Robust Face Recognition via Sparse Representation

Authors: John Wright, Allen Y. Yang, Arvind Ganesh, S. Shankar Sastry, and Yi Ma

Presented by
Hu Han
Jan. 30 2014

For CSE 902 by Prof. Anil K. Jain: Selected Topics in Recognition by Machine
Outline

• Background
• Formulation of face recognition via sparse representation
• Experimental results
• Summary and discussion
What is automatic face recognition (AFR)?

• Automatically identify or verify a person from still images, video sequences or even sketches.
Formulation of face identification

• Traditional formulation: Nearest neighbor classifier
  
  – $\text{Identity}(I_p) = \arg \max_{i=1 \ldots n} \{ S_{p,i} \}$
  – $S_{p,i}$ is a similarity measure between the probe image $I_p$ and a gallery image $I_i$
  – Image $I$ is of vectorized form $(m \times 1)$

The traditional formulation is to assign the identity of the most similar face image in the gallery set to a probe face image.
Formulation of face identification

• Can we formulate the face identification problem based on spare representation?

\[(\ell^0) : \hat{x}_0 = \arg\min \|x\|_0 \quad \text{subject to} \quad Ax = y,\]

\[(\ell^1) : \hat{x}_1 = \arg\min \|x\|_1 \quad \text{subject to} \quad Ax = y.\]

– where a signal \(y\) is coded over a dictionary \(A\) as \(y = Ax\), and \(x\) is a sparse coefficient vector
Formulation of face identification

• We can, provided a probe face image can be represented using a linear combination of the gallery face images

\[ I_p = x_1 I_1 + x_2 I_2 + \cdots + x_n I_n \]

\[(l^1): \hat{x}_1 = \text{arg min} \|x\|_1 \quad \text{subject to} \quad Ax = y\]
Mathematical fundamentals

• Mean (the average face)
  \[ -\mu = \frac{1}{n} \sum_{i=1}^{n} I_i \]

• Covariance matrix \( C \) of the gallery set
  \[ -C = \frac{1}{n} \sum_{i=1}^{n} (I_i - \mu)(I_i - \mu)^T = \frac{1}{n} AA^T \]
  \[ -A = [I_1 - \mu, I_2 - \mu, \ldots, I_n - \mu] \] (The normalized gallery set)
Mathematical fundamentals

• Rank
  – The maximum number of linearly independent column or row vectors in a matrix

\[
A = \begin{bmatrix}
1 & 1 & 0 & 2 \\
-1 & -1 & 0 & -2
\end{bmatrix}
\]

  – One of the important properties of real matrices

\[
\text{rank}(A^T A) = \text{rank}(AA^T) = \text{rank}(A) = \text{rank}(A^T)
\]

\[A^T A\] has the same number \((n)\) of non-zero eigenvalues as \(C\)
Mathematical fundamentals

• Eigenvectors and eigenvalues of the gallery set

\[ \mathbf{Cv}_k = \lambda_k \mathbf{v}_k, \quad k = 1, 2, \ldots, n \]

• Principal component analysis (PCA)

\[ \mathbf{I}_p = \mu + \sum_{i=1}^{K} \alpha_i \mathbf{v}_i \quad K \leq n \]

The first term is a linear combination of all the gallery samples \( \mathbf{A} \)

Is the second term also a linear combination of all the gallery samples \( \mathbf{A} \)?
Rationality analysis

• $I_p = x_1 I_1 + x_2 I_2 + \cdots + x_n I_n$ is reasonable or not?

