A Novel Non-negative Shrinkage Technique Based on The Normal Inverse Gaussian Density Model

Li Shang[1,2], De-Shuang Huang[1], Tat-Ming Lok[3] and Michael R. Lyu[4]

[1] Hefei Institute of Intelligent Machines, Chinese Academy of Sciences, P. O. Box 1130, Hefei, Anhui 230031, China
{shangli, dshuang}@iim.ac.cn
[2] Automation Dept., University of Science and Technology of China, Hefei, Anhui 230026
tmlok@ie.cuhk.edu.hk
[3] Information Engineering Dept., The Chinese University of Hong Kong, Shatin, Hong Kong
[4] Computer Science & Engineering Dept., The Chinese University of Hong Kong, Shatin, Hong Kong
lyu@cse.cuhk.edu.hk

Abstract. This paper proposes a novel image denoising technique based on the normal inverse Gaussian (NIG) density model using an extended non-negative sparse coding (NNSC) algorithm. Here, we demonstrate that the NIG density provides a very good fitness to the non-negative sparse data. In denoising process, by exploiting a NIG-based maximum a posteriori estimator (MAP) of an image corrupted by additive Gaussian noise, the noise can be reduced successfully. This shrinkage technique, also referred to as NNSC shrinkage technique, is self-adaptive to the statistical properties of image data. The experimental results show that the NNSC shrinkage approach is indeed efficient and effective in image denoising. In addition, we also compare the effectiveness of the NNSC shrinkage method with methods of standard sparse coding shrinkage, wavelet-based shrinkage and the Wiener filter. The simulating results show that the NNSC shrinkage method indeed outperforms the three kinds of denoising approaches mentioned above.

1 Introduction

More recently, more and more data-adaptive image denoising techniques have been explored, such as the ones based on principal components analysis (PCA), independent component analysis (ICA) and sparse coding (SC) shrinkage, etc.. All of these methods can denoise an image successfully by using different skills. SC algorithm is appropriate to multi-dimension mixed data. Therefore, this technique has been used widely in image denoising field, especially by a "shrinkage" [1] technique. However, SC technique is unrealistic as a model of V1 simple-cell behavior [2]. Therefore, P. O. Hoyer [2] introduced the concept of non-negative sparse coding (NNSC), and it had been used to model successfully receptive fields of V1 in the mammalian primary visual cortex. Thus the non-negative property has caused to different representations and applications, such as image reconstruction, data compression, image denoising,

pattern recognition, and so on. In fact, the basic principle of sparse coding shrinkage is very simple. Small amplitude values, which are thought to originate from zero-valued components influenced by noise, are suppressed while large values are preserved. Generally speaking, it is very necessary for this technique to perform a parameterized probability density function (pdf) estimate for sparse components in the transform domain. It is well known that the Laplacian density is a classical sparse density with one-parameter [1], but, it cannot be used to model different degrees of kurtosis for a given variance. In addition, the other sparse models proposed by Hyvärinen in [1], referred to as mildly sparse and strongly sparse models, are also two-parameter, zero mean and symmetric models. The parameters are related to the second order moment, the expected absolute value, and the peak value of the density. However, a proper statistical model should be flexible enough to provide a good fitness to the data by modeling various degrees of sparseness, and taking into account a possible skewness. In addition, it should be possible for ones to estimate the model parameters readily from the noisy observation. So, in this paper, we exploit the recent normal inverse Gaussian (NIG) density, which is four-parameter model [2], to model the non-negative sparse components. The NIG density has the flexibility that makes it capable of satisfying the above requirements. In addition, we can use very fast cumulant based estimators to estimate the four parameters of the density [1]. In the symmetric case, this method can model data ranging from zero normalized kurtosis, i.e., the Gaussian distribution, to any positive valued kurtosis. Referring to the model of NNSC introduced by P. O. Hoyer's [2], we propose an extended NNSC algorithm based on the model of NIG density. In particular, using a maximum a posteriori (MAP) estimator (i.e. a shrinkage operator), we can denoise a noisy image successfully, which is sparsely coded and contaminated by additive Gaussian noise.

