Locality-constrained Group Sparse Representation for Robust Face Recognition

Yu-Wei Chao, Yi-Ren Yeh, Yu-Wen Chen, Yuh-Jye Lee and Yu-Chiang Frank Wang

Research Center for Information Technology Innovation
Academia Sinica, Taipei, Taiwan
Outline

1. Introduction

2. Sparse Coding for Image Representation and Classification

3. Locality and Group-Sensitive Sparse Representation

4. Experimental Results

5. Conclusion
Introduction

- Sparse Representation
  - $\ell_1$-norm (Lasso)
  - $\ell_{1,2}$ mixed-norm (Group Lasso)

- Locality
  - Typically seen in problems of
    - Classification (e.g. NN, kNN)
    - Dimension reduction (e.g. Isomap, LLE)
    - Data representation (e.g. LLC $^1$)
  - Sparse representation does NOT guarantee locality $^1$

---

Our Proposed Method

- **Locality and Group-Sensitive Sparse Representation (LGSR)**
  - Learn data sparse representation by integrating both group sparsity and data locality into a unified formulation
  - Classification based on reconstruction error (e.g. MSE)
  - Possible extension to nonlinear versions by kernels
Different Coding Schemes

- (a) Sparse coding (SC)
- (b) SC + group lasso
- (c) SC + locality
- (d) Our method

**Figure:** Data representation with different coding strategies
Outline

1. Introduction

2. Sparse Coding for Image Representation and Classification

3. Locality and Group-Sensitive Sparse Representation

4. Experimental Results

5. Conclusion
Image Sparse Representation for Classification

- Sparse coding (SC) aims to linearly reconstruct a data instance using an over-complete dictionary:

\[
\min_w \|x - Bw\|_2^2 + \lambda \|w\|_1. \quad (1)
\]

- \(B\) is the codebook (a set of visual words) and \(\lambda\) penalizes the \(\ell_1\)-norm regularizer, which controls the sparsity of \(w\).

- For classification, one can \(^2\)
  - Use each \(w\) as a new presentation of each \(x\)
  - Train a discriminative classifier such as SVM

---

Another use of SC for Classification

- Based on data reconstruction, Wright et al. proposed the sparse representation classification (SRC) method for face recognition. \(^3\)
  - \(A = [A_1, A_2, \ldots, A_c] \in \mathbb{R}^{m \times n}\): entire training set
  - \(A_j = [x_{j1}, x_{j2}, \ldots, x_{jn_j}] \in \mathbb{R}^{m \times n_j}\): training images from \(j\)th class

The SRC minimizes the image reconstruction error:

\[
\min_w \|x_t - Aw\|_2^2 + \lambda \|w\|_1, \tag{2}
\]

where \(w = [w_{11}, w_{12}, \ldots, w_{cn_c}]^\top \in \mathbb{R}^{n \times 1}.

- Classification by minimizing MSE:

\[
j^* = \arg \min_j \|x_t - A_jw_j\|_2^2 \tag{3}
\]

- Small MSE = Good Reconstruction = Good Recognition?

---

Outline

1. Introduction
2. Sparse Coding for Image Representation and Classification
3. Locality and Group-Sensitive Sparse Representation
4. Experimental Results
5. Conclusion
SC with the Group Lasso Constraint

- A joint sparsity representation:  

\[ \min_w \|x_t - Aw\|_2^2 + \lambda \sum_{j=1}^{c} \|w_j\|_2. \]  

(4)

where \( w_j = [w_{j1}, w_{j2}, \ldots, w_{jn_j}]^\top \in \mathbb{R}^{n_j \times 1} \) be the associated coefficients of \( A_j \).

- Data representation with group sparsity is achieved by posing a \( \ell_{1,2} \) mixed-norm regularization.

---

\[ ^4 \text{X.-T. Yuan and S. Yan. Visual classification with multi-task joint sparse representation. CVPR, 2010.} \]
A locality-constrained linear coding (LLC) \(^1\)

LLC uses a distance regularization when minimizing MSE:

\[
\min_w \|x - Aw\|_2^2 + \lambda \|d \odot w\|_2^2, \tag{5}
\]

where \(d \in \mathbb{R}^{n \times 1}\) is the measurement of distance between \(x\) and each visual word in \(A\).

---

Our LGSR Algorithm

- Integration of the $\ell_{1,2}$ mixed-norm regularization and locality constraint:

$$\min_w \|x_t - Aw\|_2^2 + \lambda_1 \sum_{j=1}^{c} \|w_j\|_2 + \lambda_2 \|d \odot w\|_2^2,$$

where $\lambda_1$ and $\lambda_2$ penalize the group sparsity and locality constraints, respectively.

