High-performance Computing and Application of Zero-norm

LiYing Lang

Hebei University of Engineering, Handan,056038，China
Email:langliying@126.com

XueKe Jing

College of Information and Electrical Engineering, Hebei University of Engineering, Handan,056038，China
Email:jingxueke8681@126.com

Abstract--**Whether sparseness can be effectively controlled is one of the key elements to measure the merits of the sparse coding algorithm. One-norm is primarily used in the sparse coding algorithm to control its sparseness currently, as well as by sparse approximation to control the sparseness of sparse coding model, but all these methods have led to slow convergence and low efficiency ultimately. In order to enhance the effectiveness of sparse coding algorithms, this paper selects zero-norm to control sparseness of the sparse coding model, and calculate after continuously extended at the discontinuous point of the model. We propose a highly efficient zero-norm sparse coding algorithm. This paper not only theoretically proves feasibility and efficiency of the algorithms which is capable of effectively controlling model sparseness, but also verifies the theoretical correctness of inference through experiments. These prove the operational efficiency of the algorithm is more efficient and stronger than existing algorithms.**

Index Terms--**sparse matrix, non-negative sparse coding, sparseness, zero-norm, one-norm**

I. INTRODUCTION

Sparse coding theory comes from the studies of optic network, which can simulate the receptive fields of simple cells in the primary visual cortex V1 area, is a method on artificial neural network.

Therefore, many scholars proposed some algorithm models on sparse code, and the method was successfully applied in pattern recognition, data compression, data denoising, etc. [1].

However, sparse coding model can be better used in some areas, but the building and solving of sparse coding model restrict the development of the sparse coding. After Lee and Seung proposed the non-negative matrix factorization [2], some scholars proposed a sparse coding algorithm based on one norm. But this model is difficult to control the sparseness of sparse matrix. Later, Hoyer proposed non-negative sparse coding algorithms (NMFs) [3-4], this algorithm has been widely concerned by people. Although the sparseness of

the algorithm can be controlled effectively by optimal sparse approximation, the algorithm has slow convergence.

This paper attempts use another sparse coding model to solve, the zero norm be used to build model. Although the model is discontinuous function, we can to solve the problem after continuous development in the discontinuity point.The experimental results show that our algorithm can greatly improve convergence rate on the basis of the sparseness of sparse matrix can be controlled effectively.

II. SPARSE CODING MODEL

The standard sparse coding was proposed by Olshausen and Field in 1996. Given a data matrix V of size $M \times N$, find the sparse matrices W and H of sizes $M \times L$ and $L \times N$, where $M \times N >> (M+N) \times L$. The sparse model defined as:

$$\min f(W,H) + \lambda g(H) \tag{1}$$

Where $f(W,H)$ is the reconstruction error function between V and WH, and the $g(H)$ is the sparse function of H. The commonly used model of sparse coding defined as:

$$\min \quad \|V - WH\|_F^2 + 2\lambda \sum_{i,j} \|H_{i,j}\|_0 \tag{2}$$

$$\|H_{ij}\|_0 = \begin{cases} 0 & H_{ij} = 0 \\ 1 & H_{ij} \neq 0 \end{cases}$$

Where , this is the intuitive description of the sparseness of the coefficient matrix. But the standard sparse model is discontinuous function. Mathematical method is more difficult to solve directly. So scholars turn to solve another kind of model, the Zibulevsky[5] proposed a model can write as:

$$\min \quad \|V - WH\|_F^2 + 2\lambda \sum_{i,j} |H_{i,j}| \tag{3}$$

The disadvantage of this model is that it will be equal to the original sparse coding model when the coefficient matrix is very sparse, and the model is not a continuous function, so model to solve difficultly remain unresolved. The model uses one norm to control the sparse degree of

the sparse matrix, but the sparse degree is difficulty to control, this model will Lead to the eventual trade-off can not be effectively minimize the reconstruction error and maximization of matrix sparseness.

