ICIC Express Letters
Part B: Applications
Volume **7**, Number **11**, November **2016**
ICIC International ©2016 ISSN 2185-2766
pp. **2283–2290**

A SAMPLE WEIGHTED SPARSE REPRESENTATION APPROACH TO IMAGE CLASSIFICATION

Yongxin Wang[1,2] and Huaxiang Zhang[1,2,*]

[1]Department of Computer Science
Shandong Normal University
[2]Shandong Provincial Key Laboratory for Novel Distributed Computer Software Technology
No. 88, Wenhua East Road, Lixia District, Jinan 250014, P. R. China
[*]Corresponding author: wangyongxin1992@126.com

Received May 2016; accepted August 2016

Abstract. *Sparse representation for classification (SRC) has achieved a big success for face recognition. It utilizes a sparsely linear combination of the training samples to construct a test sample, and classifies the test sample based on the reconstruction error associated with each class. Since SRC weights each training sample equally important, which may not hold for real applications, we propose a novel sample weighted sparse representation classification approach (SWSRC) by weighting each training sample differently. We first employ the representation ability of each training sample to construct a weight matrix and then solve a weighted l_1 minimization problem to obtain the sparse reconstruction coefficients. Experimental results on AR, YaleB and USPS image datasets demonstrate its effectiveness.*

Keywords: Sparse representation, Sample weighted SRC (SWSRC), Image classification, Face recognition

1. **Introduction.** Due to its good recognition capabilities, sparse representation for image recognition has become very popular in machine learning and pattern recognition domains, and many works have been done in this branch [1, 2, 3, 4, 5, 6, 7]. All the approaches first learn a construction dictionary and the corresponding sparsely linear reconstruction parameters, and utilize reconstruction errors for classification. They got a big success in some pattern recognition problems, especially in face recognition [8, 9, 10]. Based on the lasso optimization problems, many parameters will be zero, and thus leads to the sparse construction of a test sample. SRC [1] is robust to occlusion, noise and illumination. Motivated by its good performance, a lot of algorithms of sparse representation for face recognition have been proposed.

Considering that different training samples may contribute differently to a test sample, Lu et al. [11] proposed a weighted sparse representation classification (WSRC) method by adding a weighted sparsity regulation item in the optimization problem. Fan et al. [12] proposed a weighted sparse representation algorithm by straightforwardly assigning weights on the training samples. Gao et al. [13] proposed a kernel sparse representation for image classification and face recognition, and Yang et al. [14] proposed a robust sparse coding (RSC) model which seeks for the maximum likelihood estimation solution and is more robust to outliers. In order to improve face recognition performance, Yang et al. [15] proposed a method which combines SRC with Metaface learning. Xu et al. [10] proposed a method which divides the sparse representation into two parts (TPTSR). The approach first finds some nearest neighbors for a test sample and then uses the nearest neighbors to represent the test sample.

In pattern recognition and machine learning fields, samples usually have different importance in representation, recognition and classification. Many methods have been proposed to consider sample or feature weights. In order to reduce the image dimension, Zhu et al.

[16] proposed a feature selection algorithm by learning a weight matrix, which is computed by self-representation. A weighted version of principal component analysis was proposed in [17], which sets different weights to different images.

It can be accepted that, different training samples contribute differently to the representation of a test sample. So by considering each training sample differently, we propose a new sample weighted sparse representation classification scheme, namely SWSRC. SWSRC introduces a sample weight matrix into SRC and aims to improve the performance of SRC. When all training samples are used to construct a test sample, a coefficient vector can be obtained by solving a predefined optimization problem. Each element of the coefficient vector reflects the importance of each training sample. We use the coefficient vector to construct a weight matrix and introduce it into the regularization item of SRC. If a training sample has a big weight, the corresponding representation coefficient of SRC should be nonzero. Otherwise, if a training sample has a small weight value, the corresponding coefficient should be zero. Compared with SRC, SWSRC can enhance the classification effectiveness of SRC in principle, since SWSRC exploits weight information in representing the test samples whereas SRC does not exploit it. Compared with other existing WSRC methods, SWSRC utilizes construction coefficient to weight samples instead of distance. Obviously, the construction coefficients have more direct ability than distance when using the weight to represent samples. Finally, the coefficients computed by SWSRC are more effective and take more local structures of data into account. Extensive experiments are conducted on human face images and digit images. The experimental results show that the proposed algorithm is superior to several relative algorithms.

