A parameterized fuzzy adaptive K-SVD approach for the multi-classes study of pursuit algorithms

Xiaoning Song a,b,c,n, Zi Liu b,c, Xibei Yang a,b,c, Jingyu Yang b

A R T I C L E   I N F O

Article history:
Received 15 September 2012
Received in revised form
20 June 2013
Accepted 22 June 2013
Available online 5 July 2013

Keywords:
Sparse representation
Parameterized fuzzy set
Optimization recovery
K-SVD
Image recognition

A B S T R A C T

Sparse representations using over complete dictionaries has concentrated mainly on the study of pursuit algorithms that decompose signals with respect to a given dictionary. Designing dictionaries to better fit the above model can be done by either selecting one from a pre-specified set of linear transforms or adapting the dictionary to a set of training signals. The K-SVD algorithm is an iterative method that alternates between sparse coding of the examples based on the current dictionary and a process of updating the dictionary atoms to better fit the data. However, the existing K-SVD algorithm is employed to dwell on the concept of a binary class assignment, which means that the multi-classes samples are assigned to the given classes definitely. The work proposed in this paper provides a novel parameterized fuzzy adaptive way to adapting dictionaries. In order to achieve the fuzzy sparse signal representations, the update of the dictionary columns is combined with an update of the sparse representations by embedding a new mechanism of fuzzy set, which is called parameterized fuzzy K-SVD. Experimental results conducted on the ORL, Yale and FERET face databases demonstrate the effectiveness of the proposed method.

1. Introduction

Sparse signal representations using over complete dictionaries are used in a variety of fields such as pattern recognition, image and video coding [1–6]. Over-completeness of a set means that it has more members than the dimensionality of the members. The goal of an over complete set of basis signals (i.e. the dictionary) is to express input signals as sparse linear combinations of the dictionary members. The advantage of over-completeness of a dictionary is its robustness in case of noisy or degraded signals. Also, it introduces greater variety of shapes in the dictionary, thus leading to sparse representations of a variety of input signals.

Applications that can benefit from the sparseness and over-completeness concepts (together or separately) include compression, regularization in inverse problems, feature extraction, etc. Indeed, the success of the JPEG2000 coding standard can be attributed to the sparseness of the wavelet coefficients of natural images [7]. In de-noising, wavelet methods and shift-invariant variations that exploit over-complete representation are among the most effective known algorithms for this task [8]. Sparseness and over-completeness have been successfully used for dynamic range compression in images, separation of texture and cartoon content in images [9], image restoration, etc.

Evidently, extraction of the best sparse representation is a hard problem that has been extensively investigated in the past few years. The K-SVD algorithm is an iterative method. It alternates between sparse coding of the examples based on the current dictionary and a process of updating the dictionary atoms to better fit the data, which has been one of the effective algorithms due to its power of optimization update rule for the dictionary. However, the existing K-SVD algorithm is employed to dwell on the concept of a binary class assignment, which means that the multi-classes samples are assigned to the given classes definitely. Evidently, as the samples are significantly affected by numerous environmental conditions (such as illumination, expression, etc.), it is advantageous to investigate these factors and quantify their impact on their “internal” class assignment [10].

In the present study, we propose a novel dictionary learning technique that discovers a parameterized fuzzy adaptive way to adapting dictionaries. In order to achieve the fuzzy sparse signal representations, the update of the dictionary columns is combined with an update of the sparse representations by embedding a new mechanism of fuzzy set. Using PF-K-SVD approach for updating dictionary coefficients helps prune seldom used elements. If there are many similar-looking elements, this approach helps strengthen...
only those elements that are frequently used or can be used in place of the others.

As discussed above, once the sparse coding task is done, a second stage is performed to search for a better dictionary [11]. The proposed method will show best among the other evaluated algorithms, which validates the effectiveness of the learning mode that is incorporated fuzzy manner. The superiority of proposed method stems from two aspects. On the one hand, the fuzzy membership function based on the parameter estimation is redefined to be a reformatory one, which is more effective and robust compared with that of conventional fuzzy learning approaches. On the other hand, since the objective of the sparse representation stage is to find out the best representation for each member in the training signals, the supervised PF-K-SVD algorithm is developed on the basis of the grade of membership, and then applied to perform the pattern classification.

The rest of the paper is organized as follows: Section 2 describes the K-SVD algorithm. PF-K-SVD algorithm is presented in Section 3. Experimental results are reported in Section 4 such as ORL, Yale and FERET. Section 5 gives several discussions to the experiments. Finally, concluding comments are included in Section 6.