2   The Normal Inverse Gaussian Density (NIG) Model

The NIG density is a variance-mean mixture of a Gaussian density with an inverse Gaussian. A stochastic variable u can be said to be normal inverse Gaussian if it has a probability density of the following form [2]:

$$p(u) = \frac{\alpha\delta}{\pi} \cdot \frac{\exp[m(u)]}{q(u)} K_1[\alpha q(u)] \; . \tag{1}$$

where $K_1(u)$ is the modified Bessel function of the second kind with index 1 and subject to the constrain of $|u \to \infty|$, and is defined as:

$$K_1(u) = \sqrt{\frac{\pi}{2u}} \cdot \exp(-u) \; . \tag{2}$$

as well as, $m(u)$ and $q(u)$ are respectively defined as:

$$m(u) = \delta\sqrt{\alpha^2 - \beta^2} + \beta(u - \mu) \; . \tag{3}$$

2

$$q(u) = \sqrt{(u-\mu)^2 + \delta^2} \quad . \tag{4}$$

subject to the constrains: $0 \le |\beta| < \alpha$, $\delta > 0$, and $-\infty < \mu < \infty$.

According to the definition in Eqn. (1), the shape of the NIG density is specified by the four-parameter vector $[\alpha, \beta, \mu, \delta]^T$. The $\alpha-$parameter controls the steepness or pointiness of the density. A larger value of the parameter α implies lighter tails. The rightmost panel in Fig. 1 shows the dependency on α for $\beta = \mu = 0$, and $\delta = 1$. Distinctly, it is easy to see that the tails become heavier and heavier as the value of α decreases greatly. The $\beta-$parameter controls the skewness. For $\beta < 0$, the density is skewed to the left, for $\beta > 0$, the density is skewed to the right, while $\beta = 0$ implies a symmetric density (see the left panel in Fig. 1) around μ, which is a centrality parameter. The right panel in Fig. 1 also shows the dependency on the parameter β. It can be noted that the skewness increases as β increases. Lastly, the $\delta-$parameter is a scale-like parameter in the sense that the rescaled parameters $\alpha \to \alpha\delta$, and $\beta \to \beta\delta$ are invariant under location-scale changes of μ.
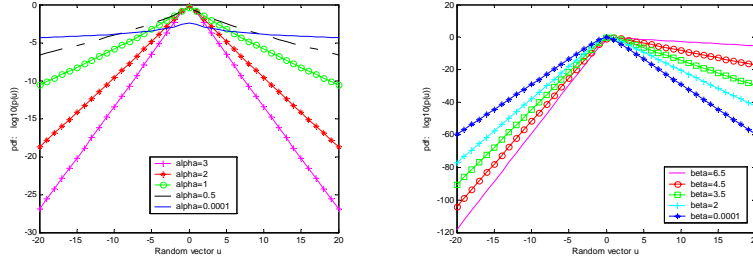

Fig. 1. NIG density (logarithmic scale) for different values of α and β. Left: α varies, and $\beta = \mu = 0$, $\delta = 1$. Right: β varies, and $\alpha = 7$, $\mu = 0$, $\delta = 1$

In [3], for the NIG parameter, Hanssen and Øigård derived a cumulant-based estimator. By estimating the first four lowest cumulants $k^{(1)}$, $k^{(2)}$, $k^{(3)}$, and $k^{(4)}$ from the sample data, and using the first cumulants to estimate the skewness $r_3 = k^{(3)} / [k^{(2)}]^{3/2}$ and normalized kurtosis $r_4 = k^{(4)} / [k^{(2)}]^2$, we can obtain the auxiliary variables:

$$\zeta = 3\left(r_4 - \frac{4}{3}r_3^2\right)^{-1}, \quad \rho = \frac{r_3}{3}\sqrt{\zeta} \quad . \tag{5}$$

Thereafter, the parameter estimators can easily be derived as follows [3]:

$$\delta = \sqrt{k^{(2)}\zeta(1-\rho^2)}, \quad \alpha = \frac{\zeta}{\delta\sqrt{1-\rho^2}}, \quad \beta = \alpha\rho, \quad \mu = k^{(1)} - \rho\sqrt{k^{(2)}\zeta} \quad . \tag{6}$$

3 The Extended NNSC Model And Algorithm

3.1 Modeling NNSC of Natural Images

Any natural image can be modeled as a linear superposition of some features a_i:

$$x = \sum_i^m a_i s_i + n \ . \tag{7}$$

where x denotes the natural image data, a_i are called basis vectors, s_i are mutually independent sparse variables, and n is Gaussian noise. The image model of NNSC is the same as that shown in Eqn. (7). The significant point is here that the input matrix X, basis vectors A and latent sparse coefficients S are non-negative in NNSC model. The fact that each unit s_i is either positively or negatively active means that every feature contributes to representing the stimuli of opposing polarity. This poses a contrast to the behavior of simple-cells in V1. Furthermore, V1 receives the visual data from the lateral geniculate nucleus (LGN) in the form of separated ON-channel and OFF-channel, and each channel's input data are positive.

