantages of our proposed methods and provided theoretical explanation, which were not discussed in our previous work.

The rest of this paper is organized as follows. We start by introducing related work in Section 2. Then in Section 3, we present the proposed local structure based methods: local structure based sparse representation classification (LS_SRC), local structure based collaborative representation classification (LS_CRC) and their extensions based on confusion matrix. Section 4 demonstrates experiments and results. Finally, we conclude in Section 5 by highlighting key points of our work.

2. RELATED WORK

Within the last two decades, face recognition systems were known to rely heavily on discriminative feature extraction. Therefore, how to extract robust and discriminative features is the key to face recognition. As face images usually lie on a lower dimensional subspace [Belhumeur et al. 1997; Basri and Jacobs 2003], subspace analysis method that aims to project high dimensional data to a lower dimensional subspace, has become one of the most popular methods. Among various subspace analysis based FR methods, the classical Eigenfaces [Zhang et al. 1997] and Fishfaces [Belhumeur et al. 1997] algorithms are the most representative ones based on principal component analysis (PCA) and linear discriminant analysis (LDA) respectively. However, PCA and LDA consider only the global scatter of training samples and they fail to reveal the essential data structures nonlinearly embedded in high dimensional space.

To overcome these limitations, a number of manifold learning methods (e.g., ISOMAP [Tenenbaum et al. 2000], LLE [Roweis and Saul 2000], LPP [He et al. 2005] and Laplacian Eigenmap [Belkin and Niyogi 2003]) were proposed by assuming that the data lie on a low dimensional manifold of the high dimensional space. Yan et al. [Yan et al. 2007] proposed a general framework called graph embedding, which generalizes the above-mentioned methods to a unified model within this common framework. As opposed to holistic features extracted by the aforementioned methods like PCA and LDA, local features have certain advantages and have been proved to be more robust against variations [Martínez 2002]. Gabor [Liu and Wechsler 2002] and local binary patterns (LBP) [Ojala et al. 2002] are two representative local features, which are predefined in a hand-crafted way. Recently, Lei et al. [Lei et al. 2014] proposed a method to learn the most discriminant local features in a data-driven way. In addition, deep learning techniques have also been applied to learn effective features for face representation in recent years [Sun et al. 2013; Kan et al. 2014; Taigman et al. 2014]. For example, [Sun et al. 2013] learned features with the verification signal, while [Taigman et al. 2014] learned feature with identification signal.

In spite of the tremendous achievements, the significance of feature extraction has still been debated. Wright et al. [Wright et al. 2009] have demonstrated that the choice of feature space is no longer critical as long as the test image can be approximated by a sparse linear combination of the training images. In Wright et al.'s pioneer work, a testing sample is first coded as a sparse linear combination of all the training samples via $l_1$-norm minimization. Then the testing face image is classified to the class which yields the least representation error. The experimental results show 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 Fisherface and Laplacianfaces-based features. However, SRC
requires a rich set of training samples and the correct sparse solution can only be recovered when the number of training samples is sufficiently larger than the dimensionality of features. To fulfill this requirement, Wagner et al. [Wagner et al. 2009] designed a system that acquires tens of images of each subject to cover all possible illumination changes. Recently, Zhang et al. showed that it is the collaborative representation (CR) mechanism rather than $l_1$-norm sparsity that truly improves the FR accuracy [Zhang et al. 2011]. Consequently, they proposed CR based classification with regularized least squares (CRC$_{RLS}$), which has significantly less complexity than SRC but leads to very competitive results. In addition, one new method was proposed to learn data representation by jointly considering the structure information and sparsity in [Li et al. 2014].

The performance of the above mentioned methods is heavily affected by the number of training samples for each person. Specifically, if there is only one training sample per person, some methods even fail to work because the intra-personal variations cannot be estimated at all. This is the so called single sample per person (SSPP) problem [Tan et al. 2006] in face recognition. In order to address SSPP problem, many methods have been developed during the last two decades. Shan et al. [Shan et al. 2003] presented a face-specific subspace method based on PCA which firstly generates a few virtual samples from single gallery image of per subject and then uses PCA to build a projection subspace for each person. But strong correlation between virtual samples decreases the representativeness of training samples and accordingly limits the performance of this method. In order to make LDA suitable for SSPP problem, Gao et al. [Gao et al. 2008] applied SVD decomposition to the only face image of a person and the obtained non-significant SVD basis images were used to estimate the within-class scatter matrix of this person approximately. However, the optimal number of non-significant SVD basis images is face-specific and should not be determined equally for all face images as they did. Considering the similarity of face images across individuals, a few generic learning methods have been proposed to solve SSPP problem, which use a generic training set to extract discriminatory information. For example, Su et al. proposed an Adaptive Generic Learning (AGL) [Su et al. 2010] method, which adapts a generic discriminant model to better distinguish the persons with single sample. Yang et al. proposed the spare variation dictionary learning (SVDL) [Yang et al. 2013] scheme by using the relationship between the gallery set and the external generic set. Recently, Deng et al. [Deng et al. 2014] proposed a novel generic learning method by mapping the intra-class facial difference of the generic faces to the zero vectors to further enhance the generalization capability of their proposed linear regression analysis (LRA). They also proposed the extended sparse representation-based classifier (ESRC) [Deng et al. 2012] to solve SSPP problem, which applies an auxiliary intra-class variant dictionary to represent possible variation between the training and testing images.

All the above-mentioned methods for SSPP problem treat the whole image as a high-dimensional vector and belong to holistic representation based methods. However, some other schemes favor local representation, in which a face image is divided into blocks and vector representation of information is conducted block by block other than globally. For example, Chen et al. [Chen et al. 2004] proposed BlockFLD method which generates multiple training samples for each person by partitioning each face image into a set of same sized blocks and then applies FLD-based methods with these blocks. However, the great differences between the appearances of long-distance blocks from one image may go against the compactness of the within-class scatter after projection. Recently, Zhu et al. [Zhu et al. 2012] proposed patch based CRC (PCRC) for small sample size (SSS) problem. They also proposed a local generic representation (LGR) [Zhu et al. 2014] based framework for SSPP problem which takes advantages of patch based local representation and generic variation representation.
al. [Kumar et al. 2011] proposed the patch based n-nearest classifier to improve the stability and generalization ability. Lu et al. [Lu et al. 2013] proposed a discriminative multi-manifold analysis (DMMA) method by learning discriminative features from image patches. Generally speaking, these patch based methods have a commonality that they just consider each patch independently. Hence, they will lose the correlation information between patches, which may be very important for classification.

3. THE PROPOSED LOCAL STRUCTURE BASED METHODS FOR FACE RECOGNITION

In this section, we firstly describe the local structure, and then propose the local structure based sparse representation classification (LS_SRC). Then local structure based collaborative representation classification (LS_CRC) will be presented to enhance the efficiency. In the third part of this section, we present Baysian inference based on confusion matrix to further improve the performance of LS_SRC and LS_CRC. At last, we analyze the advantages of our proposed methods.