Hoyer proposed the NMFs[6-7]the model defined as:

$$\min \quad \|V - WH\|_F^2$$
$$s.t. \quad spareness(W_{.i}) = S_W$$
$$spareness(H_{i.}) = S_H \qquad (4)$$

Although the model can solve the sparse control problem by optimal sparse approximation, it makes slow convergence.

III. ZERO-NORM SPARSE CODING (ZNSC)

Standard sparse coding model has the characteristics of discontinuous functions, so the solution is difficulty. We used one norm to describe the sparseness of sparse matrix, which make model is defective in mathematics construction. So in this paper, we try to use zero norm instead of one norm controlling the sparse degree to implement sparse coding algorithm, in order to illustrate the use of zero norm to solve which can better control the sparseness of sparse matrix.

We suppose matrix $A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$ and $B = \begin{bmatrix} 7 & 0 \\ 0 & 0 \end{bmatrix}$,

so $\sum_{i,j}\|A_{ij}\|_1 = 4, \sum_{i,j}\|B_{ij}\|_1 = 7$. The result is $\sum_{i,j}\|A_{ij}\|_1 < \sum_{i,j}\|B_{ij}\|_1$, but because the sparseness of matrix A less than B, so it is contradict that the one norm of the smaller matrix should obtain great sparseness. However, when we use the zero norm, $\sum_{i,j}\|A_{ij}\|_0 = 4, \sum_{i,j}\|B_{ij}\|_0 = 1$, thus $\sum_{i,j}\|A_{ij}\|_0 > \sum_{i,j}\|B_{ij}\|_0$, This accurate measure the situation of two sparse matrices. If $\sum_{i,j}\|A_{ij}\|_0 > \sum_{i,j}\|B_{ij}\|_0$, it shows that the number of nonzero in the matrix A is greater than that of B. So the sparseness of matrix A is less than that of B. This is consistent with the model

Since the original sparse coding model is discontinuous function, so it is difficulty to solve using the optimization algorithm directly. In this paper, we can to solve the problem after continuous development in the discontinuity point, then linear comparing to complete a sparse matrix solution.

In order to demonstrate this method, we first prove Lemma.

Lemma: if $\dfrac{dg(x)}{dx} = 0$ has a unique solution in the real number field R and $g(x)$ has a minimum, then

function $f(x) = g(x) + c\|x\|_0$ which obtain the minimum in $\begin{cases} x_0 & g(x_0) + c\|x_0\|_0 < g(0) \\ 0 & g(x_0) + c\|x_0\|_0 \geq g(0) \end{cases}$,

Where x_0 is the solution of $\dfrac{dg(x)}{dx} = 0$ and $\|x\|_0 = \begin{cases} 1 & x \neq 0 \\ 0 & x = 0 \end{cases}$

Prove: Let $f(x)$ continuous development when $x=0$, then set $f'(x) = g(x) + c$. Because $g(x)$ Differentiable and has a unique minimum In the real number field R, so $f'(x)$ differentiable and has a unique minimum, the minimum is the solution of $\dfrac{df'(x)}{dx} = 0$ $\dfrac{dg(x)}{dx} = 0$ has a unique minimum, set $x= x_0$.so $\dfrac{df'(x)}{dx} = \dfrac{dg(x)}{dx}$,then $f'(x)$ obtain the minimum in x_0

When $x = 0$, the value of function $f(x)$ and $f'(x)$ is different. So we can only compare the value of $f(0)$ and $f'(x_0)$, whichever is the minimum of function $f(x)$ of two minimum, we can be seen from Fig 1,the minimum corresponding point is the minima. If $x_0 \neq 0$, set $f'(x_0) = g(x_0) + c\|x_0\|_0$,and $f(0) = g(0)$,the conclusion can be obtained. If $x_0 = 0$, no matter $g(x_0) + c\|x_0\|_0 < g(0)$ or $g(x_0) + c\|x_0\|_0 \geq g(0)$ the minimum corresponding to the point is 0.this satisfy the conclusion of lemma. Proof is completed.
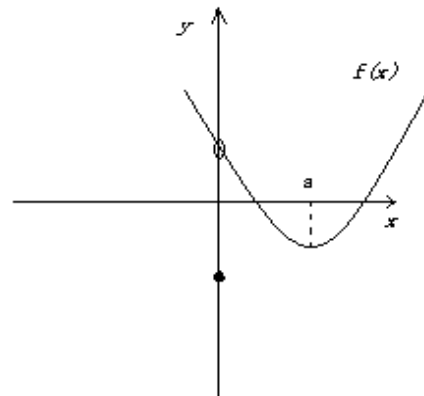