The rest of the paper is organized as follows. In Section 2, we review the related work of SRC and WSRC. In Section 3, we describe the proposed approach. Experimental results and comparisons on three real-world datasets are demonstrated in Section 4. Finally, the conclusion is given in Section 5.

2. Related Work.

2.1. The SRC algorithm.

Let $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n] \in \mathbb{R}^{m\times n}$ be a training matrix, where m is the number of features and n is the number of training samples. Let $\mathbf{D}_i \in \mathbb{R}^{m\times n_i}$ be the dataset of the ith class, where n_i is the number of samples belonging to class i and each column is a sample. So we get $\mathbf{D} = [\mathbf{D}_1, \mathbf{D}_2, \dots, \mathbf{D}_K] \in \mathbb{R}^{m\times n}$ to be the dictionary, where K is the number of total classes and $n = n_1 + n_2 + \cdots + n_K$.

Then for a test image $\mathbf{y} = [y_1; y_2; \dots; y_m]$, we use $\mathbf{D}$ to represent it. It is $\mathbf{y} = \mathbf{D}\alpha$, where $\alpha = [c_1(\hat{\alpha}), c_2(\hat{\alpha}), \dots, c_K(\hat{\alpha})]$ and $c_i(\hat{\alpha})$ is the coefficients in α associated with class i. If $\mathbf{y}$ is from class i, $\mathbf{D}_i c_i(\hat{\alpha})$ will construct $\mathbf{y}$ well, so we hope the entries of α are zero except $c_i(\hat{\alpha})$ that is associated with class i.

The SRC algorithm [1] can be summarized in Table 1.

Table 1. Description of SRC

Input: training samples $\mathbf{X}$, a test sample $\mathbf{y}$ and λ Output: the class label of $\mathbf{y}$
1. Make a big dictionary $\mathbf{D}$ by the entire training samples $\mathbf{X}$. 2. For a test image $\mathbf{y}$, solve the following l_1 minimization problem: $\hat{\alpha} = \arg\min_{\alpha} (\|\|\mathbf{y} - \mathbf{D}\alpha\|\|_2^2 + \lambda\|\|\alpha\|\|_1)$ 3. Compute the error for $i = 1, \dots, K$: $e_i(\mathbf{y}) = \|\|\mathbf{y} - \mathbf{D}c_i(\hat{\alpha})\|\|_2$ 4. Output the class label of $\mathbf{y}$ as $class(\mathbf{y}) = \arg\min_{i} e_i(\mathbf{y})$

2.2. **Weighted sparse representation.** Obviously, the classical SRC considers all training samples as equally important. So a test sample may be constructed by some samples which are far from the test sample. And this may lead to non-robust results. That means the classical SRC ignores the local structures of data [18]. Sometimes locality is more important than sparsity, so a weighted SRC (WSRC) [11] was presented. WSRC is given by the following model:

$$\hat{\alpha} = \arg\min_{\alpha} \left(||\mathbf{y} - \mathbf{D}\alpha||_2^2 + \lambda ||\mathbf{w}\alpha||_1 \right) \tag{1}$$

where $\mathbf{w} = diag[w_1, w_2, \ldots, w_n]$ is a diagonal matrix with $w_i = ||\mathbf{y} - \mathbf{x}_i||_2$.

$w_i = ||\mathbf{y} - \mathbf{x}_i||_2$ is the distance between $\mathbf{y}$ and $\mathbf{x}_i$, and reflects the similarity between the test sample $\mathbf{y}$ and the training sample $\mathbf{x}_i$. The larger w_i is, the less $\mathbf{x}_i$ contributes to $\mathbf{y}$. So WSRC imposes more discriminative information into SRC to improve the performance of classification.