3.1. Local Structure

To describe the local structure, we illustrate three kinds of neighborhood in Fig. 1. The $P$ neighbor pixels on a square of radius $R$ form a squarely symmetric neighbor sets. Suppose there are $N$ pixels in an image. For the $i$-th pixel in the image, its $P$ neighbor pixels can be denoted by $\Omega^p_i = \{ j \mid j = 1, \ldots, P \}$.

![Fig. 1. Squarely symmetric neighbor sets for different $R$.](image)

For the $i$-th pixel in the image, we select a $S \times S$ local patch (e.g. $S=3, 5$) centered at it. All the $S^2$ pixels within the patch form a $m$ dimensional local patch vector $x^i_0$, where $m = S^2$. Similarly, the neighbor pixel $i_j$ of the $i$-th also corresponds to a same sized local patch, whose patch vector is denoted by $x^j_i, j = 1, \ldots, P$. Then, the center patch and its neighbor patches determine a local block centered at the $i$-th pixel. Fig. 2 shows an example of a local block containing a central patch and 16 neighbor patches. The size of patch is $3 \times 3$ and the size of the block is $7 \times 7$. For a pixel on the margin of an image, we use the mirror transform first and then determine its local block.

As the patches are overlapped and gather in a small block, they are strongly similar. Therefore, we assume that all patches in a local block lie in a linear subspace. Based on
this subspace assumption, we can characterize the local structure between a central patch and its neighbor patches. Specifically, the central patch can be approximately represented by a linear combination of its neighbor patches. This idea has already successfully been used for feature extraction in [Qian et al. 2013].

3.2. Local Structure based Sparse Representation Classification (LS_SRC)

As shown in Fig. 3, the query image $y$ is firstly divided into a set of overlapped blocks $\{y^1, y^2, \ldots, y^N\}$. The $i$-th pixel of the image corresponds to the local block $y^i$, which consists of several overlapped patches, where $y^i = [y^i_0, y^i_1, \ldots, y^i_P] \in R^{m \times (P+1)}$. Similarly, we also decompose the training samples into blocks. Suppose that there is only one training image per person, the $i$-th block of the $k$-th person is denoted by $B^i_k = [x^i_{k,0}, x^i_{k,1}, \ldots, x^i_{k,P}] \in R^{m \times (P+1)}$, where $x^i_{k,j}$ ($j = 0, \ldots, P$) is the local patch vector in the block. All the $i$-th blocks from $K$ classes form a local dictionary $B^i$, where $B^i = [B^i_1, B^i_2, \ldots, B^i_K]$. According to the description of local structure, $B^i_k = [x^i_{k,0}, x^i_{k,1}, \ldots, x^i_{k,P}]$ is supposed to lie on a subspace $\Psi$ and $y^i = [y^i_0, y^i_1, \ldots, y^i_P]$ belongs to a subspace $\Phi$. If the test image $y$ is also from the $k$-th class and all the training images and test images are well aligned, $\Psi$ and $\Phi$ are theoretically very close. Therefore, the central patch vector $y^i_0$ will approximately lie in the linear span of all the local patch vectors from $B^i_k = [x^i_{k,0}, x^i_{k,1}, \ldots, x^i_{k,P}]$, which can be written as:

$$y^i_0 = \sum_{j=0}^{P} \alpha^i_{k,j} x^i_{k,j} = B^i_k \alpha^i_k$$

(1)

If there are $n_k$ training samples for the $k$-th subject, (1) will be written as:

$$y^i_0 = \sum_{j=0}^{P} \alpha^i_{k,j} x^i_{k,j} + \sum_{j=P+1}^{P+n_k} \alpha^i_{k,j} x^i_{k,j} = B^i_k \alpha^i_k$$

(2)

ACM Transactions on Intelligent Systems and Technology, Vol. V, No. N, Article 39, Publication date: January YYYYY.
where $B_k^i = [x_{k,0}^i, x_{k,1}^i, \ldots, x_{k,P \times N}^i]$. Since $B^i = [B_1^i, B_2^i, \ldots, B_K^i]$ is a patch dictionary which includes all the patches in $i$-th blocks of images from $K$ classes, the linear representation of $y_0^i$ can be rewritten in the form as below:
\[
y_0^i = B^i \mathbf{\alpha}^i
\]
where $\mathbf{\alpha}^i = [0, \ldots, 0, \alpha_{k,0}^i, \alpha_{k,1}^i, \ldots, \alpha_{k,P \times N}^i, 0, \ldots, 0]^T$ is a coefficient vector whose entries are zero except those associated with the $k$-th class. As $y_0^i$ can be sufficiently represented using only the training block of the same subject, the representation of $\mathbf{\alpha}^i$ is naturally sparse if the number of subjects $K$ is reasonably large. Therefore, finding the identity of $y_0^i$ equals finding the sparse solution of equation (3). This is the same as solving the following optimization problem ($l_0$-minimization):
\[
\mathbf{\alpha}^i = \arg \min \| \mathbf{\alpha}^i \|_0 \quad \text{s.t.} \quad y_0^i = B^i \mathbf{\alpha}^i
\]
However, solving the $l_0$-minimization of an underdetermined linear system is NP-hard. But it has been proved that as long as the $\mathbf{\alpha}^i$ sought is sparse enough, the solution of the $l_0$-minimization problem in (4) is equivalent to the following $l_1$-minimization problem.
\[
\hat{\mathbf{\alpha}}^i = \arg \min \| \mathbf{\alpha}^i \|_1 \quad \text{s.t.} \quad y_0^i = B^i \mathbf{\alpha}^i
\]
For the block $y^i$, the sparse representation of its central patch $y_0^i$ to the patch dictionary $B^i = [B_1^i, B_2^i, \ldots, B_K^i]$ is computed via (5). Ideally, the nonzero entries in the representation coefficient $\mathbf{\alpha}^i$ will be associated with the columns of $B^i$ from a single object class $k$. Then using only the coefficients associated with the $k$-th class, one can approximate the central patch $y_0^i$ of the test block $y^i$ as $\hat{y}_0^i = B^i \delta_k(\hat{\mathbf{\alpha}}^i)$, where $\delta_k$ is a function selecting the coefficients associated with $k$-th class. And $y_0^i$ can be finally classified by minimizing the residual between $y_0^i$ and $\hat{y}_0^i$:
\[
\min_k r_k(y_0^i) = \| y_0^i - B^i \delta_k(\hat{\mathbf{\alpha}}^i) \|_2
\]
The classification outputs of all blocks can then be aggregated. Majority voting is used for making the final decision, which means that the test sample is finally classified into the class with the largest number of votes. The whole procedure can be summarized as “divide-conquer-aggregate” procedure. Although there are $N$ blocks corresponding to $N$ pixels of the image, we do not need to use all the $N$ blocks for classification. We set a parameter $\kappa$ to represent the interval between two blocks. The larger the value of $\kappa$, the fewer the number of blocks.

