Robust Face Recognition via Sparse Representation

Allen Y. Yang
<yang@eecs.berkeley.edu>

April 18, 2008, NIST
Face Recognition: “Where amazing happens”
Robust Face Recognition via Sparse Representation

Face Recognition: “Where amazing happens”

Figure: Steve Nash, Kevin Garnett, Jason Kidd.
Sparse Representation

Sparsity

A signal is sparse if most of its coefficients are (approximately) zero.
Sparse Representation

Sparsity

A signal is sparse if most of its coefficients are (approximately) zero.

1. Sparsity in frequency domain

   Figure: 2-D DCT transform.

2. Sparsity in spatial domain

   Figure: Gene microarray data.

Allen Y. Yang <yang@eecs.berkeley.edu>  Robust Face Recognition via Sparse Representation
- **Sparsity in human visual cortex** [Olshausen & Field 1997, Serre & Poggio 2006]

1. **Feed-forward**: No iterative feedback loop.
2. **Redundancy**: Average 80-200 neurons for each feature representation.
3. **Recognition**: Information exchange between stages is not about individual neurons, but rather how many neurons as a group fire together.
Problem Formulation

Notation
- Training: For $K$ classes, collect training samples $\{v_{1,1}, \cdots, v_{1,n_1}\}, \cdots, \{v_{K,1}, \cdots, v_{K,n_K}\} \in \mathbb{R}^D$.
- Test: Present a new $y \in \mathbb{R}^D$, solve for label$(y) \in [1, 2, \cdots, K]$.
Problem Formulation

1. Notation
   - Training: For $K$ classes, collect training samples $\{v_{1,1}, \cdots, v_{1,n_1}\}, \cdots, \{v_{K,1}, \cdots, v_{K,n_K}\} \in \mathbb{R}^D$.
   - Test: Present a new $y \in \mathbb{R}^D$, solve for $\text{label}(y) \in [1, 2, \cdots, K]$.

2. Data representation in (long) vector form via stacking

![Diagram](image)

**Figure:** Assume 3-channel $640 \times 480$ image, $D = 3 \cdot 640 \cdot 480$. 
**Problem Formulation**

1. **Notation**
   - Training: For $K$ classes, collect training samples $\{v_{1,1}, \cdots, v_{1,n_1}\}, \cdots, \{v_{K,1}, \cdots, v_{K,n_K}\} \in \mathbb{R}^D$.
   - Test: Present a new $y \in \mathbb{R}^D$, solve for label($y$) $\in [1, 2, \cdots, K]$.

2. **Data representation** in (long) vector form via stacking

   
   ![Figure](image)

   **Figure:** Assume 3-channel $640 \times 480$ image, $D = 3 \cdot 640 \cdot 480$.

3. **Mixture subspace model** for face recognition [Belhumeur et al. 1997, Basri & Jacobs 2003]
Classification of Mixture Subspace Model

Assume $y$ belongs to Class $i$

\[ y = \alpha_{i,1}v_{i,1} + \alpha_{i,2}v_{i,2} + \cdots + \alpha_{i,n_i}v_{i,n_i}, \]

\[ = A_i\alpha_i, \]

where $A_i = [v_{i,1}, v_{i,2}, \cdots, v_{i,n_i}]$. 
Classification of Mixture Subspace Model

1. Assume $y$ belongs to Class $i$

\[
y = \alpha_{i,1}v_{i,1} + \alpha_{i,2}v_{i,2} + \cdots + \alpha_{i,n_i}v_{i,n_i},
\]

\[
= A_i\alpha_i,
\]

where $A_i = [v_{i,1}, v_{i,2}, \cdots, v_{i,n_i}]$.

2. Nevertheless, Class $i$ is the unknown variable we need to solve:

Sparse representation

\[
y = [A_1, A_2, \cdots, A_K] \begin{bmatrix}
\alpha_1 \\
\alpha_2 \\
\vdots \\
\alpha_K
\end{bmatrix} = Ax \in \mathbb{R}^{3.640.480}.
\]
Classification of Mixture Subspace Model

1. Assume $y$ belongs to Class $i$
\[
y = \alpha_{i,1}v_{i,1} + \alpha_{i,2}v_{i,2} + \cdots + \alpha_{i,n_i}v_{i,n_i},
\]
\[
= A_i \alpha_i,
\]
where $A_i = [v_{i,1}, v_{i,2}, \cdots, v_{i,n_i}]$.

2. Nevertheless, Class $i$ is the unknown variable we need to solve:

Sparse representation
\[
y = [A_1, A_2, \cdots, A_K] \begin{bmatrix}
\alpha_1 \\
\alpha_2 \\
\vdots \\
\alpha_K
\end{bmatrix} = Ax \in \mathbb{R}^{3,640 \cdot 480}.
\]

3. $x_0 = [0 \cdots 0 \alpha_i^T 0 \cdots 0]^T \in \mathbb{R}^n$.