3. **Sample Weighted Sparse Representation.** Motivated by the weakness of SRC and the good improvement of WSRC, we propose a new method of weighted SRC by introducing a new weighted strategy. Our approach is divided into two steps: we first compute each sample weight, and then solve a weighted l_1 minimization problem.

3.1. **Sample weighted representation.** In this part, we introduce how to compute the weight matrix of all training samples in a different way. Of course, the matrix to be learned should reflect the importance of each sample.

We think that each test sample can be well constructed by all the training samples and the contribution of each sample is different. So we want to get the contributions of all training samples when representing a test sample and use the contributions to construct a weight matrix. We use all the training samples as a dictionary and each test sample is represented by it.

For a test sample $\mathbf{y}$, we represent it as a linear combination of all training samples [10] (TPTSR):

$$\mathbf{y} = \mathbf{Dm} + \mathbf{r} \tag{2}$$

where $\mathbf{m} = [m_1, m_2, \ldots, m_n]$ is the representation coefficient matrix and $\mathbf{r}$ is the residual matrix.

When using all training samples to represent $\mathbf{y}$, if $\mathbf{D}$ is a nonsingular square matrix, we define

$$\mathbf{m} = \mathbf{D}^{-1}\mathbf{y} \tag{3}$$

Otherwise, we define

$$\mathbf{m} = \left(\mathbf{D}^T\mathbf{D} + \mu\mathbf{I} \right)^{-1} \mathbf{D}^T\mathbf{y} \tag{4}$$

where μ is a small positive number. Each training sample has its own contribution to representing a test sample, and the contribution of the ith training sample to the test sample is the reconstruction coefficient m_i. Obviously, the important samples should well represent the test sample; in other words, the residual matrix $\mathbf{r} = \mathbf{y} - \mathbf{Dm}$ should be small. We define $r_i = ||\mathbf{y} - m_i\mathbf{x}_i||_2$ to reflect the contribution of the ith training sample $\mathbf{x}_i$ to the test sample. The smaller the r_i is, the larger the contribution is. We construct a weight matrix as follows:

$$\frac{1}{\mathbf{W}} = diag[r_1, r_2, \ldots, r_n] \in \mathbb{R}^{1\times n}, \quad r_i = \|\mathbf{y} - m_i\mathbf{x}_i\|_2 \tag{5}$$

3.2. **Sample weighted sparse representation for classification.** In order to fully utilize the sample locality property, we learn the sparse parameters as

$$\hat{\alpha} = \arg\min_{\alpha} \left(\|\mathbf{y} - \mathbf{D}\alpha\|_2^2 + \lambda \left\| \frac{\mathbf{1}}{\mathbf{W}}\alpha \right\|_1 \right) \quad (6)$$

where $\mathbf{W}$ is a diagonal matrix with $\mathbf{W} = diag[W_1, W_2, \ldots, W_n]$ and can be learned from Formula (5).

W_i reflects the importance of the ith training sample. We use the form of reciprocal merely to make the model look intuitive. The smaller W_i is, the larger $\frac{1}{W_i}$ will be, and the corresponding coefficient α_i of the training sample will approach zero. In this way, when learning the sparse coefficient α, the relationship among data is taken into account, and thus, SWSRC can learn more accurate sparse coefficients than SRC.

From another point of view, our method is the extension of WSRC algorithm. In fact, if $\mathbf{m}$ is set to $\mathbf{1}_{n\times 1}$, then our method is reduced to WSRC. From the definition of $r_i = \|\mathbf{y} - m_i\mathbf{x}_i\|_2$ in SWSRC and $w_i = \|\mathbf{y} - \mathbf{x}_i\|_2$ in WSRC, we can say that, SWSRC can be regarded as a weighted WSRC which puts the first step reconstruction coefficient m_i as a weight, and thus, more local information is introduced into SWSRC than WSRC.

We summarize the overall optimization of the above model in Table 2, and we adopt the implementation in SPAMS package[1].

