22nd INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION

Toward Sparse Coding on Cosine Distance

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

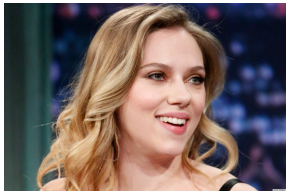
UMIACS | University of Maryland, College Park
[#]Stanford University

Visual Recognition or Classification

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

Image

Dalmatian

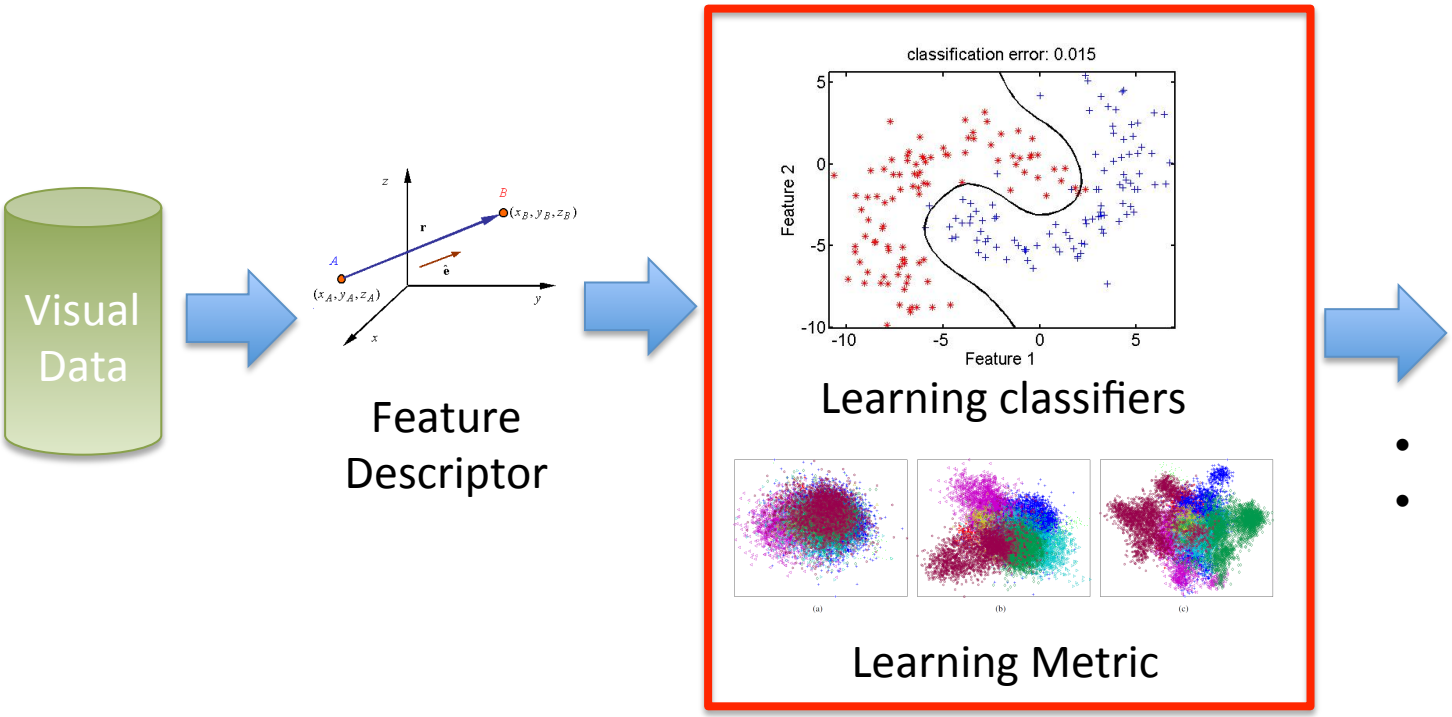


Scarlet Johansson

Video

Ballet

Typical Pipeline



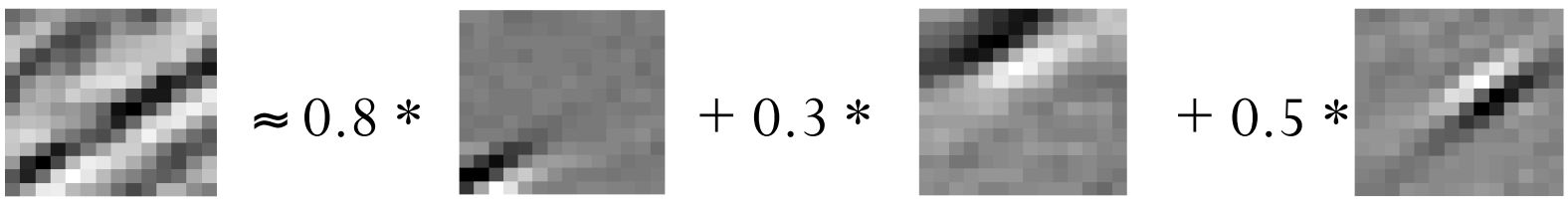
Output

- Label probability
- Nearest neighbor

- Support vector machine
- Deep convolutional neural net
- **Sparse representation**
- Metric learning

Sparse Representation

Test example



$x \approx 0.8 * \phi_{36} + 0.3 * \phi_{42} + 0.5 * \phi_{63}$

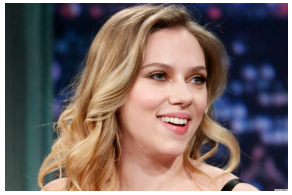
[0, 0, …, 0, **0.8**, 0, …, 0, **0.3**, 0, …, 0, **0.5**, …]

= [a_1, …, a_{64}] (feature representation)

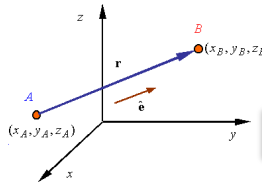
Compact & easily interpretable

[Example by Andrew Ng]

Sparse Representation based Classification (SRC) Pipeline



Input Image



Feature Descriptor

$$\hat{x} = \arg\min_{x} \|y - Dx\|_2^2 + \gamma \|x\|_p,$$

Generating a sparse code

DB

NN matching

Conventional Sparse Coding

- Regularized least-square solution
