Are Sparse Representation and Dictionary Learning Good for Handwritten Character Recognition?

Chi Nhan Duong, Kha Gia Quach, and Tien D. Bui

Concordia University, Montréal, Canada

Email: {c duon, k_q, bui}@encs.concordia.ca

14th International Conference on Frontiers in Handwriting Recognition (ICFHR) September 1- 4, 2014 in Crete, Greece.

UNIVERSITÉ Concordia UNIVERSITY

Contents

- Motivation
- Contributions
- Related works
- Sparse representation based recognition
- Experimental results
- Conclusions

Motivation

Why Sparse Representation & Dictionary Learning ?

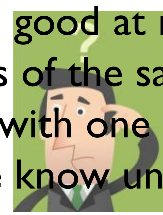
Image classification

Face recognition

Image denoising and inpainting

Medical Imaging

Human vision is good at recognizing different objects of the same kind, for example chairs with one leg or many legs, or someone we know under occlusions. So human visual system tends to retain certain sparse information that is common among objects of the same kind. And sparse representation has become a hot topic of investigation over the last few years.



Handwritten character recognition

Letter D

Could these theories produce good results for handwritten character recognition as in the case of other applications?

Contributions

- Developing a sparse representation based system for handwriting character recognition.
- Analyzing different factors that affect the SR based system such as: the choice of input data, the size of dictionary, and computation time of this method in three benchmark databases.
- Experimental results show that using this framework, the choice of feature space is less important comparing to other methods.

UNIVERSITÉ Concordia UNIVERSITY

Related works

Character image

Feature extraction

Classification

Recognized character

Statistical [6]
(Zoning, projections, crossing,…)

Structural [1]
(Topological and geometrical information)

Transformation-based [7][8]
(Gabor features, Wavelet transform coefficients)

K-Nearest Neighbor [9]

Support Vector Machine [9]

Neural network, Convolution Neural Network (LeNet) [2]

Investigate the representations of character images that are invariant to various writing styles

Investigate the classification methods

Université Concordia University

Related works

- **Hybrid approach**
  - Combining SVM & Convolution Neural Network [10].
  - Combining different features & different classifiers [1]