2. Outline of the K-SVD algorithm

A different update rule for the dictionary can be proposed, in which the atoms (i.e., columns) in dictionary are handled sequentially. This leads to the K-SVD algorithm, as developed by Aharon et al. [12]. Keeping all the columns fixed apart from the \( j_0 \)-th one, \( \alpha_{j_0} \), this column can be updated along with the coefficients that multiply it in \( X \). We isolate the dependency on \( \alpha_{j_0} \) by rewriting minimization as

\[
\|Y - AX\|^2 = \left\| \left( Y - \sum_{j \neq j_0} \alpha_j x_j^T \right) - \alpha_{j_0} x_{j_0}^T \right\|^2 _p
\]

In this description, \( x_{j_0}^T \) stands for the \( j \)-th row of \( X \). The update step targets both \( \alpha_{j_0} \) and \( x_{j_0}^T \), and refers to the term in parentheses,

\[
E_{j_0} = Y - \sum_{j \neq j_0} \alpha_j x_j^T
\]

as a known pre-computed error matrix.

The optimal \( \alpha_{j_0} \) and \( x_{j_0}^T \) minimizing Eq. (1) are the rank-1 approximation of \( E_{j_0} \), and can be obtained via an SVD, but this typically would yield a dense vector \( x_{j_0}^T \) implying that we increase the number of non-zeros in the representations in \( X \).

In order to minimize this term while keeping the cardinalities of all the representations fixed, a subset of the columns of \( E_{j_0} \) should be taken, those that correspond to the signals from the example-set that are using the \( j_0 \)-th atom, namely those columns where the entries in the row \( x_{j_0}^T \) are non-zero. This way, we allow only the existing non-zero coefficients in \( x_{j_0}^T \) to vary, and the cardinalities are preserved. Therefore, we define a restriction operator \( P_{j_0} \), that multiplies \( E_{j_0} \) from the right to remove the non-relevant columns. The matrix \( P_{j_0} \) has \( M \) rows (the number of overall examples), and \( M_{j_0} \) columns (the number of examples using the \( j_0 \)-th atom). We define \( (x_{j_0}^T)^T = x_{j_0}^T \), as the restriction on the row \( x_{j_0}^T \), choosing the non-zero entries only.

For the sub-matrix, \( E_{j_0} \), a rank-1 approximation via SVD can be applied, updating both the atom \( \alpha_{j_0} \) and the corresponding coefficients in the sparse representations \( x_{j_0}^T \). This simultaneous update may lead to a substantial speedup in the convergence of the training algorithm.

3. A parameterized fuzzy adaptive K-SVD learning method

The K-SVD algorithm has been one of the effective algorithms due to its power of optimization update rule for the dictionary. However, the existing K-SVD algorithm is employed to dwell on the concept of a binary class assignment, that is, the multi-classes samples are assigned to the given classes definitely. Evidently, as the samples are significantly affected by numerous environmental conditions, it is advantageous to investigate these factors and quantify their impact on their “internal” class assignment.

In this section, we propose a novel dictionary learning technique that discovers a parameterized fuzzy adaptive way to adapting dictionaries. In order to achieve the fuzzy sparse signal representations, the update of the dictionary columns is combined with an update of the sparse representations by embedding a new mechanism of fuzzy set. Moreover, the need for such a novel fuzzy model construction is reduced to parameter estimation when the structure is given beforehand, that is, the parameter estimation method must recursively process the measured data as they become available. Using the novel parameterized fuzzy K-SVD (PF-K-SVD) approach for updating dictionary coefficients helps prune seldom used elements. If there are many similar-looking elements, this approach helps strengthen only those elements that are frequently used.

3.1. Outline of conventional fuzzy multi-classes LDA

As researched in Ref. [13], Kwak et al. proposed to incorporate a gradual level of assignment to class, which is regarded as a membership grade with anticipation that such discrimination helps improve classification results.

Suppose we have \( C \) known patterns \( w_1, w_2, ..., w_C \), and \( X = (x_1, x_2, ..., x_N) \) is a set of samples with \( m \)-dimension. Each sample in \( X \) belongs to one of the known classes \( w_j \), i.e. \( x_i \in w_j, i = 1, ..., N, j = 1, ..., C \). Using the FKNN [14] algorithm, the computations of the fuzzy membership grades can be realized through a sequence of steps.

The results of the FKNN are used in the computations of the statistical properties of the patterns, such as mean value and scatter covariance matrices. Taking into account the fuzzy membership grades (i.e., the membership grade to class “\( i \”) for \( j \)-th pattern), the mean vector of each class \( f_j \), is

\[
f_m = \sum_{i=1}^{N} \alpha_{ij} x_i
\]

Therefore, the class center matrix \( FM \) and the fuzzy membership matrix \( U \) can be achieved with the result of FKNN.

\[
U = [\alpha_{ij}], \quad i = 1, 2, ..., C; \quad j = 1, 2, ..., N
\]

\[
FM = [f_m], \quad i = 1, 2, ..., C
\]

