An Enhanced Sparse Representation Strategy for Signal Classification

Yin Zhou, Jinglun Gao and Kenneth E. Barner
Department of Electrical and Computer Engineering
University of Delaware, Newark, DE, USA, 19716

ABSTRACT
Sparse representation based classification (SRC) has achieved state-of-the-art results on face recognition. It is hence desired to extend its power to a broader range of classification tasks in pattern recognition. SRC first encodes a query sample as a linear combination of a few atoms from a predefined dictionary. It then identifies the label by evaluating which class results in the minimum reconstruction error. The effectiveness of SRC is limited by an important assumption that data points from different classes are not distributed along the same radius direction. Otherwise, this approach will lose their discrimination ability, even though data from different classes are actually well-separated in terms of Euclidean distance. This assumption is reasonable for face recognition as images of the same subject under different intensity levels are still considered to be of same-class. However, the assumption is not always satisfied when dealing with many other real-world data, e.g., the Iris dataset, where classes are stratified along the radius direction. In this paper, we propose a new coding strategy, called Nearest-Farthest Neighbors based SRC (NF-SRC), to effectively overcome the limitation within SRC. The dictionary is composed of both the Nearest Neighbors and the Farthest Neighbors. While the Nearest Neighbors are used to narrow the selection of candidate samples, the Farthest Neighbors are employed to make the dictionary more redundant. NF-SRC encodes each query signal in a greedy way similar to OMP. The proposed approach is evaluated over extensive experiments. The encouraging results demonstrate the feasibility of the proposed method.

Keywords: Pattern recognition, sparse representation, face recognition, action recognition, computer vision

1. INTRODUCTION
Sparse representation has achieved state-of-the-art results in many fields, such as image compression and denoising, texture analysis, face recognition, objection recognition, video-based action classification, etc. The success of this technique is partially due to its robustness to noise and missing data. For example, sparse representation-based classification (SRC) has obtained impressive results on face recognition, which encodes a query face image over the entire set of training template images and identifies the label of the query sample by evaluating which class yields the minimum reconstruction error.

With a fixed dictionary, sparse representation is essentially an optimization process combining atom selection and weight assignment. The effectiveness of sparse representation based classification (SRC) is limited by a fundamental assumption that data points from different classes are not distributed along the same radius direction. Otherwise, this method will lose their discrimination ability, even though data of different classes are actually well-separated in terms of Euclidean distance. This assumption is reasonable for face recognition as images of the same subject under different intensity levels are still considered to be of same-class. In other words, the magnitudes of feature vectors are not considered as discriminative information in face recognition. However, the assumption is not always satisfied when dealing with many other real-world data. For example, as shown in Fig. 1(a), the Iris dataset contains three classes of data (points in red, green and blue), which are distributed closely along the same radius direction. Here, we have extracted 2D features, i.e., pedal length and pedal width. Such 2D features are sufficiently informative for traditional classifiers, e.g., k nearest neighbors (kNN) and support vector machine (SVM). On the other hand, SRC yields atoms that are located on the unit circle.
after normalization and have severe overlappings in the middle of the point scatter. SRC can accurately classify test samples outside the overlapped region but within that region, the accuracy is close to random guess.

Note that in the previous discussion, we have only focused on the case where all atoms are normalized to unit $\ell_2$ norm, which is a commonly used constraint in SRC\textsuperscript{4} and many other sparse representation based methods such as.\textsuperscript{1,3,5,8} Nevertheless, under constraints that do not require atoms with unit gain, the above observation is still valid for SRC to maintain discrimination ability. The reason is that among a group of atoms with same direction but different gains, $\ell_1$ minimization solvers\textsuperscript{9-13} will always favor the atom with the largest gain, because to achieve same contribution in reconstruction, the atom with the largest gain yields the smallest $\ell_1$ cost or penalty compared with its counterparts. Therefore, SRC suffers the drawback of losing discrimination ability when classifying data distributed along the same radius direction.

