Multi-view object recognition based on sparse representation

Jin-jin Cai*, Bo Liu, Wei Yao

College of computer science and technology, Agriculture University of Hebei, Baoding, China

Received 1 October 2014, www.cmnt.lv

Abstract

In recent years, sparse representation has emerged as a powerful data representative model to draw much attention. In term of the intrinsic structured characteristic of signal itself, this model decomposes a signal as a linear combination of a few atoms from an over-completed dictionary. As it turns out, we obtain the parsimonious representation of signal with regularization by different sparsity-inducing norms. Through the adaptivity and robustness of sparse representation, it is well applied to the field of signal and image processing. In this paper, the problem of recognizing an object from its multi-view images with unconstrained poses and context was considered. A novel framework called metasample-based supervised dictionary learning for multi-view object recognition exploiting the sparse property of intrinsic information was proposed. Experimental results demonstrate that the proposed algorithm exhibits better performance than the recent state-of-the-art methods.

Keywords: multi-view object recognition, sparse representation, supervised dictionary learning algorithm

1 Introduction

The sparsity of the signal [1, 2] refers to the signal has a concise representation on a group of bases. Take JPEG2000 [3] standard for example, it utilizes wavelet bases to expand the image signals. In the transform domain, the corresponding coefficient is zero or close to zero while a few larger coefficients retained most of the information of the original image. The sparsity in question reflects the characteristic of the signal itself, and the base selected is just the carrier of this characteristic. Actually, to represent the original signal via as few linear combinations of the bases as possible also shows that the original signal is usually distributed in the low dimensional subspace. In order to increase the adaptability of the original orthogonal basis, so as to obtain more sparse representation, Mallat [4, 5] et al. put forward the thought of signal decomposition in a redundant dictionary. The corresponding sparse solution can be expressed as:

$$P_0 : MIN\|X\|_0 \, s.t. b = Ax, \tag{1}$$

where, $b \in R^m$ is the original input signal; $A \in R^{m \cdot n}$ is the dictionary; $x \in R^n$ is the code of b in dictionary *A*. As of m<<n, it's also known as the over-complete dictionary or the redundant dictionary. Due to the over-complete nature, each column of a_i in A is called the atomic rather than the basis. ℓ_0 represented by $\|\cdot\|_0$ is used to count the number of the non zero elements in x. When the number of non zero elements is k, we call it k-sparsity. On the account of m<<n, the Equation (1) has infinite solutions and we hope to find the simplest solution namely the minimum number of non zero elements, therefore it is a natural choice to deem ℓ_0 as the objective function.

Introduced by Olshausen [6, 7] and Field [8, 9], recently compressive sensing based on the sparsity of signals has become a research hotspot, which is also greatly promoted the development of sparse representation. The traditional Nyquist-Shannon sampling theorem [10, 11] points out that in order to restore or describe a signal undistortedly, we should at least sample at the rate of twice the bandwidth. In order to facility the storage and transmission of the signal, we should compress the sampling signal, which will lead to the loss of some secondary information's. As a result of this, the sampling information with high sampling rate will be wanted. Compressed sensing theory [12, 13] points out that as long as the signal is sparse or compressible, it can be sampled in a low frequency and recovered the original signal with the help of an optimization algorithm.

At present, as an effective data representation, the sparsity has characteristics of flexibility, robustness. Therefore the sparsity has extensive application in the machine learning, image processing and pattern recognition. Unlike RIP(restricted isometry property) sensing matrix structure [14,15], the sparsity emphasizes the structure of the over complete dictionary, and through the sparse representation of data in the dictionary dredges the structural characteristics of the data.

