A Fuzzy Adaptive K-SVD Dictionary Algorithm for Face Recognition

Xiaoning Song, Zi Liu
Post-Doctoral Research Center
Jiangsu Sunboon Information Technology Co., Ltd
Wuxi, 214072, P.R. China
E-mail: xnsong@yahoo.com.cn

Abstract—Sparse representations using overcomplete 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 prespecified 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 meaning that the multi-classes samples are assigned to the given classes definitely. The work proposed in this paper provides a novel 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 incorporated a new mechanism of fuzzy set, which is called fuzzy K-SVD. Experimental results conducted on the ORL and Yale face databases demonstrate the effectiveness of the proposed method.

Keywords—Sparse representation; Fuzzy sets; K-SVD; Image recognition

1. INTRODUCTION

Sparse signal representations using overcomplete dictionaries are used in a variety of fields such as pattern recognition, image and video coding [1-2]. Overcompleteness of a set means that it has more members than the dimensionality of the members. Given an overcomplete set of basis signals, called the dictionary, the goal is to express input signals as sparse linear combinations of the dictionary members. The advantage of overcompleteness 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 sparser representations of a variety of input signals.

Applications that can benefit from the sparsity and overcompleteness concepts (together or separately) include compression, regularization in inverse problems, feature extraction, and more. Indeed, the success of the JPEG2000 coding standard can be attributed to the sparsity of the wavelet coefficients of natural images [3]. In denoising, wavelet methods and shift-invariant variations that exploit overcomplete representation are among the most effective known algorithms for this task [4]. Sparsity and overcompleteness have been successfully used for dynamic range compression in images, separation of texture and cartoon content in images [5], inpainting, and more.

Evidently, extraction of the sparsest representation is a hard problem that has been extensively investigated in the past few years. 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, it 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 meaning 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 face images are affected by illumination, expression, etc.), it is advantageous to investigate these factors and quantify their impact on their “internal” class assignment [6].

In the present study, we propose a novel dictionary learning technique that discovers a 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 fuzzy K-SVD (F-K-SVD). Using F-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 [7]. 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 F-K-SVD algorithm is developed on the basis of the grade of membership, and then applied in order to perform the pattern classification.

The rest of the paper is organized as follows: Section 2 describes the K-SVD algorithm. F-K-SVD algorithm is presented in Section 3. Experimental results are reported in
Section 4 and Section 5 concludes the paper.

II. 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 [8]. Keeping all the columns fixed apart from the j0-th one, $a_{j0}$, this column can be updated along with the coefficients that multiply it in $X$. We isolate the dependency on $a_{j0}$ by rewriting minimization as

$$\|Y-AX\|^2=\|Y-\sum_{j=1}^{n} a_j x_j\|^2=\left(\sum_{j=1}^{n} a_j x_j\right)-a_{j0} x_{j0}$$

(1)

In this description, $x_{j0}^T$ stands for the j-th row of X. The update step targets both $a_{j0}$ and $x_{j0}^T$, and refers to the term in parentheses,

$$E_{j0}=Y-\sum_{j=1}^{n} a_j x_j$$

(2)

as a known pre-computed error matrix.

The optimal $a_{j0}$ and $x_{j0}^T$ minimizing Equation (1) are the rank-1 approximation of $E_{j0}$, and can be obtained via an SVD, but this typically would yield a dense vector $x_{j0}$, 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 $E_{j0}$ should be taken – those that correspond to the signals from the example-set that are using the j0-th atom, namely those columns where the entries in the row $x_{j0}$ are non-zero. This way, we allow only the existing non-zero coefficients in $x_{j0}$ to vary, and the cardinalities are preserved. Therefore, we define a restriction operator, $P_{j0}$, that multiplies $E_{j0}$ from the right to remove the non-relevant columns. The matrix $P_{j0}$ has M rows (the number of overall examples), and M columns (the number of examples using the j0-th atom). We define $(x_{j0})^T = x_{j0}^T P_{j0}$ as the restriction on the row $x_{j0}^T$, choosing the non-zero entries only.