3.2 The Cost Function and Updating Rules of NNSC

On the basis of the Hoyer's NNSC model [2], we propose an extended NNSC model. Here, we also use the minimum reconstruction error and the sparseness like Hoyer, but the prior distribution of the receptive field and the sparse shape of hidden components are also considered. Then, the cost function can be constructed as:

$$J(A,S) = \frac{1}{2} \sum_{x,y} \left[X(x,y) - \sum_i a_i(x,y)s_i \right]^2 + \lambda \sum_i f\left(\frac{s_i}{\sigma_i}\right) + \eta \sum_i a_i^T a_i \ . \tag{8}$$

subject to the constraints: $X(x,y) \geq 0$, $\lambda > 0$, $\eta > 0$, $\forall_i: a_i \geq 0$, $s_i \geq 0$, and $\|a_i\| = 1$. Where $\sigma_i^2 = \langle s_i^2 \rangle$, $X(x,y)$ denotes an image, a_i and s_i denotes respectively the *ith* column of A and the *ith* row of S, λ is the tradeoff between sparseness and accurate reconstruction, and η has to do with the variance of the prior distribution of a_i. Here, the sparse measure function $f(\cdot)$ is chosen as the form of the NIG density, as shown in Eqn. (1) (see Section 2). According to the estimations of the four-parameter vector $[\alpha, \beta, \delta, \mu]^T$, the function $f(\cdot)$ can be selected as definite function.

According to Eqn. (11), we can obtain the derivatives of $\dot{a}_i$ and $\dot{s}_i$, shown as follows:

$$\dot{s}_i = a_i^T(x,y)\left[X(x,y) - \sum_{i=1}^n a_i(x,y)s_i \right] - \frac{\lambda}{\sigma_i} f'\left(\frac{s_i}{\sigma_i}\right) = a_i^T e - \frac{\lambda}{\sigma_i} f'\left(\frac{s_i}{\sigma_i}\right) \ . \tag{9}$$

$$\dot{a}_i = \left[X\left(x,\,y\right) - \sum_{i=1}^{n} a_i\left(x,\,y\right) s_i \right] s_i^T - \gamma\, a_i\left(x,\,y\right) = e\, s_i^T - \gamma\, a_i \ . \tag{10}$$

where $\sigma_i = \sqrt{\langle s_i^2 \rangle}$, $e = X - AS$ is the residual error between the original image and the reconstructed image of this model. In experiment, we exploited conjugate gradient algorithm to update basis vectors A and Eqn. (9) to update S.

4   The Denoising Algorithm of NNSC Shrinkage

4.1   MAP Estimator and NIG Shrinkage Function

Now considering a single noisy component denoted by y, which can be written as:

$$y = s + v \ . \tag{11}$$

where $v \sim N\left(0, \sigma^2\right)$, s is the original non-Gaussian random variable, and v is the Gaussian noise of zero mean and variance σ^2. We want to estimate the original s given y by $\hat{s} = g\left(y\right)$. Denoting by $p\left(s\right)$ the density of s, and by $f\left(\hat{s}\right) = -\ln p\left(\hat{s}\right)$ the negative log-density of $\hat{s}$. For an unimodal, differentiable posteriori density, $\hat{s}$ can be obtained by solving the following equation:

$$\frac{\hat{s} - y}{\sigma^2} + f'\left(\hat{s}\right) = 0 \ . \tag{12}$$

where $f\left(\hat{s}\right)$ is assumed to be convex and differentiable, and $f'\left(\hat{s}\right) = d\left(f\left(\hat{s}\right)\right)/ds$ is the score function of $\hat{s}$. Then, the following first-order approximation of the MAP estimator (with respect to noise level) is always possible:

$$\hat{s}^* = y - \sigma^2 f'\left(y\right) \ . \tag{13}$$

where the problem with this estimator in Eqn. (13) is that the sign of $\hat{s}^*$ is different from the sign of y even for symmetrical zero-mean densities. Such counterintuitive estimates are possible because $f'\left(\cdot\right)$ is often discontinuous or even singular at 0, which implies that the first-order approximation is quite inaccurate near 0. To alleviate this problem of "overshrinkage", the following approximation to the MAP estimator of a nongaussian random variable corrupted by Gaussian noise may be applied:

$$\hat{s} = g\left(y\right) = sign\left(y\right) \max\left(0, |y| - \sigma^2 |f'\left(y\right)|\right) \ . \tag{14}$$

