A NOVEL GENERALIZED ASSIGNMENT FRAMEWORK FOR THE CLASSIFICATION OF HYPERSPECTRAL IMAGE

Ding Ni and Hongbing Ma

Department of Electronic Engineering, Tsinghua University, Beijing, 100084, China
Emails: nid12@mails.tsinghua.edu.cn, hbma@mail.tsinghua.edu.cn

ABSTRACT

Recently, sparse representation based classification has been widely used in pattern recognition. Most of existing methods exploit the recovered representation coefficients to reconstruct the inputs, and the classwise reconstruction errors are used to identify the class of the sample based on the subspace assumption. Different from the reconstruction pipeline, an assignment framework is built on the representation coefficients in this paper. More specifically, we treat the representation coefficients as soft assignments of the class labels, and the distribution of the assignments reveals the class of the sample. Under this framework, we can easily generalize it to multi-sample and/or multi-feature scenarios, where multiple assignment instances can be directly fused to stabilize the distribution estimation. As such, the estimated distribution pattern can be used as a new discriminative feature for classification. Experiments on the classification of hyperspectral image demonstrate that the generalized assignment framework can effectively combine neighboring samples and multiple features for collaborative classification, which could achieve significantly better results than several state-of-the-arts.

Index Terms— Hyperspectral image classification, sparse representation, soft assignments, multi-sample, multi-feature

1. INTRODUCTION

Hyperspectral imaging, which is capable of capturing hundreds of continuous narrow spectral bands spanning the visible to infrared spectrum, has found wide applications in agriculture, environment, military, and many other fields. As a basic task of many applications, hyperspectral image (HSI) classification has become a hot research issue in recent years. Despite the plenty of spectral bands, it is usually hard or expensive to collect enough labeled samples, making it a great challenge to realize high-accuracy classification with limited training samples.

Recently, sparse representation (SR) has drawn great attention in machine learning and pattern recognition, and achieved impressive performance in many classification problems [1, 2, 3]. Sparse representation based classification (SRC) assumes that samples of the same class lie in a class-specific low-dimensional subspace, and that a test sample can be expressed as a sparse linear combination of all the training samples [1]. SRC has also been introduced to HSI classification community, and achieved promising results [4, 5, 6, 7, 8].

Nevertheless, most of existing SR-based methods take the classwise reconstruction errors induced by the recovered representation coefficients as the classification criterion. Unfortunately, SRC suffers from the instability of the recovered coefficients due to the high coherence of training samples in real applications [9, 10]. The instability implies that the recovered coefficients might differ significantly between similar input features and that the nonzero entries might spread across multiple classes, thus weakening the discriminability. Kinds of strategies have been proposed in the literature to alleviate such a weakness. A natural way is to combine multiple similar samples to be represented jointly, such as joint SRC [4], Laplacian regularized SRC [7], SRC in tangent space [11], etc. These remedies could improve the stability of the model to some extent, but at the expense of the complexity of the algorithms.

In our previous work, we have found that the SR coefficients follow a class-specific distribution, although they manifest instability in a single representation [12]. Moreover, SR coefficients can also be viewed as some measure of the similarities between the dictionary atoms and the test sample to be represented [13, 14], where larger coefficient indicates higher similarity. Intuitively, the more similar the test sample and a dictionary atom, the more likely they belong to the same class. Therefore, we can treat the SR coefficients as soft assignments of the test sample belonging to the classes of the corresponding atoms. As such, SR actually plays the role of automatic atom selection and similarity measure. The benefits of such a treatment are as follows. First, when multiple samples of the same class are available, their assignments can be directly fused by accumulation. Moreover, the assignments of multiple features of each sample can also be accumulated if the dictionary atoms of each feature follow the same arrangement. Then multi-sample and multi-feature classification can be unified into the same framework. Second, as SR coefficients follow a class-specific distribution inherently, multi-sample and multi-feature collaborative classification can be performed automatically through assignment accumulation. Last but not least, assignment accumulation can also make the model robust to few outliers from a statistic perspective.

In summary, the soft assignment framework takes no effort to stabilize the representation model itself as what conventional methods do, but exploits the class-specific distribution and the similarity measure property of SR coefficients in a statistic way. If the instances participating the statistic are sufficient, the stability can also be guaranteed. Experiments on the classification of HSIs demonstrate that soft assignment based classification significantly outperforms several state-of-the-arts, and could achieve high-accuracy classification with very limited training samples.