Figure 1. Function $f(x)$

We can be drawn from lemma that if we want to solve the most value of the discontinuous function which only has finite discontinuity point.First, we can do continuous development in the break point, and obtain the most

value of function. Then compare the most value with the value in the break point.

At last, we can obtain the most value.

Theorem $\forall k \in \{1,2,\dots, L\}$, set $F_{ab} = [W^TW]_{ab}(1-\delta_{ak}\delta_{kb})$, $i = k$, then $\delta_{ik} = 1$, otherwise $\delta_{ik} = 0$, set $HMIN = \dfrac{[W^TV]_{k\cdot} - F_{k\cdot}H}{[W^TW]_{kk}}$

and $f_H(H_{kl}) = H_{kl}^2[W^TW]_{kk} + 2\lambda\|H_{kl}\|_0 + 2H_{kl}(\sum\limits_{q\neq k}[W^TW]_{kq}H_{ql} - [V^TW]_{lk})$, the minimum of the objective function $\|V - WH\|_F^2 + 2\lambda\sum\limits_{i,j}\|H_{i,j}\|_0$ will

obtain in

$Hnew_l = \begin{cases} HMIN_l & f_H(HMIN_l) < f_H(0) \\ 0 & f_H(HMIN_l) \geq f_H(0) \end{cases}$

Prove:

$\|V - WH\|_F^2 + 2\lambda\sum\limits_{i,j}\|H_{i,j}\|_0$

$= tr\{(V-WH)(V-WH)^T\} + 2\lambda\sum\limits_{i,j}\|H_{i,j}\|_0$

$= tr\{VV^T - 2HV^TW + H^TW^TWH\} +$

$= 2\lambda\sum\limits_{i,j}\|H_{i,j}\|_0 = tr\{VV^T\} - 2\sum\limits_j[V^TW]_{j\cdot}H_{\cdot j} +$

$\sum\limits_j H_{\cdot j}^TW^TWH_{\cdot j} + 2\lambda\sum\limits_{i,j}\|H_{i,j}\|_0$

$\forall l \in \{1,2,\dots,N\}$, fixed the W, then the objective function of $H_{\cdot l}$ can be simplified :

$f_H(H_{\cdot l}) = H_{\cdot l}^TW^TWH_{\cdot l} - 2[V^TW]_{l\cdot}H_{\cdot l} + 2\lambda\sum\limits_i\|H_{il}\|_0$

$= \sum\limits_{m,n}H_{ml}[W^TW]_{mn}H_{nl} - 2\sum\limits_p[V^TW]_{lp}H_{pl} + 2\lambda\sum\limits_i\|H_{il}\|_0$

Further, $\forall k \in \{1,2,\dots,L\}$, fixed the W and $\{H_1, H_2,\dots,H_{k-1}, H_{k+1},\dots,H_{L\cdot}\}$, then the objective function of H_{kl} [8] can be simplified：

$f_H(H_{kl}) = H_{kl}^2[W^TW]_{kk} + 2\lambda\|H_{kl}\|_0 + 2H_{kl}(\sum\limits_{q\neq k}[W^TW]_{kq}H_{ql} - [V^TW]_{lk})$

(5)

According to (5) we can know that $f_H(H_{k1}), f_H(H_{k2}),\dots f_H(H_{kN})$ have no associated , and fixed W and $\{H_1, H_2,\dots,H_{k-1}, H_{k+1},\dots,H_{L\cdot}\}$, then the objective function of $H_{k\cdot}$ can be simplified :