In this paper, to overcome the aforementioned limitation of SRC, we propose a novel coding strategy, called Nearest-Farthest Neighbors based SRC (NF-SRC), based on Orthogonal Matching Pursuit.\textsuperscript{14} The performance of sparse representation based algorithms relies on the quality of the dictionary $D$.\textsuperscript{8} IN NF-SRC, the dictionary is adaptive per query sample and is composed of the nearest neighbors ($D_N$) and the farthest neighbors ($D_F$). While the nearest neighbors are used to narrow the selection of candidate samples, the farthest neighbors are employed to provide more supports for reconstruction. Then, in a greedy manner similar to OMP, a query sample will be coded based on the dictionary $[D_N, D_F]$ (rather than the entire training set). We evaluate the proposed NF-SRC over extensive experiments and the encouraging results verify the feasibility of the proposed method.

### 2. BACKGROUND

In this section, we briefly introduce sparse representation based classification.\textsuperscript{4} Consider a face recognition problem with $N$ training images collected from $K$ subjects. Denote $n_k$ as the number of training images of subject $k$, which yields $N = \sum_{k=1}^{K} n_k$. Define a matrix $A = [A_1, A_2, \ldots, A_K] \in \mathbb{R}^{m \times N}$ for the entire training set, where $A_i = [x_{i,1}, x_{i,2}, \ldots, x_{i,n_i}] \in \mathbb{R}^{m \times n_i}$ is the $i$-th sub-matrix consisting of $n_i$ samples of the $i$-th subject. To avoid trivial solution during sparse coding, atoms (columns) in $A$ are normalized to have unit $\ell_2$-norm. Given a query image $y \in \mathbb{R}^m$, its class label is identified as following. First, solve sparse representation of $y$ as

$$\alpha = \arg \min_{\alpha} \|\alpha\|_1 \quad \text{subject to} \quad \Phi A \alpha = \Phi y,$$

or solve the problem alternatively in the noisy case as

$$\alpha = \arg \min_{\alpha} \|\alpha\|_1 \quad \text{subject to} \quad \|\Phi y - \Phi A \alpha\|_2 \leq \varepsilon$$
for a given error tolerance \( \varepsilon \). \( \Phi \in \mathbb{R}^{d \times m} \) satisfying \( d \ll m \) is employed to ensure the sparsity in \( \alpha \) and can be generated randomly obeying Gaussian or Bernoulli distribution. Then, evaluate which class yields the minimum reconstruction error, as

\[
    r_i(y) = \| \Phi y - \Phi A \delta_i(\alpha) \|_2,
\]

where \( \delta_i(\alpha) = [0, \ldots, \alpha_i, 1, \ldots, \alpha_i, 0, \ldots] \). The class label of \( y \) is determined as \( \text{label}(y) = \arg \min_{i \in \{1, \ldots, K\}} r_i(y) \).

3. PROPOSED METHOD

3.1 Related Work

Li et al.\(^{15}\) proposed local sparse representation based classification (LSRC) to improve the efficiency of SRC by simply performing \( \ell_1 \) minimization over the local neighborhood of a query sample. However, the dictionary is usually an over-determined system with atoms highly coherent with each other. Thus the approach may not yield faithful reconstruction due to the lack of supports. Consequently, the classification will be unstable. Moreover, as shown in,\(^{16}\) the core power of SRC should be attributed to the collaborative representation. Therefore, it is desired to represent query samples over a redundant dictionary.

3.2 The New Coding Strategy NF-SRC

Suppose we want to compute the sparse reconstruction code \( \alpha \) for \( y \) over \( A \). Intuitively, large coefficients (in absolute value) in \( \alpha \) correspond to the nearest neighbors of \( y \), while small coefficients are associated with training samples that are less similar to \( y \). Compared with nearest neighbors, those farthest samples contribute to a very small portion of energy in reconstruction. However, this small portion of energy is critical for faithful reconstruction and robust classification.\(^{16}\) Therefore, it is desired to incorporate both the nearest neighbors and the farthest neighbors into the dictionary. Selecting the nearest neighbors can not only find the most effective atoms in reconstruction but can also exclude those training samples from being coded, which are similar to \( y \) in direction but are far away from \( y \) in Euclidean distance. Thus the drawback of SRC can be solved. The employment of farthest neighbors is to provide more supports for faithful reconstruction and robust classification. The proposed coding strategy Nearest-Farthest Neighbors based SRC (NF-SRC) is presented as following.

