Toward Sparse Coding on Cosine Distance

Jonghyun Choi, Hyunjong Cho, Jungsuk Kwak#, Larry S. Davis

UMIACS | University of Maryland, College Park
#Stanford University
Visual Recognition or Classification

- Input: visual data
- Output: semantic label of the visual data

Image
- Dalmatian

Video
- Scarlet Johansson
- Ballet
Typical Pipeline

- Visual Data
- Feature Descriptor
- Learning classifiers
- Learning Metric
- Output
  - Label probability
  - Nearest neighbor
  - Support vector machine
  - Deep convolutional neural net
  - **Sparse representation**
  - Metric learning
Sparse Representation

Test example

\[ x \approx 0.8 \times \phi_{36} + 0.3 \times \phi_{42} + 0.5 \times \phi_{63} \]

\[ [0, 0, ..., 0, 0.8, 0, ..., 0, 0.3, 0, ..., 0, 0.5, ...] \]

\[ = [a_1, ..., a_{64}] \quad \text{(feature representation)} \]

Compact & easily interpretable

[Example by Andrew Ng]
Sparse Representation based Classification (SRC) Pipeline

\[ \hat{x} = \arg \min_x \|y - Dx\|_2^2 + \gamma \|x\|_p, \]

Input Image → Feature Descriptor → Generating a sparse code → NN matching

DB
Conventional Sparse Coding

• Regularized least-square solution
  – on Euclidean metric

\[
\hat{x} = \arg \min_x \|y - Dx\|_2^2 + \gamma \|x\|_p,
\]
Better Metric on Visual Features

• Visual features are often based on histogram
  – Bag of words
  – SIFT
  – Histogram of Oriented Gradients (HOG)
  – Local Binary Pattern (LBP)

• Angular distance is better at histogram\(^1\)

---

**Table 3.** Holistic-based recognition rates (%) of different feature descriptors on CMU-PIE

<table>
<thead>
<tr>
<th>Similarities</th>
<th>Raw</th>
<th>EOH</th>
<th>Gabor</th>
<th>LBP</th>
<th>M LBP</th>
</tr>
</thead>
<tbody>
<tr>
<td>L1</td>
<td>40.32</td>
<td>40.32</td>
<td>66.13</td>
<td>51.61</td>
<td>69.35</td>
</tr>
<tr>
<td>L2</td>
<td>41.94</td>
<td>43.55</td>
<td>70.97</td>
<td>56.45</td>
<td>72.58</td>
</tr>
<tr>
<td>Cosine</td>
<td>48.39</td>
<td>61.29</td>
<td>74.19</td>
<td>72.58</td>
<td>87.10</td>
</tr>
</tbody>
</table>

\(^1\) Exploring Feature Descriptors for Face Recognition
S. Yan, H. Wang, X. Tang and T. Huang, ICASSP 2007
Cosine-Distance Based Sparse Coding

• Naïve formulation

\[ \hat{x} = \arg \min_x \left( 1 - \frac{y^T D x}{\|y\|_2 \|D x\|_2} \right) + \gamma \|x\|_p, \]

Non convex on x: hard to optimize
If $|y|=1$ and $|Dx|=1$

• Observation\textsuperscript{[1]}

$$||y - Dx||^2_2 = y^T y + (Dx)^T (Dx) - 2y^T Dx$$

$$= 1 + 1 - 2 \frac{y^T Dx}{1} = 2 - 2 \frac{y^T Dx}{||y||_2 \cdot ||Dx||_2}$$

$$= 2 \cdot (1 - \text{cosine similarity}(y, Dx))$$

$$= 2 \cdot \text{cosine distance}(y, Dx)$$

\textsuperscript{[1]} Angular Decomposition

D. Sun, C. H. Q. Ding, B. Luo, and J. Tang, IJCAI 2011
Our Formulation

• Approximated Cosine distance-based sparse coding objective with constraints on $|y|$ and $|Dx|$

$$\min_x ||\hat{y} - Dx||_2^2 + \alpha \left| 1 - ||Dx||_2^2 \right| + \gamma ||x||_1,$$

s.t. $0 < \alpha < 1,$

$$||\hat{y}||^2 = 1$$

• The normalization constraints are sometimes good or sometimes bad for classification accuracy\(^[1]\)

---

\(^{[1]}\) Dictionary learning algorithms for sparse representation
Without Constraints on $|y|$ and $|Dx|$ 

**Theorem 2.** If the norms of two vectors are both $n$, then the square of the Euclidean distance is proportional to their cosine distance (with a factor of $2n^2$).

**Proof:** The proof follows the proof of Theorem 1 with $\|y\|^2 = \|Dx\|^2 = n^2$. 

- More general formulation without the normalization constraint

$$
\min_x \|y - Dx\|^2 + \alpha \left( \|y\|^2 - \|Dx\|^2 \right) + \gamma \|x\|_1, \\
\text{s.t. } 0 < \alpha < 1.
$$
Approx. Cosine vs. Euclidean Distance

\[
\min_x \|y - Dx\|^2_2 + \alpha \|\|y\|_2^2 - \|Dx\|_2^2\| + \gamma \|x\|_1,
\]

s.t. \(0 < \alpha < 1\).