2. GENERALIZED SOFT ASSIGNMENTS

SRC is based on the observation that high-dimensional input features usually lie in a class-specific low-dimensional subspace in many real-world problems. A test sample can be sufficiently reconstructed by the training samples of the same class, which is naturally sparse

This work is supported by the National High Technology Research and Development Program of China under Grant 2007AA12Z149.

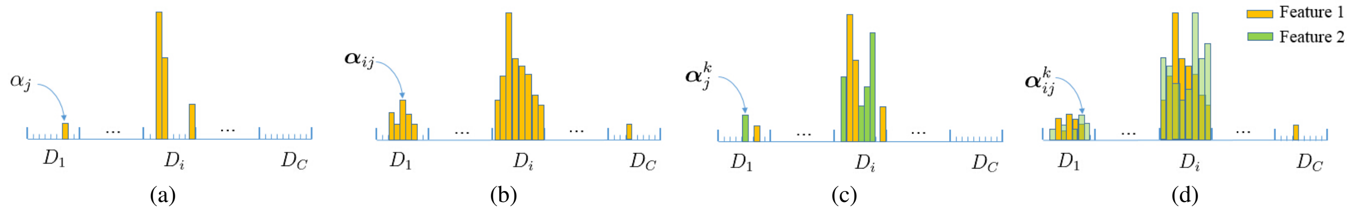


Fig. 1. Illustration of soft assignment framework. (a) Class label assignments of sparse representation. (b) Accumulated assignments of multiple samples. (c) Accumulated assignments of multiple features. (d) Accumulated assignments of both multiple samples and multiple features. For visualization purposes, only two kinds of features are illustrated here.

under the dictionary consisting of training samples from all classes. Formally, a structured dictionary $D = [D_1, D_2, \cdots, D_C] \in \mathbb{R}^{d\times N}$ is constructed by the training samples from C classes, where $D_i \in \mathbb{R}^{d\times N_i}$ is the subdictionary consisting of training samples from class $i\,(i = 1, 2, \cdots, C)$ and $N = \sum_{i=1}^{C} N_i$ is the total number of training samples. A test sample $\boldsymbol{x} \in \mathbb{R}^d$ can be represented as a sparse linear combination of all the training samples as

$$\hat{\boldsymbol{\alpha}} = \arg\min_{\boldsymbol{\alpha}} \frac{1}{2}\|\boldsymbol{x} - D\boldsymbol{\alpha}\|_2^2 + \lambda\|\boldsymbol{\alpha}\|_1 \tag{1}$$

where $\|\cdot\|_1$ is the $\mathcal{L}_1$-norm encouraging the sparseness of representation coefficients. The first term in (1) is data fidelity and the second is sparsity penalty, and the parameter λ balances the two terms.

According to the subspace assumption, the nonzero entries of the recovered representation coefficients would mainly concentrate on the subdictionary of the true class [1], meanwhile the amplitudes indicate the similarities between the test sample and the corresponding dictionary atoms [13, 14]. It is reasonable to assume that the higher the similarity, the more likely the test sample belongs to the class of the corresponding atom. Thus, we can treat the recovered coefficients as soft assignments of the test sample having the class labels of the corresponding atoms, as illustrated in Fig. 1(a). With this treatment, we can easily generalize it to multi-sample, multi-feature and both multi-sample and multi-feature scenarios.

2.1. Generalized soft assignments with multiple samples

If the SR coefficients are treated as soft assignments about the class of a test sample, when multiple samples from the same class are available, their accumulated assignments would represent the overall distribution of the class labels. For example in HSI where spatially neighboring pixels usually belong to the same class, the accumulated assignments actually reflect the total likelihood about their common category. Specifically, given a set of n input samples from the same class, $\mathcal{X} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n\}$, whose associated SR coefficients are $\mathcal{A} = \{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_n\}$ respectively, their accumulated assignments can be defined as

$$\begin{aligned} h_j &= \sum_{i=1}^{n} |\boldsymbol{\alpha}_{ij}|\,, \quad j = 1, 2, \cdots, N \\ \boldsymbol{h}(\mathcal{X}) &= \frac{1}{T}\cdot[h_1, h_2, \cdots, h_N]^{\mathrm{T}} \in \mathbb{R}^N \end{aligned} \tag{2}$$

where $\boldsymbol{\alpha}_{ij}$ is the jth entry of the coefficient vector $\boldsymbol{\alpha}_i \in \mathbb{R}^N$, and $T = \sum_{i=1}^{N} h_i$ is a constant to normalize the summation to be 1.