$f_H(H_k) = f_H(H_{k1}) + f_H(H_{k2}) + \dots + f_H(H_{kN})$

$= H_k\{[W^TW]_{kk}I\}H_k^T - 2([W^TV]_{k\cdot} - F_{k\cdot}H)H_{k\cdot}^T$

$2\lambda\|H_{k\cdot}\|_0$

（6）

Following，we prove the (5) can be solved by the lemma, The key is to proof the $\dfrac{dg(x)}{dx} = 0$ has the only solution in the real field R. Set

$g(x) = H_{kl}^2[W^TW]_{kk} + 2H_{kl}(\sum\limits_{q\neq k}W^TW]_{kq}H_{ql} - [V^TW]_{lk})$ then

$\dfrac{dg(x)}{dx} = 2H_{kl}[W^TW]_{kk} + 2(\sum\limits_{q\neq k}[W^TW]_{kq}H_{ql} - [V^TW]_{lk})$

,we can obtain the only solution of $\dfrac{dg(x)}{dx} = 0$ is

$\dfrac{[V^TW]_{lk} - \sum\limits_{q\neq k}[W^TW]_{kq}H_{ql}}{[W^TW]_{kk}}$. We can know from the

lemma, the minimum of (5) can be calculated by continuous development. Because $f_H(H_{k1}), f_H(H_{k2}),\dots,f_H(H_{kN})$ have no association, so we can operate continuous development for (6) to obtain the most.

Then operate continuous development for the function $f_H(H_{k\cdot})$ when $H_{k\cdot} = 0$. $\forall l$, set $\|H_{kl}\|_0 = 1$, then $\|H_{k\cdot}\|_0 = N$ and

$f_H'(H_k) = H_{k\cdot}\{[W^TW]_{kk}I\}H_{k\cdot}^T + 2\lambda N - 2([W^TV]_{k\cdot} - F_{k\cdot}H)H_K^T$

set $f_H'(H_{k\cdot}) = 0$ to obtain the minimum when

$HMIN = \dfrac{[W^TV]_{k\cdot} - F_{k\cdot}H}{[W^TW]_{kk}}$, because among $f_H(H_{k1}), f_H(H_{k2}),\dots,f_H(H_{kN})$ have no association, so $\forall l$, according to the lemma we can obtain the minimum of the $f_H(H_{kl})$ at

$\begin{cases} HMIN_l & f_H(HMIN_l) < f_H(0) \\ 0 & f_H(HMIN_l) \geq f_H(0) \end{cases}$

Proof is completed.

From the above theorem we can know that it is practical to solve sparse coding by zero norm, and the problem of control sparseness can be solved effectively.

In this paper, we use stochastic gradient method to solve the matrix W.ZNSC can be summarized as:

1.Random initialization matrix $W(0)$, $H(0)$, $s \leftarrow 0$;

2.Repeat the following operation until you meet the conditions so far;

3. $W(s+1) \leftarrow W(s) - \mu[W(s)H(s) - V]H^T(s)$;

4. $w_i \leftarrow w_i / \|w_i\|_2$

5. $k = 1 \rightarrow L$

(1). $F_{ab}(s) = [W^T(s+1)W(s+1)]_{ab}(1 - \delta_{ak}\delta_{kb})$

(2).Calculate

HMIN =

$$\frac{[W^T(s+1)V]_{k\cdot} - F_{k\cdot}(k)[H_{1\cdot}^T(s+1),\cdots;H_{k-1\cdot}^T(s+1),H_{k\cdot}^T(s),\cdots H_{L\cdot}^T(s)]^T}{[W^TW]_{kk}}$$

(3).

$$\forall l \quad H_{kl} = \begin{cases} \text{HMIN}_l & f_H(\text{HMIN}_l) < f_H(0) \\ 0 & f_H(\text{HMIN}_l) \geq f_H(0) \end{cases}$$

6. $s \leftarrow s+1$

IV. ZNSC ALGORITHM ANALYSIS

In order to illustrate the effectiveness of the algorithm, we compare that with the NMFs and NSC algorithm proposed by Hoyer.

