Discriminative locality-constrained sparse representation for robust face recognition

Meng Huang¹,², Guifang Shao¹,², Keqi Wang*,¹,², Tundong Liu¹,² and Hao Lu¹,²

¹Department of Automation, Xiamen University, Xiamen, Fujian, China
²Xiamen Key Laboratory of Big Data Intelligent Analysis and Decision-making, Fujian, China

Abstract. In this paper, a new joint sparse representation method called discriminative locality-constrained sparse representation (DLSR) is proposed for robust face recognition. DLSR incorporates locality and label information of training samples into the framework of sparse representation. Locality information can distinguish dissimilarity between samples and plays an important role in image classification. Compared with the existing methods, DLSR contains more discriminative information of samples and can obtain more discriminative recognition results. Due to the use of l2-norm regularization, DLSR can obtain a closed-form solution. This makes it computationally very efficient. Experimental results based on the benchmark face databases ORL have shown that DLSR can achieve more promising performance than some state-of-the-art methods.

1. Introduction
Sparse representation has received much attention in recent years and been widely applied in fields of signal processing, image processing, computer vision and pattern recognition [1,2]. In many image processing applications, such as face recognition [3-10], super-resolution [11], image denoising[12], image segmentation[13] and visual tracking [14,15], sparse representation has shown an attractive performance. Although the method based on deep learning gets a better performance[16-18] for these problems, it has poor interpretability. For face recognition, numerous different sparse representation methods have been proposed, such as weighted group sparse representation, etc.[3-10]. Based on a sparse representation computed by $l_1$-minimization, Wright et al.[3] proposed the sparse representation based classification (SRC) scheme for face recognition. In SRC, a testing sample was represented as a sparse linear combination of all the training samples, and then was classified by identifying which class yielded the minimal reconstruction residual. SRC had achieved great success and received remarkable attention in face recognition. It was widely believed that the $l_1$-norm sparsity constraint was crucial to the success of SRC. Zhang et al.[4] analyzed deeply the working mechanism of SRC and indicated that the collaborative mechanism played a more essential role than the $l_1$-norm sparsity constraint. They presented a collaborative representation based classifier (CRC) with $l_2$-norm regularization and obtained the competitive results. CRC had very competitive classification results, and it had significantly less computational cost than SRC due to the use of $l_2$ norm. Despite the success of both SRC and CRC, they just used all training samples as dictionary and didn’t consider the label information of training samples. Xu et al.[5] proposed a novel idea to design a discriminative

*E-mail: wjwkq@126.com
sparse representation method (DSR) with $l_2$-norm regularization by enhancing the discriminant of different classes. Simultaneously, Xu et al. [5] gave an insight into the rationale of the proposed method and showed that the decorrelation effect for different classes could distinguish the class really nearest to the test sample from the other classes, which enabled different classes to be more discriminative. They suggested that in addition to sparsity and collaboration playing important roles in the sparse representation methods, it was also helpful to reduce the correlation between the representation results of training samples from different classes. DSR achieved a noticeable performance for face recognition. It had not only remarkable classification accuracy but also very computationally efficient.

Although the aforementioned methods achieved excellent performance for image classification, they generally did not consider locality information of samples. The aforementioned methods all used a dictionary composed of all the training samples. Thus, the test sample might be represented by training samples far from it. To address this shortcoming, locality information of samples was introduced in some different methods [6-10]. The locality structure of individual sample is important in revealing the true geometry of feature space [19,20]. Locality information of samples reflects the dissimilarity between the samples. Locality is more essential than sparsity, as locality leads to sparsity but not necessary vice versa [21]. Wang et al. [6] proposed a locality-constrained linear coding (LLC) algorithm and showed that the image classification performance can be improved by enforcing the sparsity and locality.

Inspired by DSR, we present a new joint sparse representation method with $l_2$-norm regularization, which is called discriminative locality-constrained sparse representation (DLSR). By integrating locality and label information of training samples into the framework of sparse representation, DLSR contains more discriminative information of samples and improves recognition performance for face recognition. Due to the use of $l_2$ norm, the objective function of DLSR is convex and differentiable. So it is mathematically tractable and computationally efficient. A closed-form solution which makes the calculation very convenient for face recognition can be obtained by the detailed mathematical deduction.