3.3. Local Structure based Collaborative Representation Classification (LS_CRC)
Although SRC [Wright et al. 2009] has shown interesting results in FR and has been widely studied, the $l_1$-minimization makes it time consuming. In addition, some researchers have questioned the use of sparsity in classification such as [Berkes et al. 2009; Zhang et al. 2011]. In [Zhang et al. 2011], it is shown that the use of collaborative representation is more crucial than the $l_1$-sparsity to FR. Then they proposed collaborative representation based classification (CRC) for FR. By using $l_2$-regularized least square, CRC has much less computational cost than SRC. Inspired by this idea, we propose the local structure based collaborative representation classification (LS_CRC) to lessen the computing burden of LS_SRC.
For the test block $y^i$, the collaborative representation of its central patch $y_0^i$ to the patch dictionary $B^i$ is computed by the following $l_2$-regularized least square method:
\[
\hat{\mathbf{\alpha}}^i = \arg \min \{ || y_0^i - B^i \mathbf{\alpha}^i ||_2^2 + \lambda || \mathbf{\alpha}^i ||_2^2 \}
\]
The solution of (7) can be easily derived as:

\[ \hat{\alpha}^i = \left( (B^i)^T B^i + \lambda I \right)^{-1}(B^i)^T y_0^i \]  

(8)

Let \( P = (B^i)^T B^i + \lambda I \). Obviously, \( P \) is independent of \( y_0^i \) so that it can be pre-calculated as a projection matrix. This makes LS_CRC very fast. According to the classification rule of CRC, \( y_0^i \) can be classified by minimizing the following regularized residuals:

\[ \min_k r_k(y_0^i) = \| y_0^i - B^i \delta_k(\hat{\alpha}^i) \|_2 / \| \delta_k(\hat{\alpha}^i) \|_2 \]  

(9)

where \( \delta_k(\hat{\alpha}^i) \) is the coefficient vector associated with class \( k \) and the \( l_2 \) norm of \( \delta_k(\hat{\alpha}^i) \) can also bring some discrimination for classification. After classifying each block, the classification outputs of all blocks are then combined by majority voting.

### 3.4. Bayesian Inference based on Confusion Matrix

For LS_SRC and LS_CRC, the majority voting scheme that combines the results of each block is only based on the output label of each classifier. Each classification result that corresponds to each block is equally treated as one vote without considering the error of the classifier itself. But the errors of these local classifiers can be suppressed by the final majority voting. This is because only a few part of local blocks cannot be correctly classified due to the influence of expression, illumination, occlusion etc. Most local blocks will get correct classification results, thus make sure that the final voting result is correct. However, the majority voting classifier cannot guarantee that all the samples can be classified correctly because of the error of the voting classifier itself. Since the error of the classifier is usually described by its confusion matrix, we propose to use confusion matrix to further improve the performance of LS_SRC and LS_CRC.

Suppose that the final voting classifier is denoted by \( e_G \) and its confusion matrix is \( Conf_G \). The confusion matrix \( Conf_G \) can be given by

\[ Conf_G = \begin{pmatrix}
  p_{1,1}^G & p_{1,2}^G & \cdots & p_{1,K}^G \\
p_{2,1}^G & p_{2,2}^G & \cdots & p_{2,K}^G \\
  \vdots & \vdots & \ddots & \vdots \\
p_{K,1}^G & p_{K,2}^G & \cdots & p_{K,K}^G
\end{pmatrix} \]  

(10)

where each row corresponds to the input class label and the each column corresponds to the output class label of the classifier. Thus, an element \( p_{i,k}^G \) denotes the conditional probability that samples from the input class \( C_i \) is classified into the output class \( C_k \) via the final voting classifier \( e_G \). From this point of view, the confusion matrix also can be regarded as the prior knowledge of the classifier.

The confusion matrix \( Conf_G \) of the voting classifier \( e_G \) is obtained by using \( e_G \) to classify a training sample set and the conditional probability \( p_{i,k}^G \) is computed by:

\[ p_{i,k}^G = p(C_k|C_i) \]

\[ = \int_x p(x, C_k|C_i) \]

\[ = \int_x p(x|C_i)p(C_k|x) \]

\[ = \frac{1}{|C_i|} \sum_{x \in C_i} P(C_k|x) \]  

(11)
where \( |C_t| \) is the number of training samples from the input class \( C_t \). The output class probability \( p(C_k|x) \) can be computed by the following equation:

\[
P(C_{(k)}|x) = \frac{V^{C_{(k)}}}{\sum_j V^j}
\]

(12)

where \( V^{C_k} \) is the votes to the class \( C_k \), \( \sum_j V^j \) is the sum votes to all classes. According to Bayes rule, the probability of an input class conditioned on an output class is computed by:

\[
P(C_{(t)}|C_k) = \frac{p(C_k|C_{(t)})p(C_{(t)})}{p(C_k)} = \frac{p(C_k|C_{(t)})p(C_{(t)})}{\sum_k p(C_k|C_{(t)})p(C_{(t)})}
\]

(13)

As the output probability is probably wrong due to the inaccuracy of the classifier, we utilize the Bayesian inference to infer the true probability which is not observable. Then the probability of a hypothesized true class \( C_t \) is computed by:

\[
P(C_{(t)}|x) = \sum_{C_k} p(C_{(t)}, C_k|x) = \sum_{C_k} p(C_k|x)p(C_{(t)}|C_k)
\]

(14)

Then, the resulting probabilities \( p(C_{(t)}|x) \) for each class \( C_t, C_t \in [1, \ldots, K] \) are sorted in descending order and the top one in the rank is regarded as the final classification result.

### 3.5. Analysis of Our Proposed Methods

The merits of the proposed methods lie in several factors. First, by adopting the “divide-conquer-aggregate” strategy, we divide the high-dimensional face image into several sub-images with lower dimensions and input them into respective classifiers. By this way, the dilemma of high data dimensionality and small sample size can be alleviated. Furthermore, each sub-image classification problem can be solved quickly and in parallel. Second, as patch-based methods, our proposed methods can eliminate or lower the effects of illumination changes, occlusion and expression changes by analyzing face images locally since face local information is less sensitive to those changes. Moreover, compared with global methods, local features contain more crucial information for face representation. Third, the subspace assumption of the overlapped patches not only describes the relationship of the patches in a local block but also guarantees the robustness of SRC and CRC to SSPP problem. Finally, the majority voting with all the classification results of different blocks further improves the performance. It has already been found that majority voting is by far the simplest, yet as effective as many complicated schemes in improving the recognition results. This point has already been proved by the following theorems [Lam and Suen 1997]. The theorem 1 was first presented for majority voting involving odd numbers of classifiers while the theorem 2 was proved in the case of even numbers of classifiers.