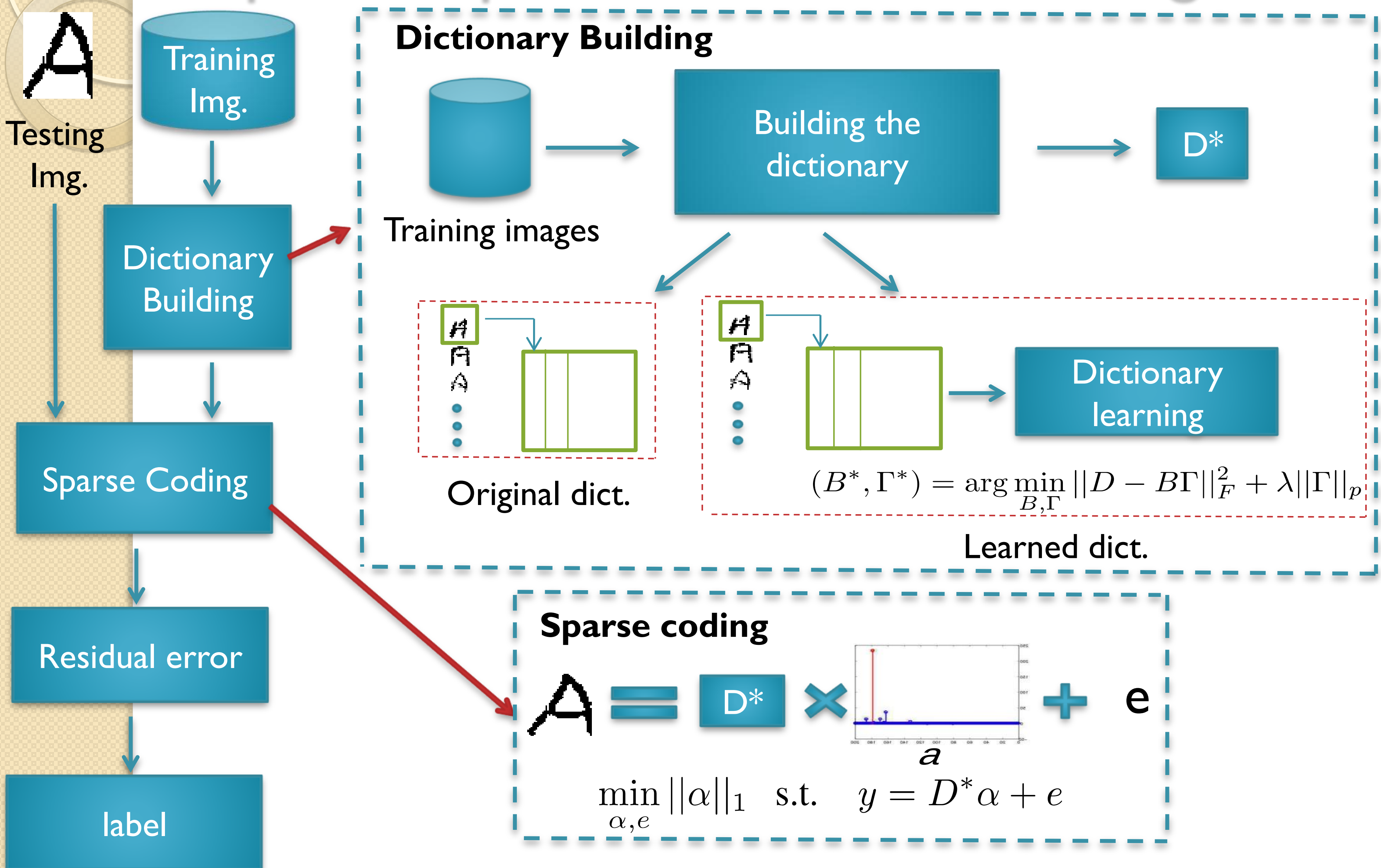
=> can exploit the strengths of features and classifiers, but expensive to decide which architecture is good for specific data.

- **Zhang et al. [11]**: decomposed image into three parts: low-rank component, sparse component and error (i. e. noise) ➔ mainly focus on *handwriting recovery*.

=> Testing with 240 images/digit and achieving 91.24% for MNIST.

- **Wei et al. [12]** took into account local information for dictionary learning and then using the learned dictionary to improve the performance.

Sparse representation based recognition



Sparse representation based recognition

Algorithm 1: Sparse representation based handwritten character recognition

Input:

- Set of training images of k classes
- Testing sample $y \in R^N$

(1) Stack the images of each class as columns of matrix $D_i, i = 1, .., k$.

(2) **Building the dictionary:**

(a) Use the original matrix D; or

(b) Use the learned matrix B^*.

$$(B^*, \Gamma^*) = \arg\min_{B,\Gamma} ||D - B\Gamma||_F^2 + \lambda||\Gamma||_p$$

(3) **Sparse Coding:** Solving (5) to obtain the sparse representation α of y

$$\min_{\alpha,e} ||\alpha||_1 \quad \text{s.t.} \quad y = D^*\alpha + e$$

(4) Compute the residuals and classify y

$$r_i = ||y - D^*\alpha_i||_2, i = 1..k$$

Output label of $label_y \leftarrow \arg\min r_i$

Experimental results

- Databases

Database	# Training	# Testing	Image size
MNIST	60000	10000	28 × 28
US Portal Service (USPS)	7291	2007	16 × 16
CEDAR – upper case	11454	1367	32 × 32
CEDAR – lower case	7691	816	32 × 32

Evaluations

- Effects of dimensional reduction and feature spaces.
- Dictionary learning for character recognition.
- Effect of dictionary sizes.
- Computational time.
- Comparison with other methods