2 The sparse representation model

The solution to compressed sensing is similar to P_0, in which A is called sensing matrix [1], *b* as the observation signal. How to quickly solve *x* to restore the original signal is the main problem faced by compressed sensing. Donoho

* ***Corresponding author's*** e-mail: jin_jin_cai@163.com

[16] et al. proved that there has only a sparse solution of the equation in (1), *spark (A)* symbolizing the minimum number of the linear correlation in column of *A* at the equation $\|x\|_0 < spark(A)/2$. As the norm of ℓ_0 is NP completeness, so it often substitutes norm of ℓ_1 for norm of ℓ_0, P_0 redefined as:

$$P_1 : \underset{x}{MIN}\|x\|_1 s.t. b = Ax, \tag{2}$$

where $\|\cdot\|_1$ stands for the sum of elements absolute value in *x*. Tao and Candès[17] proves: given that sensing matrix A agrees RIP (Restricted Isometry Property), norm of $\ell_{1\_}$ and norm of $\ell_{0\_}$ have the same solution. As norm of $\ell_{1\_}$ is convex optimization, it can be converted to linear programming problem to be solved.

In $\ell_{1\_}$ norm, the sparse coding of each component is independent each other, or that the each atom has nothing to do with each other in the dictionary. Some components in structured sparse model assumes that the code is relevant and present the prior knowledge through component grouping to make components within one group zero at the same time or not zero at the same time. As for $A \in R^{m \cdot n}$, G stands the index set of variables in each group, $G \subseteq \{1,2,...n\}$, $g = \{G_1, G_2 ... G_{|g|}\}$ stands for collection of each group. Bach [18] defined group sparse constraint as follows:

$$g(x) = \lambda_r \sum_{r \in g} \|x_r\|_q, \tag{3}$$

λ_r is defined the weight within each group, group sparse as the expansion of norm $\ell_{1\_}$, usually $q \in \{2, \infty\}$. Group sparse degrades to $\ell_{1\_}$ norm. When it has only a component in the group. The involvement of group sparse makes atoms divide into several subsets. Atoms within each subset selected in at the same time or selected out at the same time during sparse representation. Group sparse is used for a covariate choice in Group Lasso [19], and is applied in multi-task learning [20] and multi-core learning [21].

Zhao [22] et al. put forward the hierarchical structure between variables. This structure of distribute the variable to different nodes in the tree. It is only when the variable in the parent node is selected; the variables in a sub-node can be selected. Hierarchy depicts the overlapping characteristics of variables in different groups. Jenatton [23] et al. used this constraint in dictionary learning to construct the theme model and image restoration.

In addition, Ding [24] et al. put forwards $\ell_{2,1}$ norm acting on the matrix, defined as:

$$\|X\|_{2,1} = \sum_{i=1}^{n} \sqrt{\sum_{j=1}^{k} x_{ij}^2} = \sum_{i=1}^{n} \left\|x^i\right\|^2, \tag{4}$$

x^i as the *i* row of in *X*. $\ell_{2,1}$ norm is the variations of sparse set to make the matrix using row as unit to sparse. $\ell_{2,1}$ norm can be used in tasks of features selection.

3 The application of the sparse representation in image processing

According to the definition in P_2, when b represents one image, P_2 is used to solve the sparse representation of the image in the dictionary. By adjusting the reconstruction error and sparse representation, the linear combination of the sparse component in the dictionary can reconstruct the original image. Elad [25] points out that the model can be used to solve some common image At present, the additional feature of sparse representation, the model can also be used for face recognition, image classification task processing, such as image compression, image denoising, image restoration.

3.1 SPARSE REPRESENTATION APPLICATIONS IN IMAGE RESTORATION

In image restoration tasks, the first mission is the choice of dictionary. The structure of the dictionary can be divided into two categories: one is analytical method, which is using a simple mathematical function to provide the signal modeling and the dictionary generated by the analytic function. Methods commonly used are Fu Liye transform, wavelet transform and a series of multiscale analysis tools such as: Curvelet transform, Contourlet transform [26] and Bandelet etc. The other dictionary construction method is based on the training data, therefore the dictionary can adapt to the signal changes. This paper mainly relates to the latter kind of dictionary. The dictionary learning algorithm MOD (Method of Optimal Directions), K-SVD (K-Singular Value Decomposition), Mairal proposed the online dictionary learning algorithm.

