International Journal of Signal Processing, Image Processing and Pattern Recognition
Vol. 7, No. 6 (2014), pp. 303-312
http://dx.doi.org/10.14257/ijsip.2014.7.6.26

A New Automatic Target Recognition Scheme Based on Model Simulation and Structured Learning

Bo Sun, Xuewen Wu and Jun He

School of Information Science and Technology, Beijing Normal University, Beijing 100875, China
tosunbo@bnu.edu.cn, xuewenwu@hotmail.com, hejun@bnu.edu.cn

Abstract

In recent years, more and more researchers' attention has been drawn to the sparse representation-based classification (SRC) method and its application in image analysis and pattern recognition, due to its good characteristics of high recognition rate, robustness to corruption and occlusion, and little dependence on the features selection etc. However, sufficient training samples are always required by the sparse representation method for the effective recognition. In practical applications, it is generally difficult to obtain sufficient training samples of the test targets, especially non-cooperative targets. So the key issues in the effective automatic target recognition (ATR) based on the sparse representation are to obtain sufficient training samples in different scales, angles, and different illumination conditions, and to construct an overcomplete dictionary with discriminative ability. In this paper, a novel sparse representation-based scheme is proposed for the automatic target recognition in the real environment, in which the training samples are drawn from the simulation models of real targets and the overcomplete dictionary is trained using structured sparse learning method. The experimental results show that the proposed method is effective for the automatic target recognition in the practical application, especially, where the desired features of the sparse representation method are kept.

Keywords: *sparse representation; automatic target recognition; model simulation; structured learning*

1. Introduction

Automatic target recognition (ATR) based on image processing technology has been extensively researched [1]-[2]. The ATR systems must have the ability of detecting and identifying the objects of interest in the images. Recently the algorithmic problem of computing sparse linear representations of signals on an overcomplete dictionary has received great progress [3]-[5]. The sparse representation theory has been increasingly used in the fields of images processing and pattern recognition, such as image denoising [6][7], image restoration [8][9], super-resolution reconstruction [10], face recognition [11], automatic target recognition [12], *etc.*

Traditionally, feature extraction in combination with a classifier (KNN [13], SVM [14], Adaboost [15], etc.) is a popular approach utilized in recognition systems. In the traditional research paradigm the feature extraction plays significantly critical role to the recognition performance and computation performance. Recently, on the basis of sparse representation and compressed sensing theory, Wright *et al.* [11] propose a new object recognition framework which is called Sparse Representation-based Classification (SRC). In SRC method, a test sample is represented as a linear combination of the elements of an

ISSN: 2005-4254 IJSIP

overcomplete dictionary which consists of several classes of training samples. The experimental results demonstrate that the sparse representation method used for face recognition hold two major advantages: (a) feature extraction is no longer critical. The random projections or downsampled images could perform as well as any other carefully engineered features. The number of features for a given class and whether the sparse solution is computed correctly become the critical issues; (b) SRC is robust to corruption, occlusion, disguise and etc. However, there also exist some disadvantages in the approaches proposed by Wright [11] and Estabridis [12]. The training samples and testing samples used in [11] are both obtained in a controlled lab environment that only includes frontal views (one aspect angle) of individuals. Work in [12] expands the concepts in terms of detection and rotational invariance, and explores the ideas from the ATR perspective in a real world environment. The author utilizes an overcomplete dictionary that holds rotational invariance by including training data at various horizontal angles. The data used for the training and testing include infrared and visible vehicles imagery captured from 1, 2 and 3 kilometer ranges for all horizontal angles. Overall detection and recognition rates are above 95%. However, the approach presented in [12] is difficult to gain the training samples of ground vehicles which need a large field with radius of several kilometers. Besides, it is inconvenient to collect new training samples of new targets in the practical applications.