According to the NIG density model (see Eqn. (1)), the score function of the NIG density is found to be the following formula:

5

$$f'_{NIG}(u) = \frac{\alpha(u-\mu)}{q(u)}\left(\frac{K_0[\alpha q(u)]}{K_1[\alpha q(u)]} + \frac{2}{\alpha q(u)}\right) - \beta \; . \tag{15}$$

where $K_0(\cdot)$ is the Bessel function of the first kind with index 1. And it is clear to see that the form of $|f'_{NIG}(u)|$ depends on the four-parameter vector $[\alpha,\beta,\delta,\mu]^T$ estimated by the sample data. In fact, the function in (14) is a shrinkage function that reduces the absolute value of its argument by the score function $f'(u)$.

4.2 NNSC Shrinkage Rules Based on The NIG Density Model

The model of NNSC has the same transformation equation as linear sparse coding (SC), i.e., $S = WX$. Here, $X = (x_1, x_2, \ldots, x_n)^T$ denotes a $n-$dimensional set of noise-free random vectors, $S = (s_1, s_2, \ldots, s_m)^T$ denotes the $m-$dimensional hidden components ($m \le n$), and W is the weight matrix with the size of $m \times n$. The distinct difference between NNSC and SC is that X, S and W are all non-negative in NNSC, but they are all signed in SC. Each non-negative sparse independent component is input to the cumulant based NIG parameter estimator, which determines a very good fitness of the NIG density of the noise-free components, and we can calculate the corresponding shrinkage function. Here, the NNSC shrinkage algorithm is briefly summarized as follows:

1. Using a noise-free set of data Z that has the same statistical properties as the $n-$dimensional input data $\tilde{X}$, estimate the non-negative feature basis vectors $A1$ by our extended NNSC algorithm. In terms of $A1$, compute the basis vectors difference of ON-channel minus OFF-channel, denoted by A. Thus, the NNSC transformation matrix W can be found, which is the inverse or pseudoinverse of A, and it should be orthogonalized in practical.

2. For every $i = 1,2,\cdots,m$, estimate a NIG density model for the non-negative sparse components $s_i = w_i Z$, where w_i is the *ith* row of W. Determine the four-parameter vector $[\alpha,\beta,\delta,\mu]^T$ in terms of Eqns. (6) and find the corresponding NIG shrinkage function g_i according to Eqn. (14).

3. Observing a noisy version $\tilde{X}$, which has been beforehand centered and normalized in order to make $\tilde{X}$ have zero-mean and unit variance, compute the projections on the sparsifying basis by the transformation of $Y = W\tilde{X}$.

4. Appling the shrinkage operator g_i to every component y_i of Y, to obtain $\hat{s}_i = g_i(y_i)$, therefore, $\hat{S} = (\hat{s}_1, \hat{s}_2, \cdots, \hat{s}_m)$.

5. Do the inverse transformation to obtain estimates $\hat{X}$ of the noise-free data X, i.e., $\hat{X} = W^{-1}\hat{S} = W^T\hat{S}$.

5 Experimental Results

5.1 Application to Natural Image Data

All test images used in our experiment can be available on the Internet *http:// www.cis.hut.fi/projects/ica/data/images*. Firstly, selecting randomly 10 noise-free natural images with 256×512_pixels. Then, we sampled patches of 8×8_pixels 10000 times from each original image, and converted every patch into one column. Thus, the input data set X with the size of 64×100000 is acquired. Considering the non-negativity, we separate X into ON-channel and OFF-channel, denoted respectively by Y and Z. So, the non-negative matrix $I = (Y; Z)$ with the size of 2×64×100000 is obtained. And then, using the updating rules of A and S in turn, we minimized the objective function given in Eqn. (8).

5.2 Estimating the NIG density model and Shrinkage Function

Note that the NIG density is indeed suitable for the super-Gaussian data. For the purpose of illustrating how close the NIG density models the NNSC transformed data for the "grasshopper" image, the first non-negative sparse vector s_{1j} ($j = 1, 2, \cdots, T$. T is the sample number of images patches) was used. The estimated kurtosis of this sparse vector was $r_4 = 17.11$, the estimated skewness was $r_3 = 0.165$, and the estimated four parameters of the NIG density modeling the underlying probability density function of s_{1j} were found to be $\hat{\alpha} = 2.17$, $\hat{\beta} = 0.05$, $\hat{\delta} = 0.081$, and $\hat{\mu} = 0.022$. Furthermore, the NIG density model of s_{1j} calculated according to Eqn. (1) was:

$$p(s) = 0.0476 \ \frac{\exp(F1 - F2)}{F3} \ . \tag{16}$$

Where $F1$, $F2$ and $F3$ are respectively calculated as: $F1 = 0.1757 + 0.05(s - 0.022)$, $F2 = 2.17 \left[(s - 0.022)^2 + 0.0066 \right]^{1/2}$, and $F3 = \left[(s - 0.022)^2 + 0.0066 \right]^{3/4}$. The resulting NIG density of the shrinked sparse components of s_{1j} was shown in the left of Fig. 3 in a log-plot (solid line). It has a negligible skewness and is centered close to the origin. For comparison, the NIG density plots of noisy s_{1j} (the noisy level added is 0.5) and the noise-free s_{1j} were also shown in Fig. 2. It can be readily seen that the NIG density of the shrinked s_{1j} approaches highly to that of the noise-free s_{1j}. The noise has been reduced effectively, and the shrinked components are concentrated around zero to a much higher degree, compared to the noisy components. The shrinkage result also showed indirectly that the estimates $\hat{X}$ of the given image are very close to the original image data X.
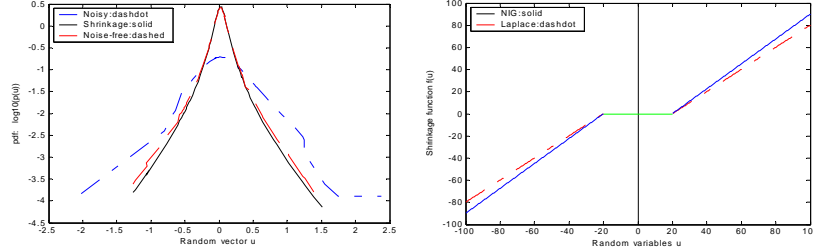
Fig. 2. Results of denoising noisy sparse components of s_{1j} by the NIG-based NNSC shrinkage function. Left: NIG density corresponding to the given sparse vector s_{1j}. Dashed: the noise-free s_{1j}. Dash-dotted: the noisy s_{1j}. Solid: the shrinked s_{1j}. Right: the in-out property of the shrinkage function. Solid: NIG shrinkage function. Dashed: Laplace shrinkage function

Based on the NIG pdf calculated in Eqn. (14), the shrinkage function for the noisy sparse components of s_{1j} can be calculated. It was shown in the right panel in Fig. 2 as the solid non-linearity. For comparison, we assumed that the classical Laplacian density, has modeled the given sparse data. This shrinkage function is given by

$$p(s) = sign(s) \max(0, |s| - \frac{\sqrt{2}\,\sigma^2}{d}) \;. \tag{17}$$

where d is the standard deviation of the density model. In this case, the two shrinkage functions have almost identical thresholds. But the large components are shrinked less by the NIG model than by the Laplacian model. The reason for this is that the estimated NIG density has heavier tails than the estimated Laplacian density.

5.3 The Result of Denoising

Here, the quality of denoised images is evaluated by the values of normalized SNR, which is defined as follows:

$$SNR_n = \frac{1}{MN} \frac{\sum\limits_{i=1}^{N} \sum\limits_{j=1}^{M} \left(X_{ij} - \overline{X}_{ij}\right)^2}{\sum\limits_{i=1}^{N} \sum\limits_{j=1}^{M} \left(X_{ij} - \hat{X}_{ij}\right)^2} \;. \tag{18}$$

where M and N denote the size of the image data, X denotes the input image data set, $\overline{X}$ denotes the mean value of X and $\hat{X}$ denotes the denoised image data. The calculated SNR_n value of the denoised image is 18.9864, and the SNR_n value of the noisy image is 1.1701. Clearly, the SNR_n value of the former is larger than that of

the later, which indicates that the visual effect has been enforced greatly and the noise has been effectively reduced. The denoised results of the noisy grashopper image, which were obtained by our algorithm, were shown in the rightmost panel in Fig. 3.