3.2 IMAGE DENOISING

Traditional image denoising methods mainly focus on the space and frequency domain. In the space domain, it takes the advantage of mean filters to weightily equalize the local information of images. This smoothing process is simple, but it is likely to lose some de tail image information, which can make images become vague. In the frequency domain, taking the advantage of different frequency distributions of the original information and noise information, it will selectively filter the high frequency information represented by noises with the use of frequency filters. But the noise information and the

original information are usually overlapped in frequency band, so the noising effect is not ideal.

Given noise model $y = y_0 + v$, where y represents the observed noise image; y_0 is the original image; v is additive-white Gaussian noise. Sparse denoising model utilizes sparse decomposition of y in a certain basis or over-complete dictionary, namely the sparse code refactoring of original images. And the reconstructed error is the noise.

Elad adopts the method of dictionary study to denoise. Dictionary training can be achieved when people get the sparse representation of images. Under the consideration of dictionary training, this method makes image blocks as the basis denoising unit. The consistency of images in local features and global features has been taken into account at donoising. The following is the given denosing model:

$$\min_{x_{i,j},y_0} \frac{1}{2}\lambda\|y - y_0\|_2^2 + \sum_{ij}\mu_{ij}\|x_{ij}\|_0 + \sum_{ij}\|Ax_{ij} - R_{ij}y_0\|_2^2 . \quad (5)$$

The first variable in this equation ensures the global consistency. And for image block denosing, it also overcomes the unnatural problem of visual effect on each block boundary in late mosaic period. The last two variables are to get the sparse representation of every image block and also to minimize reconstruction error. $R_{ij}y$ refers to extracted image blocks corresponding to labels. OMP algorithm is applied to sparse solution. K-SVD algorithm is used for dictionary study. The thesis points out that by taking advantage of the noise robustness of K-SVD and by focusing on noise images themselves, the training of more suitable dictionary has also achieved better denoising effect.

3.3 SUPER-RESOLUTION

Super resolution was generated according to a corresponding high-resolution image or the sub-resolution image. Since a large number of data loss from high resolution to low resolution, so that the reverse process with great uncertainty. Degradation model is given as $y_l = SHy_h + v$, where $y_l \in R^m$ is a low resolution image wherein, $y_h \in R^n$ is the corresponding high resolution image ($m < n$), H is a linear filter, such as fuzzy matrix. S is the sample matrix, v is the noise term. To solve this underdetermined problem, Yang introduced a priori sparse image. A_l high resolution dictionary with A_h a low resolution dictionary are constructed, y_h is expressed as in a sparse A_h: $y_h \approx A_h x$, $y_l \approx SHA_h x + v$.

This indicates the approximate consistency representation of the image in high and low resolution dictionary. After calculating image sparse representation in the low-resolution dictionary, use the representation to reconstruct image in high resolution dictionary. In order to achieve consistency with the b image on a sparse representation, while dictionary learning simultaneously generating two dictionaries, the following equation:

$$\min_{\{A_h,A_L,X\}} \frac{1}{N}\|Y_h - A_h X\|_2^2 + \frac{1}{M}\|Y_l - A_l X\|_2^2 + \lambda\left(\frac{1}{N} + \frac{1}{M}\right)\|X\|_1 , (6)$$

where y_h and y_l represent the high and low resolution images of training data $Y = \{y_1, y_2, ..., y_t\}$. Each column as image block expressed in vector form, X as sparse coding matrix. N and M were high and low resolution image blocks dimension.

4 Multi view object recognition based on sparse representation

Currently, the sparse representation has been widely used in the field of image classification and achieved good results. Based on BoF (bag-of-feature) and SPM (spatial pyramid matching) image classification method in vector quantization (vector quantization, VQ) stage, which is usually formed by using k-means clustering feature dictionary, when the local feature of an image block is assigned to use hard assignment mode coded, resulting in excessive coding error.