In SRC, the construction of overcomplete dictionary is the critical issue to the recognition rate and computational complexity. In practice, however, it is usually difficult to obtain sufficient training data of target to build the sparse representation dictionary. And the dictionary is inefficient if it is formed directly from the original training samples without any optimization as proposed in [11]-[12]. In this paper, we address these two problems by proposing a modified scheme based on model simulation and structured learning, which improves the basic SRC model mainly in two aspects: (1) simulation models of the real world targets (such as cars, boats, planes and etc.) are utilized to obtain sufficient training data under different illumination conditions at various angles and scales. (2)The structured dictionary learning algorithms are used to train the dictionary through the training data to gain better recognition rate and computational efficiency. Experiments conducted on the practical object datasets demonstrate that the proposed ATR scheme is effective for real non-cooperative target recognition in real environment, as well remains the comparative recognition performance and robustness.

The outline of this paper is organized as follows. In Section II, we briefly introduce the principles of sparse representation based classification and dictionary learning, and present the improved algorithms. Then we discuss the proposed ATR scheme in Section III. In Section IV, the experimental results are given in different set-ups. Conclusions and future works are presented in the final section.

2. The Improved Algorithms

2.1. Sparse Representation based Classification

The sparse representation based classification method was proposed in work [11] which shows that the SRC is little dependent on the extracted features and robust to occlusion and corruption with competitive recognition performance. In the SRC method, the atoms of the overcomplete dictionary are formed by stacking the columns of their corresponding two-dimensional training images. Given sufficient training samples of the ith object class $D_i = \{d_{i,1}, d_{i,2}, \cdots, d_{i,n_i}\} \in \mathrm{R}^{m \times n_i}$, as shown in Figure 1, all samples constitute a new matrix $D = [D_1, D_2, \cdots, D_k] = [d_{1,1}, d_{1,2}, \cdots, d_{k,n_k}] \in \mathrm{R}^{m \times n}$ for the entire n training samples of all k object classes which is used for the representation of the test sample, where each training sample image $d_{i,j}$ is associated with a class label l_i .

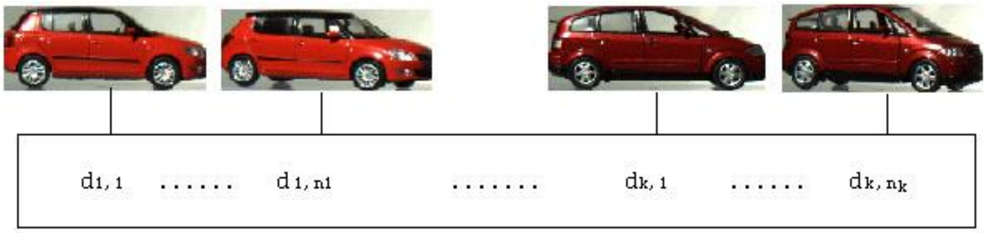

Figure 1. The Training Images of Each Object Form the Dictionary

We now briefly describe the basic process of the SRC algorithm.

1) Input the test sample $y \in \mathrm{R}^m$ and the training samples matrix D.

2) Normalize the columns of D to unit $l_2 -$ norm.

3) Solve the representation error constrained $l_1 -$ norm minimization optimization problem with error tolerance (ε >0).

$$\hat{x} = \arg\min_x \|x\|_1, \; s.t. \|y - Dx\|_2 \leq \varepsilon \tag{1}$$

Or solve the following regularization form. The trade-off constant $\lambda \geq 0$ is to balance between the representation error and the l_1 regularization term.

$$\hat{x} = \arg\min_x \frac{1}{2}\|y - Dx\|_2^2 + \lambda\|x\|_1 \tag{2}$$

4) Compute the representation residual $r_i(y) = \|y - D\delta_i(\hat{x})\|_2$, for each class $i = 1, \cdots, k$, let $\delta_i(x): \mathrm{R}^n \rightarrow \mathrm{R}^n$ be the characteristic function that only selects the components in x that are associated with class i.

5) Output $L(y) = \arg\min_i r_i(y)$, where $L(y)$ represent the identified class label of the test sample.