Fig. 3. Denoising experiment on the grashopper image with 256×512. Leftmost: the original image; Middle: the noisy image with the noise level: $\sigma = 0.5$; Rightmost: the denoised image obtained by the method of NIG-based NNSC shrinkage

Fig. 4. Comparison results of denoising obtained by different denoising algorithms. Leftmost: Wiener filtered; Middle: Wavelet-based soft shrinkage; Rightmost: Sparse coding shrinkage

Table 1. Values of normalized SNR obtained by different denoising algorithms

Algorithm	SNR_n of denoised images	SNR_n of the noise image ($\sigma = 0.5$)
Wiener filter	4.7728	
Wavelet-based soft shrinkage	5.6533	1.1701
Sparse coding shrinkage	11.3024	
NIG-based NNSC shrinkage	18.9864	

Furthermore, we compared our algorithm with other denoising methods: the usual Wiener filter, the wavelet-based soft shrinkage, and the standard SC shrinkage. As a result, the denoised images and the values of corresponding normalized SNR were respectively shown in Fig. 4 and Table 1. According to the experimental results, it can be concluded that our NIG-based NNSC shrinkage method is the best denoiser than other denoising methods considered here. The Wiener filter is the worst denoiser, and the wavelet-based soft shrinkage method is better than the Wiener filter but worse than the SC shrinkage algorithm. Moreover, it can be also very easily to tell the denoised effects of the different methods with naked eyes only.

6   Conclusions

In this paper, we proposed the NIG-based extended NNSC neural network model for denoising natural images. The NIG density is a flexible, four-parameter density, highly suitable for molding possibly skewed super-Gaussian data. In the NIG case, to yield accurate results for fairly large datasets, very fast and simple cumulant based parameter estimators can be obtained. We obtained sparsely coded image data by applying our extended NNSC algorithm to natural images selected. The experimental results demonstrated that the NIG density is a very good fitness to the NNSC transformed data. In denoising process, we performed the NIG-based NNSC shrinkage technique on the "grashopper" image contaminated by additive Gaussian noise. The results showed that this technique is highly efficient in reducing noises. Compared with the methods of the Wiener filter, the wavelet-based soft shrinkage and the SC shrinkage, the NIG-based NNSC shrinkage method is also the best denoiser.

References

1. Hyvärinen, A.: Sparse Coding Shrinkage: Denoising of Nongaussian Data by Maximum Likelihood Estimation. Neural Computation. Vol. 11, No. 7, (1997) 1739-1768
2. Hoyer, P. O.: Modeling Receptive Fields with Non-negative Sparse Coding. In E. D. Schutter (ed.), Computational Neuroscience: Trends in Research 2003. Elsevier, Amsterdam 3. Hanssen, A., Øigård, T. A.: The Normal Inverse Gaussian Distributions as A Flexible Model for Heavy Tailed Processes. Proc. IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing. June 3-6, Baltimore, Maryland, USA, 2001
4. D.S.Huang: Systematic Theory of Neural Networks for Pattern Recognition. Publishing House of Electronic Industry of China, Beijing (1996)
5. D.S.Huang, Horace H.S.Ip, Law Ken C K, Zheru Chi: Zeroing Polynomials Using Modified Constrained Neural Network Approach. IEEE Transactions on Neural Networks. Vol. 16, No. 3, (2005) 721-732.
6. D.S.Huang, Horace H.S.Ip, Zheru Chi: A Neural Root Finder of Polynomials Based on Root Moments. Neural Computation. Vol. 16, No. 8, (2004) 1721-1762
7. D.S.Huang: A Constructive Approach for Finding Arbitrary Roots of Polynomials by Neural Networks. IEEE Transactions on Neural Networks. Vol. 15, No. 2, (2004) 477-491
8. D.S.Huang and S.D.Ma: Linear and Nonlinear Feedforward Neural Network Classifiers: A comprehensive understanding. Journal of Intelligent Systems. Vol. 9, No. 1, (1999) 1-38
9. D.S.Huang: Application of Generalized Radial Basis Function Networks to Recognition of Radar Targets. International Journal of Pattern Recognition and Artificial Intelligence, Vol. 13, No. 6, (1999) 945-962
10. D.S.Huang: Radial Basis Probabilistic Neural Networks: Model and Application. International Journal of Pattern Recognition and Artificial Intelligence, Vol. 13, No. 7, (1999) 1083-1101
11. D.S.Huang: The Local Minima Free Condition of Feedforward Neural Networks for Outer-Supervised Learning. IEEE Trans. on Systems, Man and Cybernetics. Vol. 28B, No. 3, (1998) 477-480
12. Li Shang, D. S. Huang, Chun-Hou Zheng and Zhan-Li Sun, "Natural Image Compression Using An Extended Non-negative Sparse Coding Neural Network Technique," IJCNN2005, Montreal, Quebec, Canada, July 31-August 4, 2005 (accepted)（EI）.