Effects of dimensional reduction

Table I
EFFECT OF DIMENSIONAL REDUCTION

Input Data	MNIST	USPS	CEDAR	
			LWR	UPPR
Raw Img.	97.4 (784)	95.67 (256)	92.65(1024)	93.56(1024)
PCA(t=80)	97.2(47)	95.22(20)	91.05(67)	92.17(95)
PCA(t=70)	97.1(29)	94.27(15)	91.05(31)	91.66(46)
PCA(t=60)	95.82 (18)	90.13(10)	89.83(17)	91(25)

accuracy

dimension

Table II
EFFECT OF FEATURE SPACES

Input Data	MNIST	USPS	CEDAR	
			LWR	UPPR
Raw Img.	**97.4**	95.67	**92.65**	**93.56**
Gradient	97.35	**95.96**	88.36	92.68
Gabor	91.22	91.48	70.35	75.2

Gabor feature [7] is mainly designed for digits rather than for character images

Performance of this feature:

- Use k-nearest neighbor (k=3)
- MNIST: 90.45%
- USPS: 89.74%
- CEDAR:
  - Upper case: 49.63 %
  - Lower case: 52.38 %

Effects of dictionary learning

Table III
DICTIONARY LEARNING FOR SR

	MNIST	USPS	CEDAR	
			LWR	UPPR
SR-RAW	97.4	95.67	92.65	93.56
SR-RAW + Dict. Learning	97.66	96.26	89.09	89.61
SR-PCA	97.2	95.07	91.05	92.17
SR-PCA + Dict. Learning	97.25	95.52	87.26	87.2

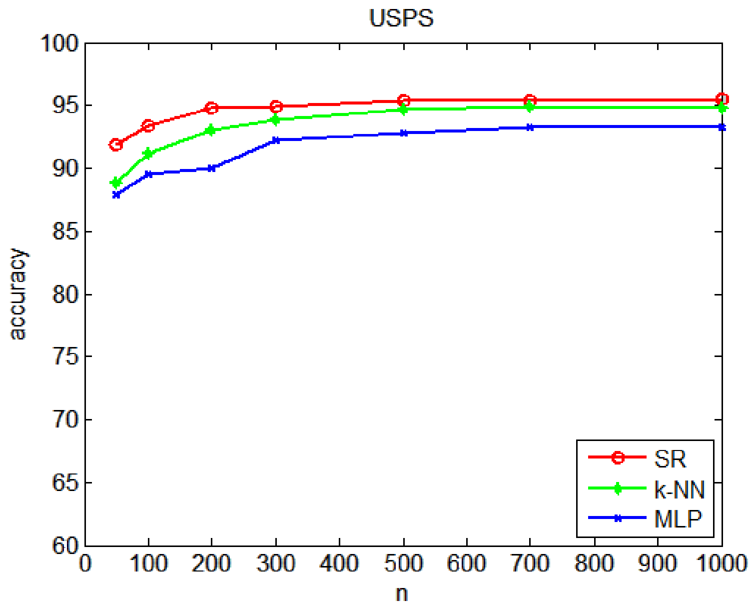
- Dictionary learning → boosting the accuracy of SR based system
- UPPR & LWR: reduce bout 3%