Yang et al. put forward ScSPM (Sparse code SPM) methods. Using linear SVM classification result is better than the nonlinear SPM. The paper also pointed out that the size of dictionary impact on the classification accuracy is too small, and feature representation loses judgment ability; if it is too large, it will bring representation inconsistencies.

Sparse represents the potential instability, and making sparse coding is not suitable for classification tasks directly. The instability is embodied in similar characteristics with no similarity after encoded. Even if the same feature is attached to a small perturbation, it may lead to an entirely different encoding atoms select ion from the dictionary. A direct method is to use the similarity between the training examples, a dictionary for each class learning:

$$P_3 : \min_{A,X} \frac{1}{2}\|B - AX\|_F^2 + \lambda\|X\|_0 + \mu_i \sum_i \Omega_i(B, A, X) . \quad (7)$$

Then the discrimination item group should be added according to the algorithm in order to increase dictionary or the encoding discriminated property. According to the different of discriminate items, such algorithms that are divided into similarity based on discriminant of unsupervised dictionary learning algorithms and supervised learning algorithm based on a consideration of the classification error.

4.1 SUPERVISED DICTIONARY LEARNING ALGORITHM

The algorithm makes use of such supervision information to enhance the distinguish ability of sparse representation. Huang etc., for the problem, introduced the Fisher linear discriminant criterion, at the expense some of the data reconstruction accuracy are of the premise, to maximize the linear discriminant ability of sparse representation.

The method adopted the form of class OMP in sparse solving. Dictionaries use wavelet basis analytical model dictionaries, so there is no dictionary learning process, and coded representation limits capacity. Yang, [22] such as a combination of [24] the reconstruction error minimum classification criterion and [25] the Fisher linear discriminant criterion, proposed FDDL (fish discrimination dictionary learning) method, While it takes into account their determining abilities of dictionary itself and sparse representation, the kind of sample is with small reconstruction error. To other categories, the sample reconstruction error are corresponding large. Assuming that there are C categories of training examples, the following constraints for the *i* class example of the dictionary:

$$\Omega_i(B,A,X)=\left\|Y_i-AX_i\right\|_F^2+\left\|Y_i-A_iX_i^j\right\|_F^2+\sum_{\substack{j=1\\ j\neq i}}^{c}\left\|A_jX_i^j\right\|, \quad (8)$$

where Y_i is a class i sample, A is for the entire dictionary, A_i is set according to category *i* sub-sample dictionary to learn, X_i^j represents the encoding coefficient Y_i in the A_i. So the first item is the entire dictionary for the reconstruction error of class I example. The second term represents the sample class i belonging to dictionary A_i for the sample reconstruction error.

Minimizing the third objective is to reduce other types of dictionaries for Class i sample representation ability. By testing, FDDL in face recognition result is better than SRC method. Similar methods FDDL, Mairal [17] et al. adopted soft-max discriminant cost function, which is defined as follows:

$$C_i^{\lambda}(y_1,y_2,...,y_n)\equiv\log\left(\sum_{j=1}^{N}e^{-\lambda(y_j-y_i)}\right), \quad (9)$$

where Y_i belongs to reconstruction error of category *i*, the smaller value of C_i^{λ}, the stronger the identification of the corresponding category dictionary. This method is used for texture classification. Meanwhile Mairal [18] extended the method to multi-scale dictionaries for edge detection.

