Sparse Models in Image Understanding And Computer Vision

Jayaraman J. Thiagarajan
Arizona State University

Collaborators
Prof. Andreas Spanias
Karthikeyan Natesan Ramamurthy
Sparsity

- Sparsity of a vector means that only a small number of its elements differ significantly from zero.
  - Modeling natural signals [Elad2010].
  - Supervised classification [Figueiredo 2003].
  - Understanding natural image statistics [Olshausen & Field 1996].

- Object recognition in the “cortex” [Mutch and Lowe 2008], [Riesenhuber and Poggio 2005]
  - Sparse, localized features
  - Hierarchical approach - Simple low-level features having are pooled and combined into complex, higher-level features.

Sparse methods lead to parsimonious (and interpretable) models in addition to being efficient for large scale learning.
Sparse Coding

- Solve an underdetermined (overcomplete) system of equations

\[ y = \Psi x \]

- Signal/data - \( y \in \mathbb{R}^M \)
- Dictionary - \( \Psi \in \mathbb{R}^{M \times K} \)
- Coefficient vector - \( x \in \mathbb{R}^K \)

\[
\begin{align*}
\min_x & \quad \|x\|_0 \quad \text{subject to} \quad \|y - \Psi x\|_2 \leq \epsilon \\
\min_x & \quad \|x\|_1 \quad \text{subject to} \quad \|y - \Psi x\|_2 \leq \epsilon
\end{align*}
\]

- **Algorithms** - Matching Pursuit, Orthogonal Matching Pursuit, Iterated shrinkage, LARS, Feature sign search and many others.

Combinatorial Complexity

Basis Pursuit [Chen 2001]
Dictionary Design

- The dictionary $\Psi$ can be constructed as a
  - Pre-defined set of basis functions
  - Union of orthonormal bases
  - Overcomplete set of features adapted to the data

- **Algorithms** – Conjugate gradient descent based methods, K-SVD, FOCUSS, MOD and many others…

- Adapting dictionaries to the training data – Generalization of data clustering [Thiagarajan et.al. 2010]
  - A data sample can be associated to more than one cluster and an activation value is computed for each cluster.
  - K-lines clustering – Special case of K-subspace clustering with $K = 1$ and the additional constraint that the subspace passes through origin.
Learning Global Dictionaries

- Training a dictionary for every new set of training samples is not feasible.

- A learning algorithm is a map from the space of training samples to the hypothesis space of functional solutions.

- Can the learning algorithm recover the underlying global dictionary “stably”?
  - A stable algorithm will depend only on the probability space to which the training samples belong.

- Given that the training error is small, can we ensure that the “Expected” test error is also small?
  - Need to obtain an upper bound for the difference between the empirical and expected errors.
Designing Global Dictionaries

- Natural image patches exhibit **redundancy** and hence can be efficiently coded.
- Image patches contain either **low dimensional geometric patterns** or **stochastic textures**, or a combination of both.
- Energy hierarchy in the patterns can be exploited.

**Multilevel Dictionary Learning**
An Example Dictionary

- Geometric patterns in the first few levels and stochastic patterns in the last few levels.
Compressed Sensing

- Recovery of images using compressed measurements
- Measurement system: Random (or) Optimized to the dictionary.

Perform sparse recovery using observations ($Z$)

Using 25% Noisy Measurements

KSVD (PSNR = 30.45 dB)

Proposed (PSNR = 32.58 dB)
Sparse Coding in Recognition

Challenges:
- No single descriptor can describe the whole dataset.
- Diverse nature and high dimensionality of the descriptors – vectors, histograms, matrices and tensors.

Proposed solution:
- Employ kernel methods to learn models using the similarities between data samples.
- Perform sparse coding in the feature space obtained by fusing multiple kernels (MKSR).
- Low-dimensional compact representation for recognition.

Learning dictionaries in the ensemble feature space.
Learn a separate dictionary for each descriptor and obtain ensemble kernel matrices for sparse coding.