**Theorem 1**: Suppose odd numbers \( (2n + 1) \) of classifiers are combined. If \( p > \frac{1}{2} \), \( P_c(2n + 1) \) is monotonically increasing in \( n \) and \( P_c(2n + 1) \rightarrow 1 \) as \( n \rightarrow \infty \).

**Theorem 2**: Suppose even numbers \( (2n) \) of classifiers are combined. If \( p > \frac{n}{2n+1} \), \( P_c(2n) \) is is monotonically increasing in \( n \) and \( P_c(2n) \rightarrow 1 \) as \( n \rightarrow \infty \).

In both theorems, \( n \) is a positive number and \( p \) represents the probability of individual classifier being correct. \( P_c \) is the probability of the classifier being correct after majority voting. As the individual classifiers are independent and their probabilities of correct classification are regarded to be greater than 1/2, the probability after majority voting is monotonically increasing with the number of individual classifiers increasing. This conclusion has also been demonstrated by our following experiments.
Even though there are still some samples cannot be correctly classified by LS_SRC or LS_CRC, the majority voting can statistically guarantee that the true class always lies in the top of voting rank. This point has been proved by the following experiments. We conduct LS_CRC on the Extended Yale B, PIE and AR databases respectively and sort the votes to each class in descending order. In the experiments, we only use one image from each subject for training and randomly select one image from each subject for testing. Fig. 4 shows the accuracy under different ranks, where rank 1 means the top 1 match and rank 20 is the top 20 matches. It can be observed that there are big differences of the recognition rates being captured between the top 20 matches and top 1 match. It demonstrates that the true class always lies in the top of voting rank. In other words, the error of the voting classifier is very little and easy to be corrected. Therefore, the Bayesian inference based on confusion matrix can further improve the performance of LS_SRC and LS_CRC because the error of the classifier is described by its confusion matrix and rectified by Bayesian inference. From the perspective of probability, the confusion matrix also can be regarded as the prior knowledge of the classifier. By taking use of the prior information of the classifier, we can deduce the more accurate class probability distribution according to Bayesian inference. Therefore, the rank position of the true class is finally promoted and the recognition performance is improved.

4. EXPERIMENTAL RESULTS

In this section, we use Extended Yale B [Georghiades et al. 2001], PIE [Sim et al. 2003], AR [Martinez 1998] and LFW [Huang et al. 2007] databases to evaluate the proposed methods and compare them with the SRC, CRC, patch based SRC (PSRC) and patch based CRC (PCRC). To demonstrate the robustness of our proposed methods to SSPP problem, we also compare them with several state-of-the-art methods dealing with SSPP problem, which include BlockFLD [Chen et al. 2004], patch based nearest neighbor (PNN) classifier [Kumar et al. 2011], FLDA_single [Gao et al. 2008], AGL [Su et al. 2010], ESRC [Deng et al. 2012], SVDL [Yang et al. 2013], DMMA [Lu et al. 2013], LGR [Zhu et al. 2014],and LRA [Deng et al. 2014].

In the following experiments, except the experiment testing the impact of different neighbor sets, the neighbor sets are always fixed at $R = 1$. And the patch size is fixed...
at $11 \times 11$. In addition, except the experiments discussing the impact of the parameter $\kappa$, we always use $\kappa = 4$ to save computing time. The $l_1$-minimization in this paper is solved by Homotopy [Malioutov et al. 2005] due to its good accuracy and fast speed in the application of FR. The sparse parameter is 0.0001 and the iteration numbers are 200, 200, 100, 100 respectively for Extended Yale B, PIE, AR and LFW databases. All the experiments are conducted on a 2GHz machine with Xeon CPU and 32G RAM. Since the classification of each local block can be solved independently before combing decisions, we open 12 Matlab workers for parallel computation to improve the efficiency.

4.1. Extended Yale B database

The Extended Yale B face database [Georghiades et al. 2001] contains 38 human subjects under 9 poses and 64 illumination conditions. The 64 images of a subject in a particular pose are acquired at camera frame rate of 30 frames per second, so there is only small change in head pose and facial expression for those 64 images. However, its extreme lighting conditions still make it a challenging task for most face recognition methods. All frontal-face images marked with P00 are used in our experiment. The cropped and normalized $192 \times 168$ face images are captured under various laboratory-controlled lighting conditions [Lee et al. 2005] and resized to $80 \times 80$ in our experiments.

<table>
<thead>
<tr>
<th>Method</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>SRC</td>
<td>40.8 ± 5.8</td>
<td>56.5 ± 4.3</td>
<td>59 ± 3.1</td>
<td>61.2 ± 3</td>
<td>62.2 ± 4.4</td>
</tr>
<tr>
<td>CRC</td>
<td>41 ± 4.6</td>
<td>61.5 ± 4.2</td>
<td>65.6 ± 5.1</td>
<td>69.5 ± 3.9</td>
<td>72.5 ± 5.8</td>
</tr>
<tr>
<td>LS_SRC</td>
<td>86 ± 2.5</td>
<td>86.5 ± 2.5</td>
<td>87.4 ± 1.6</td>
<td>84.7 ± 1.9</td>
<td>84.7 ± 2.0</td>
</tr>
<tr>
<td>LS_CRC</td>
<td>90.8 ± 1.3</td>
<td>86.8 ± 1.0</td>
<td>86.2 ± 1.8</td>
<td>82.3 ± 3.3</td>
<td>83.5 ± 2.9</td>
</tr>
<tr>
<td>LS_SRC+Bayes</td>
<td>94.4 ± 2.3</td>
<td>96.5 ± 0.8</td>
<td>97.2 ± 1.1</td>
<td>97 ± 1.8</td>
<td>97 ± 1.6</td>
</tr>
<tr>
<td>LS_CRC+Bayes</td>
<td>95 ± 1.1</td>
<td>96.6 ± 0.9</td>
<td>96.4 ± 0.9</td>
<td>96.3 ± 0.4</td>
<td>96.3 ± 0.8</td>
</tr>
</tbody>
</table>

