Local Sparse Representation Based Classification

3 authors, including:

Chun-Guang Li
Beijing University of Posts and Telecommunications
64 PUBLICATIONS   **1,118** CITATIONS

Jun Guo
Beijing University of Posts and Telecommunications
391 PUBLICATIONS   **4,052** CITATIONS

Local Sparse Representation based Classification

Chun-Guang Li, Jun Guo and Hong-Gang Zhang
School of Information and Communication Engineering
Beijing University of Posts and Telecommunications
Beijing, China
{*lichunguang, guojun, zhhg*}*@bupt.edu.cn*

Abstract—**In this paper, we address the computational complexity issue in Sparse Representation based Classification (SRC). In SRC, it is time consuming to find a global sparse representation. To remedy this deficiency, we propose a Local Sparse Representation based Classification (LSRC) scheme, which performs sparse decomposition in local neighborhood. In LSRC, instead of solving the l_1-norm constrained least square problem for all of training samples we solve a similar problem in the local neighborhood of each test sample. Experiments on face recognition data sets ORL and Extended Yale B demonstrated that the proposed LSRC algorithm can reduce the computational complexity and remain the comparative classification accuracy and robustness.**

Keywords-**LSRC; SRC; k-nn; Sparse Representation;**

I. INTRODUCTION

Recently the algorithmic problem of computing sparse linear representations with respect to an over-complete dictionary of base elements or signal atoms has received a great deal of attentions in statistical signal processing community [1], [2], [3], [4], [5]. The sparse representation is widely used for different applications, such as signal separation [6], denoising [7], [8], image inpainting [9], robust classification [10], inducing similarity measurement [11] and shadow removal [12].

In this paper, we focus on the issue of calculating sparse representation for classification, which proposed by Wright et al.[10], called Sparse Representation based Classification (SRC). In SRC, the sparse representation is used for robust face recognition to cope with noise corruption, occlusion, outlier detection and etc. The basic idea of SRC is that: (a) to compute the sparse decomposition of test sample on training data set at first; and then (b) to calculate the reconstruction residual errors which reconstruct test sample by sparse decomposition coefficients that related to training samples of each class respectively; finally (c) to classify test sample into the class which remains the minimum reconstruction residual error. In SRC, the step of finding sparse representation is fast in theory if the sparsest solution is found (i.e. few non-zero coefficients). In practice, however, it is very time consuming to calculate a sparse representation. This is caused by the fact that the existing sparse representation algorithms [1], [2], [3], [4], [5] have difficulty to produce exactly zero-valued coefficients in finite time. As a result the calculated sparse

representation has a long tail (i.e., many small non-zero reconstruction coefficients) which leads to time consuming.

The expensive computational cost in SRC hamper its applications in pattern recognition as a general reconstructive classifier. To alleviate this deficiency, we propose a modified scheme, called Local Sparse Representation based Classification (LSRC) in which one step to determine the local neighborhood is added before calculating sparse representation. Experiments conducted on face recognition data sets ORL and Extended Yale B demonstrated that the proposed LSRC algorithm can reduce the computational complexity and remains the comparative classification accuracy and robustness.

II. A BRIEF REVIEW OF SPARSE REPRESENTATION BASED CLASSIFICATION(SRC)

SRC formulated in [10] belongs to the reconstructive classification approach which aims at tackling the classification problem on data with corruption (i.e. noise, missing data and outliers)

Given training data set $\{\mathbf{a}_j \in R^m, j = 1, 2, ..., n\}$, where each sample $\mathbf{a}_j$ relates a class label l_j and the number of classes is c. Let matrix $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_n] \in R^{m \times n}$. Given an optional error tolerance $\epsilon > 0$ and an unseen test sample $\mathbf{y} \in R^m$, the SRC algorithm can be summarized as below:

1) (Preprocessing) Normalize each column of $\mathbf{A}$ to unit l_2- norm;
2) Solve the l_1- norm minimization problem

$$\hat{\mathbf{x}}_1 = arg \min_{\mathbf{x}} \quad \|\mathbf{x}\|_1, \ s.t. \ \mathbf{A}\mathbf{x} = \mathbf{y} \qquad (1)$$