The within-class fuzzy scatter matrix \( FS_w \) and between-class fuzzy scatter matrix \( FS_b \) incorporate the membership values in their calculations

\[
FS_w = \sum_{i=1}^{C} \left( \sum_{x_i \in w_j} (x_i - \bar{m}_j)(x_i - \bar{m}_j)^T \right)
\]

\[
FS_b = \sum_{i=1}^{C} (f_m - \bar{m})(f_m - \bar{m})^T
\]

where \( \bar{m} \) stands for the mean of all vectors (images). The optimal fuzzy projection \( \phi_{LDA} \) and the feature vector transformed by the F-LDA method follows the expressions

\[
\phi_{LDA} = \arg \max_{\phi} \left[ \frac{\phi^T FS_b \phi}{\phi^T FS_w \phi} \right] = \left[ \phi_1, \phi_2, ..., \phi_m \right]
\]
where \( \{ \phi_i | i = 1, 2, ..., m \} \) is the set of generalized eigenvectors (discriminant vectors) of \( F_{S_b} \) and \( F_{S_w} \) corresponding to the \( C-1 \) largest generalized eigenvalues \( \{ \lambda_i | i = 1, 2, ..., m \} \), that is
\[
F_{S_b} \phi_i = \lambda_i F_{S_w} \phi_i, \quad i = 1, 2, ..., m
\]

(9)

3.2. Establish a parameterized fuzzy mechanism

Up to now, the F-LDA method combined with KNN algorithm is considered to regain the statistical properties of the patterns. However, after investigating the membership allocation formula, we find that the method attempts to “fuzziness” only by fuzzy regulating the each class center. How can we make full use of the distribution information of each sample to the redefinition of scatter matrices? In our previous research in Ref. [13], we extended the F-LDA presented by Kwak and included complete fuzziness in the calculation of between-class scatter matrix and within-class scatter matrix. By this means, a relaxed normalized fuzzy membership condition is presented to achieve the distribution information of each sample.

In this method, the first key step of the reformatory F-LDA (RF-LDA) method is how to address the problem coming under the influence of the outliers in the patterns. As shown there, the membership matrix denoted by \( U = [u_{ij}] \) for \( i = 1, 2, ..., C \) and \( j = 1, 2, ..., N \) satisfies the obvious property \( \sum_{i=1}^{C} u_{ij} = 1 \). Taking into account the fact that the outliers may have some inverse influence to the feature extraction performance, a relaxed normalized condition in the fuzzy membership degrees is proposed as follows:
\[
\sum_{i=1}^{C} \sum_{j=1}^{N} \mu_i(x_j) = N
\]

(10)

By condition (10), we can redefine the fuzzy membership as follows:
\[
\rho_j = \begin{cases} 
\gamma + (1-\gamma) \cdot (n_{ij}/k) & \text{if class}(j) \\
(1-\gamma) \cdot (n_{ij}/k) & \text{otherwise}
\end{cases}
\]

(11)

where, \( \gamma = (N-C)/(2^mN) \), \( n_{ij} \) stands for the number of the neighbors of the \( j \)-th pattern that belong to the \( i \)-th class, and \( \gamma \) are constants that ultimately control the value of \( \rho_j \), and satisfy the constraints \( \rho_j(0,1), \gamma(0,1) \). Since the normalized condition of the membership degree has been relaxed, all the samples coming with use of membership matrix \( U \) are insensitive to the class center of each pattern. Thus, it shows high stability of this normalized condition compared to others due to its power of regulating the fuzzy membership grades.

The second key step of this method is how to incorporate the contribution of each training sample into the redefinition of scatter matrices. More specifically, the membership grade of each sample (contribution to each class) should be considered and the corresponding fuzzy within-class scatter matrix can be redefined as follows:
\[
RFS_w = \sum_{i=1}^{C} \left( \sum_{x_j \in S_i} u_{ij}^p (x_j - m_i)(x_j - m_i)^T \right)
\]

(12)

where \( p \) is a parameter that controls the influence of fuzzy membership grade in the redefinition of within-class scatter matrix.

Similarly, the fuzzy between-class scatter matrix can be redefined as follows:
\[
RFS_b = \sum_{i=1}^{C} \left( 1 - \sum_{j=1}^{N} u_{ij}^p \right) \sum_{j=1}^{N} u_{ij}^p (m_j - m)(m_j - m)^T
\]

(13)

where \( p \) is the parameter that controls the influence of fuzzy membership grade in the redefinition of between-class scatter matrix, \( m \) stands for the mean of all vectors (images). Therefore, the RF-LDA shows high stability compared with others due to its power of regulating the fuzzy membership grades.

3.3. Apply HNN to parameter estimation of RF-LDA

As discussed above, in our previous research, we proposed a reformatory version of F-LDA (RF-LDA). Nevertheless, how can we dynamically assign the particular value of offset in the calculation of the grade of membership? There is no reason reported in our previous research. In this section, in the line of previous arguments, we approach the problem of control parameter estimation of the RF-LDA model by considering the formulation of a HNN [15–17], which is named HRF-LDA. By this means, as a control parameter in formula (11), \( m \) is dynamically assigned the particular value of offset in the calculation of the grade of membership. Therefore, the next work is as follows.