Perceptually, each sample $\boldsymbol{x}_i$ can be viewed as a random sampling from the class it belongs to. As the SR coefficients follow a class-specific distribution [12], an individual assignment would also be a random instance of the class-specific overall assignments. When the cardinality of the sample set is large enough, the accumulated assignments defined in (2) would be a good estimate of the class distribution statistically, as illustrated in Fig. 1(b). Another advantage is that accumulated assignments would not be impacted much by few outliers, and thus the robustness can be guaranteed.

2.2. Generalized soft assignments with multiple features

Similarly, we can generalize the soft assignment framework to multi-feature scenario. Since different features might provide complementary information among each other, multi-feature classification has drawn great attention in HSI classification recently [15, 16]. When K kinds of features are extracted, we can construct a dictionary set $\{D^1, D^2, \cdots, D^K\}$, where D^k is the dictionary corresponding to the kth feature and the dictionary atoms are all arranged in the same order. Intuitively, the SR coefficients of different features of a sample would follow a similar distribution. When treated as soft assignments of class labels, the accumulation of these coefficients would be the total assignments of the sample. Formally, for a test sample $\boldsymbol{x}$ with K features $\mathcal{F}(\boldsymbol{x}) = \{\boldsymbol{x}^1, \boldsymbol{x}^2, \cdots, \boldsymbol{x}^K\}$, we first obtain the SR coefficients of each feature by

$$\hat{\boldsymbol{\alpha}}^k = \arg\min_{\boldsymbol{\alpha}^k} \frac{1}{2}\|\boldsymbol{x}^k - D^k\boldsymbol{\alpha}^k\|_2^2 + \lambda\|\boldsymbol{\alpha}^k\|_1 \tag{3}$$

Then, the accumulated assignments of multiple features can be defined as

$$\begin{aligned} h_j &= \sum_{k=1}^{K} w^k|\boldsymbol{\alpha}_j^k|\,, \quad j = 1, 2, \cdots, N \\ \boldsymbol{h}(\mathcal{F}) &= \frac{1}{T}\cdot[h_1, h_2, \cdots, h_N]^{\mathrm{T}} \in \mathbb{R}^N \end{aligned} \tag{4}$$

where w^k is the weight of the kth feature. The introduction of feature weights provides the model more flexibility, where we can emphasize more on features with better discriminability. The weights might be provided in priori or learned adaptively. In this paper, we adopt the sparsity concentration index (SCI) [1] as the adaptive weights

$$\mathrm{SCI}(\boldsymbol{\alpha}^k) = \frac{C\cdot\max_i\|\delta_i(\boldsymbol{\alpha}^k)\|_1/\|\boldsymbol{\alpha}^k\|_1 - 1}{C-1} \tag{5}$$

where $\delta_i(\boldsymbol{\alpha}^k)$ indicates the entries associated with the ith class. The larger the value of SCI, the more the nonzero entries concentrate on a specific subdictionary. Features with higher concentration are more reliable, and hence larger weights should be assigned.

Different from the multi-sample scenario which exploits the diversity of multiple samples, multi-feature scenario tries to exploit the complementarity of different features. With the assumption that the SR coefficients of different features follow a similar distribution, multi-feature assignment accumulation is also try to estimate their shared distribution pattern in statistic (see Fig. 1(c)).