Sparse representation encodes membership!
Construct linear projection $R \in \mathbb{R}^{d \times D}$, $d$ is the **feature dimension**.

$$\tilde{y} = Ry = RAx_0 = \tilde{A}x_0 \in \mathbb{R}^{d}.$$ 

$\tilde{A} \in \mathbb{R}^{d \times n}$, but $x_0$ is unchanged.
Dimensionality Reduction

1. Construct linear projection $R \in \mathbb{R}^{d \times D}$, $d$ is the feature dimension.

\[ \tilde{y} = Ry = RAx_0 = \tilde{A}x_0 \in \mathbb{R}^d. \]

$\tilde{A} \in \mathbb{R}^{d \times n}$, but $x_0$ is unchanged.

2. Holistic features
   - Eigenfaces [Turk 1991]
   - Fisherfaces [Belhumeur 1997]
   - Laplacianfaces [He 2005]

3. Partial features

4. Unconventional features
   - Downsampled faces
   - Random projections
$\ell^1$-Minimization

1. Ideal solution: $\ell^0$-Minimization

\[
(P_0) \quad x^* = \arg \min_x \| x \|_0 \quad \text{s.t.} \quad \hat{y} = \tilde{A} x.
\]

$\| \cdot \|_0$ simply counts the number of nonzero terms. However, generally $\ell^0$-minimization is *NP-hard*. 
\( \ell^1 \)-Minimization

1. **Ideal solution:** \( \ell^0 \)-Minimization

\[
(P_0) \quad x^* = \arg \min_x \| x \|_0 \ \text{s.t.} \ \tilde{y} = \tilde{A} x.
\]

\( \| \cdot \|_0 \) simply counts the number of nonzero terms. However, generally \( \ell^0 \)-minimization is *NP-hard*.

2. **Compressed sensing:** Under mild condition, \( \ell^0 \)-minimization is equivalent to

\[
(P_1) \quad x^* = \arg \min_x \| x \|_1 \ \text{s.t.} \ \tilde{y} = \tilde{A} x,
\]

where \( \| x \|_1 = |x_1| + |x_2| + \cdots + |x_n| \).
\[ (P_0) \quad x^* = \arg \min_x \|x\|_0 \text{ s.t. } \tilde{y} = \tilde{A}x. \]

\[ \| \cdot \|_0 \text{ simply counts the number of nonzero terms.} \]

However, generally \( \ell^0 \)-minimization is \( NP \)-hard.

**Compressed sensing:** Under mild condition, \( \ell^0 \)-minimization is equivalent to

\[ (P_1) \quad x^* = \arg \min_x \|x\|_1 \text{ s.t. } \tilde{y} = \tilde{A}x, \]

where \( \|x\|_1 = |x_1| + |x_2| + \cdots + |x_n| \).

**\( \ell^1 \)-Ball**

- \( \ell^1 \)-Minimization is convex.
- Solution equal to \( \ell^0 \)-minimization.
\(\ell^1\)-Minimization Routines

- **Matching pursuit** [Mallat 1993]
  1. Find most correlated vector \(v_i\) in \(A\) with \(y\): \(i = \text{arg max} \langle y, v_j \rangle\).
  2. \(A \leftarrow A^{(i)}, x_i \leftarrow \langle y, v_i \rangle, y \leftarrow y - x_i v_i\).
  3. Repeat until \(\|y\| < \epsilon\).

- **Basis pursuit** [Chen 1998]
  1. Start with number of sparse coefficients \(m = 1\).
  2. Select \(m\) linearly independent vectors \(B_m\) in \(A\) as a basis
     \[x_m = B_m^\dagger y.\]
  3. Repeat swapping one basis vector in \(B_m\) with another vector not in \(B_m\) if improve \(||y - B_m x_m||\).
  4. If \(||y - B_m x_m||_2 < \epsilon\), stop; Otherwise, \(m \leftarrow m + 1\), repeat Step 2.

- **Quadratic solvers**: \(y = Ax_0 + z \in \mathbb{R}^d\), where \(\|z\|_2 < \epsilon\)
  \[x^* = \text{arg min}\{\|x\|_1 + \lambda \|y - Ax\|_2\}\]
  [LASSO, Second-order cone programming]: Much more expensive.

Matlab Toolboxes for \(\ell^1\)-Minimization

- **\(\ell^1\)-Magic** by Candes
- **SparseLab** by Donoho
- **cvx** by Boyd
Sparse Representation Classification

Solve \((P_1) \Rightarrow x_1\).