Consider a network with \( N \) neurons indexed in neurons indexed in \( (N) = \{1, ..., N\} \), let \( p_i \) denote the total input to neuron \( i \), defined as
\[
dp_{in}(t) = -\left( \sum_{j=1}^{N} W_{ij} s_j(t) + I_i(t) \right)
\]

(14)

where \( s_j \) represents the state (or output) of neuron \( j \), \( W_{ij} \) the weight associated with the connection from neuron \( j \) to neuron \( i \), and \( I_i \) the bias (or external input) of neuron \( i \). The total input state relation for neuron \( i \) is given by
\[
s_i(t) = a \tan\left( \frac{p_i(t)}{\beta} \right)
\]

(15)

where \( a, \beta > 0 \), and therefore
\[
\forall i \in (N), \quad t \geq t_0, \quad s_i(t) \in [-a, a]
\]

(16)

Using matrix notation, this HNN is thus given by
\[
\begin{cases} 
\dot{\mu}_{ij}(t) = -\left( W(t) s_j(t) + I_i(t) \right) \\
s_j(t) = a \tan\left( \frac{p_j(t)}{\beta} \right)
\end{cases}
\]

(17)

where \( p(t), s(t), I(t) \in \mathbb{R}^{m+1} \) and \( W(t) \in \mathbb{R}^{N \times N} \), yielding the state space representation,
\[
ds_{in}(t) = -\frac{1}{a^2} D_{in}(s(t))(W(t)s(t) + I(t))
\]

(18)

where
\[
D_{in}(s(t)) = \text{diag}(\alpha^2 - s_j(t)^2)_{j \in (N)}
\]

(19)

is positive definite and invertible by Eq. (16). Henceforth, we shall refer to Eq. (18) as a HNN.

As researched in Ref. [17], motivated by the methodology adopted for the time-invariant case, we start by considering a HNN whose state at time \( t \), \( s(t) \), is taken as the estimate \( \hat{h}(t) \) of the sought parameterization \( \theta \), i.e., \( \hat{h}(t) = s(t) \). Implicitly, we are considering a network with as many neurons as the parameters to estimate, i.e. \( N = n \). Therefore, the present work mentioned above aims at improving the control parameter of the membership function of RF-LDA by setting up the problem for optimizing the parameter \( m \) as shown in Eq. (11). In fact, the value of \( m \) found by HNN depends on the different search ranges in various feature spaces of data sources. The optimal values of \( m \) that result in different experiments indicate that the error surface has multimodal minima.

3.4. How to incorporate the fuzzy mechanism into dictionary-update stage of K-SVD

Now, the most important key step of our method is how to incorporate the contribution of each training sample into objective function of the updating dictionary coefficients. The algorithm consists of a training phase (that can be done off-line) and a reconstruction phase, and it performs the reconstruction on the
test images using the trained model from the training phase. Each of patches undergoes a pre-processing stage that extracts features from the parameterized RF-LDA. Dimensionality reduction is also applied on the features, which makes the dictionary training step much faster. By utilizing the parameterized fuzzy algorithm discussed above, each sample can be classified into multi-classes under the membership degrees of the labeled patterns. Thus, a novel dictionary learning technique that discovers a fuzzy adaptive way to adapting dictionaries is proposed. The optimal dictionary design algorithm is redefined by following the objective function:

\[
J_k = \sum_{i=1}^{N} \sum_{j=1}^{C} [\mu_i(X_j)^2] ||Y - AX||^2
\]

\[
= \sum_{i=1}^{N} \sum_{j=1}^{C} [\mu_i(X_j)^2] \left[\left(Y - \sum_{j=1}^{C} h_j X^*_j \right) - h_k X^*_k \right]^2
\]

\[
= \sum_{i=1}^{N} \sum_{j=1}^{C} [\mu_i(X_j)^2] ||E_k - h_k X^*_k ||^2
\]

(20)

Since the normalized condition of the membership grade has been proposed, all the samples coming with use of membership matrix \(U\) are insensitive to the class center of each pattern. Thus, it shows high stability of this improved normalized condition compared to others due to its power of regulating the fuzzy membership grades. Also, the matrix \(E_k\) stands for the error for all the examples when the \(k\)-th atom is removed. Here, it would be tempting to suggest the use of the parameterized fuzzy adaptive K-SVD (PF-K-SVD) to find alternative \(h_k\) and \(\theta_j\). The PF-K-SVD utilizes an update of the sparse representations by embedding the proposed mechanism of fuzzy set. It finds the closest rank-1 matrix that approximates \(E_k\), which will effectively minimize the error as defined in formula (20).

3.5. The detailed PF-K-SVD algorithm

The detailed PF-K-SVD algorithm is described as follows.

**PF-K-SVD Task:** Train a dictionary \(A\) to sparsely represent the data \(y_i\), by approximating the solution to the optimization problem.