We first randomly divide all samples into two sessions, where each session has 32 images. Then we randomly select 1-5 samples from the first session and use their corresponding local blocks to construct local dictionary. 5 additional samples from the same session are selected for computing confusion matrix. Then 5 samples from the second session are randomly selected for testing. We conduct the testing for 5 times and the average result is reported. The experimental results are shown in Table I. It can be seen that our proposed methods always achieve the highest recognition rates with the dictionary sample size increasing from 1 to 5. Compared to LS_SRC and LS_CRC, LS_SRC+Bayes and LS_CRC+Bayes obtain much better results, which can fully validate the effectiveness of Bayesian inference based on confusion matrix. Different from SRC and CRC, our proposed methods do not rely on the dictionary sample size. Even when there is only one training sample for dictionary, they still get much better results. On the contrary, their performance may have a little decrease when the dictionary sample size is increasing. This is mainly because the intra-class variation also increases with the dictionary sample size increasing. If two images from the same subject have too drastic changes, some of their corresponding blocks will also have large variations and cannot be regarded as to be from the same subject. In fact, these blocks with large variations are regarded as noise rather than clean training samples for classification. In other words, the local dictionary includes more and more noise with the dictionary sample size increasing. This is why the performance of LS_SRC and LS_CRC will decrease when the dictionary sample size increases.
To further evaluate the performance of our proposed methods to SSPP problem, we compare them with a few specially designed methods for SSPP problem. For each subject, we use the image under the best illumination condition for training, whose azimuth and elevation are both 0 degree. The remaining 63 images are used for testing and the average results are reported. The average results of the five subsets of the Extended Yale B are also reported respectively. The details of the five subsets are provided in Table II. The experimental results are shown in Table III. As there is only one sample per person, the proposed Bayesian inference based on confusion matrix cannot be conducted. From the table, we can see that our proposed LS_SRC and LS_CRC still achieve the best results. They outperform FLDA, BlockFLD, PNN and AGL by 35.1%, 24.4%, 18.4% and 29.2% respectively. They are also superior to DMM, ESRC, SVTL and LGR, which respectively reach to 61.7%, 67.9%, 85% and 86.6% in [Zhu et al. 2014]. Although LRA also obtains a high recognition rate of 90% in [Deng et al. 2014], it’s still lower than LS_SRC and LS_CRC. Moreover, it’s only tested on two subsets of Extended Yale B (Subset2 and Subset3) with better illumination condition. Comparing with SRC and CRC, our proposed LS_SRC and LS_CRC achieve nearly 50% improvement. Although patch based SRC and CRC also obtain much better results than SRC and CRC, they are still not superior to LS_SRC and LS_CRC. For the two proposed methods (LS_SRC and LS_CRC), LS_CRC achieves the highest accuracy but also with significantly less time. In terms of CPU time, LS_SRC consumes more time than the other methods because they need to solve $l_1$-minimization for each block. However, we can significantly improve their computational efficiency by decreasing the number of blocks for classification. Table IV shows the recognition rates and time of LS_SRC and LS_CRC under different values of $\kappa$, in which the time of recognizing one test sample is reported. It can be noticed that they still output the best results but the time drops significantly. Since the number of individual classifiers increases with the value of $\kappa$ decreasing, Table IV also demonstrates the above-mentioned theorem that the correct classification probability after majority voting is monotonously increasing with the number of individual classifiers increasing.
Table IV. The impact of different $\kappa$ to LS_SRC and LS_CRC

<table>
<thead>
<tr>
<th>interval</th>
<th>LS_SRC</th>
<th></th>
<th>LS_CRC</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Accuracy(%)</td>
<td>Time(s)</td>
<td>Accuracy(%)</td>
<td>Time(s)</td>
</tr>
<tr>
<td>$\kappa = 2$</td>
<td>93.7 ± 7.6</td>
<td>2.1334</td>
<td>93.4 ± 7.6</td>
<td>0.043</td>
</tr>
<tr>
<td>$\kappa = 4$</td>
<td>92.2 ± 9.1</td>
<td>0.5285</td>
<td>92.5 ± 8.4</td>
<td>0.0095</td>
</tr>
<tr>
<td>$\kappa = 6$</td>
<td>91.7 ± 9.3</td>
<td>0.2590</td>
<td>91.7 ± 9.3</td>
<td>0.051</td>
</tr>
<tr>
<td>$\kappa = 8$</td>
<td>88.7 ± 11.8</td>
<td>0.1295</td>
<td>88.1 ± 12.1</td>
<td>0.0032</td>
</tr>
<tr>
<td>$\kappa = 10$</td>
<td>83.4 ± 17.2</td>
<td>0.0875</td>
<td>85.7 ± 14.1</td>
<td>0.0027</td>
</tr>
<tr>
<td>$\kappa = 12$</td>
<td>81.6 ± 18.3</td>
<td>0.0679</td>
<td>82.5 ± 17.6</td>
<td>0.0022</td>
</tr>
</tbody>
</table>

4.2. PIE database

The CMU PIE face database contains 68 subjects with 41368 face images as a whole [Sim et al. 2003]. Images of each person were taken across 13 different poses, under 43 different illumination conditions, and with 4 different expressions. All images have been cropped and resized to be 64 x 64 pixels.

Table V. Recognition rates (%) on PIE database

<table>
<thead>
<tr>
<th>Method</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>SRC</td>
<td>33.2 ± 11.9</td>
<td>32.7 ± 11.9</td>
<td>63.2 ± 20.0</td>
<td>66.1 ± 19.2</td>
<td>84.6 ± 11.6</td>
</tr>
<tr>
<td>CRC</td>
<td>32.5 ± 12.1</td>
<td>34.2 ± 12.5</td>
<td>62.2 ± 20.2</td>
<td>63.9 ± 18.3</td>
<td>82.3 ± 9.1</td>
</tr>
<tr>
<td>LS_SRC</td>
<td>34.8 ± 12.6</td>
<td>33 ± 11.8</td>
<td>81.9 ± 12.5</td>
<td>84.8 ± 11.1</td>
<td>91.6 ± 8.6</td>
</tr>
<tr>
<td>LS_CRC</td>
<td>30.3 ± 10.8</td>
<td>24 ± 9.3</td>
<td>71.3 ± 11.6</td>
<td>71.9 ± 11.1</td>
<td>78.9 ± 11.5</td>
</tr>
<tr>
<td>LS_SRC+Bayes</td>
<td>68.4 ± 9.8</td>
<td>69.3 ± 7.8</td>
<td>87.4 ± 8.9</td>
<td>89.7 ± 7.4</td>
<td>93.7 ± 7.9</td>
</tr>
<tr>
<td>LS_CRC+Bayes</td>
<td>63.4 ± 8.7</td>
<td>61.8 ± 9.0</td>
<td>80.8 ± 10.2</td>
<td>80.4 ± 7.9</td>
<td>85.8 ± 8.9</td>
</tr>
</tbody>
</table>

The images under five near frontal poses (C05, C07, C09, C27 and C29) are used in our experiment. We select 6 images with different illumination under each pose, thus we get 30 images for each individual. And we randomly divide all samples into two sessions, where each session has 15 images. Then, we first randomly select 1-5 training samples to construct dictionary and 5 additional samples for training confusion matrix from the first session. 5 samples from session 2 are randomly selected for testing. The testing experiment is conducted 5 times and the average result is reported. The classification results are shown in Table V. From the table, we can see that our proposed LS_SRC, LS_CRC, LS_SRC+Bayes, LS_CRC+Bayes still achieve the best results, and Bayesian inference based on confusion matrix bring further improvement for LS_SRC and LS_CRC.