• Denote $L = A^T A$; $L$ is real and symmetric, and thus it has $n$ non-zero eigenvalues

\[ Lu_k = \gamma_k u_k, \quad k = 1, 2, \ldots, n \]
Rationality analysis

\[
C = \frac{1}{n} \sum_{i=1}^{n} (I_i - \mu)((I_i - \mu))^T = \frac{1}{n} AA^T
\]

\[nCA = n \frac{1}{n} AA^T A = AL\]

\[nCAu_k = ALu_k = A\gamma_k u_k = \gamma_k Au_k\]

\[CAu_k = \frac{1}{n} \gamma_k Au_k\]

\[CAu_k = \frac{\gamma_k}{n} Au_k\]

\[Cv_k = \lambda_k v_k\]

\[v_k = Au_k\]
Rationality analysis

- The eigenvectors $\mathbf{v}_k$ of a set of gallery face images
  $\mathbf{A}$ is the linear combination of its original data $\mathbf{A}$, and $\mathbf{u}_k$ is the coefficients for linear combination.

**PCA model**

\[
I_p = \mu + \sum_{i=1}^{K} \alpha_i \mathbf{v}_i
\]

The first term is a linear combination of all the gallery samples $\mathbf{A}$

The second term is also a linear combination of all the gallery samples $\mathbf{A}$
Formulation of face identification

• Therefore, it is reasonable to represent a probe face image using a simple linear combination of gallery face images

\[ I_p = x_1 I_1 + x_2 I_2 + \cdots + x_n I_n \]

• We can also reversely solve the linear combination coefficients based on a sparse representation formulation

\[ \hat{x} = (x_1, x_2, \ldots, x_n) = \arg \min_{x} ||x||_1 \quad s.t. \quad Ax = I_p \]
Formulation of face identification

• The sparse representation itself is not a classifier

• But a classifier can be built based on the solved sparse coefficients $\hat{x} = (x_1, x_2, \ldots, x_n)$
Classification using the sparse coefficients

- One classification scheme that we can naturally think of

\[ \text{Identity}(I_p) = \arg \max_{i=1 \ldots n} |x_i| \]

The extended YaleB database: ~1200 training images of 38 subjects
Classification using the sparse coefficients

- The classification scheme proposed in Wright et al.'s paper:

\[
\text{Identity}(I_p) = \arg \min_{i=1...m} e_i = \arg \min_{i=1...k} \|I_p - A\delta_i(\hat{x})\|_2
\]

\(\delta_i(\cdot)\) selects the coefficients associated with the i-th subject (totally \(k\) subjects)
Sparse Representation-based Classification (SRC) Algorithm

1: **Input:** a matrix of training samples
   \[ A = [A_1, A_2, \ldots, A_k] \in \mathbb{R}^{m \times n} \text{ for } k \text{ classes, a test sample} \]
   \[ y \in \mathbb{R}^m, \text{ (and an optional error tolerance } \varepsilon > 0.) \]
2: Normalize the columns of \( A \) to have unit \( \ell^2 \)-norm.
3: Solve the \( \ell^1 \)-minimization problem:
   \[ \hat{x}_1 = \arg \min_x \| x \|_1 \text{ subject to } Ax = y. \]
   (Or alternatively, solve
   \[ \hat{x}_1 = \arg \min_x \| x \|_1 \text{ subject to } \| Ax - y \|_2 \leq \varepsilon. \])
4: Compute the residuals \( r_i(y) = \| y - A \delta_i(\hat{x}_1) \|_2 \)
   for \( i = 1, \ldots, k. \)
5: **Output:** identity(\( y \)) = \( \arg \min_i r_i(y). \)

- In SRC, the training set is also the gallery set. All face images must be aligned.
- I am using \( l_p \) instead of \( y \) for a probe image.
Why sparse representation?

• In practice, the linear system $I_p = Ax$ is typically underdetermined, because there are fewer equations ($m$) than unknowns ($n$).

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} = [I_1 \ I_2 \ \cdots \ I_n]x = \begin{bmatrix} a_{1,1} & a_{2,1} & \cdots & a_{n,1} \\ a_{1,2} & a_{2,2} & \cdots & a_{n,2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1,m} & a_{2,m} & \cdots & a_{n,m} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

$m = 12 \times 10 = 120$ downsampling

$n = 1200$
Why sparse representation?