2.3. Generalized soft assignments with both multiple samples and multiple features

Following the same idea, we can further generalize the soft assignment framework to both multi-sample and multi-feature scenario. It can be observed that whatever multi-sample or multi-feature, the process of the assignment accumulation is actually a distribution estimation. With this common objective in mind, we can combine multi-sample and multi-feature to boost the estimation. Specifically, given a sample set $\mathcal{X} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n\}$ and the associated multiple features of each sample $\mathcal{F}(\boldsymbol{x}_i) = \{\boldsymbol{x}_i^1, \boldsymbol{x}_i^2, \cdots, \boldsymbol{x}_i^K\}$, we first obtain the SR coefficients $\boldsymbol{\alpha}_i^k$ of $\boldsymbol{x}_i^k$ by solving (3), then the accumulated assignments of both multiple samples and multiple features can be defined as

$$\begin{aligned} h_j &= \sum_{k=1}^{K} w^k \sum_{i=1}^{n} |\boldsymbol{\alpha}_{ij}^k|, \quad j = 1, 2, \cdots, N \\ \boldsymbol{h}(\mathcal{X}, \mathcal{F}) &= \frac{1}{T} \cdot [h_1, h_2, \cdots, h_N]^{\mathrm{T}} \in \mathbb{R}^N \end{aligned} \tag{6}$$

where w^k is the weight of the kth feature too. Here, we define w^k similarly to (5) but with $\boldsymbol{\alpha}^k$ replaced by $\boldsymbol{h}^k$ calculated by (2), as we believe that the feature with more concentrated distribution is more discriminant, and thus we should assign larger weight.

Compared to multi-sample or multi-feature scenario, their combination increases the number of instances participating in the estimation of the assignment distribution, which could enhance the robustness and accuracy in turn, as illustrated in Fig. 1(d). In essence, the combination jointly exploits the diversity of multiple samples and the complementarity of multiple features.

3. HSI CLASSIFICATION USING THE GENERALIZED ASSIGNMENT FEATURE

According to the homogeneity of land covers' distribution, spatially neighboring pixels usually come from the same class and hence they actually form a set of similar samples. Besides the spectral feature, kinds of spatial features have been proposed in the literature to include the contextual information in HSI classification [17, 18, 19]. In principle, any reasonable spectral-spatial features can be applied to the proposed framework. In this paper, we employ the widely used spatial mean of spectral features, the extended morphological profiles (EMPs) [18] and the extended multi-attribute profiles (EMAPs) [20] to conduct the multi-feature classification. The spatial neighbors and/or multiple features of each pixel meet the requirements of multi-sample and/or multi-feature classification and the proposed framework can be utilized.

It can be noted that the generalized soft assignments can be viewed as a kind of histogram reflecting the assignment distribution. A simple way to determine the class label is to count the total assignments associated with each subdictionary, but it only exploits the local information of the distribution and ignores the overall pattern. Considering that support vector machine (SVM) with histogram intersection kernel (HIK) is very effective in the classification of histogram-type feature [21], we use the assignment result as a new input feature to the HIK-based SVM for classification. The histogram intersection between two histograms is defined as

$$K(\boldsymbol{h}(\boldsymbol{x}_i), \boldsymbol{h}(\boldsymbol{x}_j)) = \sum_{k=1}^{N} \min(\boldsymbol{h}_k(\boldsymbol{x}_i), \boldsymbol{h}_k(\boldsymbol{x}_j)) \tag{7}$$

where $\boldsymbol{h}(\boldsymbol{x}_i)$ is the multi-sample and/or multi-feature assignment feature of sample $\boldsymbol{x}_i$, and N is the number of histogram entries.

4. EXPERIMENTAL RESULTS

Two popular benchmarks in HSI classification are used to validate the effectiveness of the proposed method. The first is the Indian Pines data set, which is of size 145×145 pixels with 200 spectral bands. There are 16 ground-truth classes available, with 10366 labeled samples in total. The second is the University of Pavia data set, which is of size 610×340 pixels with 103 spectral bands. Nine ground-truth classes, with 42776 labeled samples in total, are available for this scene. Both of the data sets are available online. [1]