Table VI. Recognition rates(%) PIE database for SSPP problem

<table>
<thead>
<tr>
<th>Method</th>
<th>SRC</th>
<th>CRC</th>
<th>PNN</th>
<th>BlockFLD</th>
<th>FLDA_single</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>61.9 ± 25.7</td>
<td>62.7 ± 25.1</td>
<td>72.2 ± 21.4</td>
<td>61.3 ± 22</td>
<td>50 ± 28</td>
</tr>
<tr>
<td>Method</td>
<td>PSRC</td>
<td>PCRC</td>
<td>AGL</td>
<td>LS_SRC</td>
<td>LS_CRC</td>
</tr>
<tr>
<td>Accuracy</td>
<td>62.7 ± 23.8</td>
<td>67.7 ± 21.3</td>
<td>81.4 ± 13.6</td>
<td>90 ± 13</td>
<td>86.4 ± 14.1</td>
</tr>
</tbody>
</table>

To further evaluate the performance of our proposed LS_SRC and LS_CRC, we also conduct experiments with single sample per person. The single image with good illumination from frontal pose C27 is used for training. The remaining images of each subject are used for testing. The experimental result is shown in Table VI. We can see that LS_SRC and LS_CRC are superior to AGL, PNN, BlockFLD and FLDA_single. According to the results published in [Deng et al. 2014], the recognition rate of LRA
is close to that of LS_CRC, but still lower than LS_SRC's recognition rate. Moreover, it's only tested with the frontal subset C27; while our methods are tested with all near frontal subsets (C05, C07, C09, C27 and C29). Compared to SRC, CRC, PSRC and PCRC, they lead to at least 18% improvement. The experimental results show that as long as the single training sample is good enough, our proposed methods can still perform well under minor pose variation.

4.3. AR database

The AR face database [Martinez 1998] contains over 4,000 color face images of 126 people (70 men and 56 women), including frontal views of faces with different facial expressions, lighting conditions and occlusions. The pictures of 120 individuals (65 men and 55 women) were taken in two sessions (separated by two weeks) and each session contains 13 color images. These 120 individuals are selected to use in our experiment. The images are resized to $32 \times 32$ and converted into gray scale.

Table VII. Recognition rates on AR database

<table>
<thead>
<tr>
<th>Method</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>SRC</td>
<td>50.5 ± 6.6</td>
<td>59.9 ± 6.1</td>
<td>64.5 ± 1.3</td>
<td>71.6 ± 3.0</td>
<td>65.1 ± 4.5</td>
</tr>
<tr>
<td>CRC</td>
<td>48.5 ± 7.6</td>
<td>52.8 ± 5.9</td>
<td>53.4 ± 5.1</td>
<td>52.9 ± 4.3</td>
<td>57.9 ± 1.0</td>
</tr>
<tr>
<td>LS_SRC</td>
<td>57.7 ± 9.1</td>
<td>79.9 ± 1.2</td>
<td>81.4 ± 1.4</td>
<td>82.9 ± 0.8</td>
<td>86 ± 1.0</td>
</tr>
<tr>
<td>LS_CRC</td>
<td>38.4 ± 10.2</td>
<td>61.0 ± 2.6</td>
<td>60.1 ± 3.3</td>
<td>61.2 ± 2.5</td>
<td>65.3 ± 2.5</td>
</tr>
<tr>
<td>LS_SRC+Bayes</td>
<td>71.7 ± 6.0</td>
<td>82.1 ± 1.2</td>
<td>83.7 ± 0.7</td>
<td>84.5 ± 1.2</td>
<td>86.6 ± 0.6</td>
</tr>
<tr>
<td>LS_CRC+Bayes</td>
<td>63.1 ± 5.8</td>
<td>71.8 ± 1.5</td>
<td>72.5 ± 1.8</td>
<td>74.9 ± 0.9</td>
<td>73.1 ± 0.6</td>
</tr>
</tbody>
</table>

For each subject, we randomly select 1-5 samples from session 1 to construct dictionary and 5 samples from session 2 for computing confusion matrix. Then 5 samples from session 2 are randomly selected for testing. The testing experiment is conducted five times and the average result is reported. The experimental results are shown in Table VII. We can find that our proposed methods still achieve the best results. However, when there is only one sample for constructing dictionary, LS_CRC gets the lowest accuracy. The reason is that the single training sample is randomly selected from AR database and the selected image may be with expression, bad illumination or occlusion.