Project \(x_1\) onto face subspaces:

\[
\delta_1(x_1) = \begin{bmatrix}
\alpha_1 \\
0 \\
\vdots \\
0
\end{bmatrix}, \quad \delta_2(x_1) = \begin{bmatrix}
0 \\
\alpha_2 \\
\vdots \\
0
\end{bmatrix}, \quad \cdots, \quad \delta_K(x_1) = \begin{bmatrix}
0 \\
0 \\
\vdots \\
\alpha_K
\end{bmatrix}.
\] (1)
Sparse Representation Classification

Solve $\left( P_1 \right) \Rightarrow x_1$.

1. Project $x_1$ onto face subspaces:

$$\delta_1(x_1) = \begin{bmatrix} \alpha_1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \delta_2(x_1) = \begin{bmatrix} 0 \\ \alpha_2 \\ \vdots \\ 0 \end{bmatrix}, \quad \cdots, \quad \delta_K(x_1) = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ \alpha_K \end{bmatrix}. \quad (1)$$

2. Define residual $r_i = \| \tilde{y} - \tilde{A}\delta_i(x_1) \|_2$ for Subject $i$:

- $id(y) = \arg \min_{i=1, \ldots, K} \{ r_i \}$
### Partial Features on Extended Yale B Database

<table>
<thead>
<tr>
<th>Features</th>
<th>Nose</th>
<th>Right Eye</th>
<th>Mouth &amp; Chin</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dimension</td>
<td>4,270</td>
<td>5,040</td>
<td>12,936</td>
</tr>
<tr>
<td>SRC [%]</td>
<td>87.3</td>
<td>93.7</td>
<td>98.3</td>
</tr>
<tr>
<td>nearest-neighbor [%]</td>
<td>49.2</td>
<td>68.8</td>
<td>72.7</td>
</tr>
<tr>
<td>nearest-subspace [%]</td>
<td>83.7</td>
<td>78.6</td>
<td>94.4</td>
</tr>
<tr>
<td>Linear SVM [%]</td>
<td>70.8</td>
<td>85.8</td>
<td>95.3</td>
</tr>
</tbody>
</table>

SRC: sparse-representation classifier
Extension I: Outlier Rejection

- $\ell^1$-Coefficients for invalid images
Extension I: Outlier Rejection

- $\ell^1$-Coefficients for invalid images

Outlier Rejection

When $\ell^1$-solution is not sparse or concentrated to one subspace, the test sample is invalid.

Sparsity Concentration Index: $\text{SCI}(x) \doteq \frac{K \cdot \max_i \|\delta_i(x)\|_1/\|x\|_1 - 1}{K - 1} \in [0, 1]$. 

Allen Y. Yang <yang@eecs.berkeley.edu>  
Robust Face Recognition via Sparse Representation
Figure: ROC curve on Eigenfaces and AR database.
Extension II: Occlusion Compensation

![Face Recognition Example](image)

Robust Face Recognition via Sparse Representation

Allen Y. Yang <yang@eecs.berkeley.edu>
Extension II: Occlusion Compensation

1. Sparse representation + sparse error

\[ y = Ax + e \]

2. Occlusion compensation

\[ y = (A | I) \begin{pmatrix} x \\ e \end{pmatrix} = Bw \]
Figure: Training samples for Subject 1.
Figure: Training samples for Subject 1.

(a) random corruption

(b) occlusion
Figure: Training samples for Subject 1.

(a) random corruption

(b) occlusion

<table>
<thead>
<tr>
<th>sunglasses</th>
<th>scarves</th>
</tr>
</thead>
<tbody>
<tr>
<td>97.5%</td>
<td>93.5%</td>
</tr>
</tbody>
</table>
Future Directions

Open problems:

1. Pose variation
2. Scalability to $>1000$ subjects
Future Directions

Open problems:
1. Pose variation
2. Scalability to $> 1000$ subjects

Other databases:
1. Multi-PIE (about 350 subjects)
2. Chinese CASPEAL (about 1000-3000 subjects)
Future Directions

Open problems:
1. Pose variation
2. Scalability to > 1000 subjects

Other databases:
1. Multi-PIE (about 350 subjects)
2. Chinese CASPEAL (about 1000-3000 subjects)

Wish list: Because few algorithm succeed under all-weather conditions (illumination, expression, pose, disguise), we are looking forward to a comprehensive database
1. large number of subjects
2. carefully controlled subclasses
Acknowledgments

Collaborators

- **Berkeley**: Prof. Shankar Sastry
- **UIUC**: Prof. Yi Ma, John Wright, Arvind Ganesh

Funding Support

- ARO MURI: Heterogeneous Sensor Networks (HSNs)

References

- http://www.eecs.berkeley.edu/~yang