Table 2. Description of SWSRC

Input: training samples $\mathbf{X}$, a test sample $\mathbf{y}$, μ and λ Output: the class label of $\mathbf{y}$
1. Make a big dictionary $\mathbf{D}$ by the entire training samples $\mathbf{X}$. 2. Learn a weight matrix $\mathbf{W}$ of all training samples as Formula (5). 3. For a test image $\mathbf{y}$, solve the follow weighted l_1 minimization problem: $\hat{\alpha} = \arg\min_{\alpha} \left(\|\mathbf{y} - \mathbf{D}\alpha\|_2^2 + \lambda \|\frac{\mathbf{1}}{\mathbf{W}}\alpha\|_1 \right)$ 4. Compute the error for $i = 1, \ldots, K$: $e_i(\mathbf{y}) = \|\mathbf{y} - \mathbf{D}c_i(\hat{\alpha})\|_2$ 5. Output the class label of $\mathbf{y}$ as $class(\mathbf{y}) = \arg\min_i e_i(\mathbf{y})$

4. **Experiments.** In this section, we have conducted three experiments on the popular datasets to demonstrate the effectiveness of SWSRC. We use principal component analysis (PCA) [19] to implement image dimensionality reduction and set 0.1 and 0.01 to μ and λ respectively in SWSRC. For the parameters used in SRC, WSRC and TPTSR, we follow the author's original settings in [1, 10, 11] respectively.

4.1. **The AR face image database.** The AR database [20] consists of over 4000 frontal face images of 126 individuals with different facial expressions, occlusions and lighting conditions. Figure 1 shows some samples of this dataset. We choose a subset of the dataset consisting of 50 female and 50 male subjects. For each subject, we choose 7 images that only illumination and expressions change from Session 1 for training and Session 2 for testing separately.

The recognition rates are reported in Table 3, and the values in bold face are the highest recognition rates. It shows that, SWSRC outperforms the other methods under all five different dimensions, and its recognition rate increases as the dimension increases. The time costs of three methods related to SRC are shown in Table 4, and the values in bold face represent the lowest time costs. The results also demonstrate that SWSRC costs far

[1] http://spams-devel.gforge.inria.fr.

Figure 1. Sample face images from AR database

Table 3. Recognition rate on AR database under different feature dimensionality

	50	100	200	400	2580
KNN	0.6529	0.6743	0.6829	0.6886	0.6871
SRC	0.7457	0.7886	0.8329	0.8357	0.8286
WSRC	0.7743	0.8143	0.8186	0.8171	0.7586
TPTSR	0.7286	0.7914	0.8200	0.8343	0.8386
SWSRC	**0.7829**	**0.8257**	**0.8529**	**0.8557**	**0.8629**

Table 4. Time cost (second) on AR database under different feature dimensionality

	50	100	200	400	2580
SRC	619.7	633.1	738.0	1404.0	1980.3
WSRC	8.8	15.4	30.2	103.0	353.5
SWSRC	**6.3**	**8.8**	**11.8**	**18.0**	**41.0**

Figure 2. Sample face images from YaleB database

less time than SRC. SRC uses the truncated Newton interior-point method [21] to solve the l_1-regularized least squares problem. In SWSRC, we use the LARS algorithm [22] to solve the optimization problem. The difference between the two algorithms may be the reason why the time costs are different. SWSRC costs less time than WSRC. Because SWSRC and WSRC use the same optimization algorithm, their time cost difference may be caused by the weight matrix learning process.

4.2. **The YaleB face image database.** The YaleB database [23] consists of 2414 frontal face images of 38 individuals under various lighting conditions. Figure 2 gives some samples of this dataset. For each individual, we randomly select half of them for training and the left for testing.

The recognition rates are reported in Table 5, and the values in bold face are the best ones. From Table 5, we can see that our method performs the best under the last three high dimensions and performs almost equally well compared with SRC under the other dimensions. The time costs of the three methods related to SRC are shown in Table 6, and the values in bold face represent the lowest time costs. SWSRC outperforms SRC and WSRC based on the time cost.