Other supervision dictionary learning algorithm directly introduced linear classifier cost function [19]. During the dictionary learning, alternating update classification parameters to minimize the classification errors. As Pham [20] and other introduced classifier constraints is:

$$\Omega(B,A,X)=\left\|L-W^TX\right\|_F^2, \quad (10)$$

where *L* is a standard class, *W* is a classification parameter, the method also considered unlabeled data suitable for semi-supervised learning. Jiang [21] et al based on this basis, add amark label consistent constraint. Consistency mark approximate an optimal determination of sparse coding. Referring to the dictionary A, a subset A_i is generated by the class i samples. Referring to a dictionary, a subset of Class I R is generated by samples, and samples belongs to the class i sparse decomposition on A, only the selected atoms in A. For example, a dictionary atoms a_1 and a_2 belong to the class c_1, atom a_3 belongs to class c_2, then for samples y_1, y_2 in c_1, consistency annotation of samples y_3 in c_2 are represented by the matrix *Q*.

$$Q\equiv\begin{vmatrix}1&1&0\\1&1&0\\0&0&1\end{vmatrix}, \quad (11)$$

So the label consistency constraint is expressed as:

$$\Omega(X)=\left\|Q-RX\right\|_F^2, \quad (12)$$

where *R* is a linear transformation, responsible for the transformation of the original sparse coding to the most discriminate sparse space R^k, and *k* is as the dictionary capacity.

4.2 MULTI VIEW OBJECT RECOGNITION BASED ON METASAMPLES

Generally speaking, data metasample defines the linear element combination and can capture the essential structure of the data. Then, the different perspectives of the same data can be viewed as a linear combination of metasample. Mathematically, the data matrix *A* with the same goal of multi view is decomposed into two matrices: A~CD. Matrix *A* represents the multi view data set *m*·*n*, and each column is data of a view, each row represents the features of the data. Matrix *W* represents M·P, where each column is defined as a metasample. Matrix *H* represents *N*·*P*, the coefficients of each row represent expression pattern of the metasample, shown in Figure 1.

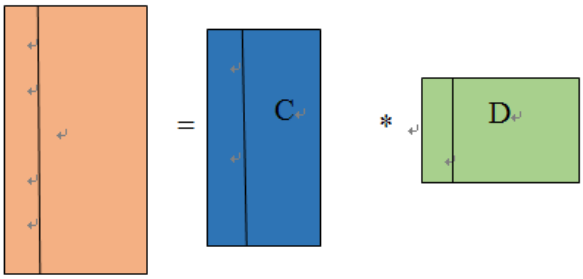

FIGURE 1 Metasample model of multi-view data

Due to the large number of shared features from the different perspective of the same object they share a common public dictionary. I.e. in the same dictionary atom there is not a factor of 0. However, because of the different angles, in public dictionary, sparse representation coefficients are not the same, as is shown in Figure 2. Therefore, the multi view sample is sparse.

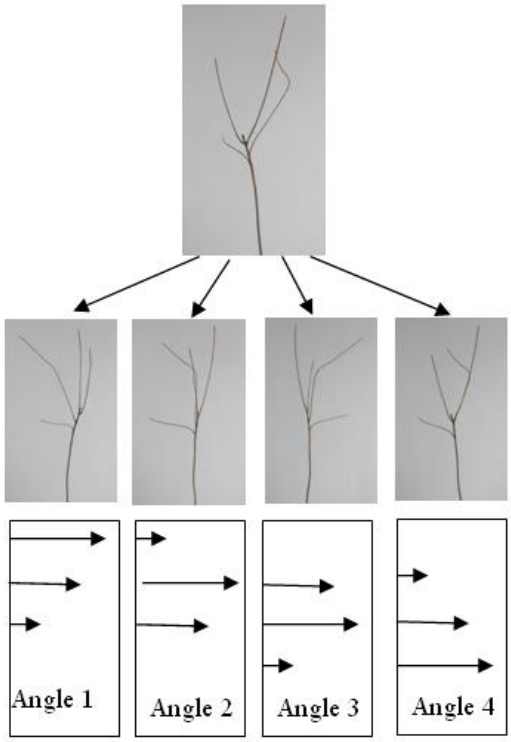

FIGURE 2 Sparse representation of multi-view data

5 Experimental results and analysis

5.1 BRANCH TEST

We make use of a lot of pictures of fruit trees in different illumination, perspective, involving samples of 4 sequences, only consider 7 different angle, θ={0°, 15°, 30°, 45°, 60°, 75°, 90°}.