Or alternatively, solve

$$\hat{\mathbf{x}}_1 = arg \min_{\mathbf{x}} \quad \|\mathbf{x}\|_1, \ s.t. \ \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2 \le \epsilon \qquad (2)$$

3) Compute the residual $r_i(\mathbf{y}) = \|\mathbf{y} - \mathbf{A}\delta_i(\hat{\mathbf{x}}_1)\|_2$, for $i = 1, ..., c$, where $\delta_i : R^n \to R^n$ is the characteristic function that selects the coefficients associated with the ith class;
4) Identify $\mathcal{I}(\mathbf{y}) = arg \min_{i} r_i(\mathbf{y})$, where $\mathcal{I}(\mathbf{y})$ stands for finding the class label of $\mathbf{y}$.

Note that each column of $\mathbf{A}$ is required to be unit l_2- norm (or bounded norm) in order to avoid trivial solutions that are due to the ambiguity of the linear reconstruction.

The second step which is used to calculate the sparse decomposition is the core of SRC algorithm. Theoretically, suppose that $\mathbf{A}$ can offer an over-complete basis, and then to find the sparse representation, we need to solve the following l_0-norm minimization problem:

$$\hat{\mathbf{x}}_1 = arg \min_{\mathbf{x}} \quad \|\mathbf{x}\|_0, \ s.t. \ \mathbf{Ax} = \mathbf{y} \qquad (3)$$

where $\|\mathbf{x}\|_0$ is l_0- norm which is equivalent to the number of non-zero components in the vector $\mathbf{x}$. Notice that the linear system in Eq. 3 is under-determined since $m \ll n$. Finding the exact solution to Eq. 3 is NP-hard due to its nature of combinational optimization. An approximate solution is obtained by replacing the l_0 norm in Eq. 3 with the l_1 norm as in Eq. 1. It can be proved that the solution of Eq. 3 is equivalent to the solution of Eq. 1 if a certain condition on the sparsity is satisfied, i.e. the solution is sparse enough[13]. To tolerate to a certain degree of noise, the problem in Eq. 1 can be further generalized as Eq. 2, which can be cast into an l_1- norm constrained least square problem:

$$\hat{\mathbf{x}}_1 = arg \min_{\mathbf{x}} \quad \|\mathbf{Ax} - \mathbf{y}\|_2^2 + \lambda \|\mathbf{x}\|_1, \qquad (4)$$

where $\lambda > 0$ is a scalar regularization parameter that balances the tradeoff between reconstruction error and sparsity.

In theory, the computational complexity to obtain the sparse representation is about $O(t^2n)$ per test sample, where t is the number of nonzero entries in reconstruction coefficients and n is the number of training samples. In practice, however, the obtained solution is far more from sparse vector due to many very small non-zero reconstruction coefficients. As a result, the computation complexity to find the 'sparse' solution tends to about $O(n^3)$ per test sample due to t tends to n.

III. THE PROPOSED LSRC ALGORITHM

To alleviate the deficiency in SRC, we propose to solve the sparse decomposition problem in a local manner to obtain an approximate solution. Instead of solving the $l_1 - norm$ constrain least square problem for all of training samples, LSRC solves the same problem in the local neighborhood of each test sample.

The proposed LSRC algorithm is summarized as follows:
1) Normalize the columns of $\mathbf{A}$ to have unit l_2- norm;
2) Find the k neighbors for test samples $\mathbf{y}$ from training samples using *k-nn* rules;
3) Solve the l_1- norm constrained least square problem in neighborhood of test sample $\mathbf{y}$

$$\hat{\mathbf{x}}_1 = arg \min_{\mathbf{x}} \quad \|\mathbf{A}_{\mathcal{N}(\mathbf{y})}\mathbf{x} - \mathbf{y}\|_2^2 + \lambda \|\mathbf{x}\|_1, \qquad (5)$$

where $\mathbf{A}_{\mathcal{N}(\mathbf{y})} \in R^{m \times k}$ stands for data matrix which consists of the k samples in neighborhood of test sample $\mathbf{y}$, $\hat{\mathbf{x}}_1 \in R^k$ is the best reconstruction coefficient to represent the test sample $\mathbf{y}$;

4) Compute the residual $r_i(\mathbf{y}) = \|\mathbf{y} - \mathbf{A}_{\mathcal{N}(\mathbf{y})}\delta_i(\hat{\mathbf{x}}_1)\|_2$, for $i = 1, ..., c$, where $\delta_i : R^k \rightarrow R^k$ is the characteristic function that selects the coefficients associated with the ith class;
5) Identify $\mathcal{I}(\mathbf{y}) = arg \min_i r_i(\mathbf{y})$, where $\mathcal{I}(\mathbf{y})$ stands for finding the class label of the test sample $\mathbf{y}$.