**Step 1.** Each of patches undergoes a pre-processing stage that extracts features from the parameterized RF-LDA. PCA dimensionality reduction is also applied on the features, making the dictionary training step much faster.

**Step 2.** Initialization: initialize \(k = 0\), and initialize dictionary. Build \(A_0 = R^{C \times N}\), either by using random entries, or using \(m\) randomly chosen examples.

**Step 3.** With the proposed parameterized fuzzy algorithm, the computations of the fuzzy membership grade of each sample can be realized through a sequence of steps described in Section 3. Thus, the fuzzy membership matrix \(\mu_i(X_j)\) can be achieved with the result of the method.

**Step 4.** Normalization: normalize the columns of \(A_0\).

**Step 5.** Sparse Coding Stage: use a pursuit algorithm to approximate the solution of

\[
\hat{x}_i = \arg \min_x ||y_i - A_k x||^2 \quad \text{subject to } ||x||_0 \leq k_0
\]

obtaining sparse representations \(\hat{x}_i\) for \(1 \leq i \leq M\). These form the matrix \(X_k\).

**Step 6.** Fuzzy K-SVD Dictionary-Update Stage: use the following procedure to update the columns of the dictionary and obtain \(A_k\):

repeat for \(j_0 = 1, 2, ..., m\).

Define the group of examples that use the atom \(a_{j_0}\).

\(\Omega_{j_0} = \{i|1 \leq i \leq M, x_i[j_0] \neq 0\}\)

Compute the residual matrix

\(E_{j_0} = \sum_{i=1}^{N} \sum_{j=1}^{C} [\mu_i(X_j)^2] \left(Y - \sum_{j=1}^{C} a_{j_0} X^*_j \right)\)

where \(x_i^j\) are the \(j\)-th rows in the matrix \(X_k\).

Restrict \(E_{j_0}\) by choosing only the columns corresponding to \(\Omega_{j_0}\), and obtain \(E_{j_0}^R\).

Apply SVD decomposition \(E_{j_0}^R = U \Sigma V^T\). Update the dictionary atom \(a_{j_0} = u_1\), and the representations by \(x_i^j = \bar{a}([1, 1] v_1)\).

**Step 7.** Stopping Rule: if the change in \(J_k = \sum_{i=1}^{N} \sum_{j=1}^{C} [\mu_i(X_j)^2] ||E_k - h_k X^*_k ||^2\) small enough, stop. Otherwise, apply another iteration.

**Output:** the desired result is \(A_k\).

4. Experimental results

The proposed method is used for face recognition and tested on the ORL [18], Yale [19] and FERET [20] face image database. To evaluate the proposed method properly, we also include experimental results for the D-LDA [21], C-LDA [22], R-LDA [23], 2D-LDA [24], LPP [25], NPE [26]. For its simplicity, the \(k\) nearest neighbor (\(k\)-NN) classifier with Euclidean distance is employed for the classification, here, the parameter of \(k\)-NN is fixed as \(k = 3\). In this work, the score of the proposed method does slightly better in accuracy on some face experimental databases. At this level of accuracy, the advantage of a method over another one cannot be easily justified. To make full use of the available data and to evaluate the generalization power of algorithms more accurately, the figures of merit are success rates averaged over 20 runs, and the error margin for both methods (the mean and standard deviations) is provided in the following experiments.

4.1. On ORL database

The ORL contains a set of faces taken between April 1992 and April 1994 at the Olivetti Research Laboratory in Cambridge. It contains 40 distinct persons with 10 images per subject. The images were taken at
different time instances, with varying lighting conditions, facial expressions, and facial details. All persons are in the up-right, frontal position, with tolerance for some side movement. Some sample images from the ORL database are shown in Fig. 1.

We randomly choose \( \theta (\theta = 3, 4, 5) \) images per individual for training, and the rest images are used for testing. To make full use of the available data and to evaluate the generalization power of algorithms more accurately, 10 experiments were performed. The final result was the average recognition rate over the 10 random training sets. The proposed algorithm is implemented in MATLAB using fuzzy optimized implementation for K-SVD and OMP algorithms.

Moreover, as described in Section 3.3, how to dynamically tune the parameter \( m \) is a first crucial problem for the RF-LDA approach. More specifically, \( m \) is a parameter which controls the influence of fuzzy membership grade in the redefinition of each scatter matrix, which satisfies the constraint \( m \in (0, 1) \). Thus, with variation of \( m \), the discriminatory capability of RF-LDA is gradually changed due to properly emphasizing or weakening the membership grade to class “I” for \( j \)th pattern. Now, we approach the problem of control parameter estimation of the RF-LDA model by considering the formulation of a HNN. By this means, as a control parameter in formula (11), \( m \) is dynamically assigned the particular value of offset in the calculation of the grade of membership. After executing the evaluation algorithm HRF-LDA on the ORL database, we have found the optimal value for the parameter \( m = 0.3 \). Also, in the experiments of this section, the k-NN classifier with the Euclidean distance is employed for the classification, in this work, we choose the parameter of k-NN as \( k = 3 \). In the following, Fig. 2 indicates the convergence time of the HNN estimator as a function of the adjustable parameter \( m \).