Algorithm: NMF with sparseness constraints [7]

1. Initialize W and H to random positive matrices

2. If sparseness constraints on W apply, then project each column of W to be non-negative, have unchanged $L2$ norm, but $L1$ norm set to achieve desired sparseness

3. If sparseness constraints on H apply, then project each row of H to be non-negative, have unit $L2$ norm, and $L1$ norm set to achieve desired sparseness

4. Iterate

(a) If sparseness constraints on W apply,

i. Set $W := W - \mu_W(WH-V)H^T$

ii. Project each column of W to be non-negative, have unchanged $L2$ norm, but $L1$norm set to achieve desired sparseness

Else take standard multiplicative step $W := W \otimes (VH^T) \oslash (WHH^T)$

(b) If sparseness constraints on H apply,

i. Set $H := H - \mu_H W^T(WH-V)$

ii. Project each row of H to be non-negative, have unit $L2$ norm, and $L1$ norm set to achieve desired sparseness;

Else take standard multiplicative step $H := H \otimes (W^TV) \oslash (W^TWH)$

Above, $\otimes$ and $\oslash$ denote element wise multiplication and division, respectively.

Algorithm: NSC [9]

1.random initialization matrix $W(0) \geq 0$, $H(0) \geq 0$, $\forall i$, $W_{\bullet i} = W_{\bullet i}/\|W_{\bullet i}\|_2$, $s \leftarrow 0$;

2. Repeat the following operation until you meet the conditions so far;

3. $W \leftarrow W - \mu(WH-V)H^T$

4. $H \leftarrow \max\{H,0\}$

5. $W \leftarrow W \otimes [VH^T] \oslash [WHH^T]$

6. $\forall i$, $W_{\bullet i} = W_{\bullet i}/\|W_{\bullet i}\|_2$

7. $s \leftarrow s+1$

The main difference between ZNSC algorithm and NMFs algorithm are to solve sparse matrix. For ZNSC algorithm, when solve the sparse matrix, each column of sparse matrix is ordered , and no cycle to determine, however the NMFs algorithm should need to go through the cycle many times and it is not linear, slow convergence rate. Thus, the convergence rate of ZNSC algorithm is superior to the NMFs or NSC algorithm.

V. EXPERIMENT AND ANALYSIS

The convergence speed can not be reflected accurately through comparing the size of each model function value, because of the difference between ZNSC and NMFs algorithm. so the two algorithm can be as the feature extraction algorithm in face recognition, then we can compare their recognition rate by classification algorithm(PSVM)[10], when the recognition rate reaches stability state, the solution of the sparse model can reach convergence state. Therefore we can judge the algorithm's convergence speed through the time of the recognition rate reaches stability state.

In this paper, we use the ZNSC algorithm, NSC algorithm and NMFs algorithm to do feature extraction on the ORL face database, the proximal support vector machine (PSVM)[11] is used as classifier. The first image of each of the forty kinds of face in ORL face database are chosen respectively to be trained as training samples and we can calculate the basis matrix, where L = 40, L is the column length of coefficient matrix H, and we select the iteration step μ=10-9 in the ZNSC algorithm and NSC algorithm, and λ =100 in the NSC algorithm, the unit of iteration time is second.

▪A. The Relationship between Regulatory Factors and Recognition Rate

Because of using the zero norm, the value of λ should be determined by the actual situation, in this paper, we choose different λ , and observe the change of recognition rate with increasing the iterate time, this result is shown in Fig. 2.

Fig.2 shows that the difference between the recognition rate is not large when the magnitude of λ is 10^3-10^4,however when the λ equals 5×10^5,the recognition rate begin to decline. From (1), we can know that when the value of λ is bigger, the second item becomes major factor, which lead to increased reconstruction error between V and WH. So the value of λ should not choose a bigger one.