The SRC method is based on the assumption that a test sample can be represented as a linear combination using only the training samples from the same object. So, when the object classes k is fairly large and the training samples are sufficient, this problem can be mathematically formulated as a sparse decomposition problem on the overcomplete dictionary D. Generally, to find the sparest solution of $y = Dx$, we need to solve the following representation error constrained l_0 −norm minimization optimization problem:

$$\hat{x} = \arg\min_x \|x\|_0, \; s.t. \|y - Dx\|_2 \leq \varepsilon \tag{3}$$

Where $\|\cdot\|_0$ denotes the $l_0 -$ norm which counts the number of nonzero components in a vector. Unfortunately the problem to find the exactly sparsest solution is NP-hard to the underdetermined linear system of Eq. 3. We can approximately solve this problem using greedy algorithms such as Matching Pursuit (MP) or Orthogonal Matching Pursuit (OMP) [20] , but the solution is usually suboptimal. Works [16] in the statistic community has proved that if the solution of Eq.3 is sparse enough, we can relax the l_0 norm to the l_1 norm. Namely, the solution of the l_0 norm minimization problem in Eq.3 and the l_1 norm minimization problem in Eq.1 are approximately equal. The problem of Eq.1 is classic lasso problem [21] in statistics which can be solved by standard linear programming methods. Furthermore, we can transfer the representation error constrained l_1 minimization optimization problem in Eq. 1 into $l_1 -$ norm regularization problem in Eq.2.

The SRC method in [11] assumes that the underlying subspace for each class is low dimensional, the sparsest representation of a test sample ideally corresponds to the training data from the identical class. The experimental results show that the SRC method offers a great advantage over many classification methods since it can effectively deal with corrupted data within the same sparse representation model. However, the SRC method looks for the sparsest representation of a test sample on a dictionary composed of all training samples across all classes without considering the structure hidden in the training data. As shown in Figure 1, the dictionary of the training data obviously has a block structure in which data from different class form individual blocks of the dictionary. In this paper, we propose a scheme to explicitly consider the inhere structure in the overcomplete dictionary.

Instead of looking for the sparsest representation of a test sample y on the dictionary D composed of all the training samples, a better criterion for classification is to look for a representation of the test sample that involves the minimum number of blocks from the dictionary [17]. We can formulate this problem using the following optimization program.

$$\hat{x} = \arg\min_{x} \sum_{i=1}^{k} I\left(\|x[i]\|_q > 0\right) s\, t \,\| y - Dx\|_2 \le \varepsilon \qquad (4)$$

Where $I(\square)$ is the indicator function, $q \ge 1$, and $x[i] \in R^{n_i}$ are the components of the sparse representation vector x corresponding to the i-th block of the dictionary, $D[i] \in R^{m \times n_i}$, as shown in Figure 1. This optimization problem seeks the minimum number of nonzero coefficient blocks that approximately represent the test sample. Note that the optimization problem Eq.4 is NP-hard since it requires searching exhaustively over all possible blocks of D and checking whether they can span the given y. In order to solve this problems efficiently, we propose a convex relaxation strategy similar with Eq. 3. A ℓ_1 relaxation of this optimization problem is given by

$$\hat{x} = \arg\min_{x} \sum_{i=1}^{k} \|x[i]\|_q \;\; s.t. \; \|y - Dx\|_2 \le \varepsilon \qquad (5)$$

2.2. The Dictionary Learning Algorithm

In the aforementioned SRC method, when the number of training data in each class is large, we can better capture the underlying distribution of data and the classification performance will increase. Nonetheless, existing sparse decomposition algorithms do not have theoretical guarantees when it comes to highly redundant dictionaries and the computation consumption is large. On the other hand, when the number of training data in each class is small, sparse decomposition methods have good theoretical guarantees. However, classification algorithms do not perform well. So in our proposed recognition scheme, we use dictionary learning method to construct a compact dictionary from a large set of training data.

It is well known that the K-SVD [18] algorithm can find the overcomplete dictionary D that yields sparse representations for a set of training samples. Mathematically, this problem can be formulated as

$$\min_{D,X}\{\|Y-DX\|_F^2\} \quad \mathit{Subject\ to}\ \forall i, \|x_i\|_0 \le T_0 \tag{6}$$

Where, $Y=\{y_i\}_{i=1}^N$ is the set of training samples and $X=\{x_i\}_{i=1}^N$ is the set of representation coefficients of the training samples. $\|X\|_F$ is the Frobenius norm defined as $\|X\|_F=\sqrt{\sum_{ij} X_{ij}^2}$.