Table 1 indicates average recognition rates of D-LDA, C-LDA, R-LDA, 2D-LDA, LPP, NPE and the proposed method vary with the number of training samples per individual on the ORL face image database. As shown in Table 1, it is therefore reasonable to believe that the proposed method is the most effective one no matter how many training samples per individual are used. As can be seen from the previous experiment, to evaluate the efficiency of the algorithm and remove the influence caused by the choice of the training set and test set, we repeated the experiment 10 runs.

Moreover, the next experiment aims to give a comprehensive comparison between the results of our method and classical bicubic interpolation algorithm. 100 training patches are collected and the proposed parameterized RF-LDA is applied to feature extraction and dimensions reduction. In this experiment, dictionary learning takes approximately 20 s for 22 iterations of the PF-K-SVD algorithm, and we experimentally estimate \( h \) atoms in the dictionary, and allocate \( L \) atoms per patch-presentation. For the evaluations of parameters, we randomly choose the former 3 images per individual for training, and the rest images are used for testing on ORL face database. Table 2 presents the recognition rates of PK-K-SVD varying with parameters \( h \) and \( L \). It can be observed that the optimal accuracy of PK-K-SVD is obtained for the parameters \( h = 196 \) and \( L = 5 \). Fig. 3 shows the training image patch dictionary.

![Fig. 2. Convergence time of the HNN estimator as a function of the adjustable parameter \( m \).](image)

![Fig. 3. The training image patch dictionary.](image)

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>( h = 147 ) ( L = 3 )</td>
<td>88.6</td>
</tr>
<tr>
<td>( h = 147 ) ( L = 5 )</td>
<td>88.9</td>
</tr>
<tr>
<td>( h = 147 ) ( L = 7 )</td>
<td>89.0</td>
</tr>
<tr>
<td>( h = 196 ) ( L = 3 )</td>
<td>89.3</td>
</tr>
<tr>
<td>( h = 196 ) ( L = 5 )</td>
<td><strong>90.2</strong></td>
</tr>
<tr>
<td>( h = 196 ) ( L = 7 )</td>
<td>89.8</td>
</tr>
<tr>
<td>( h = 245 ) ( L = 3 )</td>
<td>89.1</td>
</tr>
<tr>
<td>( h = 245 ) ( L = 5 )</td>
<td>89.5</td>
</tr>
<tr>
<td>( h = 245 ) ( L = 7 )</td>
<td>89.5</td>
</tr>
</tbody>
</table>

Table 1
The average recognition rate (%) of each method varies with number of training samples per individual on ORL face image database.

<table>
<thead>
<tr>
<th>Method</th>
<th>Number of training samples</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>D-LDA</td>
<td></td>
<td>88.27 ± 1.37</td>
<td>90.92 ± 1.42</td>
<td>92.93 ± 1.40</td>
</tr>
<tr>
<td>C-LDA</td>
<td></td>
<td>88.97 ± 1.41</td>
<td>91.81 ± 1.37</td>
<td>94.47 ± 1.32</td>
</tr>
<tr>
<td>R-LDA</td>
<td></td>
<td>88.72 ± 1.34</td>
<td>92.04 ± 1.45</td>
<td>94.37 ± 1.37</td>
</tr>
<tr>
<td>2D-LDA</td>
<td></td>
<td>89.25 ± 1.44</td>
<td>92.18 ± 1.35</td>
<td>94.75 ± 1.42</td>
</tr>
<tr>
<td>LPP</td>
<td></td>
<td>89.17 ± 1.39</td>
<td>92.08 ± 1.31</td>
<td>94.57 ± 1.37</td>
</tr>
<tr>
<td>NPE</td>
<td></td>
<td>88.72 ± 1.48</td>
<td>91.75 ± 1.45</td>
<td>94.08 ± 1.49</td>
</tr>
<tr>
<td>PF-K-SVD</td>
<td></td>
<td>89.94 ± 1.28</td>
<td>93.08 ± 1.29</td>
<td>95.26 ± 1.25</td>
</tr>
</tbody>
</table>
Fig. 4 shows the reconstruction process of an original face image on ORL database by means of the proposed algorithm. The reconstruction algorithm is tested on 280 test images (taking a few seconds on each image, using fully uncrossed patches scale) and its peak signal-to-noise ratio (PSNR) results are compared versus bicubic interpolation algorithm.

The PSNR comparison results between the proposed algorithm and classical bicubic interpolation algorithm are summarized in Table 3. On the left is the comparison strategy, followed by bicubic interpolation algorithm, and the proposed algorithm’s results on the right, at the last column. We also compare the different methods quantitatively in terms of their RMS errors for the selected image sample. As shown in Table 3, the proposed algorithm performs better than bicubic interpolation.