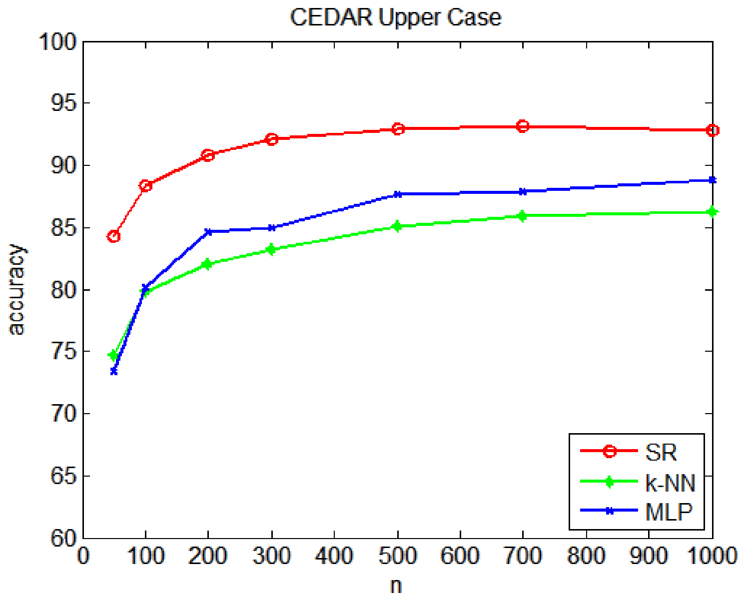
➔ increasing the number of classes (26 instead of 10)

➔ insufficient training data for some characters (only ~ 5 images/characters) ➔ reduce the quality of atoms comparing with original full images.

Effect of dictionary sizes

Evaluation system:
(1) Choose n images (per class) randomly
(2) Classification based on dictionary conducted from these images.
Ex:
For USPS (10 classes)
n = 100 => total images used for learning dictionary is N = 100 * 10 = 1000





Overall performance

Table IV
COMPUTATIONAL TIME

	MNIST	CEDAR-LWR	CEDAR-UPPR
SR-RAW	0.25	0.17	0.28
SR-PCA	0.045	0.023	0.04
VAM [15] (no feature extraction time)	0.069	0.018	0.026

Table V
COMPARING WITH OTHER METHODS

	MNIST	LWR (26 classes)	UPPR (26 classes)
SR-RAW	98.21	92.65	93.56
VAM [15]	99.03	93.5	95.9
SAB [16]	NA	84.93	79.52
BLU [17]	NA	71.52	81.58

[15] G. Vamvakas, B. Gatos, and S. J. Perantonis, "Handwritten character recognition through two-stage foreground sub-sampling," *PR*, 2010.

Conclusions

- In this paper, we have developed a sparse representation based system for handwritten character recognition.
- Representing the testing image as a combination of atoms in a dictionary makes the system more robust to the changes in feature spaces and the dimension of input data.
- Different factors that affect the performance of the system are also examined in our experiments.
- *Although the best performance of SR based system cannot beat the state-of- the-art methods, its ability to remove the effect of feature space can help to improve its flexibility and efficiency.*


Q
A


THANK
YOU

UNIVERSITÉ Concordia UNIVERSITY

References

[1] A.Rahman and B.Verma,"Effect of ensemble classifier composition on offline cursive character recognition," *Information Processing & Management*, pp. 852–864, 2013.

[2] D. Ciresan, U. Meier, and J. Schmidhuber,"Multicolumn deep neural networks for image classification," in *CVPR*. IEEE, 2012, pp. 3642–3649.

[3] J.Wright,Y. Ma, J. Mairal, G. Sapiro,T. S. Huang, and S.Yan,"Sparse representation for computer vision and pattern recognition," *Proceedings of the IEEE*, pp. 1031–1044, 2010.

[4] V. M. Patel and R. Chellappa,"Sparse representations, compressive sensing and dictionaries for pattern recognition," in *ACPR*. IEEE, 2011, pp. 325–329.

[5] J.Wright,A.Y.Yang,A. Ganesh, S. S. Sastry, and Y. Ma,"Robust face recognition via sparse representation," *PAMI, IEEE Transactions on*, pp. 210–227, 2009.

[6] D. Impedovo, G. Pirlo, and R. Modugno,"New advancements in zoning-based recognition of handwritten characters," in *ICFHR, 2012 International Conference on*. IEEE, 2012.

[7] Y. Hamamoto, S. Uchimura, M.Watanabe,T.Yasuda,Y. Mitani, and S.Tomita,"A Gabor filter-based method for recognizing handwritten numerals," *PR*, pp. 395–400, 1998.