Like the K-means algorithm, K-SVD also uses a two step processes to update D and X iteratively [19]. In the sparse coding step, D is fixed and some pursuit algorithm such as Orthogonal Matching Pursuit (OMP) [20] and Basis Pursuit (BP) [21] algorithms can be used to compute x_i in Eq. 1. In the dictionary update step, D and X are assumed to be fixed and only one column d_k of D is updated at a time. Defines the group of training samples that use d_k as:

$$\omega_k = \{i \mid 1 \le i \le N, x_i(k) \ne 0\} \tag{7}$$

Then we can compute $E_k = Y - \sum_{j\ne k} d_j x^j$ and restrict E_k by choosing the columns corresponding to ω_k , so that we obtain E_k^R . Finally, we can apply SVD decomposition $E_k^R = U\Delta V^T$ and update d_k to be first column of U, and x_k^R to be the first column of V multiplied by $\Delta(1,1)$. All dictionary columns are updated in this way. Iteration through the two steps will produce the dictionary that can approximately represent the given y_i sparsely and accurately.

Assume that we have k classes of training samples, i = 1 . . . k. The simplest strategy for using dictionary learning for classification is to learn k individual dictionaries ($D_i, i=1\ldots k$), one for each class. We can approximate the test sample using a constant sparsity L and the k different dictionaries. The k different residual errors can then be used for the classification task. This is essentially the strategy employed in [21]. Thus, the first simple method of estimating the class $L(y)$ for certain sample y can be written as blew:

$$L(y) = \arg\min_{i=1\ldots k} R(y,\ D_{\mathrm{i}}) \tag{8}$$

Where, $R(y,\ D_{\mathrm{i}})$ represents the representation error of y on $\mathrm{D_i}$ with a sparsity factor L.

In this paper, instead of this reconstruction-based approach, we propose to impose a block structured constraint described as Eq. 4 on the sparse representation coefficients vector x in the sparse coding process to simultaneously learn a block structured dictionary with better discriminative ability. The dictionary update process is same as K-SVD algorithm.

3. Our Proposed ATR Scheme

3.1. Samples

To collect sufficient training samples, the scaled models (1:43) of real world targets (such as ships, cars, etc.) are made. The models are putted on the center of the rotary table of the image acquisition system shown in Figure 2. The models images are collected in horizontal direction from 0° to 360° and in vertical direction 0° to 90°. In order to cover the illumination variations of nature environment, we change the illumination intensity and illumination direction to repeat the image acquisition procedure. The models images were cropped and normalized shown in Figure 3 to construct the training samples dataset for the next dictionary learning process. The test samples are collected from nature images of the real target in real world situations for the detection and recognition process.

3.2. Dictionary

To get an efficient sparse representation-based classification dictionary that contains various variations from the large-scale training samples. In our proposed approach, the number of the training samples is large because of the changing angle and illumination parameters. The method proposed in [11][12] to form the dictionary directly from the original training samples is inefficient. We implement the structured sparse representation learning method to build the overcomplete dictionary utilizing the large-scale training samples. Assume that we have k sets $S_i\,(i=1\cdots k)$ of training samples, belonging to k different classes. The structured dictionary $D=[D_1,D_2,\cdots,D_k]$ is learned, which has k blocks structure, one for each class. The dictionary D is used in the next detection and recognition process.

3.3. Detection

Before classifying a given test sample, we must first decide if it is a valid object from one of the classes in the training dataset. The ability to detect and then reject invalid test samples is crucial for recognition systems to work in real world situations. We use a sliding rectangle window over the test image in order to find potential target regions. In the sparse representation paradigm, the representation coefficients x of the sliding window regions are computed in terms of the representation dictionary D which is constructed in section 2.2. Then, the sparsity concentration index (SCI) of the coefficient vector $x\in\mathbf{R}^{\mathrm{n}}$ is computed as

$$\mathrm{SCI}(x)=\frac{k\times\max_i\|\delta_i(x)\|_1/\|x\|_1-1}{k-1}\in[0,1] \qquad (9)$$

Where $\delta_i(x)$ is the characteristic function that selects the coefficients of x associated with the ith class. Supposing $SCI>\tau$, then the rectangle region is accepted as a possible target region and sent to the next recognition process. Otherwise, the rectangle region is rejected and discarded.