<table>
<thead>
<tr>
<th>Evaluating indicator</th>
<th>Method</th>
<th>Bicubic interpolation</th>
<th>Proposed method</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSNR (dB)</td>
<td>31.7</td>
<td>32.8</td>
<td></td>
</tr>
<tr>
<td>RMS errors</td>
<td>4.74</td>
<td>4.57</td>
<td></td>
</tr>
</tbody>
</table>

Fig. 4. Part of the reconstruction process of one ORL face image signal.

Table 3
PSNR comparison results of each method on ORL image sample.

4.2. On Yale database

The Yale face image database contains 165 grayscale images of 15 individuals. There are 11 images per subject, one per different facial expression or configuration. We manually crop the facial portion of each face image. Each cropped face is resized to 40 × 50 pixels. Some sample images from the Yale database are shown in Fig. 5.

We randomly choose the former 5 images per individual for training, and the rest images are used for testing. Similarly, 10 experiments were performed to obtain the average recognition rate. To make full use of the available data and to evaluate the generalization power of algorithms more accurately, 10 experiments were performed. Table 4 presents the recognition accuracies of D-LDA, C-LDA, R-LDA, LPP, NPE and the proposed method. For all methods, the corresponding dimensionality of the reduced subspace is also given in Table 4. Again, the recognition accuracy of each method listed in Table 4 indicates that the proposed method is still the most effective one among the other traditional approaches.

4.3. On FERET database

The FERET face image database is a result of the FERET program, which was sponsored by the U.S. Department of Defense through the DARPA program [20]. It has become a standard database for the evaluation of state-of-the-art face recognition techniques. The HFSL algorithm was evaluated on a subset of FERET database, which includes 1400 images of 200 individuals with seven different images of each individual. In our experiment, all images are grayscale and normalized with a resolution of 40 × 40. Some sample images from the FERET database are shown in Fig. 6.

In this experiment, the FERET database is randomly partitioned into a training set and a test set with no overlap between the two.
The partition of the database into training and testing sets, which call for three images per individual randomly chosen for training, and the rest four images for testing. Thus, a training set of 600 images and a test set with 800 images are created. To make full use of the available data and to evaluate the generalization power of algorithms more accurately, the figures of merit are accurate recognition rates averaged over 20 runs, and the error margin for both methods (standard deviation) is provided. Table 5 presents the average recognition accuracies of traditional methods and PF-K-SVD. For all methods, the average CPU time consumed for training and testing is also given in Table 5. Again, the experimental results indicate that the proposed method is the most effective one among the facial feature extraction approaches. However, it is worth stressing that the proposed method needs more CPU time for the whole process because it costs more computation by using fuzzy adaptive dictionary learning for sparse representation.

5. Discussions

In this section we describe the characteristics, rationale, and potential advantages of the proposed method. The basic characteristic of this method is to use a parameterized fuzzy adaptive learning way to best adapt dictionaries, and it exploits the optimal representation result to perform classification. Here ‘best’ means that the error between the obtained reconstruction sample and test sample is almost the smallest. Moreover, we provide a more detailed analysis of the method and extensive discussion, and we make further improvement to extend this method.

1) The merit of the PF-K-SVD method is that the fuzzy membership function is adaptive represented in terms of a parameter estimation, so as to achieve efficient sparse representation and robust discriminant analysis. Concretely, we approach the problem of control parameter estimation of the RF-LDA model by considering...
the formulation of a HNN, thus, the control parameter \( m \) in formula (11) is dynamically assigned the optimal value of offset in the calculation of the grade of membership. Therefore, the PF-K-SVD utilizes an update of the sparse representations by embedding the proposed parameterized mechanism of fuzzy set. It finds the closest rank-1 matrix that approximates \( E_k \), which will effectively minimize the error as defined in formula (20). This is the main difference between the fuzzy K-SVD and PF-K-SVD, that is, the PF-K-SVD is a parameterized version of fuzzy K-SVD algorithm.

2) Recent criterion-based feature extraction algorithms are locality-oriented and generally suffer from the following issues: (1) They often suffer from the small sample size problem in process of estimation of manifold structure; (2) They model the local geometry information of data by respective optimal criterion, which generally fails to obtain sufficient structured information due to the high dimensionality of image space; (3) They do not sufficiently take advantage of parameterized mechanism of fuzzy set among high dimensional samples. Therefore, if a collection of representative samples is found for the distribution, we should expect that a typical sample has a very sparse representation with respect to such a (possibly learned) basis. In this paper, we develop PF-K-SVD algorithm to address the above problems. As discussed in Section 3.2, in our previous research in Ref. [13], we extended the F-LDA presented by Kwak et al. to be RF-LDA, which includes a complete fuzziness in the calculation of between-class scatter matrix and within-class scatter matrix, so as to achieve the distribution information of each sample. Motivated by this idea, we incorporate a parameterized version of RF-LDA to K-SVD algorithm. More specifically, (1) we reorganize all the feature sets via RF-LDA mode and estimate the dynamic parameter in terms of different patterns; (2) the parameterized fuzzy mechanism of the labeled patterns are then incorporated into the discriminant analysis; (3) we redesign the fuzzy objective function of K-SVD, which contains considerable sparse discriminative information. Thus, a novel dictionary learning technique is presented, which discovers a parameterized fuzzy mechanism to adapting dictionaries.