2. The proposed method
Suppose that there are $c$ classes in the total $n$ images, each class has $n_k$ images for $k = 1,..., c$. Let $A = [A_1 \cdots A_c] \in R^{m \times c}$ be the set of training samples, where $A_i = [a_{i1} \cdots a_{in_k}] \in R^{m \times n_k}$ is the subset of the training samples from $k$th class, $a_{ij}$ is the $i$th training sample from the $k$th class. $x = [x_1^T \cdots x_n^T] \in R^n$ is the representation vector, where $x_k = (x_{i1} \cdots x_{in_k})^T \in R^{n_k}$ denotes the representation vector of the $k$th class. That is $x = (x_{11} \cdots x_{in_k} \cdots x_{1n} \cdots x_{in})^T$. The symbol “$\odot$” means element-wise multiplication. The task of face recognition is to correctly determine the class which a test sample $y \in R^n$ belongs to by using labeled training samples $A$.

2.1. Objective function of the DLSR method
In order to capture the locality information, we present the new joint sparse representation scheme DLSR by introducing the $l_2$ regularization-based locality constraint into DSR. The objective function of DLSR is defined as follows.

$$
\text{min} \left\| y - Ax \right\|^2 + \lambda_1 \sum_{k=1}^{c} \sum_{j=1}^{n_k} \left\| A_k x_k + y - \sum_{i \neq k} A_i x_i \right\|^2 + \lambda_2 \left\| d \odot x \right\|_2
$$

(1)

where $\lambda_1$ and $\lambda_2$ are the regularization parameters which weight the second term (the correlation of the different classes) and the third term (the locality constraint), respectively. $A_k x_k$ denotes the representation of test sample by using the training samples from the $i$th class. The vector $d \in R^n$ is the locality adaptor that measures the similarity between the test sample $y$ and all the training samples. The element $d_k$ of vector $d$ is defined as follows.
where $\sigma$ is the bandwidth parameter. The larger $d_i$ is, the smaller the weight coefficient $x_i$ will be. A large $d_i$ would make the corresponding $x_i$ shrink to zero. This actually leads to sparsity.

From the Eq.(1), we can find that locality, sparsity, collaborative representation and the effect of decorrelation were all integrated into a unified framework for face recognition. As a result, the proposed scheme DLSR can include more discriminative information in order to improve recognition result.

2.2. Optimization of the objective function
Since the objective function in Eq.(1) is convex and differentiable, the optimal solution of Eq.(1) can be derived directly by taking the derivative of the objective function. Detailed calculation is as follows.

Let the objective function in Eq.(1) be $L(x)$. First, we will take the derivative of the first term $p(x)$ in $L(x)$.

\[
\frac{dp}{dx} = \frac{d}{dx} \| y - Ax \|^2 = -2A^T (y - Ax)
\]

Let $f(x) = \lambda_i \sum_{i=1}^{c} \sum_{j=1}^{k} \left\| A_i x_i + A_j x_j \right\|^2$, then the derivative of the second term in Eq.(1) $df / dx$ can be got by calculating the partial derivatives $\partial f / \partial x_k$ ($k = 1, ..., c$).

\[
f(x) = \lambda_i \left( \sum_{i=1}^{c} \sum_{j=1}^{k} \left\| A_i x_i + A_j x_j \right\|^2 + \sum_{i \neq j} \sum_{k \neq l} \left\| A_i x_i + A_j x_j \right\|^2 \right)
\]

\[
= \lambda_i \left( \sum_{i=1}^{c} \sum_{j=1}^{k} \left\| A_i x_i + A_j x_j \right\|^2 + \sum_{i \neq j} \sum_{k \neq l} \left\| A_i x_i + A_j x_j \right\|^2 \right)
\]

\[
\frac{\partial f}{\partial x_k} = 4\lambda_i \left( \sum_{i=1}^{c} \frac{\partial}{\partial x_k} \left\| A_i x_i + A_j x_j \right\|^2 \right) + 2\lambda_i \sum_{i \neq j} \frac{\partial}{\partial x_k} \left\| A_i x_i + A_j x_j \right\|^2
\]

\[
= 4\lambda_i (cA_i x_i + Ax)
\]

where $B = \begin{bmatrix} A_1^T & A_2^T & \cdots & 0 & \vdots & \vdots & 0 & \cdots & A_c^T \end{bmatrix}$. The derivative of the third term $g(x)$ in $L(x)$ can be obtained as follows.

\[
\frac{dg}{dx} = \frac{d}{dx} (\sum_{i=1}^{c} \left\| A_i x_i \right\|^2) = 2\lambda_i A^T x
\]

where $D = \text{diag}(d_1^2, \ldots, d_i^2, \ldots, d_n^2)$. From Eq.(3) to Eq.(5), we can obtain the derivative of the objective function $L(x)$.

\[
\frac{dL(x)}{dx} = -2A^T (y - Ax) + 4\lambda_i (cBx + A^T x) + 2\lambda_i D
\]

Finally, let Eq.(6) be 0, then the optimal $x$ is obtained as follows.

\[
x = ((1 + 2\lambda_i)A^T A + 2\lambda_i cB + \lambda_i D)^{-1} A^T y
\]

The classification criterion is as follows.

\[
\text{Identity}(y) = \arg \min_x \| y - A_i x_i \|_2
\]
The proposed method is summarized in Algorithm 1.