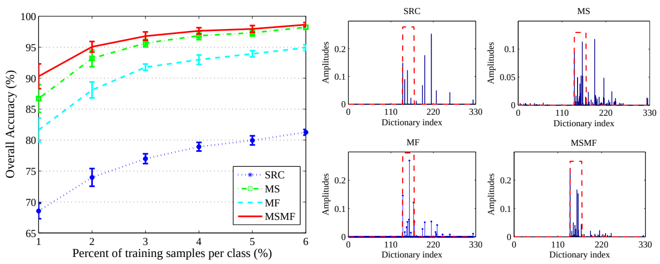
In the experiments, part of the labeled samples are randomly selected for training, and the rest are for testing. Overall accuracy (OA), average accuracy (AA) and the κ coefficient measure are used to evaluate the classification performance. In order to avoid any sampling bias, all of the experiments are repeated ten times with different training sets, and the mean results are reported. In the multi-sample classification, the neighborhood is simply set as a fixed window of size 7×7 pixels around each test sample. In the multi-feature classification, the spatial mean of spectral features is extracted by the 3×3 average filtering, the EMPs and EMAPs are extracted on the first five PCA components. The structural elements of EMPs are disks with radius ranging from 1 to 10 pixels. The EMAPs are built on the area and standard deviation attributes. The threshold values of the area attributes are chosen in the range {50,500} with a step size of 50, and the deviation attributes are chosen in the range {2.5%, 20%} of the mean of the component with a step size of 2.5%, as suggested in [17]. For simplicity, the regularization parameter λ is set to the same for all the features, and we find that $\lambda = 10^{-3}$ for the Indian Pines dataset and $\lambda = 10^{-4}$ for the University of Pavia dataset could achieve satisfactory results.

4.1. Effectiveness validation

In this experiment, we investigate the effectiveness of the proposed assignment feature. The multi-sample, multi-feature and both multi-sample and multi-feature assignment feature based classifications are denoted as MS, MF, and MSMF, respectively. Fig. 2(a) shows the classification results of each method on the Indian Pines data set with different percent of training samples per class. Under the soft assignment framework, the diverse samples involved in multi-sample classification and the complementary features involved in multi-feature classification are all treated as instances to estimate the assignment distribution. As one can see, both MS and MF achieve significantly better results than the single spectral based SRC. In addition, we know that the more instances involved in the statistic, the more accurate and stable the estimation will be. Comparing MS with MF, there are 49 (7×7) instances for each sample in MS, while there are only 4 instances in MF. Consequently, MS could get better estimate than MF, thus obtaining better results. Since MSMF inherits the merits of both MF and MS, more accurate estimation can be guaranteed and thus the best performance is achieved.

Fig. 2(b) shows the estimated assignment distribution of a test sample under different scenarios. One can note that the coefficients of SRC spread widely and are not dominant on the block of the true class due to the instability of SR. In contrast, the assignments of MS concentrate more on the right block, and large assignments of MF are located on the desired block too. The assignments of MSMF appear purer than MS, concentrating more on the right block. Therefore, assignment feature based classification could correctly classify the samples that might be misclassified by SRC.

[1] http://www.ehu.eus/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes.



(a) (b)

Fig. 2. Results on the Indian Pines data set. (a) Classification accuracy versus the percentage of training samples per class. (b) Details of the assignments of a test sample, using 3% training samples. The red dashed box indicates the block corresponding to the true class.

4.2. Performance comparison

In order to demonstrate the superiority of the proposed method in HSI classification, we compare it with several conventional approaches, including SRC, SVM, contextual SVM (CSVM) [22], Laplacian regularized SRC (LS) [7], KOMPCK [8], and JCRCMTL [16]. SRC and SVM are single spectral based methods. LS and CSVM are multi-sample based methods. KOMPCK and JCRCMTL are multi-feature and both multi-sample and multi-feature based methods, respectively. For fair comparisons, the neighborhood in multi-sample classification and the extracted features in multi-feature classification are all the same among the related methods. Best parameters specific to individual methods are obtained by cross-validation.