4.3. **The USPS handwritten digit database.** In order to show the effectiveness of SWSRC, we also conduct experiments on USPS database [24], which consists of two parts. Part 1 consists of 7291 images, and part 2 consists of 2007 images. Some sample images are shown in Figure 3. We randomly select 60% images from part 2 for training and the rest 40% for testing.

TABLE 5. Recognition rate on YaleB database under different feature dimensionality

	50	100	200	400	1024
KNN	0.5643	0.6373	0.6780	0.6971	0.7004
SRC	0.8963	0.9245	0.9427	0.9451	0.9485
WSRC	0.8996	0.9278	0.9452	0.9436	0.9394
TPTSR	**0.9029**	**0.9303**	0.9378	0.9411	0.9394
SWSRC	0.8970	0.9247	**0.9459**	**0.9462**	**0.9494**

TABLE 6. Time cost (second) on YaleB database under different feature dimensionality

	50	100	200	400	1024
SRC	1315.5	1530.2	2834.6	4411.5	8780.9
WSRC	17.7	35.3	76.2	348.8	1888.8
SWSRC	**17.4**	**31.0**	**67.7**	**247.7**	**1836.3**

FIGURE 3. Sample digit images from USPS database

TABLE 7. Recognition rate on USPS database under different feature dimensionality

	30	50	100	200	1024
KNN	0.9328	0.9402	0.9377	0.9377	0.9377
SRC	0.9427	0.9390	0.9440	0.9402	0.9402
WSRC	0.9427	0.9422	0.9402	0.9253	0.9290
TPTSR	0.9141	0.9303	0.9290	0.9365	0.9303
SWSRC	**0.9477**	**0.9440**	**0.9465**	**0.9477**	**0.9477**

TABLE 8. Time cost (second) on USPS database under different feature dimensionality

	30	50	100	200	1024
SRC	1228.0	1406.2	1687.0	3303.7	3908.0
WSRC	7.3	12.9	30.8	62.2	66.0
SWSRC	**4.2**	**5.0**	**7.2**	**8.8**	**10.1**

The recognition rates are reported in Table 7, and the values in bold face are the highest recognition rates. The time costs of the three methods related to SRC are shown in Table 8, and the values in bold face represent the lowest time costs. The results demonstrate that, our method performs better than the other methods based on the recognition rate and the time cost. We also can see that the recognition rate is not sensitive to data dimension.

The results on AR, YaleB and USPS show that, SWSRC performs well on face recognition and other pattern recognition problems such as handwritten recognition. SWSRC considers some local information when constructing samples, so it outperforms SRC on data that local structure is essential for discrimination. It is based on the reconstruction errors to evaluate the importance of each individual training sample, and obtains the weights more effectively than WSRC which uses distance to evaluate the importance of training samples.

5. **Conclusion.** In this paper, we propose a novel sample weighted approach SWSRC for learning the sparse representations of test samples based on the dictionary learning approaches. SWSRC learns the weights of training samples for a test sample based on a linear reconstruction model. We introduce the weights into a weighted lasso optimization problem to learn the sparse reconstruction coefficients. It is a modification of WSRC, and experiments on two face image datasets and a handwritten digit image dataset demonstrate its effectiveness. However, the proposed approach still has room for improvement. For instance, other proposed weighting methods with a reduced time complexity could be introduced to improve the computational efficiency. Doing so, in a principled manner, it remains an important direction for future work.

Acknowledgement. The work is partially supported by the National Natural Science Foundation of China (Nos. 61572298, 61373081, 61402268, and 61401260) and the Taishan Scholar Project of Shandong, China.

REFERENCES

[1] J. Wright, A. Y. Yang, A. Ganesh et al., Robust face recognition via sparse representation, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.31, no.2, pp.210-227, 2009.

[2] E. Elhamifar and R. Vidal, Robust classification using structured sparse representation, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.1873-1879, 2011.

[3] J. Wright, Y. Ma, J. Mairal et al., Sparse representation for computer vision and pattern recognition, *Proc. of the IEEE*, vol.98, no.6, pp.1031-1044, 2010.