▪B. The Relationship between Regulatory Factors and Sparseness

The sparseness can be defined as:

$$sp_0(x) = \frac{n - \|x\|_0}{n-1}$$

(7)

Where x may be a scalar, vector, matrix. If x is scalar ,$\|x\|_0$ show whether x is zero or not. If x is vector or matrix, $\|x\|_0$ show the number of nonzero.
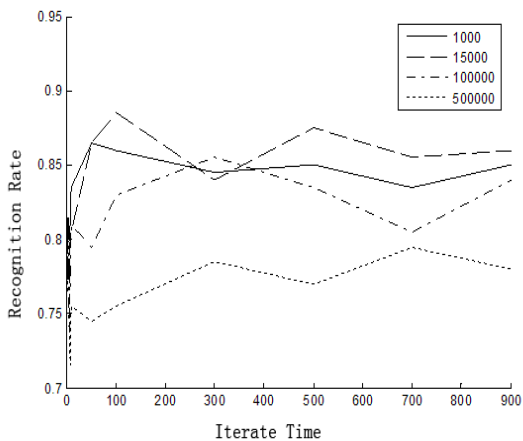
Figure 2. Recognition rate in different λ with increasing the iterate time

Next, observe the change of sparseness with increasing the iterate time. This result is shown in Fig.3 .

The Fig. 3 shows that sparseness is increased with the increase of λ, when λ is greater than 15000,the sparstiy is more than 0.5. This can meet the general sparse requirements. In addition, the function curve is a straight line basically, this illustrates iteration time has little influence on the sparseness. The sparseness can stabilize in the short period of time.

C. The Efficiency of the Algorithm

The experimental conditions are that we set λ =15000 and other parameters are same as the above experiment. We will compare the recognition rate of three algorithms. The result is shown in Fig. 4.

From Fig.4 we can see that three algorithm's recognition isn't monotone rising with the increase of iterative time, because initial matrix is selected randomly when iteration is operated every time, at last which lead to the result of convergence is different. Overall, the recognition rate can reach up to 85% within tens of seconds, and recognition rate shows a steady trend with the increase of iterative time. This shows that ZNSC algorithm can achieve a stable recognition rate in a few seconds. But the recognition rate is not good by NMFs algorithm. we can see that the recognition rate only reaches 80 percent when the iterative time in 300 seconds, this shows that its convergence speed is relatively slow.

Although the iteration time of NSC is shorter than that of NMFs, it is also more time-consuming than ZNSC algorithms.

The above has stated that ZNSC algorithm is better than NMFs on the recognition rate, and ZNSC algorithm can reach the desired results in a short time. We can see from Fig.3 the ZNSC algorithm's iteration reach the stable value at beginning, it shows that the algorithm of this paper reached the expected sparse degree at first, that is to say the algorithm in this paper indeed achieve the ideal convergence state in a shorter time.
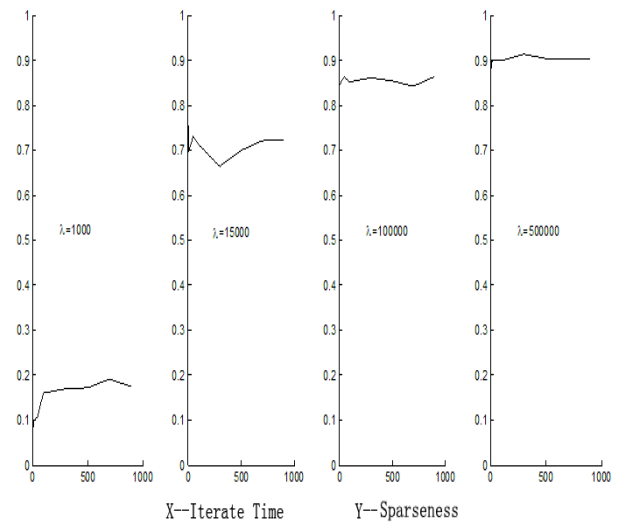