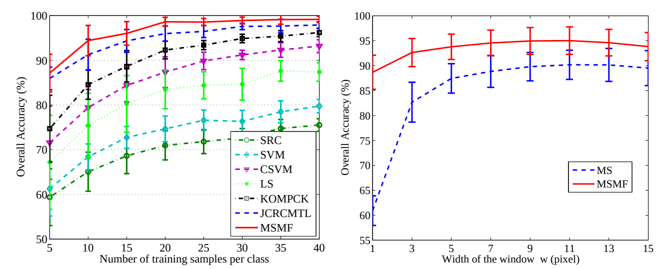
Table 1 presents the classification results of each method on the Indian Pines data set, with 3% of labeled samples used for training. It can be noted that multi-sample classifications achieve much better results than the baselines on single sample (e.g., CSVM vs. SVM, LS vs. SRC). It is because multi-sample classification exploits contextual information to strengthen the power of the classifier. Multi-feature classification enhances the distinguishability of samples by employing more discriminative features. Evidently, the combination of multi-sample and multi-feature could further improve the classification performance, as one can see from JCRCMTL. While, the proposed MSMF exploits multi-sample and multi-feature jointly to estimate a stable assignment distribution, and further use the distribution pattern as a higher-level discriminative feature for classification. Compared to JCRCMTL, MSMF is not only much simpler in the model complexity, but also obtains better results for most of the classes as well as the overall performance.

In the sequel, we examine the classification performance of various algorithms with different number of training samples. For the University of Pavia data set, we randomly select 5-40 labeled samples per class for training and the rest are for testing. The overall accuracies of all the methods in comparison are shown in Fig. 3(a). It can be noted that the proposed MSMF can consistently achieve noticeably better performance than the other counterparts. Especially, even when only 10 labeled samples per class are available, the proposed MSMF could achieve 94.39% overall accuracy, significantly better than the second best 91.25% of JCRCMTL. It demonstrates again that the proposed soft assignment framework can effectively combine multiple samples and multiple features to realize high-accuracy classification, even with very limited labeled samples.

In addition, it should be noted that since soft assignment feature is a statistic based feature, a small number of outliers would not impact seriously as long as their proportion is low. Fig. 3(b) shows

Table 1. Classification accuracy (%) for the Indian Pines data set using 3% training samples per class

No	SRC	SVM	CSVM	LS	KOMPCK	JCRCMTL	MSMF
1	65.69	69.61	68.04	70.39	86.86	**93.14**	91.76
2	67.11	75.87	90.37	78.45	91.98	91.28	**95.60**
3	57.82	65.87	91.00	67.87	91.65	91.76	**93.91**
4	51.73	50.49	85.31	72.26	75.66	**96.28**	91.81
5	81.47	88.67	89.11	81.24	87.68	87.18	**89.59**
6	95.58	92.55	96.49	**99.67**	97.17	99.10	98.84
7	47.83	91.30	91.74	26.96	96.09	96.09	**98.70**
9	98.50	92.55	97.49	99.96	99.66	100	**100**
9	87.06	85.88	97.06	33.53	100	62.95	**100**
10	57.31	67.93	85.03	71.33	89.04	91.42	**93.19**
11	82.12	76.87	93.23	95.86	95.56	96.12	**97.83**
12	63.31	70.96	87.92	79.88	90.57	91.60	**96.07**
13	99.46	96.78	97.12	**99.85**	99.32	99.76	98.05
14	96.45	93.65	96.37	99.54	99.56	99.81	**99.86**
15	42.99	44.51	76.63	62.58	93.37	97.45	**98.78**
16	88.37	84.78	95.65	**97.39**	83.59	83.91	93.91
OA	76.34	77.90	91.53	86.13	93.68	94.76	**96.54**
AA	73.99	78.01	89.97	77.30	92.36	92.36	**96.12**
κ	72.84	74.80	90.35	84.04	92.79	94.03	**96.06**



(a) (b)

Fig. 3. Results on the University of Pavia data set. (a) Classification accuracy versus the number of training samples per class. (b) Classification accuracy versus the width of the neighborhood window, using 10 training samples per class.

the classification accuracies of MS and MSMF with the variation of neighborhood window width. One can see that both of the methods behave robustly to the size of the window in a wide range. So, the window size is not crucial as long as it is in a reasonable range.

5. CONCLUSION

In this paper, we have proposed a novel soft assignment framework to combine multiple samples and multiple features in classification. Under this framework, multi-sample classification and multi-feature classification can be unified and easily fused. The accumulated assignments of multiple samples and/or multiple features can be treated as a high-level discriminative feature for classification. Experiments on HSI classification have demonstrated that the proposed method outperforms several state-of-the-art approaches, and could realize high-accuracy classification with very limited training samples. Besides, the proposed soft assignment framework is also applicable to other multi-sample and/or multi-feature classifications, such as image set based face recognition. It is our future work to test its effectiveness in other applications.