The supervised dictionary learning algorithm in this paper with 4 machine learning methods are compared, the results are as follows: support vector machine (SVM), element samples and support vector machine (META+SVM), linear discriminant analysis (LDA), sparse representation classifier (SRC). For the META+SVM, refers to the proposed element samples from multiple perspectives in the sample, and then used as training samples for SVM classification of test samples.

TABLE 1 Recognition rate of fruit branch

Method	Recognition (%)
SVM	78.5
SRC	75.2
LDA	82.7
MSRC	81.0
Supervised dictionary learning algorithm	89.7

From the data in Table 1 shows: compared with other methods, the algorithm the author proposed has higher rate. The experimental results show the effectiveness of the method in multi view object recognition, and can accurately recognition multi view object.

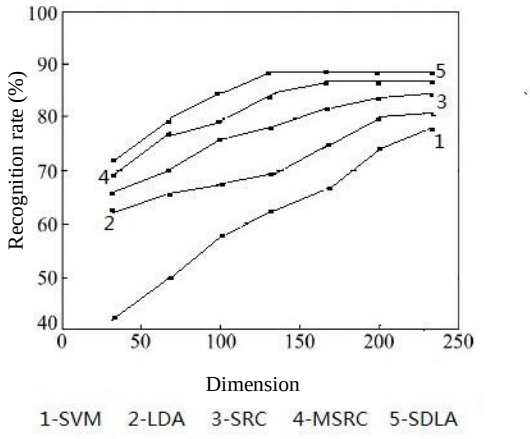

FIGURE 3 Recognition results under the condition of different dimension

In order to analyze the influence of the recognition results' dimensionality, the experiments are carried out on different feature dimension conditions, as is shown in Figure 3. The feature dimension increases at the same time, but in any dimension, the recognition rate proposed by the author is superior to other methods of recognition rate.

5.2 Belgium TSC database

Belgium TSC database is a subset of Belgium traffic sign benchmark set, including a plurality of traffic sign image of Belgium on the street, which contains 62 types of traffic signs, each have at least 10 different perspective image with different backgrounds, whose perspectives of this database is not fixed. In this experiment, 4281 images were selected as training samples, 3987 images as test sample. Each picture was cut into 40×40 in size, and the gray value feature was adopted, so that each picture had a 1600 dimensional vector. Table 2 gives the recognition results.

TABLE 2 Recognition rate of fruit branch

Method	Recognition (%)
SVM	85.2
SRC	83.1
LDA	81.7
MSRC	88.5
Supervised dictionary learning algorithm	96.7

Acknowledgment

This work was supported by Scientific research fund of Hebei Education Department (QN20131151) and by the science, technology research and development projects of Baoding in 2012(12ZG003).