Complexity: $O(MK)$. 
Multiple Kernel Sparse Representations

Approach 2

- Perform kernel dictionary learning using the ensemble kernel matrix directly.
- Complexity: \( O(MN) \).
Object Recognition Performance

- Caltech-101/Caltech-256 Object Datasets
  - SIFT, Self similarity, LBP, Gist, PHOG, Geometric Blur, C2-SWP, C2-ML.
  - Linear SVM for classifying the MKSR codes.
## Object Recognition Performance

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>5</th>
<th>10</th>
<th>15</th>
<th>20</th>
<th>25</th>
<th>30</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spatial Pyramid Matching</td>
<td>-</td>
<td>-</td>
<td>56.4</td>
<td>-</td>
<td>-</td>
<td>64.6</td>
</tr>
<tr>
<td>Sparse Coding + SPM</td>
<td>-</td>
<td>-</td>
<td>67</td>
<td>-</td>
<td>-</td>
<td>73.2</td>
</tr>
<tr>
<td>LLC + SPM</td>
<td>51.15</td>
<td>59.77</td>
<td>65.43</td>
<td>67.74</td>
<td>70.16</td>
<td>73.44</td>
</tr>
<tr>
<td>LC-KSVD</td>
<td>54</td>
<td>63.1</td>
<td>67.7</td>
<td>70.5</td>
<td>72.3</td>
<td>73.6</td>
</tr>
<tr>
<td>Multiple Kernel SC (Approach 1)</td>
<td>56.34</td>
<td>64.81</td>
<td>68.56</td>
<td>71.4</td>
<td>73.07</td>
<td>74.29</td>
</tr>
<tr>
<td>Multiple Kernel SC (Approach 2)</td>
<td>56.9</td>
<td>65.3</td>
<td>68.94</td>
<td>71.83</td>
<td>73.61</td>
<td>74.88</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>15</th>
<th>30</th>
<th>45</th>
<th>60</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sparse Coding + SPM</td>
<td>27.73</td>
<td>34.02</td>
<td>37.46</td>
<td>40.14</td>
</tr>
<tr>
<td>LLC + SPM</td>
<td>34.46</td>
<td>41.19</td>
<td>45.91</td>
<td>47.68</td>
</tr>
<tr>
<td>Multiple Kernel SC (Approach 1)</td>
<td>36.46</td>
<td>43.12</td>
<td>46.24</td>
<td>48.26</td>
</tr>
<tr>
<td>Multiple Kernel SC (Approach 2)</td>
<td>37.19</td>
<td>43.81</td>
<td>46.92</td>
<td>48.87</td>
</tr>
</tbody>
</table>
Tumor Segmentation

- Robust method to automatically segment a medical image into its constituent heterogeneous regions.
  - Active and necrotic tumor components from T1-weighted contrast enhanced MR images.

- Challenges:
  - Variability in size, shape and location.
  - Similarity in intensities of normal and abnormal brain tissue regions.
  - Intensity variations of identical tissues across volumes.
  - Avoid overestimation.
Kernel Coding for Segmentation

- Sparse coding typically applied to image patches (or) feature vectors.
  - Trivial to obtain sparse codes for pixel intensities.
  - Proposed solution: Perform coding using kernel similarities.

\[
E_T = \| \Phi(y_i) - \Phi(D_T)x_i^T \|_2
\]

\[
y_i = \begin{cases} 
    \text{Tumor}, & \text{if } E_N - E_T \geq \epsilon, \\
    \text{Non-tumor}, & \text{otherwise.}
\end{cases}
\]

\[
E_N = \| \Phi(y_i) - \Phi(D_N)x_i^N \|_2
\]
Tumor Identification

- Need to identify locally connected segments
  - Segmentation algorithms typically consider pixel locations in addition to intensities.

- Incorporation of locality information
  - Approach 1: Perform spectral clustering only on the pixels determined as tumor based on kernel codes.
  - Approach 2: Include the locality information as part of the kernel

- Ensemble kernel can be constructed as the Hadamard product of intensity and locality kernels.
  - Tumor region can be identified using linear SVM.