  - on **Euclidean metric**

Predefined dictionary

$$\hat{x} = \arg\min_{x} \|y - Dx\|_2^2 + \gamma \|x\|_p,$$

Input signal

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

Similarities	Raw	EOH	Gabor	LBP	M LBP
L1	40.32	40.32	66.13	51.61	**69.35**
L2	41.94	43.55	70.97	56.45	**72.58**
Cosine	48.39	61.29	74.19	72.58	**87.10**

[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

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

Non convex on x: hard to optimize

If |y|=1 and |Dx|=1

- Observation[1]

$$
\begin{aligned}
||y - Dx||_2^2 &= y^T y + (Dx)^T (Dx) - 2y^T Dx \\
&= 1 + 1 - 2\frac{y^T Dx}{1} = 2 - 2\frac{y^T Dx}{\|y\|_2 \cdot \|Dx\|_2} \\
&= 2 \cdot (1 - \text{cosine similarity}(y, Dx)) \\
&= 2 \cdot \text{cosine distance}(y, Dx)
\end{aligned}
$$

[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|

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

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

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

- The normalization constraints are sometimes good or sometimes bad for classification accuracy[1]

[1] **Dictionary learning algorithms for sparse representation**
K. Kreutz-Delgado, J. F. Murray, B. D. Rao, K. Engan, T. W. Lee, and T. J. Sejnowski, Neural Computation 2003