References

[1] Candès E J, Wakin M B 2008 An introduction to compressive sampling *Signal Processing Magazine* **25**(2) 21-30
[2] Zeng F, Zhang G, Jiang J 2013 Text Image with Complex Background Filtering Method Based on Harris Corner-point Detection *Journal of Software* **8**(8) 1827-34
[3] Tan J, Zhang Y 2013 *Research on Multi-Sensor Multi-Target Tracking Algorithm Journal of Networks* **8**(11) 2527-33
[4] Mallat S G, Zhang Z 1993 *IEEE Transactions on Signal Processing* **41**(12) 3397-415
[5] Candès E J, Romberg J, Tao T 2006 *IEEE Transactions on Information Theory* **52**(2) 489-509
[6] Donoho D L 2006 *IEEE Transactions on Information Theory* **52**(4) 1289-306
[7] Candes E J, Tao T 2006 *IEEE Transactions on Information Theory* **52**(12) 5406-25
[8] Donoho D L, Elad M 2003 Optimally sparse representation in general (nonorthogonal) dictionaries via *ℓ1* minimization *Proceedings of the National Academy of Sciences* **100**(5) 2197-2202
[9] Candes E J, Romberg J K, Tao T 2006 Stable signal recovery from incomplete and inaccurate measurements *Communications on pure and applied mathematics* **59**(8) 1207-23
[10] Candes E J, Tao T 2005 *IEEE Transactions on Information Theory* **51**(12) 4203-15
[11] Baraniuk R, Davenport M, DeVore R, Wakin M 2008 A simple proof of the restricted isometry property for random matrices *Constructive Approximation* 28(3) 253-63
[12] Xu Z 2011 *Deterministic sampling of sparse trigonometric polynomials Journal of Complexity* 27(2) 133-40
[13] Yang J, Zhang *Y* Alternating direction algorithms for *ell_1*-problems in compressive sensing *SIAM Journal on Scientific Computing* 33(1) 250-78
[14] Needell D, Tropp. J A 2009 CoSaMP: Iterative signal recovery from incomplete and inaccurate samples *Applied and Computational Harmonic Analysis* 26(3) 301-21
[15] Mairal J, Bach F, Ponce J, Sapro G 2009 Non-local sparse models for image restoration *2009 IEEE 12th International Conference on Computer Vision* 2272-9
[16] Jiang Z, Lin Z, Davis L S 2011 Learning a discriminative dictionary for sparse coding via label consistent K-SVD *2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2011)*
[17] Boureau Y, Ponce J, LeCun Y 2010 A theoretical analysis of feature pooling in visual recognition *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*
[18] Wagner A. Wright J, Ganesh A, Zhou Z, Ma Y 2012 *IEEE Transactions on Pattern Analysis and Machine Intelligence* **34**(2) 372-86
[19] Bach F, Jenatton R, Mairal J, Obozinski G 2012 Structured sparsity through convex optimization *Statistical Science* **27**(4) 450-68
[20] Obozinski G, Taskar B, Jordan M I 2010 Joint covariate selection and joint subspace selection for multiple classification problems *Statistics and Computing* 20(2) 231-52
[21] Bach F 2008 Consistency of the group lasso and multiple kernel learning *The Journal of Machine Learning Research* **9** 1179-225
[22] Elad M, Figueiredo M A T, Ma Y 2010 On the role of sparse and redundant representations in image processing *Proceedings of the IEEE* **98**(6) 972-82
[23] Yang J, Wrighy J, Huang T S, Ma Y 2010 *IEEE Transactions on Image Processing* **19**(11) 2861-73
[24] Xu H, Caramanis C, Mannor S 2012 *IEEE Transactions on Pattern Analysis and Machine Intelligence* **34**(1) 187-93
[25] Jenatton R, Gribonval R, Bach F 2012 Local stability and robustness of sparse dictionary learning in the presence of noise *arXiv preprint arXiv:1210.0685*

Authors

Jin-jin Cai, November 1981, Hebei, China.

Current position, grades: lecturer at the College of Mechanical & Electrical Engineering, Agricultural University of Hebei. PhD student at the Agricultural University of Hebei.
University studies: master degree in Agricultural Mechanization Engineering at the Agriculture University of Hebei in 2008.
Scientific interests: automatic control technology, information technology, sparse representation in image processing.
Publications: 6.

Bo Liu, February 1981, Hebei, China.

Current position, grades: lecturer at the College of Information Science and Technology, Agricultural University of Hebei. PhD student at Beijing Jiaotong University.
University studies: bachelor's master's and degree in computer science and technology at Hebei University in 2006.
Scientific interests: subspace learning, semi-supervised learning, sparse representation.
Publications: 4.

Wei Yao, July 1981, Hebei, China.

Current position, grades: Lecturer at the College of Information Science and Technology, Agricultural University of Hebei. PhD student at the Agricultural University of Hebei.
University studies: master's degree in computer application technology at the Agriculture University of Hebei in 2007.
Scientific interests: image processing, artificial intelligence.
Publications: 9.