The proposed coding strategy can be seen as an extension of OMP\(^{14}\) with better discrimination ability. Note that step 9 in Algorithm 1 is critical to NF-SRC, since it essentially controls the relative contribution of farthest samples in reconstruction. If a particular farthest neighbor \( d_j, j > N_1 \) is to be coded with nonzero value, it has to satisfy the requirement that its relative contribution in terms of the absolute value of coefficient is substantially smaller than that of nearest neighbors. In this paper, we uniformly set parameter ratio to be 0.01.

4. EXPERIMENTS

In this section, we evaluate the proposed method using several benchmark datasets, including five UCI Machine Learning datasets*, the Extended Yale B database,\(^{17,18}\) and the Weizmann Action Database.\(^{19}\) The performance of the proposed method is compared with SRC,\(^{4}\) k-Nearest Neighbor (kNN), and multi-class SVM.\(^{20}\) Results reported for these methods are based on our own implementation. Over each dataset, the sparsity constraint \( T \) is set consistently for sparse representation based approaches, *i.e.*, NF-SRC, SRC. The empirical parameter settings of the proposed approach for all the datasets involved are listed in Table 1.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Iris</th>
<th>Satellite</th>
<th>Segmentation</th>
<th>Letter</th>
<th>Vehicle</th>
<th>Extended YaleB</th>
<th>Weizmann</th>
</tr>
</thead>
<tbody>
<tr>
<td>( N_1 )</td>
<td>10</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>300</td>
<td>60</td>
</tr>
<tr>
<td>( N_2 )</td>
<td>6</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>60</td>
<td>30</td>
</tr>
<tr>
<td>( T )</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>30</td>
<td>5</td>
</tr>
</tbody>
</table>

*Published by UCI KDD at http://kdd.ics.uci.edu/summary.data.date.html
Algorithm 1 Coding Strategy NF-SRC

**Require:** Test sample $y \in \mathbb{R}^m$, training set $A \in \mathbb{R}^{m \times N}$ and sparsity parameter $T$

1: Find the $N_1$ nearest neighbors ($D_N \in \mathbb{R}^{m \times N_1}$) and $N_2$ farthest neighbors ($D_F \in \mathbb{R}^{m \times (N_1+N_2)}$) from $A$

2: Form the dictionary $D = [D_N, D_F] \in \mathbb{R}^{m \times (N_1+N_2)}$

3: Preprocess $d_i \leftarrow \frac{d_i}{\|d_i\|_2}$, for all $i = 1, \ldots, N_1 + N_2$

4: Initialize the residual $r_0 = y$, the index set $\Lambda_0 = \emptyset$, and set counter $t = 1$

5: Sort $| < r_{t-1}, d_i > |$ for all $i = 1, \ldots, N_1 + N_2$ in descending order and store the indices in $s$, such that $| < r_{t-1}, d_{s_1} > | > | < r_{t-1}, d_{s_2} > |, \ldots, | < r_{t-1}, d_{s_{N_1+N_2}} > |

6: Set $cur = 1$

7: Augment the index set and the collection of selected atoms

$\Lambda_t = [\Lambda_{t-1}, s_{cur}]$ and $\Phi_t = [\Phi_{t-1}, d_{s_{cur}}]$

8: Solve least square problem $\hat{x} = \arg \min_x \| y - \Phi_t x \|$,

9: Set $cur = cur + 1$ and return to step 7, if the following statement is true:

$s_{cur} > N_1 \text{ AND } |\hat{x}_t| > \max(|\hat{x}|) \cdot \text{ratio}$

where $\hat{x}_t$ is the $t$-th entry in $\hat{x}$ and the operator $| \cdot |$ computes the element-wise absolute value.