• The new formulation only differs from Euclidean distance based one by the term

• The effect of new term
  • Example from an experiment of E-Yale Dataset
  • Without normalization constraints

Fig. 1. An example of growing of \(|Dx|\) with or without the new term that enforces the norm of \(|Dx|\) to be closer to norm of \(|y|\) (Left). ‘Loss’ in the right figure means \(\|y - Dx\|^2\) (Right). The red curve is generative by the original Euclidean distance based sparse coding formulation (\(\alpha = 0\)). The blue curve is generated by our new objective function with \(\alpha = 0.001\). (must viewed in color)
Experimental Set-up

• Datasets
  – UCF 101 Action Recognition dataset
    • 101 classes, 13,320 clips from 27 hours of YouTube video footages
    • Largest and the most challenging dataset of the kind
    • SIFT+Space-time interest Point(STIP)+Dense Trajectory Feature(DTF)
  – Two face-recognition datasets
    • **AR dataset**: 2,600 frontal face images of 100 subjects (50 males and 50 females)
    • **Extended-YaleB dataset**: 2,414 frontal face images of 38 subjects (about 64 images/subject)
    • RandomFace, HOG, LBP, Gabor, SIFT
Action Recognition: UCF101

- Average accuracy

<table>
<thead>
<tr>
<th>Methods</th>
<th>Set1</th>
<th>Set2</th>
<th>Set3</th>
<th>Avg</th>
</tr>
</thead>
<tbody>
<tr>
<td>EUC-Sparse</td>
<td>66.1</td>
<td>66.1</td>
<td>69.2</td>
<td>67.1</td>
</tr>
<tr>
<td>Ours</td>
<td>68.0</td>
<td>68.3</td>
<td>70.4</td>
<td>68.9</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Methods</th>
<th>Set1</th>
<th>Set2</th>
<th>Set3</th>
<th>Avg</th>
</tr>
</thead>
<tbody>
<tr>
<td>EUC-Sparse</td>
<td>65.3</td>
<td>65.9</td>
<td>67.7</td>
<td>66.3</td>
</tr>
<tr>
<td>Ours</td>
<td>67.5</td>
<td>67.8</td>
<td>69.1</td>
<td>68.1</td>
</tr>
</tbody>
</table>

- Class-wise accuracy improvement (%): Our method – Euclidean-SparseCoding


[2] Learning a Discriminative Dictionary for Sparse Coding via Label Consistent K-SVD
### Face Recognition

#### AR Dataset

<table>
<thead>
<tr>
<th>Feature</th>
<th>Plain-Dict</th>
<th>EUC-SC</th>
<th>Ours</th>
<th>EUC-SC</th>
<th>Ours</th>
</tr>
</thead>
<tbody>
<tr>
<td>Randomface</td>
<td>66.3</td>
<td>66.3</td>
<td>94.3</td>
<td>94.3</td>
<td></td>
</tr>
<tr>
<td>HOG</td>
<td>84.8</td>
<td>86.0</td>
<td>99.7</td>
<td>99.7</td>
<td></td>
</tr>
<tr>
<td>LBP</td>
<td>89.2</td>
<td>89.5</td>
<td>99.7</td>
<td>99.8</td>
<td></td>
</tr>
<tr>
<td>Gabor</td>
<td>94.7</td>
<td>94.2</td>
<td>99.8</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>SIFT</td>
<td>91.2</td>
<td>92.2</td>
<td>99.8</td>
<td>100</td>
<td></td>
</tr>
</tbody>
</table>

#### Extended-YaleB

<table>
<thead>
<tr>
<th>Feature</th>
<th>Plain-Dict</th>
<th>EUC-SC</th>
<th>Ours</th>
<th>EUC-SC</th>
<th>Ours</th>
<th>LC-KSVD2</th>
<th>Ours</th>
</tr>
</thead>
<tbody>
<tr>
<td>Randomface</td>
<td>92.6</td>
<td>92.7</td>
<td>94.3</td>
<td>94.3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HOG</td>
<td>95.4</td>
<td>95.5</td>
<td>98.0</td>
<td>98.2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LBP</td>
<td>87.1</td>
<td>87.7</td>
<td>97.8</td>
<td>98.4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gabor</td>
<td>84.0</td>
<td>84.3</td>
<td>92.6</td>
<td>92.6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SIFT</td>
<td>96.2</td>
<td>96.3</td>
<td>99.8</td>
<td>99.9</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Comparative Face Identification Accuracy

#### Extended Yale B and AR Dataset

<table>
<thead>
<tr>
<th>Approach</th>
<th>eYaleB Acc. (%)</th>
<th>AR Acc. (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SRC (Best) [1]</td>
<td>99.0</td>
<td>97.5</td>
</tr>
<tr>
<td>LLC (Best) [38]</td>
<td>96.7</td>
<td>88.7</td>
</tr>
<tr>
<td>LC-KSVD2 (Best) [29]</td>
<td>99.0</td>
<td>97.8</td>
</tr>
<tr>
<td><strong>Ours</strong></td>
<td><strong>99.9</strong></td>
<td><strong>100</strong></td>
</tr>
</tbody>
</table>
Conclusion

• Propose a new formulation of sparse coding on cosine metric
  – Convex, easy to solve by a simply modification of the feature-sign algorithm
• Show that the new formulation changes the solution path
• Outperform the conventional formulation in three visual recognition datasets
Questions?

Thank you!

http://umiacs.umd.edu/~jhchoi/cossparse

Supported by MURI from the Office of Naval Research under the Grant N00014-10-1-0934.