• An underdetermined system either has no unique solution or has no solution
  – An example with no solution
    \[ x + y + z = 1 \]
    \[ x + y + z = 0 \] \textbf{Inconsistent}
  – But for the linear system \( I_p = Ax \) in face recognition, we assume there is at least one solution
Why sparse representation?

- If there is no sparsity constraint
  \[
  \hat{x} = (x_1, x_2, \ldots, x_n) = \arg \min_{x} \|x\|_2 \text{ s.t. } Ax = I_p
  \]

\( m = 12 \times 10 = 120 \) downsampling

\( \hat{x} \) is not discriminative, and it does not indicate any identity of the probe image;

The nonzero entries in \( \hat{x} \) are very unstable;
Experimental evaluations

• Face recognition on Extended Yale B database
  – 2,414 images (192x168) of 38 subjects
  – Half (~ 32) of the images per subject used for training (gallery), the other half for testing
  – Thus, the dictionary size (n) is ~1200
  – Features: Eigenfaces, Laplacianfaces, Randomfaces, Fisherfaces, and raw intensities with m= 30, 56, 120, and 504, respectively
  – SRC is compared with Nearest Neighbor, Nearest Subspace, and Linear SVM
A brief description of different feature extraction methods

• Eigenfaces
  – Project a face image into a low-dimensional subspace that keeps the largest the data variations

• Laplacianfaces
  – Project a face image into a low-dimensional subspace that keeps the local relationship in high-dimensional domain

• Randomfaces
  – Project a face image into a new subspace using a Gaussian random matrix

• Fisherfaces
  – Project a face image into a low-dimensional subspace that best separates individual classes
Face recognition on Extended Yale B

• Example images of Extended Yale B Extreme illumination

http://vision.ucsd.edu/~leekc/ExtYaleDatabase/ExtYaleB.html
Face recognition on Extended Yale B

**SRC**

**Nearest Neighbor**

**Nearest Subspace**

**Linear SVM**

Good feature extraction method is critical for SVM
Experimental evaluations

• Face recognition on AR
  – 14,00 images (165x120) of 100 subjects (50 males & 50 females)
  – 7 images per subject used for training (gallery), 7 images for testing
  – Thus, the dictionary size \( n \) is \( \sim 700 \)
  – Features: Eigenfaces, Laplacianface, Randomfaces, Fisherfaces, and raw intensities with \( m = 30, 56, 120, \) and 504, respectively
  – SRC is compared with Nearest Neighbor, Nearest Subspace, and Linear SVM
Face recognition on AR

- Example images of AR: Illumination, expression, occlusions

http://www2.ece.ohio-state.edu/~aleix/ARdatabase.html
Face recognition on AR

(a) SRC

(b) Nearest Neighbor

(c) Nearest Subspace

(d) Linear SVM
Experimental evaluations

• Face recognition on Extended YaleB with random pixel corruption
  – 1,170 images (96x84) of 38 subjects
  – 717 images used for training (gallery) and 453 images for testing
  – Thus, the dictionary size is ~717
  – Percentage of random pixel corruption: 0%, 10%, 20%, ..., 90%
    • Replace a pixel intensity with uniformly sampled value from [0, max-pixel intensity]
  – Feature: raw intensity (8064 dimensions)
Face recognition on Extended YaleB with random pixel corruption

• Example images with random pixel corruption

30%

50%

70%
Face recognition on Extended YaleB with random pixel corruption
Experimental evaluations

• Face recognition on Extended YaleB with random block occlusion
  – 1,170 images (96x84) of 38 subjects
  – 717 images for training (gallery) and 453 images for testing
  – Thus, the dictionary size is ~717
  – Percentage of random block size: 0%, 10%, 20%, ..., 50%
    • The location of occlusion square block is random
  – Feature: raw intensity (8064 dimensions)
Face recognition on Extended YaleB with random block occlusion

- Example images with random block occlusion

30% by LS
Face recognition on Extended YaleB with random block occlusion
Experimental evaluations

• Face recognition on Extended Yale B database using partial face (component) features