10: Calculate the new approximation of the data and the new residual

$a_t = \Phi_t \hat{x}$ and $r_t = y - a_t$

11: Increment $t$ and return to step 5, if $t < T$

12: **return** The sparse code $\alpha$ with nonzero values at entries indicated by $\Lambda_T$

### 4.1 UCI Machine Learning Datasets

The five data sets from UCI Machine Learning Archive are Iris, Satellite, Segmentation, Letter and Vehicle and their basic information are listed in Table 2. In the previously reported literature, various combinations of feature extraction, dimensionality reduction and classifier have been applied over the selected datasets. For fair comparison, we directly employ raw data samples for classification without any preprocessing or feature extraction. The parameter settings for the proposed approach are listed in Table 1. SRC uses all the available training samples for recognition. The classification results are based on 10-fold cross-validation with 30 repetitions, and are listed in Table 3. The proposed approach NF-SRC achieves the highest recognition rate over 4 datasets, namely Iris, Satellite, Segmentation and Vehicle. Over the Letter dataset, the accuracy of NF-SRC is slightly lower than SVM by 0.4%.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Total samples</th>
<th>Dimensions</th>
<th>Classes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Iris</td>
<td>150</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Satellite</td>
<td>6435</td>
<td>36</td>
<td>6</td>
</tr>
<tr>
<td>Segmentation</td>
<td>2310</td>
<td>19</td>
<td>7</td>
</tr>
<tr>
<td>Letter</td>
<td>20000</td>
<td>16</td>
<td>26</td>
</tr>
<tr>
<td>Vehicle</td>
<td>946</td>
<td>18</td>
<td>4</td>
</tr>
</tbody>
</table>

In addition, we compare the methods by only extracting information regarding pedal length and pedal width. Such 2D features are sufficiently informative for kNN and SVM, and are usually employed for visualization. As illustrated in Fig. 1, dictionary atoms of SRC collapse onto the unit-ball such that different classes overlap with
each other, which explains the extreme poor performance of SRC (3rd column in Table 3). Therefore, without proper constraint, SRC would lose their discrimination ability in classifying data distributed along the same radius direction. Our observation coincides with. On the other hand, in our approach, by performing nearest neighbor search, training samples that are similar to the query sample in direction but are far away in Euclidean distance can be eliminated effectively from the next-step coding. Thus, the proposed NF-SRC is capable of achieving robust classification.

4.2 Extended YaleB Database

We evaluate NF-SRC over the Extended Yale B Database which contains 2414 frontal face images of 38 subjects, i.e., about 64 image per person. The images are normalized to $32 \times 32$ to form 1024D feature vectors for classification. As in for each subject, we randomly select half of the images (about 32 per person) for training and the other half for testing. In pre-process stage, $\ell_2$ normalization is carried out for sparse-representation-based approaches. Prior to selecting nearest and farthest neighbors, a simple histogram equalization preprocessing is performed on images. We evaluate various methods over two random subspaces with dimension 300D and 504D and over the original feature subspace with dimension 1024D. The experiment is repeated 20 times.

<table>
<thead>
<tr>
<th>Dimension</th>
<th>SRC</th>
<th>SVM</th>
<th>kNN</th>
<th>NF-SRC</th>
</tr>
</thead>
<tbody>
<tr>
<td>300D</td>
<td>95.7%</td>
<td>92.3%</td>
<td>77.8%</td>
<td>97.1%</td>
</tr>
<tr>
<td>504D</td>
<td>97.5%</td>
<td>95.6%</td>
<td>88.1%</td>
<td>98.7%</td>
</tr>
<tr>
<td>1024D</td>
<td>94.0%</td>
<td>97.4%</td>
<td>88.6%</td>
<td>98.9%</td>
</tr>
</tbody>
</table>

From Table 4, we see that NF-SRC achieves the highest accuracy in all three scenarios. Moreover, consistent with the argument made by Shi et al., the performance of SRC degrades when the dictionary approaches to a square matrix, i.e., $m$ is close to $N$, where $\ell_1$ minimization algorithms meet trouble in finding sparse solutions. This can be observed from the 4-th column in Table 4.