3.4. Recognition

The recognition process utilizes the l_1 norm minimization algorithm as Eq. 2 to get the sparse representation coefficient vector x of the potential target region y found in Section 2.1. Then we can classify y by assigning it to the object class that minimizes the individual class representation residual, $\mathrm{identity}(y) = \arg\min_i r_i(y),$ where $\mathrm{r}_\mathrm{i}(y)$ is defined as $\mathrm{r}_\mathrm{i}(y) = \|y - D\delta_i(x)\|_2$.

4. Experiments

4.1. Model Image Acquisition System

We have designed a system that can capture images of a target in different scales, aspect angles and different illumination. A sketch of the system is shown in Figure 2: The image capture system consists of four axes that control the different moving direction. The first is the rotation axis and the target on the rotary table can rotate at all 360 degrees. The second is the horizontal axis, along which the rotary table can move front and back. The third is the vertical axis, along which the camera can move up and down. The fourth is the pitch axis which enables the camera pitch to 0-90 degree to ensure that the captured target is always located on the center of the lens. The light illumination system can reflects off of the white board and illuminates the targets indirectly. Our four-axis linkage image acquisition system has several advantages:

• The capture parameters can be modified in software, rather than hardware.

• It is easy to capture target images in different angles, scales, and illumination quickly.

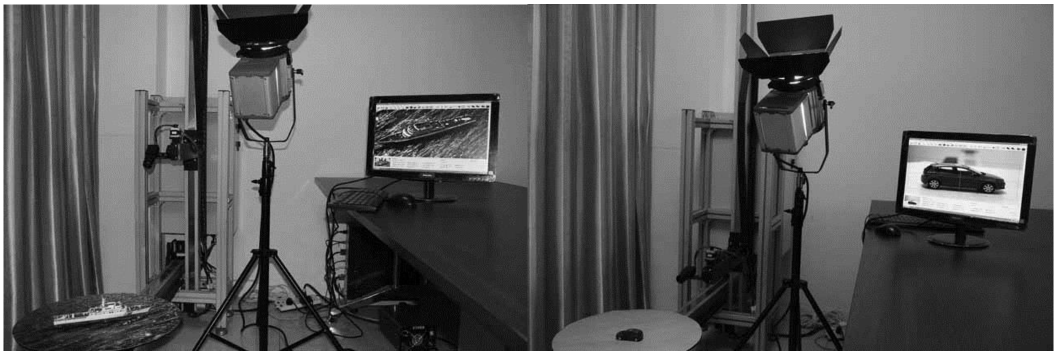

Figure 2. Four-axis Linkage Image Acquisition Platform

4.2. Tests on the Real Target Dataset

Using the image acquisition system that we describe in Section 4.1, and show in Figure 2, we collect many model images of different real car targets of new Passet, new Polo, Octavia and Tiguan shown in Figure 3 in different aspect angles and illumination. For testing our algorithm, we have also collected 40 images of these cars in real environment shown in Figure 4.