Table VIII. Recognition rates(%) on session 1 of AR database for SSPP problem

<table>
<thead>
<tr>
<th>Method</th>
<th>expression</th>
<th>illumination</th>
<th>sunglasses</th>
<th>scarves</th>
<th>average</th>
</tr>
</thead>
<tbody>
<tr>
<td>SRC</td>
<td>95.3 ± 1.9</td>
<td>94.7 ± 0.5</td>
<td>88.1 ± 5.3</td>
<td>50.6 ± 10.4</td>
<td>82.2 ± 19.9</td>
</tr>
<tr>
<td>CRC</td>
<td>95.3 ± 2.1</td>
<td>93.9 ± 0.5</td>
<td>86.7 ± 6.7</td>
<td>49.2 ± 12.3</td>
<td>81.3 ± 20.6</td>
</tr>
<tr>
<td>PNN</td>
<td>85.8 ± 10.1</td>
<td>91.4 ± 1.9</td>
<td>70.8 ± 12.9</td>
<td>36.9 ± 5.9</td>
<td>71.3 ± 23.4</td>
</tr>
<tr>
<td>BlockFLD</td>
<td>72.5 ± 12.0</td>
<td>84.4 ± 3.2</td>
<td>72.5 ± 3.6</td>
<td>41.7 ± 7.5</td>
<td>67.8 ± 17.7</td>
</tr>
<tr>
<td>FLDA,single</td>
<td>93.1 ± 3.9</td>
<td>96.1 ± 13</td>
<td>89.4 ± 2.9</td>
<td>43.1 ± 6.5</td>
<td>80.4 ± 22.9</td>
</tr>
<tr>
<td>AGL</td>
<td>91.4 ± 6.7</td>
<td>98.9 ± 0.5</td>
<td>88.3 ± 4.6</td>
<td>71.7 ± 7.1</td>
<td>87.6 ± 11.4</td>
</tr>
<tr>
<td>PSRC</td>
<td>83.9 ± 7.5</td>
<td>93.3 ± 1.4</td>
<td>87.2 ± 2.7</td>
<td>63.9 ± 7.6</td>
<td>82.8 ± 12.5</td>
</tr>
<tr>
<td>PCRC</td>
<td>84.4 ± 4.9</td>
<td>91.4 ± 2.4</td>
<td>83.6 ± 3.4</td>
<td>63.6 ± 7.9</td>
<td>80.8 ± 11.7</td>
</tr>
<tr>
<td>LS_SRC</td>
<td>96.1 ± 2.1</td>
<td>98.3 ± 1.7</td>
<td>96.7 ± 2.2</td>
<td>84.5 ± 5.4</td>
<td>93.9 ± 6.4</td>
</tr>
<tr>
<td>LS_CRC</td>
<td>89.2 ± 5.8</td>
<td>92.8 ± 4.2</td>
<td>83.6 ± 7.7</td>
<td>69.5 ± 9.1</td>
<td>83.8 ± 11.0</td>
</tr>
</tbody>
</table>
To further demonstrate the performance of our proposed methods for SSPP problem, we conduct experiments with the single image under natural expression and illumination from session 1 for training. The remaining images of two sessions are used for testing. The classification results on the two sessions are shown in Table VIII and Table IX respectively. One can see from the tables that LS\_SRC achieves the best result. It’s not only robust to expression and illumination variations, but also shows great robustness to occlusion. In the experiment of session 1, it achieves the best average recognition rate of 93.9%, which is at least 6.3% higher than those of SRC, CRC, PSRC, PCRC, PNN, FLDA\_single, BlockFLD and AGL. LS\_SRC also outperforms the special designed sparse representation based methods ESRC [Deng et al. 2012] and SVDL [Yang et al. 2013] for SSPP problem. As described in [Deng et al. 2012] and [Zhu et al. 2014], ESRC only reaches to 89.4% with better image resolution on session 1 while SVDL achieves 87.6%. In the experiment of session 2, the proposed LS\_SRC still achieves the best result and exceeds SRC, CRC, PSRC, PCRC, PNN, FLDA\_single, BlockFLD and AGL by at least 14.4%. Since samples from session 2 are collected after two weeks of session 1, the best result of LS\_SRC also demonstrates its robustness to time variation. Therefore, we can conclude once again that as long as the single training image is good enough, LS\_SRC will obtain excellent performance even under occlusion and time variation. In the contrary, patch based SRC and CRC (PSRC and PCRC) are not superior to SRC and CRC in the experiments of both sessions. This demonstrates that LS\_SRC has better generalization ability than PSRC and the local structure is more effective than simple patch based methods. In addition, different from the experiments for SSPP problem on Extended Yale B and PIE databases, LS\_SRC does not show very competitive performance on AR database. This is because CRC is sensitive to occlusion and time variation and LS\_CRC inherits these shortcomings. Therefore, only in the circumstances without occlusion and time variation, LS\_CRC will be considered. Except the methods listed in the tables, we also compare our methods with the latest published methods LRA [Deng et al. 2014] and DMMA [Lu et al. 2013]. LRA and DMMA achieve the average recognition rates of 49.4% and 55.1% respectively on both two sessions; while the average recognition rate of LS\_SRC is up to 82.3%. Although the recognition rate of LRA can be further improved to 72.4% and 90.1% by using Gabor and LBP feature respectively, their image resolutions are respectively \( 128 \times 128 \) and \( 165 \times 120 \) which are much larger than \( 32 \times 32 \) in our experiments.

<table>
<thead>
<tr>
<th>Method</th>
<th>expression</th>
<th>illumination</th>
<th>sunglasses</th>
<th>scarves</th>
<th>average</th>
</tr>
</thead>
<tbody>
<tr>
<td>SRC</td>
<td>63.5 ± 7.5</td>
<td>62.2 ± 4.1</td>
<td>46.9 ± 7.1</td>
<td>25.8 ± 6.0</td>
<td>49.7 ± 16.8</td>
</tr>
<tr>
<td>CRC</td>
<td>62.5 ± 7.1</td>
<td>60.8 ± 4.6</td>
<td>41.7 ± 6.5</td>
<td>22.2 ± 6.6</td>
<td>47.1 ± 17.5</td>
</tr>
<tr>
<td>PNN</td>
<td>52.5 ± 11.2</td>
<td>51.9 ± 4.3</td>
<td>34.2 ± 6.0</td>
<td>17.8 ± 2.1</td>
<td>39.1 ± 16.0</td>
</tr>
<tr>
<td>BlockFLD</td>
<td>33.1 ± 12.1</td>
<td>41.1 ± 4.6</td>
<td>30.7 ± 5.8</td>
<td>22.5 ± 2.2</td>
<td>31.7 ± 9.3</td>
</tr>
<tr>
<td>FLDA_single</td>
<td>57.2 ± 8.0</td>
<td>53.3 ± 1.4</td>
<td>46.1 ± 8.7</td>
<td>20.0 ± 3.6</td>
<td>44.2 ± 16.1</td>
</tr>
<tr>
<td>AGL</td>
<td>62.5 ± 8.0</td>
<td>67.2 ± 4.2</td>
<td>55 ± 1.4</td>
<td>40.3 ± 6.3</td>
<td>56.3 ± 11.7</td>
</tr>
<tr>
<td>PSRC</td>
<td>47.2 ± 7.2</td>
<td>53.1 ± 1.7</td>
<td>47.5 ± 8.3</td>
<td>38.1 ± 4.8</td>
<td>46.5 ± 7.6</td>
</tr>
<tr>
<td>PCRC</td>
<td>43.9 ± 9.4</td>
<td>48.3 ± 3.8</td>
<td>38.3 ± 8.2</td>
<td>32.5 ± 5.2</td>
<td>40.8 ± 8.6</td>
</tr>
<tr>
<td>LS_SRC</td>
<td>74.4 ± 9.6</td>
<td>76.4 ± 9.9</td>
<td>70 ± 7.5</td>
<td>61.9 ± 7.6</td>
<td>70.7 ± 9.4</td>
</tr>
<tr>
<td>LS_CRC</td>
<td>57.2 ± 11.4</td>
<td>63.9 ± 8.4</td>
<td>38.9 ± 12.5</td>
<td>34.4 ± 6.3</td>
<td>48.6 ± 15.4</td>
</tr>
</tbody>
</table>

Table IX. Recognition rates(%) on session 2 of AR database for SSPP problem
4.4. LFW database

The LFW database [Huang et al. 2007] contains images from 5749 different subjects in unconstrained environment. LFW-a is a version of LFW after alignment using commercial face alignment software [Wolf et al. 2010]. We gathered 158 subjects with no less than 10 samples from LFW-a and collected 10 near frontal face images for each subject. The images are firstly cropped to $120 \times 120$ and then resized to $32 \times 32$.