- We adopt the distance metric in LLC:

$$d_{ji} = \exp \left( \frac{\|x_t - x_{ji}\|_2}{\sigma} \right). \quad (6)$$

- Classification by minimizing MSE:

$$j^* = \arg \min_j \|x_t - A_jw_j\|_2^2 \quad (7)$$
Outline

1. Introduction
2. Sparse Coding for Image Representation and Classification
3. Locality and Group-Sensitive Sparse Representation
4. Experimental Results
5. Conclusion
Experimental Results

**ORL Face Database**

- 400 cropped face images of 40 human subjects under 10 different conditions (lighting, poses, etc.) with $56 \times 46$ pixels.
- For each subject, we randomly and equally split the data into training and test sets.
- We vary the number of Eigenfaces as features.
- For each experiment, we perform five random trials and compare our approach with SC-based methods.
Results of ORL Database (Eigenface)

![Graph showing recognition accuracy for different data dimensionality](image)

<table>
<thead>
<tr>
<th>dim</th>
<th>10</th>
<th>18</th>
<th>40</th>
<th>70</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_1$</td>
<td>0.1</td>
<td>0.05</td>
<td>0.01</td>
<td>0.1</td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td>0.5</td>
<td>0.01</td>
<td>0.001</td>
<td>0.001</td>
</tr>
</tbody>
</table>

**Figure:** Top: Recognition using different SC approaches with different data dimensionality. Bottom: Parameters of our LGSR.
Results of ORL Database (Fisherface)

We also extract Fisherfaces as features and repeat the experiments (all 39 eigenspaces for Fisherface are used).

**Table**: Recognition of the ORL database using Fisherface

<table>
<thead>
<tr>
<th>Method</th>
<th>kNN</th>
<th>SRC</th>
<th>SC+GL</th>
<th>SC+LLC</th>
<th>Ours</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>93.20±1.10</td>
<td>94.20±0.76</td>
<td>94.40±0.82</td>
<td>95.00±1.32</td>
<td><strong>95.40±1.43</strong></td>
</tr>
</tbody>
</table>
2414 frontal images of 38 human subjects with gray-level 192 \times 168 pixels, each image has up to 64 illumination variations.

We extract Fisherfaces as the features (all 37 eigenspaces for Fisherface are used) and consider different number of training images \(N_T\) per class for evaluation.

Once the training images are randomly extracted, the remaining will be test images.

For each experiment, we perform three random trials.
Experimental Results

Results of Extended Yale B Database

**Table:** Comparisons of recognition performance on Extended Yale B. $N_T$ indicates the number of training images.

<table>
<thead>
<tr>
<th>$N_T$</th>
<th>8</th>
<th>16</th>
<th>32</th>
</tr>
</thead>
<tbody>
<tr>
<td>kNN (k=3)</td>
<td>76.70 ± 1.14</td>
<td>86.99 ± 1.36</td>
<td>93.29 ± 0.42</td>
</tr>
<tr>
<td>SRC</td>
<td>84.38 ± 1.21</td>
<td>92.95 ± 0.42</td>
<td>97.83 ± 0.14</td>
</tr>
<tr>
<td>SC+GL</td>
<td>84.39 ± 1.21</td>
<td>92.95 ± 0.42</td>
<td>97.83 ± 0.17</td>
</tr>
<tr>
<td>SC+LLC</td>
<td>84.74 ± 1.24</td>
<td>93.17 ± 0.19</td>
<td>97.89 ± 0.13</td>
</tr>
<tr>
<td>Ours</td>
<td><strong>85.17 ± 1.15</strong></td>
<td><strong>93.54 ± 0.40</strong></td>
<td><strong>98.14 ± 0.21</strong></td>
</tr>
</tbody>
</table>
Figure: (a) An input test image. (b) Training images with non-zero weights using different SC methods (SC, SC+GL, SC+LLC, and our method).
Outline

1 Introduction

2 Sparse Coding for Image Representation and Classification

3 Locality and Group-Sensitive Sparse Representation

4 Experimental Results

5 Conclusion
Our LGSR balances group sparsity and locality constraint and improves classification performance.

More significant improvements were achieved when data with lower dimensionality was used.

Our LGSR is training free.

Future research directions at choosing $D$ for encoding and optimization in large-scale problems.