[4] K. Huang and S. Aviyente, Sparse representation for signal classification, *Advances in Neural Information Processing Systems*, pp.609-616, 2006.

[5] R. Ptucha and A. E. Savakis, LGE-KSVD: Robust sparse representation classification, *IEEE Trans. Image Processing: A Publication of the IEEE Signal Processing Society*, vol.23, no.4, pp.1737-1750, 2014.

[6] K. Hui, C. Li and L. Zhang, Sparse neighbor representation for classification, *Pattern Recognition Letters*, vol.33, no.5, pp.661-669, 2012.

[7] L. Zhang, W. D. Zhou, P. C. Chang et al., Kernel sparse representation-based classifier, *IEEE Trans. Signal Processing*, vol.60, no.4, pp.1684-1695, 2012.

[8] S. R. Rao, R. Tron, R. Vidal et al., Motion segmentation via robust subspace separation in the presence of outlying, incomplete, or corrupted trajectories, *IEEE Conference on Computer Vision and Pattern Recognition*, pp.1-8, 2008.

[9] M. Dikmen and T. S. Huang, Robust estimation of foreground in surveillance videos by sparse error estimation, *IEEE International Conference on Pattern Recognition*, pp.1-4, 2008.

[10] Y. Xu, D. Zhang, J. Yang et al., A two-phase test sample sparse representation method for use with face recognition, *IEEE Trans. Circuits and Systems for Video Technology*, vol.21, no.9, pp.1255-1262, 2011.

[11] C. Y. Lu, H. Min, J. Gui et al., Face recognition via weighted sparse representation, *Journal of Visual Communication and Image Representation*, vol.24, no.2, pp.111-116, 2013.

[12] Z. Fan, M. Ni, Q. Zhu et al., Weighted sparse representation for face recognition, *Neurocomputing*, vol.151, pp.304-309, 2015.

[13] S. Gao, I. W. H. Tsang and L. T. Chia, Kernel sparse representation for image classification and face recognition, *Proc. of the 11th European Conference on Computer Vision*, pp.1-14, 2010.

[14] M. Yang, D. Zhang and J. Yang, Robust sparse coding for face recognition, *IEEE Conference on Computer Vision and Pattern Recognition*, pp.625-632, 2011.

[15] M. Yang, L. Zhang, J. Yang et al., Metaface learning for sparse representation based face recognition, *IEEE International Conference on Image Processing*, pp.1601-1604, 2010.

[16] P. Zhu, W. Zuo, L. Zhang et al., Unsupervised feature selection by regularized self-representation, *Pattern Recognition*, vol.48, no.2, pp.438-446, 2015.

[17] D. Skočaj, A. Leonardis and H. Bischof, Weighted and robust learning of subspace representations, *Pattern Recognition*, vol.40, no.5, pp.1556-1569, 2007.

[18] L. Wei, F. Xu and A. Wu, Weighted discriminative sparsity preserving embedding for face recognition, *Knowledge-Based Systems*, vol.57, pp.136-145, 2014.

[19] I. Jolliffe, *Principal Component Analysis*, 2nd Edition, Springer, New York, 1986.
[20] A. Martinez and R. Benavente, The AR face database, *CVC Technical Report*, vol.24, 1998.
[21] S. J. Kim, K. Koh, M. Lustig et al., An interior-point method for large-scale l_1-regularized least squares, *IEEE Journal of Selected Topics in Signal Processing*, vol.1, no.4, pp.606-617, 2007.
[22] B. Efron, T. Hastie, I. Johnstone et al., Least angle regression, *The Annals of Statistics*, vol.32, no.2, pp.407-499, 2004.
[23] K. C. Lee, J. Ho and D. J. Kriegman, Acquiring linear subspaces for face recognition under variable lighting, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.27, no.5, pp.684-698, 2005.
[24] J. Mairal, J. Ponce, G. Sapiro et al., Supervised dictionary learning, *Advances in Neural Information Processing Systems*, pp.1033-1040, 2009.