Algorithm 1 for DLSR

Input: training sample $A$, test sample $y$, parameters $\lambda_1, \lambda_2, \sigma$.

Step1: Compute the matrix $B,D$ in Eq. (4) and (5).

Step2: Solve the representation vector $x$ by using Eq. (7).

Step3: Calculate the residuals $r_i(y) = \|y - A_i x\|_2$, $k = 1,\ldots,c$.

Output: $\text{Identity}(y) = \arg \min_i r_i(y)$.

3. Experimental results

3.1. Database

As shown in figure 1(a), the ORL database [24] contains 400 face images taken from 40 subjects. There are 10 images for each subject with varying lighting, facial details (glasses/no glasses) and facial expressions (open/closed eyes, smiling/not smiling). Each image was resized to an image with one half of the original size by using the down-sampling algorithm.

3.2. Experimental results

In this section, we conduct experiments on ORL face databases and verify the effectiveness of the proposed method by contrast with DSR [5], CRC [4] and SRC [3]. Considering the accuracy and efficiency, we choose L1LS [22] and DALM [23] to solve the $l_1$-minimization in SRC. In order to make fair comparison, the optimal parameters for different sparse representation methods on the database are selected by manual adjustment and the best results are reported.

For ORL face database, the first 2-6 face images of each subject were used as training samples and the remaining ones were used as test samples. The parameters $\lambda_1$ and $\lambda_2$ in DLSR were both set to 0.001. The setting of parameter $\sigma$ is related to the training sample size. When $\sigma$ was set to 0.2, 0.4, 0.6 respectively, the best results for different training sample size could be obtained in this experiment. Table 1 shows the best recognition rates of five methods. Figure 1(b) shows the best recognition rates versus the variation of the training sample size on ORL face database.

| Number of training samples per subject | 2     | 3     | 4     | 5     | 6     |
|----------------------------------------|-------|-------|-------|-------|-------|
| DLSR                                   | 88.12 | 89.64 | 94.58 | 95.00 | 95.63 |
| DSR                                    | 86.88 | 89.64 | 94.17 | 94.50 | 95.00 |
| CRC                                    | 83.44 | 86.07 | 89.17 | 88.50 | 91.87 |
| L1-Ls                                  | 85.25 | 88.57 | 92.08 | 92.50 | 93.75 |
| DALM                                   | 86.88 | 89.64 | 92.08 | 92.00 | 94.37 |

Figure 1. Some face images from ORL database and the recognition results for ORL face images.
3.2.1. Analysis of parameters

There are three parameters $\lambda_1, \lambda_2$ and $\sigma$ in DLSR. The parameters $\lambda_1$ and $\lambda_2$ balance the contribution of the regularization terms, which represent the weight of different regularization terms in the optimization goal and are generally used in almost all the aforementioned classification approaches. In the experiment, the parameters are elected by cross validation. We keep the values of two parameters fixed and consider the influence of the other parameter. Then the best parameter values for each experiment are decided. We select the value of $\lambda_1$ and $\lambda_2$ in the range of $[10^{-4},10^3]$. The parameter $\sigma$ is the bandwidth parameter, which is influenced by the size of training sample.

3.2.2. Discussion

From Table 1, we can see that DLSR can obtain better recognition results than the state-of-the-art methods on the face databases. Compared with DSR and CRC, which are both $l_2$ regularization-based methods and have the closed-form solution, DLSR can obtain better recognition results. Compared with SRC, DLSR not only improves the recognition results, but also greatly improves the computational efficiency. The experimental results validate the effectiveness of DLSR.

4. Conclusion

This paper has proposed a new joint sparse representation method (DLSR) for robust face recognition. DLSR incorporates locality information of samples into the framework of sparse representation, which makes it contain more discriminative information of samples and achieve a promising performance for face recognition. Through detailed mathematical derivation, a closed-form solution is obtained, which makes DLSR computationally efficient. Experiment based on the benchmarking face database has been conducted and verified that DLSR can achieve better performance than some state-of-the-art methods.

Acknowledgments

This research was supported by the Industry and University Cooperation Project of Fujian Province (No. 2018H6018) and the Science and Technology Project of Xiamen (No.3502Z20183004).