[8] S.-W. Lee, C.-H. Kim, H. Ma, and Y.Y.Tang,"Multiresolution recognition of unconstrained handwritten numerals with wavelet transform and multilayer cluster neural network," *PR*, 1996.

References

[9] Y. LeCun, L. Jackel, L. Bottou, C. Cortes, J. S. Denker, H. Drucker, I. Guyon, U. Muller, E. Sackinger, P. Simard *et al.*, "Learning algorithms for classification: A comparison on handwritten digit recognition," *Neural networks: the statistical mechanics perspective*, p. 276, 1995.

[10] X.-X. Niu and C. Y. Suen, "A novel hybrid CNN–SVM classifier for recognizing handwritten digits," *PR*, pp. 1318–1325, 2012.

[11] Z. Zhang, C.-L. Liu, and M. Zhao, "Handwriting representation and recognition through a sparse projection and low-rank recovery framework," in *IJCNN*. IEEE, 2013, pp. 1–8.

[12] C.-P. Wei, Y.-W. Chao, Y.-R. Yeh, and Y.-C. F. Wang, "Locality-sensitive dictionary learning for sparse representation based classification," *PR*, pp. 1277–1287, 2013.

[13] Y. LeCun and C. Cortes, "The MNIST database of handwritten digits," 1998.

[14] J. J. Hull, "A database for handwritten text recognition research," *PAMI, IEEE Transactions on*, pp. 550–554, 1994.

[15] G. Vamvakas, B. Gatos, and S. J. Perantonis, "Handwritten character recognition through two-stage foreground sub-sampling," *PR*, pp. 2807–2816, 2010.

References

[16] T. Saba, A. Rehman, and G. Sulong, "Improved statistical features for cursive character recognition," *IJICIC*, 2011.

[17] M. Blumenstein, X. Y. Liu, and B. Verma, "A modified direction feature for cursive character recognition," in *IJCNN '04*. IEEE, 2004, pp. 2983–2987

Algorithm 2 Algorithm for dictionary learning [14]

Input: Image data matrix X and parameter λ

Step 1: Initialize $\mathcal{D}$ randomly with unit l_2-norm for each column of $\mathcal{D}$

Step 2: Fix $\mathcal{D}$ and solve Λ

Solve the following minization problem using convex optimization technique described in [18]

$$J_\Lambda = \arg\min_\Lambda\{\|X - \mathcal{D}\Lambda\|_F^2 + \lambda\|\Lambda\|_1\}$$

Step 3: Fix Λ and update $\mathcal{D}$

We update d_j one by one while fixing all the other columns of $\mathcal{D}$, i.e. $\mathrm{d}_l, l \neq j$. We can find the update by optimizing the following problem.

$$J_\mathcal{D} = \arg\min_\mathcal{D}\|X - \mathcal{D}\Lambda\|_F^2 \ \ s.t. \ \mathrm{d}_j^T\mathrm{d}_j = 1, \forall j$$

We use Lagrange multiplier Y to convert the objective function. After that differentiating J_{d_j} w.r.t. d_j, and set it to 0. We have

$$\mathrm{d}_j = Y\alpha_j^T(\alpha_j\alpha_j^T - \lambda)^{-1}$$

$$\mathrm{d}_j = Y\alpha_j^T/\|Y\alpha_j^T\|_2$$

Step 4: Go back to step 2 until the values of $J_\mathcal{D}$ and J_Λ are converged or the maximum number of iterations is reached. Finally, output $\mathcal{D}$.

Output: $\mathcal{D}$

[14] M. Yang and L. Zhang, "Gabor feature based sparse representation for face recognition with gabor occlusion dictionary," in ECCV. Springer, 2010, pp. 448–461.

[18] K. Koh, S.-J. Kim, and S. Boyd, "An interior-point method for largescale l1-regularized logistic regression." Journal of Machine learning research, vol. 8, no. 7, 2007.