4.3 Weizmann Action Database

The Weizmann Action Database contains 90 videos of 9 different individuals. Each person performed 10 natural actions, i.e., bend, jumping jack (jack), jump forward (jump), jump in place (pjump), run, gallop sideways (side), skip, walk, wave one hand (wave1) and wave both hands (wave2). As this database is captured by a fixed camera under static background, a simple background subtraction and normalized cross-correlation based registration strategy could align human figures very well.

Obeying the same evaluation protocol in, we perform leave-one-person-out experiments to compare various methods. We utilize Motion History Image (MHI) to transform each aligned silhouette sequence into a single image by averaging all the frames. The MHI is capable of capturing both the shape and the temporal information of human actions and is very efficient to compute. Specifically, for each training sequence, we divide it into two subsequences, i.e., odd-numbered frames or even-numbered frames. Then, an MHI is obtained from each subsequence. Hence, the total number of training samples is 160, i.e., 2 MHIs (per person) $\times$ 8 (persons) $\times$ 10 (action classes). Such a dividing strategy provides classification algorithms with more training samples to better model the feature space. Similarly, the test set is generated by computing an MHI of each query sequence.

<table>
<thead>
<tr>
<th>Dimension</th>
<th>SRC</th>
<th>SVM</th>
<th>kNN</th>
<th>NF-SRC</th>
</tr>
</thead>
<tbody>
<tr>
<td>300D</td>
<td>95.7%</td>
<td>92.3%</td>
<td>77.8%</td>
<td>97.1%</td>
</tr>
<tr>
<td>504D</td>
<td>97.5%</td>
<td>95.6%</td>
<td>88.1%</td>
<td>98.7%</td>
</tr>
<tr>
<td>1024D</td>
<td>94.0%</td>
<td>97.4%</td>
<td>88.6%</td>
<td>98.9%</td>
</tr>
</tbody>
</table>
Example MHIs of actions are illustrated in Fig. 2. Finally, all the samples are mapped onto a subspace with dimension $m = 38$ by PCA for classification. The confusion matrix of the proposed approach is illustrated in Table 5. As listed in Table 6, the proposed NF-SRC and SVM achieve the same highest recognition rate 96.7% among all the competing algorithms.

Table 5: Confusion matrix of the proposed approach.

<table>
<thead>
<tr>
<th></th>
<th>Bend</th>
<th>Jack</th>
<th>Jump</th>
<th>PJump</th>
<th>Run</th>
<th>Side</th>
<th>Skip</th>
<th>Walk</th>
<th>Wave1</th>
<th>Wave2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bend</td>
<td>1.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Jack</td>
<td>0.00</td>
<td>1.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Jump</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>PJump</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Run</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Side</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Skip</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Walk</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Wave1</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Wave2</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.90</td>
</tr>
</tbody>
</table>

Table 6: Recognition results over the Weizmann Action Database.

<table>
<thead>
<tr>
<th>Methods</th>
<th>NF-SRC</th>
<th>SRC</th>
<th>SVM</th>
<th>kNN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>96.7%</td>
<td>94.5%</td>
<td>96.7%</td>
<td>93.3%</td>
</tr>
</tbody>
</table>
5. CONCLUSION AND FUTURE WORK

In this paper, we propose a Nearest-Farthest Neighbors based SRC (NF-SRC) to effectively classify various signals. The dictionary is the conjunction of nearest neighbors and farthest neighbors. While the nearest neighbors are used to narrow the selection of candidate samples, the farthest neighbors are employed to make the dictionary redundant. The query signal is then decomposed in a greedy way similar to OMP. The proposed approach is evaluated over extensive experiments. The encouraging results demonstrate the feasibility of the proposed method.

Our future work includes the refinement of the coding strategy by designing more effective strategies to control the relative contribution of farthest neighbors. Moreover, the performance of the proposed approach should be further tested over more large-scale datasets.

ACKNOWLEDGMENTS

This work is supported by National Science Foundation under Grant No.0812458.

REFERENCES