For the sub-matrix, $E_{j0}$, $P_{j0}$, a rank-1 approximation via SVD can be applied, updating both the atom $a_{j0}$ and the corresponding coefficients in the sparse representations, $x_{j0}^T$. This simultaneous update may lead to a substantial speedup in the convergence of the training algorithm.

III. A 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 meaning 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 face images are affected by illumination, expression, etc.), 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 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 incorporated a new mechanism of fuzzy set, which is called fuzzy K-SVD. Moreover, the need for such a novel fuzzy model construction is reduced to parameter estimation when the structure is given beforehand. Typically, it arises when a model is required in order to make some decisions about the system, and therefore the parameter estimation method must recursively process the measured data as they become available. Using the fuzzy K-SVD (F-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.

A. Establish a fuzzy mechanism

According to the study in Ref. [6], Kwak proposed to incorporate a gradual level of assignment to class being regarded as a membership grade with anticipation that such discrimination helps improve classification results. Now, the F-LDA method combined with FKNN [9] algorithm is considered to regain the statistical properties of the patterns such as the mean value and scatter covariance matrices. However, after investigating the membership allocation formula, we find that the method attempts to “fuzzify” or refine the membership grades of the labeled patterns only by fuzzifying 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. [10], 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 condition is presented to achieve the distribution information of each sample that represented with fuzzy membership degree.

In this method, the first key step of the reformative 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 $U = [\mu_{ij}]$ for $i = 1,2,\cdots,C$ and $j = 1,2,\cdots,N$ satisfies the obvious property $\sum_{i=1}^{C} \mu_{ij} = 1$. Taking into account the fact that the some outliers may have some harmful influence to the performance of conventional feature extraction approach, a
relaxed normalized condition in the fuzzy membership degrees is proposed as follows:

\[
\sum_{i=1}^{C} \sum_{j=1}^{N} \mu_i(x_j) = N
\]  

(3)

By the condition (10), we can define the fuzzy membership as follows:

\[
\mu_i = \begin{cases} 
\gamma + (1 - \gamma) \cdot \left( \frac{n_j}{K} \right) & \text{if } i \not\in \text{class } (j) \\
(1 - \gamma) \cdot \left( \frac{n_j}{K} \right) & \text{otherwise}
\end{cases}
\]  

(4)

where, \( \gamma = \frac{N - C}{2^n - N} \), \( n_j \) stands for the number of the neighbors of the \( j \)th pattern that belong to the \( i \)th class, and \( \gamma \) and \( \gamma \) are constants which ultimately control the value of \( \mu_i \) and satisfy the constraints \( m \in (0, 1) \), \( \gamma \in (0, 1) \). Since the normalized condition of the membership degree has been relaxed, all the examples 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.

The second key step of this method is to incorporate the contribution of each training sample into the redefinition of scatter matrices. Specifically, the membership grade of the contribution of each training sample into the redefinition of membership grades. Therefore, the RF-LDA shows high stability compared with 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. Specifically, the membership grade of the contribution of each training sample into the redefinition of membership grades. Therefore, the RF-LDA shows high stability compared with others due to its power of regulating the fuzzy membership grades.