<table>
<thead>
<tr>
<th>Method</th>
<th>SRC</th>
<th>CRC</th>
<th>PNN</th>
<th>BlockFLD</th>
<th>FLDA_single</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>34.5 ± 2.8</td>
<td>23.5 ± 1.8</td>
<td>21.1 ± 0.9</td>
<td><strong>40 ± 2.4</strong></td>
<td>25.6 ± 3.7</td>
</tr>
</tbody>
</table>

Table X. Recognition rates(%) on LFW database for SSPP problem

In the experiment, only one sample is randomly chosen for training and another 5 samples for test. The program is run for five times and the average results are reported. The results are listed in Table X. Although face alignment has been conducted, the variations in this database are still very large. Therefore, no method achieves very high accuracy. Nonetheless, LS_SRC still achieves the best result among all methods. Comparing with PSRC and PCRC, LS_SRC and LS_CRC respectively lead to 5.5% and 3.1% improvement. Although ESRC, SVDL and LGR also obtain relatively good results in [Zhu et al. 2014], they are still much lower than LS_SRC. Moreover, their image resolution ($80 \times 80$) is much larger than that of our experiments.

4.5. Effects of Different Parameters

In this section, we will discuss the influence of parameter setting of patch size, neighbor radius $R$. To fully show the influence of the two parameters, we conduct experiments on PIE and AR databases respectively.

![Fig. 5](image)

Fig. 5. The impact of different neighbor sets to LS_SRC and LS_CRC.

To indicate the impact of different neighbor sets to our methods, we also conduct an additional experiment on PIE database. In the experiment, LS_SRC and LS_CRC are tested with different neighbor sets under the patch scale $11 \times 11$. The neighbor sets vary with the radius $R$, such as $P = 8, R = 1, P = 16, R = 2, P = 24, R = 3, P = 32, R = 4$ and so on. As shown in Fig. 5, the recognition rates of LS_SRC and LS_CRC degrade with the radius $R$ increasing. The main reason is that the subspace assumption starts to...
crumble as the radius $R$ increasing. To prove this point, we reconstruct some face images under different $R$. According to the subspace assumption, given a central patch $x_i^0$ and its neighbor patches $x_i^j, j = 1, \ldots, P$ in the $i$-th pixel centralized block, the central patch can be approximately represented by a linear combination of neighbor patches. The representation coefficients can be solved by ridge regression. Then the central patch can be reconstructed by multiplying neighbor patches with the representation coefficients. Fig. 6 shows the reconstructed images, from which we can find that they become blurry with $R$ increasing. Therefore, for our proposed methods, $P = 8, R = 1$ is the best, which yields the best classification results with the least computing time.

Fig. 6. Reconstructed face images under different $R$ (Face Image courtesy CMU Pose, Illumination, and Expression (PIE) database).

For the proposed methods, the patch size will have a great impact on the recognition performance. To demonstrate this point, we conduct experiments with different patch sizes on AR database. Fig. 7 shows the FR accuracies of LS_SRC and LS_CRC under different patch sizes. It can be observed that the patch size has a great impact on the recognition performance especially when the patch size is too small. When the patch size is too small, it does not contain enough information to distinguish different person. Therefore, the performance drops significantly. When patch size is too big, the recognition rates of LS_SRC and LS_CRC under occlusion cases also degrade. This is because those images with occlusion will contain many corrupted patches whose error classification results will disturb final decision. Moreover, the number of corrupted patches increases when the patch size becomes larger. However, LS_SRC is more robust to occlusion than LS_CRC. Even when the patch size reaches up to $59 \times 59$, the recognition rate of LS_SRC with sunglasses still achieves 91.7%. Meanwhile, LS_SRC also shows greater robustness to expression and illumination variations than LS_CRC. The performance of LS_SRC changes slightly with the patch size increasing under these cases. Considering the fact that larger patch size will result to more computing time, we select $11 \times 11$ as a relatively good patch size for all databases.

Fig. 7. The impact of different patch size to LS_SRC and LS_CRC.
5. CONCLUSION
In this paper, we propose a series of methods based on local structure. Motivated by the “divide-conquer” strategy, we partition each face into a set of overlapped blocks and classify each block, then aggregate the classification results to make the final decision. To make classification on each local block, we further divide block into overlapped local patches and assume that they lie in a linear subspace. This assumption reflects the local structure relationship of the patches and makes SRC and CRC feasible even when encountering SSPP problem. Based on local structure, we first propose the local structure based sparse representation classification (LS_SRC). To lighten the computing burden of LS_SRC, we then present the local structure based collaborative representation classification (LS_CRC). To further improve the performance, we propose LS_SRC+Bayes and LS_CRC+Bayes which take advantages of the confusion matrix of the classifier by Bayesian inference. Experimental results show that confusion matrix greatly improves the performance of LS_SRC and LS_CRC. LS_SRC and LS_CRC generalize well to SSPP problem and outperform many state-of-the-art methods. Moreover, they also show strong robustness to the large variations of expression and illumination, little poses variation, occlusion and time variation. However, as LS_CRC inherits the shortcomings of CRC, it will only be suggested when there is no occlusion and time variation. In addition, the proposed methods also rely on the basis that all the training and testing images are well aligned. Therefore, drastic pose variation will severely degrade their performance. This problem is planned to be solved in our future work.

ACKNOWLEDGMENTS
This work was partially supported by the 973 Program (Project No. 2014CB347600), the Natural Science Foundation of Jiangsu Province under Grant BK2012033, the Program for New Century Excellent Talents in University under Grant NCET-12-0632, the National Nature Science Foundation of China under Grant 61202133 and the Research Fund for the Doctoral Program of Higher Education of China under Grant 20122219120027.

REFERENCES

ACM Transactions on Intelligent Systems and Technology, Vol. V, No. N, Article 39, Publication date: January YYYY.
Local Structure based Sparse Representation for Face Recognition


Athinodoros S. Georgiades, Peter N. Belhumeur, and David Kriegman. 2001. From few to many: Illumination cone models for face recognition under variable lighting and pose. Pattern Analysis and Machine Intelligence, IEEE Transactions on 23, 6 (2001), 643–660. DOI: http://dx.doi.org/10.1109/34.927464


Zechar Li, Jing Liu, Jinhui Tang, and Hanqing Lu. 2015. Robust Structured Subspace Learning for Data Representation. IEEE Transactions on Pattern Analysis and Machine Intelligence (2015), 1–1. DOI: http://dx.doi.org/10.1109/jeticas.2013.2256752

Zechar Li, Jing Liu, Yi Yang, Xiaofang Zhou, and Hanqing Lu. 2014. Clustering-Guided Sparse Structural Learning for Unsupervised Feature Selection. IEEE Transactions on Knowledge and Data Engineering 26, 9 (2014), 2138–2150. DOI: http://dx.doi.org/10.1109/tdke.2013.65


ACM Transactions on Intelligent Systems and Technology, Vol. V, No. N, Article 39, Publication date: January YYYY.


Shuicheng Yan, Dong Xu, Benyu Zhang, Hong-Jiang Zhang, Qiang Yang, and Stephen Lin. 2007. Graph embedding and extensions: a general framework for dimensionality reduction. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 29, 1 (2007), 40–51. DOI: http://dx.doi.org/10.1109/tpami.2007.250598