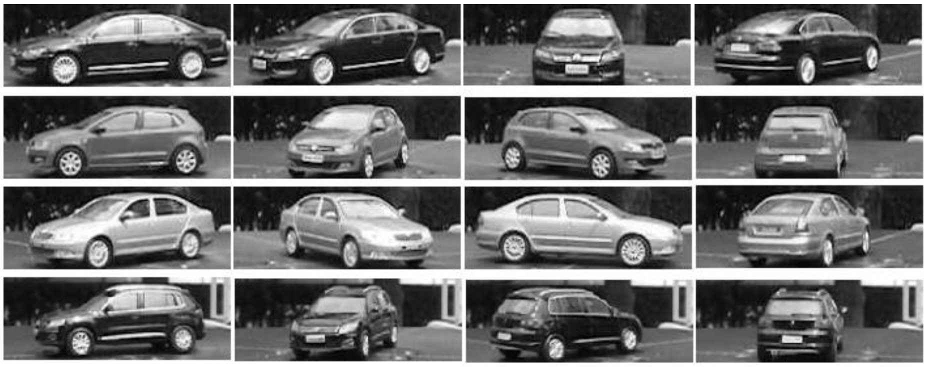

Figure 3. Model Images of New Passet, New Polo, Octavia and Tiguan

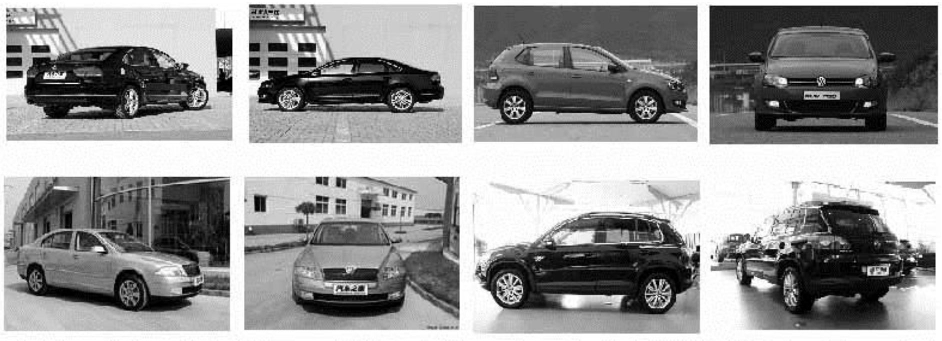

Figure 4. Real Car Images of New Passet, New Polo, Octavia and Tiguan

We compared our proposed scheme with methods in [11][12]. The target recognition performances are reported in Tables 1-3 for the practical target images. The results in Table 1 shows the recognition performance for the various number of training images which are selected to directly form the dictionary with different $\Delta\theta$ degrees of 8°, 4°, 2° and 1° for each target same as described in [12]. Table 2 shows the recognition performance for the various feature space dimensions of 80, 120, 200 and 300. The training samples and the test target regions are reduced to same dimensions using random projection. The result shows that the feature is no longer crucial as long as the dimension of the feature space surpasses certain threshold. Table 3 shows the benefit of structured dictionary learning approach in terms of the recognition rate for various dictionary sizes 45, 90, 180 and 360.

Table 1. Recognition Rate for Various Numbers of Training Images

Number of training images per class object	90	120	180	360
Recognition rate	0.75	0.75	0.75	0.81

Table 2. Recognition Rate for Various Feature Dimensions

feature dimensions	80	120	200	300
Recognition rate	0.43	0.5	0.81	0.81

Table 3. Recognition Rate for Various Dictionary Sizes

Number of dictionary elements per class object	90	120	180	360
Recognition rate	0.81	0.81	0.81	0.81

5. Conclusion and Future Work

Using a well thought-out improvement of existing ideas (model simulation, SRC, dictionary learning), we have proposed a new simulation sparse representation-based method for automatically recognizing the targets in natural images taken under real environments. The experimental results show that the proposed approach is effective for the automatic target recognition in the practical application. The system achieves stable recognition performance under some variations in illumination, misalignment, and even under small amounts of corruption and occlusion.

We achieve good recognition performance on our practical targets images in the experiments, while using only some aspect images and scale images. Our system could potentially be extended to better handle practical targets images by incorporating more model training images with various angles, scales and illumination. Another important direction for future investigation is to introduce the partition algorithm to better tackle big illumination variations and misalignment.

Acknowledgements

This work is supported by the Fundamental Research Funds for the Central Universities of China (2012YBXS10).

References

[1] S. Arivazhagan and L. Ganesan, "Automatic target detection using wavelet transform", EURASIP Journal on Applied Signal Processing, vol. 17, **(2004)**, pp. 2663-2674.

[2] D. Kumar and S. Varma, "Multiresolution framework with neural network approach for automatic target recognition", IEEE Conference on Signal Acquisition and Processing, **(2009)**, pp. 168-173.

[3] E. Candes and T. Tao, "Near-optimal signal recovery from random projections: universal encoding strategies?", IEEE Transactions on Information Theory, vol. 52, no. 12, **(2006)**, pp. 5406-5425.

[4] P. Zhao and B. Yu, "On model selection consistency of lasso", Journal of Machine Learning Research, vol. 7, **(2006)**, pp. 2541–2567.