The holistic faces are already aligned with each other. Therefore, component images are also aligned

<table>
<thead>
<tr>
<th>Features</th>
<th>Nose</th>
<th>Right Eye</th>
<th>Mouth &amp; Chin</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dimension ((d))</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>NN</td>
<td>49.2%</td>
<td>68.8%</td>
<td>72.7%</td>
</tr>
<tr>
<td>NS</td>
<td>83.7%</td>
<td>78.6%</td>
<td>94.4%</td>
</tr>
<tr>
<td>SVM</td>
<td>70.8%</td>
<td>85.8%</td>
<td>95.3%</td>
</tr>
</tbody>
</table>

Feature: raw intensity
Experimental evaluations

- Face recognition on AR database using block based features

<table>
<thead>
<tr>
<th>Algorithms</th>
<th>Rec. rate sunglasses</th>
<th>Rec. rate scarves</th>
</tr>
</thead>
<tbody>
<tr>
<td>SRC (partitioned)</td>
<td>87.0% (97.5%)</td>
<td>59.5% (93.5%)</td>
</tr>
<tr>
<td>PCA + NN</td>
<td>70.0%</td>
<td>12.0%</td>
</tr>
<tr>
<td>ICA I + NN</td>
<td>53.5%</td>
<td>15.0%</td>
</tr>
<tr>
<td>LNMF + NN</td>
<td>33.5%</td>
<td>24.0%</td>
</tr>
<tr>
<td>$\ell^2$ + NS</td>
<td>64.5%</td>
<td>12.5%</td>
</tr>
</tbody>
</table>

Feature: raw intensity

Partition is based on the aligned holistic images, and thus each block is aligned.
Extension of SRC

• Sparse representation based rejection of invalid face images
  – The subject in a probe image is not enrolled in the gallery set
  – A probe image is not a face image

• Sparsity concentration index (SCI):
  \[ \hat{x} = (x_1, x_2, ..., x_n) = \arg \min_{i=1...n} ||x||_1 \quad s.t. \quad A\hat{x} = I_p \]
  \[ SCI(\hat{x}) = \frac{k \cdot \max_i \frac{||\delta_i(\hat{x})||_1}{||x||_1} - 1}{k - 1} \quad \in [0,1] \]
Sparsity concentration index (SCI)

\[
SCI(\hat{x}) = \frac{k \cdot \max_i \frac{\|\delta_i(\hat{x})\|_1}{\|x\|_1} - 1}{k - 1} \in [0,1]
\]

• Rejection criterion: \(SCI(\hat{x}) < \tau\)

• Physical explanation: If a testing face image can be well represented by a single subject in the gallery set, the testing subject has been enrolled in the system; otherwise the subject has not been enrolled in the gallery
Experimental evaluations

• Invalid face rejection on Extended YaleB with random block occlusion
  – 1,170 images (96x84) of 38 subjects
  – 360 images of 19 subjects for training (gallery) and 453 images of 38 subjects for testing
  – Percentage of random block size: 0%, 10%, 30%, and 50%
    • The location of occlusion square block is random
  – Feature: raw intensity
Invalid face rejection on Extended YaleB with random block occlusion
Summary and discussion

• Sparsity constraint in SRC assures the discriminative characteristics of solved coefficients for face recognition
• SRC is robust to various feature extraction methods, such as Eigenfaces, Laplacianface, Randomfaces, Fisherfaces, and raw intensities. SVM achieves similar performance as SRC but SVM is sensitive to the feature extraction methods
• In the linear system $I_p = A\mathbf{x}$, the dictionary $A$ should be significantly overcomplete ($m \ll n$) in order to get good performance
• The SRC method requires that the face images are aligned
Resources

• L-1 minimization algorithms (code)
  – http://www.eecs.berkeley.edu/~yang/software/l1benchmark/
• Talks on sparse representation
  – http://www.cs.technion.ac.il/~elad/talks/
• Discussions on face recognition using SRC
Q&A
Thanks!