Figure 3. Sparseness of the sparse matrix in different λ with increasing the iterate time
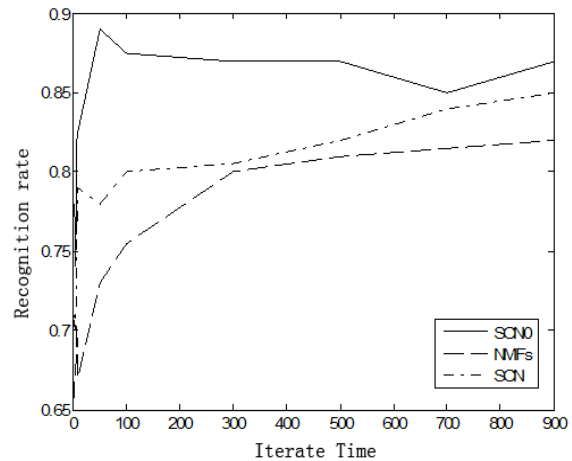

Figure 4. Recognition rate of the three algorithms with different iterate time

VI. CONCLUSION

This paper can solve controlling problem of sparse degrees effectively by using zero norm measure the sparseness of sparseness matrix to solve the problem of sparse coding. Firstly this model just do one addition, subtraction, multiplication and division method in solving each column of sparse matrix, then get the results by doing a judgment for each value of column vector in parallel, but all the columns of the matrix is linear, no cycle computing. This shows that this algorithm is more effective in speed to solve sparse matrix comparing with NMFs algorithm, and ideal sparse matrix can be taken in a short period.

Although the algorithm spends a short operation time but the adjustment factor λ , iteration step length μ and the iterative initial matrix W(0) and H(0) in the model are of no unified standard for selecting, so these need to be further study.

REFERENCES

[1] Garrigues, P., Olshausen, B.. Learning horizontal connections in a sparse coding model of natural images[J]. Advances in Neural Information Processing Systems, 2008,pp. 505–512.

[2] Ouhsain M,Hamza A B.Image watermarking scheme us-ing nonnegative matrix factorization and wavelet transform.Expert Systems with Applications, pp.2123-2129,2009.

[3] Fevotte, N. Bertin, and J. L. Durrieu. Nonnegative matrix factorization with the itakurasaito divergence: With application to music analysis. Neural Computation, PP.793–830,2009.

[4] J Mairal, F Bach, J Ponce,G Sapiro.Online Learning for Matrix Factorization and Sparse Coding.Journal of Machine Learning Research,2010,pp.19-60.

[5] Zibulevsky M and Pearlmutter B A. Blind source separation by sparse decomposition in a signal dictionary. Neural computation,2001,pp.863-882.

[6] Hoyer P O．Modeling receptive fields with non-negative sparse coding[C]IEEE Workshop on Neural Networks for Signal Processing,Toulouse，France，2003, pp. 547-552.

[7] Patrik O.Hoyer．Non-negative Matrix Factorization with Sparseness Constrains[J]．Journal of Machine Learning Research, 2004,pp.1457-1469.

[8] LI Le, ZHANG Yu-Jin．a Stable and Efficient Algorithm for Nonnegative Sparse Coding[J]．Acat Automatica Sinica, 2009,pp.1257-1271.

[9]Hoyer P O. Non-negative sparse coding[C].//IEEE Workshop on Neural Networks for Signal Processing,Valais，Switzerland，2002,pp.557-565.

[10]N. Saravanan , K.I. Ramachandran.A case study on classification of features by fast single-shot multiclass PSVM using Morlet wavelet for fault diagnosis of spur bevel gear box,India,pp.10854-10862,2009.

[11]Liu Y, Zhang H H, Park C, et al. Support vector machines with adap tive Lq penalty,J.Comput Statist and Data Anal, pp. 6380-6394, 2007.

LiYing Lang HeBei Province, China. Birthdate: September, 1974.is Optical Engineering Doctor, graduated from Tianjin University of Optical Engineering. And research interests on Face recognition and image processing.

She is a associate professor of Dept. School of Information and Electrical Engineering, Hebei University of Engineering

XueKe Jing HeBei Province, China. is Face Recognition Master, graduated from Shandong University of Technology. And research interests on Face recognition and image processing.