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

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

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

Approx. Cosine vs. Euclidean Distance

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

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

- 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

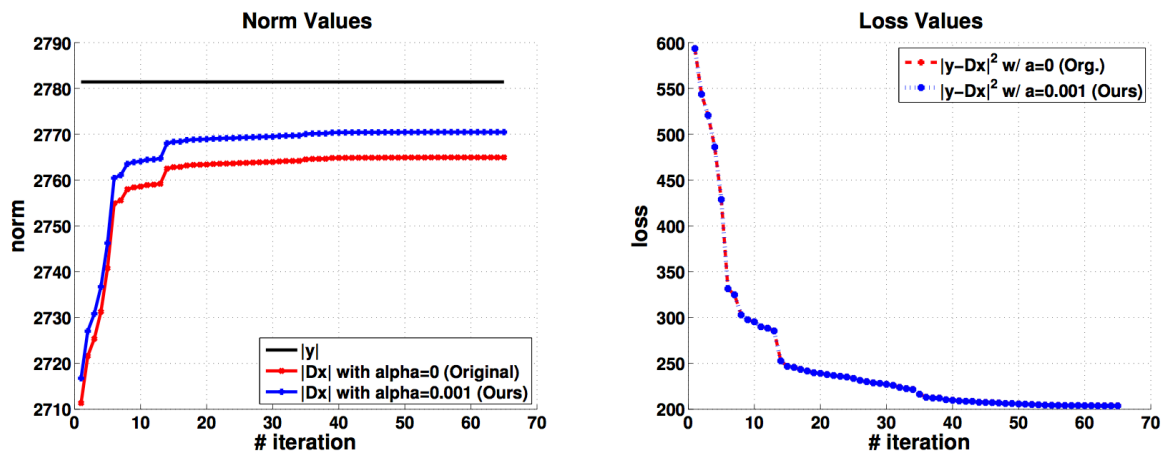
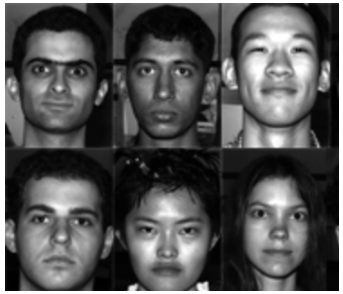


Fig. 1. An example of growing of $|Dx|$ with or without the new term that enforces the norm of $|Dx|$ to be closer to norm of $|y|$ (Left). 'Loss' in the right figure means $||y - Dx||^2$ (Right). The red curve is generative by the original Euclidean distance based sparse coding formulation ($\alpha = 0$). 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