References

[1] Zhang, Z., Xu, Y., Yang, J., et al.: 'A survey of sparse representation: algorithms and applications', IEEE Access, 2017, 3, pp490-530.
[2] Wright, J., Ma, Y., Mairal, J., et al.: 'Sparse representation for computer vision and pattern recognition', Proceedings of IEEE, 2010, 98, (6), pp 1031-44.
[3] Wright, J., Yang, A.Y., Ganesh, A., et al.: 'Robust face recognition via sparse representation', IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, 31, (2), pp 210-27.
[4] Zhang, L., Yang, M., Feng, X., et al.: 'Sparse representation or collaborative representation: Which helps face recognition?', IEEE International Conference on Computer Vision, 2012, 2011, (5), pp 471-78.
[5] Xu, Y., Zhong, Z., Yang, J., et al.: 'A new discriminative sparse representation method for robust face recognition via $l_2$ regularization', IEEE Trans Neural Netw Learn Syst, 2017, 99, pp 1-10.
[6] Wang, J., Yang, J., Yu, K., et al.: 'Locality-constrained linear coding for image classification', Computer Vision and Pattern Recognition, 2010, 119, (5), pp 3360-67.
[7] Li, Z., Lai, Z., Xu, Y., et al.: 'A locality-constrained and label embedding dictionary learning algorithm for image classification', IEEE Transactions on Neural Networks and Learning Systems, 2017, 28, (2), pp 278-92.
[8] Tang, X., Feng, G., Cai, J.: 'Weighted group sparse representation for undersampled face recognition', Neurocomputing, 2014, 145, (18), pp 402-15.
[9] Zheng, J., Yang, P., Chen, S., et al.: 'Iterative re-constrained group sparse face recognition with adaptive weights learning', IEEE Transactions on Image Processing, 2017, 26, (5), pp 2408-
23.
[10] Tan, S., Sun, X., Chan, W., et al.: ‘Robust face recognition with kernelized locality-sensitive group sparsity representation’, IEEE Transactions on Image Processing, 2017, 26, (10), pp 4661-68.
[11] Yang, J., Wright, J., Huang, T.S., et al.: ‘Image super-resolution via sparse representation’, IEEE Transactions on Image Processing, 2010, 19, (11), pp 2861-73.
[12] Dong, W., Li, X., Zhang, L., et al.: ‘Sparsity-based image denoising via dictionary learning and structural clustering’, IEEE Conference on Computer Vision and Pattern Recognition, June 2011, pp 457-64.
[13] Lu, X., Li, X.: ‘Group sparse reconstruction for image segmentation’, Neurocomputing, 2014, 136, (1), pp 41-48.
[14] Zhang, S., Yao, H., Sun, X., et al.: ‘Sparse coding based visual tracking: Review and experimental comparison’, Pattern Recognition, 2013, 46, (7), pp 1772-88.
[15] Wang, D., Lu, H., Yang, M.H.: ‘Online object tracking with sparse prototypes’, IEEE Transactions on Image Processing, 2013, 22, (1), pp 314-25.
[16] H. Wang, Y.T. Wang, Z. Zhou, et al.: ‘CosFace: Large Margin Cosine Loss for Deep Face Recognition’, The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2018.
[17] J. Zhao, Y. Cheng, Y. Xu, et al.: ‘Towards Pose Invariant Face Recognition in the Wild’, The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2018.
[18] J.K. Deng, J. Guo, N.N. Xue, et al.: ‘ArcFace: Additive Angular Margin Loss for Deep Face Recognition’, The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2019.
[19] Cui, M.S., Prasad, S.: ‘Class-dependent sparse representation classifier for robust hyperspectral image classification’, IEEE Trans. Geoscience and Remote Sensing, 2015, 53, (5), pp 2683-95.
[20] Jiang, J., Hu, R., Wang, Z., et al.: ‘Noise robust face hallucination via locality- constrained representation’, IEEE Trans. Multimedia, 2014, 16, (5), pp 1268-81.
[21] Yu, K., Zhang, T., Gong, Y.: ‘Nonlinear learning using local coordinate coding’, International Conference on Neural Information Processing Systems, 2009, pp 2223-31.
[22] Kim, S., Koh, K., Lustig, M., et al.: ‘An interior-point method for large-scale 1 1 -regularized least squares’, IEEE Journal of Selected Topics in Signal Processing, 2008, 1, (4), pp 606-17.
[23] Yang, A.Y., Zhou, Z., Balasubramanian, A., et al.: ‘Fast 11-minimization algorithms for robust face recognition’, IEEE Transactions on Image Processing, 2013, 22, (8), pp 3234-46.
[24] Samaria, F. S., Harter, A. C.: ‘Parameterisation of a stochastic model for human face identification’, Proceedings of the Second IEEE Workshop on Applications of Computer Vision, 1994, pp 138-42.