In contrast with SRC, the system of linear equations in Eq. 5 is over-determined. The computational complexity to obtain the sparse representation is about $O(t^2k)$ per test sample, where k is the number of the neighbors of the test sample, t is the number of nonzero entries in reconstruction coefficients and $t \leq k$. Compared with the original SRC, the computational cost will be saved remarkably when $k \ll n$.

In manifold learning the local linearity is used to capture the local geometric structure [14]. Especially the local nonnegative linear reconstructing coefficients are used to discover the natural class structure [15],[16]. Since the local nonnegative linear decomposition (LNLD) can also yield the sparse representation, we adopt it into LSRC framework to compare the effect of l_1- norm constraint. Therefore we call the variant version of LSRC as LNLD based Classification (LNLDC).

In LNLDC, the sparse representation is calculated by LNLD, which can be formulated as follows:

$$\begin{aligned} \hat{\mathbf{x}}_2 \quad &= arg \min_{\mathbf{x}} \|\mathbf{A}_{\mathcal{N}(\mathbf{y})}\mathbf{x} - \mathbf{y}\|_2, \\ s.t. \quad &\sum_i x_i = 1, \ x_i \geq 0, \end{aligned} \qquad (6)$$

where $\mathbf{A}_{\mathcal{N}(\mathbf{y})}$ stands for the data matrix which consists of samples in neighborhood of test sample $\mathbf{y}$ and x_i denotes the $i-$th component of $\mathbf{x}$;

The computational complexity for finding the sparse representation in LNLDC is about $O(k^2)$ per test sample.

In general sparse signal representation needs an over-complete basis (dictionary). In LSRC, however, the condition of over-complete basis may be unsatisfied. A natural question arose is that whether LSRC can inherit the merits (i.e. robustness to data corruption) of SRC? In other words whether we can balance the trade-off between the merits of SRC and its computational complexity by turning the local neighborhood scale parameter k? We prepared some experiments to answer this question.

IV. EXPERIMENTS

We conducted experiments for face recognition on ORL data set and Extended Yale B data set.

1) Solvers: In SRC and LSRC, we used the approach proposed in [5] to solve the l_1-norm constraint least square minimization problem[1] since it is a specialized interior-point method for solving large scale problem. For LNLD, we solve it as equality and inequalities constrained linear least square problem[2].

---

[1]We use the solver for l_1 constraint minimization *l1_ls.m* which is downloaded from http://www.stanford.edu/~boyd/software.html
[2]We use $lsqlin.m$ in Matlab

Figure 1. Experiments on ORL data set

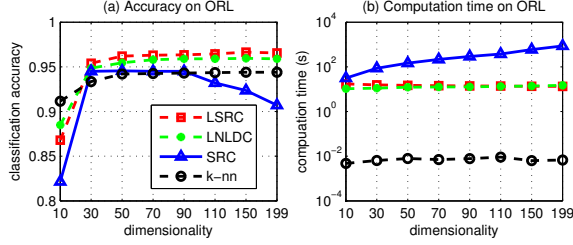Figure 2. Experiments on Extended Yale B data set

Figure 3. Experiments on Extended Yale B data set

2) Parameters Setting: At first we fixed the number of nearest neighbors used in LSRC and LNLDC algorithms as $k = 50$ for classification and anti-noise experiments on data set ORL and Extended YaleB. And then we also prepared classification experiments to investigate the effect of the parameter k of local neighborhood size in LSRC. For the regularization parameter λ and the relative target duality gap tolerance ϵ in the solver for SRC and LSRC, we take the default setting where $\lambda = 0.01$ and $\epsilon = 0.01$.