6. REFERENCES

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

[2] S. F. Cotter, "Sparse representation for accurate classification of corrupted and occluded facial expressions," in *ICASSP*, March 2010, pp. 838–841.

[3] Z. Wang, J. Yang, N. Nasrabadi, and T. Huang, "A max-margin perspective on sparse representation-based classification," in *ICCV*, Dec 2013, pp. 1217–1224.

[4] Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Hyperspectral image classification using dictionary-based sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 10, pp. 3973–3985, Oct 2011.

[5] Q. S. ul Haq, T. Linmi, F. C. Sun, and S. Q. Yang, "A fast and robust sparse approach for hyperspectral data classification using a few labeled samples," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 6, pp. 2287–2302, Jan 2012.

[6] Z. Wang, N. Nasrabadi, and T. Huang, "Discriminative and compact dictionary design for hyperspectral image classification using learning VQ framework," in *ICASSP*, May 2013, pp. 3427–3431.

[7] X. Sun, Q. Qu, N. M. Nasrabadi, and T. D. Tran, "Structured priors for sparse-representation-based hyperspectral image classification," *IEEE Geosci. Remote Sens. lett.*, vol. 11, no. 7, pp. 1235–1239, July 2014.

[8] Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Hyperspectral image classification via kernel sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 1, pp. 217–231, Jan 2013.

[9] S. Gao, I. W. H. Tsang, and L. T. Chia, "Laplacian sparse coding, hypergraph laplacian sparse coding, and applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 92–104, Jan 2013.

[10] L. He, Y. Li, X. Li, and W. Wu, "Spectral-spatial classification of hyperspectral images via spatial translation-invariant wavelet-based sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2696–2712, May 2015.

[11] D. Ni and H. Ma, "Classification of hyperspectral image based on sparse representation in tangent space," *IEEE Geosci. Remote Sens. lett.*, vol. 12, no. 4, pp. 786–790, April 2015.

[12] D. Ni and H. Ma, "Hyperspectral image classification via sparse code histogram," *IEEE Geosci. Remote Sens. lett.*, vol. 12, no. 9, pp. 1843–1847, Sept 2015.

[13] S. Xiang, S. Chen, and L. Qiao, "Sparse representation: Extract adaptive neighborhood for multilabel classification," in *PRICAI 2010: Trends in Artificial Intelligence*, Aug 2010, pp. 304–314.

[14] E. Elhamifar and R. Vidal, "Sparse subspace clustering: Algorithm, theory, and applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 11, pp. 2765–2781, Nov 2013.

[15] L. Zhang, L. Zhang, D. Tao, and X. Huang, "On combining multiple features for hyperspectral remote sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 3, pp. 879–893, March 2012.

[16] J. Li, H. Zhang, L. Zhang, X. Huang, and L. Zhang, "Joint collaborative representation with multitask learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 9, pp. 5923–5936, Sept 2014.

[17] B. Song, J. Li, M. Dalla Mura, et al., "Remotely sensed image classification using sparse representations of morphological attribute profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 8, pp. 5122–5136, Aug 2014.

[18] M. Fauvel, J. A. Benediktsson, J. Chanussot, and J. R. Sveinsson, "Spectral and spatial classification of hyperspectral data using svms and morphological profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 11, pp. 3804–3814, Nov 2008.

[19] W. Li, C. Chen, H. Su, and Q. Du, "Local binary patterns and extreme learning machine for hyperspectral imagery classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 7, pp. 3681–3693, July 2015.

[20] M. Dalla Mura, J. A. Benediktsson, B. Waske, and L. Bruzzone, "Extended profiles with morphological attribute filters for the analysis of hyperspectral data," *International Journal of Remote Sensing*, vol. 31, no. 22, pp. 5975–5991, Nov 2010.

[21] J. Wu, "Efficient HIK SVM learning for image classification," *IEEE Transactions on Image Processing*, vol. 21, no. 10, pp. 4442–4453, Oct 2012.

[22] P. Gurram and H. Kwon, "Contextual SVM using Hilbert space embedding for hyperspectral classification," *IEEE Geosci. Remote Sens. lett.*, vol. 10, no. 5, pp. 1031–1035, Sept 2013.