[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, **(2007)**, pp. 606-617.

[6] M. Protter and M. Elad, "Image sequence denoising via sparse and redundant representations", IEEE Transactions on Image Processing, vol. 18, no. 1, **(2009)**, pp. 27-35.

[7] R.. Yan, L. Shao and Y. Liu, "Nonlocal hierarchical dictionary learning using wavelets for image denoising", IEEE Transactions on Image Processing, vol. 22, no. 12 , **(2013)**, pp. 4689-4698.

[8] Z. Xu and J. Sun, "Image inpainting by patch propagation using patch sparsity", IEEE Transactions on Image Processing, vol. 19, no. 5, **(2010)**, pp. 1153-1165.

[9] J. Marial, M. Elad and G. Sapiro, "Sparse representation for color image restoration", IEEE Transactions on Image Processing, vol. 17, no. 1, **(2008)**, pp. 53-69.

[10] J. Yang, Z. Wang, Z. lin, S. Cohen and T. Huang, "Coupled dictionary training for image super-resolution", IEEE Transactions on Image Processing, vol. 21, no. 8, **(2012)**, pp. 3467-3478.

[11] J. Wright, A. Yang, A. Ganesh, S. Sastry and Y. Ma, "Robust face recognition via sparse representation", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 31, no. 2, **(2009)**, pp. 210-227.

[12] K. Estabridis, "Automatic target recognition via sparse representations", SPIE, vol. 7696, 76960O, **(2010)**, pp. 1-9.

[13] K. Q. Weinberger and L. K. Saul, "Distance metric learning for large margin nearest neighbor classification", Journal of Machine Learning Research, vol. 10, no. 2, **(2009)**, pp. 207-244.

[14] C. Cortes and V. Vapnik, “Support vector networks”, Machine Learning, vol. 20, **(1995)**, pp. 273-297.

[15] Y. Freund and R. E. Schapire, “A short introduction to boosting”, Journal of Japanese Society for Artificial Intelligence, vol. 14, no. 5, **(1999)**, pp. 771-780.

[16] D. L. Donoho and X. Huo, “Uncertainty principles and ideal atomic decomposition”, IEEE Transactions on Information Theory, vol. 47, no. 7, **(2001)**, pp. 2845-2862.

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

[18] M. Aharon, M. Elad and A. Bruckstein, “K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation”, IEEE Transactions on Signal Processing, vol. 54, no. 11, **(2006)**, pp. 4311-4322.

[19] J. Z. Feng, L. Song, X. K. Yang and W. J. Zhang, “Sub clustering K-SVD: Size variable Dictionary learning for Sparse Representations”, IEEE International Conference on Image Processing, **(2009)**, pp. 2149-2152.

[20] J. A. Tropp, “Greed is good: algorithmic results for sparse approximation”, IEEE Transactions on Information Theory, vol. 50, no. 10, **(2004)**, pp. 2231-2242.

[21] S. S. Chen, D. L. Donoho and M. A. Saunders, “Atomic decomposition by basis pursuit”, SIAM Review, vol. 43, no. 1, **(2001)**, pp. 129-159.

[22] K. Skretting and J. H. Husoy, “Texture classification using sparse frame-based representations”, EURASIP Journal on Applied Signal Processing, vol. 1, **(2006)**, pp. 1-11.

Authors

Bo Sun, He received the Ph. D. degree in information science and technology from Beijing Normal University, Beijing, P. R. China in 2003. He is now a professor at Beijing Normal University. His current research interests include image processing, pattern recognition, machine learning, intelligent computing.

Xuewen Wu, He received the M. S. degree in computer software and theory from Tsinghua University, Beijing, P. R. China in 2003. Now he is pursuing Ph. D. degree in School of Information Science and Technology at Beijing Normal University. His current research interests include image processing, pattern recognition, sparse representation, machine learning.

Jun He, She received the Ph. D. degree in physical electronics from Beijing Institute of Technology, Beijing, P. R. China in 2003. She is now an associate professor at Beijing Normal University. Her current research interests include signal processing, pattern recognition, Optical imaging, artificial intelligence.