Thus, it shows high stability of this improved 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. Specifically, the membership grade of the contribution of each training sample into the redefinition of membership grades. Therefore, the RF-LDA shows high stability compared with 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. Specifically, the membership grade of the contribution of each training sample into the redefinition of membership grades. Therefore, the RF-LDA shows high stability compared with 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. Specifically, the membership grade of the contribution of each training sample into the redefinition of membership grades. Therefore, the RF-LDA shows high stability compared with 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. Specifically, the membership grade of the contribution of each training sample into the redefinition of membership grades. Therefore, the RF-LDA shows high stability compared with 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. Specifically, the membership grade of the contribution of each training sample into the redefinition of membership grades. Therefore, the RF-LDA shows high stability compared with 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. Specifically, the membership grade of the contribution of each training sample into the redefinition of membership grades. Therefore, the RF-LDA shows high stability compared with 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. Specifically, the membership grade of the contribution of each training sample into the redefinition of membership grades. Therefore, the RF-LDA shows high stability compared with 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. Specifically, the membership grade of the contribution of each training sample into the redefinition of membership grades. Therefore, the RF-LDA shows high stability compared with 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. Specifically, the membership grade of the contribution of each training sample into the redefinition of membership grades. Therefore, the RF-LDA shows high stability compared with 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. Specifically, the membership grade of the contribution of each training sample into the redefinition of membership grades. Therefore, the RF-LDA shows high stability compared with 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. Specifically, the membership grade of the contribution of each training sample into the redefinition of membership grades. Therefore, the RF-LDA shows high stability compared with 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. Specifically, the membership grade of the contribution of each training sample into the redefinition of membership grades. Therefore, the RF-LDA shows high stability compared with 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. Specifically, the membership grade of the contribution of each training sample into the redefinition of membership grades. Therefore, the RF-LDA shows high stability compared with 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. Specifically, the membership grade of the contribution of each training sample into the redefinition of membership grades. Therefore, the RF-LDA shows high stability compared with 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. Specifically, the membership grade of the contribution of each training sample into the redefinition of membership grades. Therefore, the RF-LDA shows high stability compared with 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. Specifically, the membership grade of the contribution of each training sample into the redefinition of membership grades. Therefore, the RF-LDA shows high stability compared with 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. Specifically, the membership grade of the contribution of each training sample into the redefinition of membership grades. Therefore, the RF-LDA shows high stability compared with others due to its power of regulating the fuzzy membership grades.
for 1 ≤ i ≤ M. These form the matrix $X_j$.

**Step6.** Fuzzy K-SVD Dictionary-Update Stage: Use the following procedure to update the columns of the dictionary and obtain $A_j$: Repeat for $j_0 = 1, 2, \ldots, m$

1. Define the group of examples that use the atom $a_{j_0}$,
   $$\Omega_{j_0} = \{ i \mid 1 \leq i \leq M, X_i[j_0, i] \neq 0 \}$$
2. Compute the residual matrix
   $$E_{j_0} = \sum_{i=1}^{N} \sum_{j=1}^{J} \mu_i(X_j) \left( Y - \sum_{j \neq j_0} a_{j_0} x_{j_0}^T \right),$$
   where $x_{j_0}^T$ are the j'th rows in the matrix $X_j$.
3. Restrict $E_{j_0}$ by choosing only the columns corresponding to $\Omega_{j_0}$, and obtain $E_{j_0}^\Omega$.
4. Apply SVD decomposition $E_{j_0}^\Omega = UD\Sigma V^T$. Update the dictionary atom $a_{j_0} = u_l$, and the representations by $x_{j_0}^\Omega = \Delta [1, 1] v_l$.
5. **Step7.** Stopping Rule: If the change in $J_\ell = \sum_{i=1}^{N} \sum_{j=1}^{J} \mu_i(X_j) \left\| E_i - h_{j_0} X_j \right\|_2^2$ is small enough, stop. Otherwise, apply another iteration.

**Step8.** Output: The desired result is $A_j$.

IV. EXPERIMENTAL RESULTS

The proposed method is used for face recognition and tested on the ORL [14] and Yale face image database. To evaluate the proposed method properly, we also include experimental results for the D-LDA [15], C-LDA [16], R-LDA [17], 2D-LDA [18], LPP [19], NPE [20]. For its simplicity, the $k$ nearest neighbor (k-NN) classifier with Euclidean distance is employed for the classification, in which the parameter $m \in (0, 1)$. Thus, with variation of $m$, the discriminatory capability of RF-LDA is gradually changed due to the properly emphasizing or weakening the membership grade to class “i” for $j$ th pattern. Now, we have found the optimal value for the parameter $m = 0.3$. Also, in the experiments of this section, the k-NN classifier with Euclidean distance is employed for the classification, in this work, we choose the parameter of k-NN as $k = 3$.