A. Classification Accuracy

In ORL data set, all the samples were resized into 28×23. We selected five samples per class randomly to consist of training data set and the other half were used for testing. We computed the recognition accuracy with different feature dimensions 10, 30, 50, 70, 90, 110, 150 and 199, that was computed by principal component analysis (PCA). Each experiment is repeated for ten trials and the averaged results are kept. The classification results are given in Fig. 1 panel (a). The LSRC and LNLDC both overmatched SRC and k-NN(i.e. 1-NN classifier). The computational times are presented in Fig. 1 panel (b). The proposed LSRC and LNLDC both are faster than SRC.

In Extended Yale B database, each image was resized into 32×32 and we selected half of the images for training (i.e., about 32 images per subject) randomly and the other half were used for testing. We computed the face recognition accuracy with the different feature dimensions 10, 60, 110, 160, 210, 260, 310, 410, 510 and 610, that was extracted by PCA. The classification results are given in Fig. 2 panel (a). Although the LSRC and LNLDC cannot overmatch SRC, the computation time was greatly reduced about two orders, see Fig. 2 panel (b).

B. Classification on Data with Noisy Corruption

We corrupted a percentage of randomly chosen pixels from each test sample, replacing their pixel values with independent and identically distributed noise from a uniform distribution. The corrupted pixels were randomly chosen for each test sample. The percentage of corrupted pixels was changed from 0% to 90%.

To check the robustness of LSRC to noisy corruption, we conducted experiments on the pixel noisy corrupted
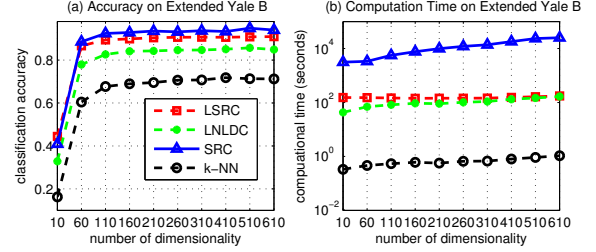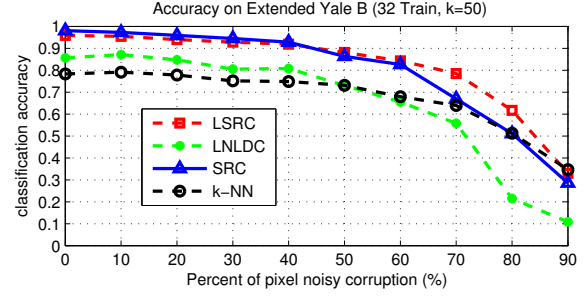
Extended Yale B data set. In the experiments, each image of 32×32 were concatenated to 1024 dimensional feature vector. There was no further extra feature extraction step. The experimental results for comparison of classification accuracy under different level of random pixel noisy corruption are given in Fig. 3. As can be seen from the experimental results, the proposed LSRC can obtain the comparative results with SRC when the percent of noisy corruption less than 60%, whereas LSRC can overmatch SRC when the percent of noisy corruption higher than 60%. However the LNLDC produced the inferior results compared with both SRC and LSRC. These results confirm that: (a) LSRC does inherit the merits of SRC; (b) Although LNLDC can also yield the sparse decomposition, it is different from LSRC. In other words, these results suggest that the merits which LSRC inherited from SRC are endowed by the $l_1 - norm$ constraints.

C. Change the local neighborhood size k

Finally we prepared experiments to evaluate the effect on the neighborhood size parameter k in LSRC. We fixed the dimension of feature vector which extracted by PCA at 510 for Extended Yale B data set and repeated the classification experiments with different neighborhood size k. For each k, we repeated 10 times and the averaged results are reported in Fig. 4. The results validated that LSRC can approximate the original SRC when the size of local neighborhood increases. This means that there is a trade-off between the classification accuracy of LSRC and its computational complexity, which can be turned by the local scale parameter k.