- Complexity reduction can be achieved by allowing user to initialize the tumor region.
Experiment Results
Experiment Results

<table>
<thead>
<tr>
<th>Image</th>
<th>Acc (%)</th>
<th>CR</th>
<th>Acc (%)</th>
<th>CR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Slice 1</td>
<td>93</td>
<td>0.9</td>
<td>94</td>
<td>0.93</td>
</tr>
<tr>
<td>Slice 2</td>
<td>96</td>
<td>0.95</td>
<td>97</td>
<td>0.95</td>
</tr>
<tr>
<td>Slice 3</td>
<td>92</td>
<td>0.9</td>
<td>95</td>
<td>0.94</td>
</tr>
<tr>
<td>Slice 4</td>
<td>90</td>
<td>0.86</td>
<td>90</td>
<td>0.86</td>
</tr>
<tr>
<td>Slice 5</td>
<td>94</td>
<td>0.85</td>
<td>94</td>
<td>0.87</td>
</tr>
<tr>
<td>Slice 6</td>
<td>92</td>
<td>0.82</td>
<td>92</td>
<td>0.81</td>
</tr>
<tr>
<td>Slice 7</td>
<td>94</td>
<td>0.76</td>
<td>95</td>
<td>0.72</td>
</tr>
<tr>
<td>Slice 8</td>
<td>98</td>
<td>0.95</td>
<td>98</td>
<td>0.95</td>
</tr>
<tr>
<td>Slice 9</td>
<td>98</td>
<td>0.92</td>
<td>98</td>
<td>0.92</td>
</tr>
<tr>
<td>Slice 10</td>
<td>92</td>
<td>0.81</td>
<td>92</td>
<td>0.84</td>
</tr>
</tbody>
</table>
Image Retrieval

- Local descriptors from small patches – Object recognition.
- For general image retrieval tasks – Typical to consider a heterogeneous combination of multiple features.
- Assumption: Well annotated tags available for a sample set.
- Is it possible to use this supervised information in coding?
Locality in Sparse Models

- **Moving away from VQ** – Relative importance of the different bases are not considered.
- **Sparse Coding** – Lesser reconstruction error, but loses correlation between the codes.
- **Consistency** - Similar features must have similar codes.
- **Salient patterns in the neighborhood** – Local linear model.
- By adding a suitable regularization for locality, sparse coding can provide improved recognition performance.

\[
\min_{x} \sum_{k=1}^{K} w(k)|x_k| \quad \text{subject to} \quad \|y - \Psi x\|_2 \leq \epsilon
\]

\[
w(k) = \|y - \psi_k\|_2^2 \quad \quad w(k) = \|y - (y^T \psi_k)\psi_k\|_2^2
\]
Supervised Coding

- Using heterogeneous features from large regions of an image.
- Provides enough variability to understand the interactions between different entities.
- Learning a dictionary for sparse coding these features – Bag of Visual Phrases.

- Can be further improved by performing supervised coding
  - Simultaneous sparse coding of features within a group (tag/label).
    \[
    \hat{X}(g) = \min_X \|Y(g) - \Psi W^{-1}_{(g)}X\|_F^2 \text{ s.t. } \|X\|_{row-0} \leq L
    \]
  - One image can be part of several groups.
  - Design dictionaries to “optimize” this representation.
Algorithm

**TRAINING**

1. **Group #1**: Divide each image into large overlapping regions (Sub-images)
2. **Group #2**: Extract multiple global/local features from each sub-image and build the heterogeneous feature
3. **Dictionary update using the proposed algorithm**
4. **Perform supervised local sparse coding of all sub-images in a group**
5. **Aggregate Max-pooled sparse codes at multiple spatial levels**

**TEST IMAGE**

1. **Build the heterogeneous feature for each sub-image**
2. **Perform supervised local sparse coding of all sub-images**
3. **Aggregate Max-pooled sparse codes at multiple spatial levels**
4. **Nearest Neighbor Retrieval**

**Hashing**
Simulations

- Bag of Words
- Sparse Coding
- Supervised Coding
Other Research Problems

- Sparse coding on Riemannian manifolds for activity recognition
- Dictionary Learning with graph embedding constraints
- Discriminative clustering in ambient and feature spaces
- Combined sparse representations
  - Derived conditions for unique recovery using convex and greedy methods.
- Wavelet domain statistical models for template learning
  - Fast image registration using non-stationary GMRF templates.
- Example based coding for Image recovery
- Shift-invariant sparse representations
- Transform domain features for ion-channel signal classification
References