TABLE I indicates average recognition rates of D-LDA, C-LDA, R-LDA, 2D-LDA, LPP, NPE and proposed method vary with the number of training samples per individual on the ORL face image database. As shown in Table I, 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.

<table>
<thead>
<tr>
<th>Method</th>
<th>Number of training samples</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3</td>
</tr>
<tr>
<td>D-LDA</td>
<td>88.27</td>
</tr>
<tr>
<td>C-LDA</td>
<td>88.97</td>
</tr>
<tr>
<td>R-LDA</td>
<td>88.72</td>
</tr>
<tr>
<td>2D-LDA</td>
<td>89.25</td>
</tr>
<tr>
<td>LPP</td>
<td>89.17</td>
</tr>
<tr>
<td>NPE</td>
<td>88.72</td>
</tr>
<tr>
<td>F-K-SVD</td>
<td>89.51</td>
</tr>
</tbody>
</table>

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. The each cropped face is resized to $40 \times 50$ pixels. Some sample images from the Yale database are shown in Fig. 2.

<table>
<thead>
<tr>
<th>Method</th>
<th>Number of training samples</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3</td>
</tr>
<tr>
<td>D-LDA</td>
<td>88.27</td>
</tr>
<tr>
<td>C-LDA</td>
<td>88.97</td>
</tr>
<tr>
<td>R-LDA</td>
<td>88.72</td>
</tr>
<tr>
<td>2D-LDA</td>
<td>89.25</td>
</tr>
<tr>
<td>LPP</td>
<td>89.17</td>
</tr>
<tr>
<td>NPE</td>
<td>88.72</td>
</tr>
<tr>
<td>F-K-SVD</td>
<td>89.51</td>
</tr>
</tbody>
</table>

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. The each cropped face is resized to $40 \times 50$ pixels. Some sample images from the Yale database are shown in Fig. 2.
more accurately, ten experiments were performed. Table II presents the recognition accuracies of D-LDA, C-LDA, R-LDA, LPP, NPE and proposed method.

<table>
<thead>
<tr>
<th>Method</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>D-LDA</td>
<td>97.71</td>
</tr>
<tr>
<td>C-LDA</td>
<td>98.04</td>
</tr>
<tr>
<td>R-LDA</td>
<td>97.44</td>
</tr>
<tr>
<td>LPP</td>
<td>98.05</td>
</tr>
<tr>
<td>NPE</td>
<td>97.82</td>
</tr>
<tr>
<td>F-K-SVD</td>
<td>98.47</td>
</tr>
</tbody>
</table>

Again, as shown in Table II, it is therefore reasonable to believe that the F-K-SVD method is more than effective than others.

V. CONCLUSIONS

In this paper, a 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 incorporated a new mechanism of fuzzy set, which is called F-K-SVD. The strength of the technique is that it successfully utilizes fuzzy adaptive learning scheme as a feature analysis tool, while quantifying those factors which exert influence on the sparse signal representations, by means of K-SVD method. We believe this kind of fuzzy adaptive dictionary learning method, which nowadays is not being commonly used, can successfully replace popular representation methods both in image enhancement and in compression. Future work is required to enable such a trend. As mentioned by Aharon in Ref. [8], among the many possible research directions we mention two: 1) exploration of the connection between the chosen pursuit method in the 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.

ACKNOWLEDGEMENTS

This work was supported by the Natural Science Foundation of Jiangsu Province (Grant no BK2012700), the China Postdoctoral Science Foundation (Grant no. 2011M500992), the Jiangsu Postdoctoral Science Foundation (Grant no. 1102063C) and the Foundation of Artificial Intelligence Key Laboratory of Sichuan Province (Grant no. 2012RZY02).

REFERENCES