3) As demonstrated in Section 3.2, as a control parameter in formula (11), \( m \) is dynamically assigned the particular value of offset in the calculation of membership grade of RF-LDA via HNN. Therefore, the value of control parameter \( m \) found by HNN is optimal. Moreover, another two parameters \( h \) and \( L \) are experientially estimated by K-SVD update. Concretely, \( h \) specifies the number of dictionary atoms to train, if it is specified without the “initial dictionary”, the dictionary is initialized with dictionary size randomly selected training signals. Here, “initial dictionary” specifies the initial dictionary for the training, it should be either a matrix of size \( N \times L \), where \( N \) size (data,1), or an index vector of length \( L \), specifying the indices of the examples to use as initial atoms. If “dictionary size” and “initial dictionary” are both present, \( L \ll \) dictionary size, and in this case the dictionary is initialized using the first dictionary size columns from initial dictionary. If only initial dictionary is specified, dictionary size is set to \( L \). As demonstrated in Section 4.1, we randomly choose the former 3 images per individual for training, and the rest images are used for testing. Table 2 presents the recognition rates (%) of PK-K-SVD varying with parameters \( h \) and \( L \) on ORL face database, it can be observed that the optimal accuracy of PK-K-SVD is obtained for the parameters \( h = 196 \) and \( L = 5 \).

6. Conclusions

In this paper, a parameterized fuzzy adaptive way to adapting dictionaries in order to achieve the fuzzy sparse signal representations has been developed, the update of the dictionary columns is combined with an update of the sparse representations by embedding a new mechanism of fuzzy set. The strength of the technique is that it successfully utilizes the fuzzy adaptive learning scheme as a feature analysis tool, while quantifying those factors that exert influence on the sparse signal representations, by means of the K-SVD method. We believe that this kind of fuzzy adaptive dictionary learning method can successfully replace popular representation methods both in image enhancement and in reconstruction. Future work is required to enable such a trend. As mentioned by Aharon in Ref. [12], among the many possible research directions we mention two: (1) exploration of the connection between the chosen pursuit method in the PF-K-SVD and the method used later in the application; (2) a study of the effect of introducing weights to the atoms, allowing them to get varying fuzzy degrees of popularity.

Acknowledgments

The authors would like to thank the anonymous reviewers for their constructive advice. This work was supported by the Natural Science Foundation of China (Grant no. 61100116), the Natural Science Foundation of Jiangsu Province (Grant nos. BK2012700 and BK2011492), the China Postdoctoral Science Foundation (Grant no. 2011M500926), the Jiangsu Postdoctoral Science Foundation (Grant no. 1102063C) and the Foundation of Artificial Intelligence Key Laboratory of Sichuan Province (Grant no. 2012ZRZY02).

References

Xiaoning Song received his B.S. degree in Computer Science and Technology from Southeast University, Nanjing, PR China, in 1997. He received the M.S. degree in Computer Science from Jiangsu University of Science and Technology, PR China, in 2005 and the Ph.D. degree in Pattern Recognition and Intelligence System at Nanjing University of Science and Technology, PR China in 2010. Since 2005, he has been teaching in the School of Computer Science and Engineering, Jiangsu University of Science and Technology, PR China. His current research interests include Pattern Recognition, Sparse Representation, Image Recognition and Fuzzy Systems.

Zi Liu received his B.S. degree in Computer Science and Technology from Nanjing University, Nanjing, PR China, in 2008. Presently, he is a Ph.D. candidate in Nanjing University of Science and Technology, Nanjing, PR China. His current research interests include Pattern Recognition and Intelligence System, Sparse Representation, Image Recognition and Fuzzy Systems.

Jingyu Yang received his B.S. degree in Computer Science from Nanjing University of Science and Technology, Nanjing, PR China. From 1982 to 1984, he was a visiting scientist at the Coordinated Science Laboratory, University of Illinois at Urbana-Champaign. From 1993 to 1994, he was a visiting professor in the Department of Computer Science, Missouri University. And, in 1998, he acted as a visiting professor at Concordia University in Canada. He is currently a professor in the Department of Computer Science at Nanjing University of Science and Technology. He is the author of more than 400 scientific papers in Computer Vision, Pattern Recognition, and Artificial Intelligence. He has won more than 20 provincial and national awards. His current research interests are in the areas of Pattern Recognition, Robot Vision, Image Processing, Data Fusion, and Artificial Intelligence.