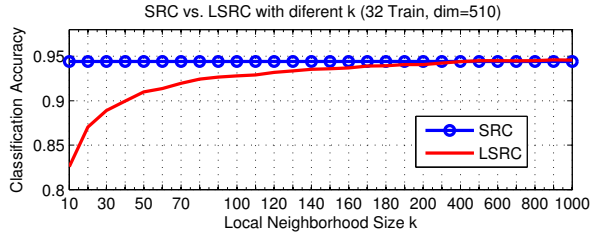In experiments we also observed that the size of local

Figure 4. Experiments on Extended Yale B with different k

neighborhood in LSRC to correctly classify a test sample is different. There are some 'easy' test samples which a small local neighborhood can also yield the correct results; meanwhile there are also some 'hard' test samples which need a larger local neighborhood to correctly classify. For most test samples a small or middle size of local neighborhood is enough to yield the correct results. Only a few test samples need a large local neighborhood even the entire training data. Therefore it would be interesting to design an adaptive scheme to adjust the local neighborhood size in LSRC. This issue will be further investigated in future.

V. CONCLUDING REMARKS AND DISCUSSIONS

In this paper we proposed LSRC to address the issue of computational complexity in SRC. In LSRC, the sparse representation is calculated in the local neighborhood of each test sample. Experiments on face recognition data sets ORL and Extended Yale B demonstrated that the proposed LSRC algorithm can reduce the computational complexity and remain the comparative classification accuracy and anti-noise merits. Therefore we can balance the classification accuracy of LSRC and its computational complexity by turning the local scale parameter k. In addition the comparison experiments between LSRC and LNLDC also indicated that l_1-norm constraint does offer some superiority on noise corruption.

When we face the task to classify data with corruption, i.e. noise, missing data and outliers, the reconstructive classification methods which based on reconstruction error, such as SRC and LSRC, will be the better choice. Compared with SRC, LSRC is more practical owning to its lightweight computational demand.

ACKNOWLEDGMENT

REFERENCES

[1] D. Donoho, "For most large underdetermined systems of linear equations the minimal l1-norm solution is also the sparsest solution," *Comm. Pure and Applied Math.*, vol. 59, no. 6, pp. 797–829, 2006.

[2] E. J. Candès, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Comm. Pure and Applied Math.*, vol. 59, no. 8, pp. 1207–1223, 2006.

[3] E. J. Candès and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?" *IEEE Trans. Information Theory*, vol. 52, no. 12, pp. 5406–5425, 2006.

[4] P. Zhao and B. Yu, "On model selection consistency of lasso," *J. Machine Learning Research*, no. 7, pp. 2541–2567,, 2006.

[5] S.-J. Kim, K. Koh, M. Lustig, S. Boyd, and D. Gorinevsky, "A method for large-scale l1-regularized least squares," *IEEE Journal on Selected Topics in Signal Processing*, vol. 1, no. 4, pp. 606–617, 2007.

[6] Y. Li, A. Cichocki, and S. Amari, "Analysis of sparse representation and blind source separation," *Neural Computation*, vol. 16, no. 6, pp. 1193–1234, 2004.

[7] M. Elad and M. Aharon, "Image denoising via learned dictionaries and sparse representation," in *CVPR*, 2006.

[8] M. Elad, B. Matalon, and M. Zibulevsky, "Image denoising with shrinkage and redundant representation," in *CVPR*, 2006.

[9] J. Starck, M. Elad, and D. Donoho, "Image decomposition via the combination of sparse representation and a variational approach," *IEEE Trans. on Image Processing*, vol. 14, no. 10, pp. 1570–1582, 2005.

[10] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. on PAMI*, vol. 31, no. 2, pp. 210–227, Feb 2009.

[11] H. Cheng, Z. Liu, and J. Yang, "Sparsity induced similarity measure for label propagation," in *ICCV*, 2009.

[12] X. Mei, H. Ling, and D. W. Jacobs, "Sparse representation of cast shadows via l_1-regularized least squares," in *ICCV*, 2009.

[13] D. L. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *IEEE Trans. on Information Theory*, vol. 47, no. 7, pp. 2845–2862, 2001.

[14] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, Dec. 2000.

[15] J. Lim, J. Ho, M. hsuan Yang, and K. chih Lee, "Image clustering with metric, local linear structure and affine symmetry," in *ECCV*, vol. 1, 2004, pp. 456–468.

[16] F. Wang and C. Zhang, "Label propagation through linear neighborhoods," in *ICML*. New York, NY, USA: ACM, 2006, pp. 985–992.