UCF101



Extended-Yale B

AR

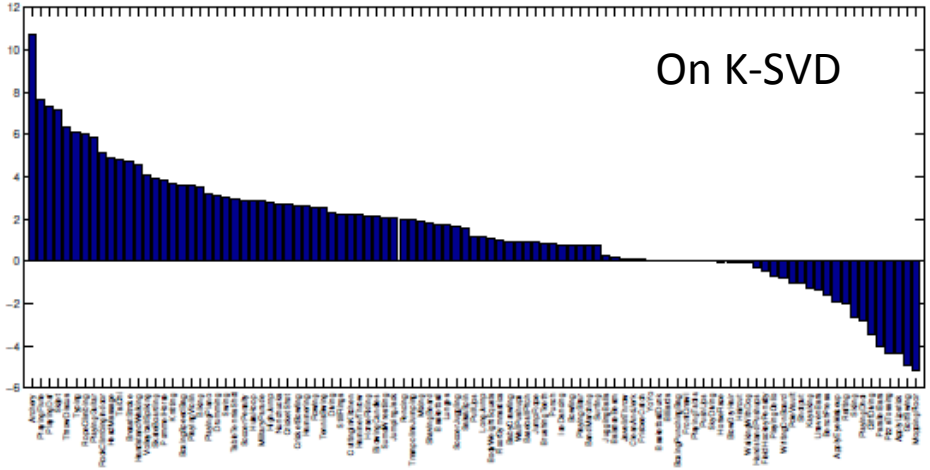
Action Recognition: UCF101

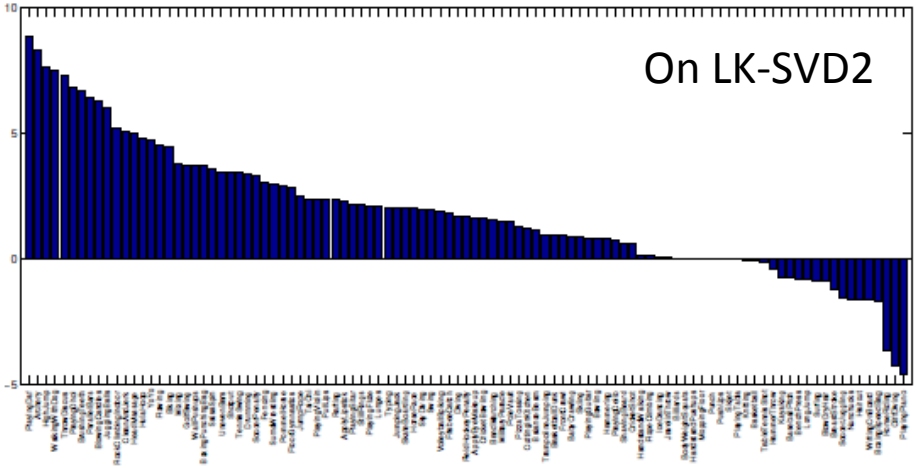
- Average accuracy

TABLE I. AVERAGE RECOGNITION ACCURACY (%) ON UCF101 DATASET. WE COMPARE OUR METHOD (OURS) TO THE CONVENTIONAL EUCLIDEAN DISTANCE BASED SPARSE CODING (EUC-SPARSE) ON DICTIONARIES LEARNED BY K-SVD AND LC-KSVD2.

Methods	Set1	Set2	Set3	Avg
on a dictionary by K-SVD [1]				
EUC-Sparse	66.1	66.1	69.2	67.1
Ours	**68.0**	**68.3**	**70.4**	**68.9**
on a dictionary by LC-KSVD2 [2]				
EUC-Sparse	65.3	65.9	67.7	66.3
Ours	**67.5**	**67.8**	**69.1**	**68.1**

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





[1] **K-SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation**
M. Aharon, M. Elad, and A. Bruckstein, IEEE T-Signal Processing, 2006

[2] **Learning a Discriminative Dictionary for Sparse Coding via Label Consistent K-SVD**
Z. Jiang, Z. Lin, and L. S. Davis, in CVPR, 2011.

Face Recognition

- AR Dataset

TABLE III. FACE IDENTIFICATION ACCURACY ON THE AR DATASET BY SRC APPROACH [1]. WE USE 20 IMAGES PER SUBJECT AS A TRAINING SET (DICTIONARY ATOM), TESTING WITH 6 IMAGES PER SUBJECT.

Feature	Accuracy (%)			
	Plain Dict		LC-KSVD2	
	EUC-SC	**Ours**	EUC-SC	**Ours**
Randomface	66.3	66.3	94.3	94.3
HOG	84.8	**86.0**	99.7	99.7
LBP	89.2	**89.5**	99.7	**99.8**
Gabor	**94.7**	94.2	99.8	**100**
SIFT	91.2	**92.2**	99.8	**100**

- Extended-YaleB

TABLE II. FACE IDENTIFICATION ACCURACY ON THE EXTENDED YALEB DATASET BY SRC [1] APPROACH. WE USE A PLAIN DICTIONARY AND A DICTIONARY LEARNED BY LC-KSVD2. EUC-SC AND OURS STAND FOR EUCLIDEAN DISTANCE BASED SPARSE CODING OBJECTIVE FUNCTION AND OURS, RESPECTIVELY.

Feature	Accuracy (%)			
	Plain Dict		LC-KSVD2	
	EUC-SC	**Ours**	EUC-SC	**Ours**
Randomface	92.6	**92.7**	94.3	94.3
HOG	95.4	**95.5**	98.0	**98.2**
LBP	87.1	**87.7**	97.8	**98.4**
Gabor	84.0	**84.3**	92.6	92.6
SIFT	96.2	**96.3**	99.8	**99.9**

TABLE IV. COMPARATIVE FACE IDENTIFICATION ACCURACY ON THE EXTENDED YALE B AND AR DATASET. **OURS** REFERS TO OUR BEST RESULT (WITH SIFT DESCRIPTOR).

Approach	eYaleB	AR
	Acc.(%)	Acc.(%)
SRC (Best) [1]	99.0	97.5
LLC (Best) [38]	96.7	88.7
LC-KSVD2 (Best) [29]	99.0	97.8
Ours	**99.9**	**100**

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.