

A novel shrinkage technique based on the normal inverse Gaussian density model

Li Shang^{1,2}, Kang Li³, Tat-Ming Lok⁴ and Michael R. Lyu⁵

¹Hefei Institute of Intelligent Machines, Chinese Academy of Sciences, P.O. Box 1130, Hefei, Anhui 230031, China

²Automation Department, University of Science and Technology of China, Hefei, Anhui 230026

³School of Electrical & Electronic Engineering Queen's University Belfast

⁴Information Engineering Department, The Chinese University of Hong Kong, Shatin, Hong Kong

⁵Computer Science & Engineering Department, The Chinese University of Hong Kong, Shatin, Hong Kong

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 proposed by us. This algorithm can converge to feature basis vectors, which behave in the locality and orientation in spatial and frequency domain. Here, we demonstrate that the NIG density provides a very good fitness to the non-negative sparse data. In the 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 the NNSC shrinkage technique, is self-adaptive to the statistical properties of image data. This denoising method is evaluated by values of the normalized signal to noise rate (SNR). Experimental results show that the NNSC shrinkage approach is indeed efficient and effective in denoising. Otherwise, we also compare the effectiveness of the NNSC shrinkage method with methods of standard sparse coding shrinkage, wavelet-based shrinkage and the Wiener

Address for correspondence: Li Shang, Hefei Institute of Intelligent Machines, Chinese Academy of Sciences, P.O. Box 1130, Hefei, Anhui 230031, China. E-mail: shangli@iim.ac.cn

filter. The simulation results show that our method outperforms the three kinds of denoising approaches mentioned above.

Key words: feature basis function; image denoising; non-negative sparse coding; normal inverse Gaussian; shrinkage.

1. Introduction

Removing noise from data can be considered a process of producing the optimal estimates of unknown signals from the available noise data. In general, the problem of how to select a suitable denoising algorithm is dependent on a specific application. Spatial filters based on the principle of filtering have long been used as the traditional means of removing noise from images and signals (Weeks, 1996). These filters can usually smooth the data and reduce the noise, but also blur the data to some extent (Bovik, 2000). More recently, more and more data-adaptive image denoising techniques have been explored, such as those based on principal components analysis (PCA) (Akkarakaran and Vaidyanathan, 1999; Mika *et al.*, 1999; Oja, 1992) independent component analysis (ICA) (Hyvärinen, 1998, 1999a,b; Hyvärinen and Oja, 2001) and sparse coding (SC) shrinkage (Hyvärinen, 1997), and so on. All of these methods can denoise an image successfully by using different skills. The SC algorithm is appropriate to multi-dimension mixed data. Therefore, this technique has been used widely in the image denoising field, especially by a 'shrinkage' (Hyvärinen, 1997) technique. However, the SC technique is unrealistic as a model of V1 simple-cell behaviour (Hoyer, 2003). Therefore, Hoyer introduced the concept of non-negative sparse coding (NNSC) (Hoyer, 2002, 2003), and it has been used to model successfully receptive fields of V1 in the mammalian primary visual cortex. Thus the non-negative property can cause different representations and applications, such as image reconstruction, data compression, image denoising, pattern recognition, and so on. In fact, the basic principle of SC 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 essential 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 (Hyvärinen, 1997), but, it cannot be used to model different degrees of kurtosis for a given variance. In addition, the other sparse models proposed by Hyvärinen (1997), 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 modelling various degrees of sparseness, and taking into account a possible skewness. In addition, it should be possible 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 (Bandorff-Nielsen, 1997; Hanssen and Øigård,

2001), 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 (Bandorff-Nielsen, 1997; Hanssen and Øigård, 2001). In the symmetric case, this method can model data ranging from zero normalized kurtosis, ie, Gaussian distribution, to any positive valued kurtosis. Referring to the model of NNSC introduced by Hoyer (2002, 2003), we propose an extended NNSC algorithm based on the model of NIG density. In particular, using a maximum a posteriori (MAP) estimator (ie, a shrinkage operator), we can successfully denoise a noisy image, 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 (Bandorff-Nielsen, 1997; Hanssen and Øigård, 2001):

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

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

$$K_1(v) = \sqrt{\frac{\pi}{2u}} \cdot \exp(-v) \quad (2)$$

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

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

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

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

According to the definition in Equation (1), the shape of the NIG density is specified by the four-parameter vector $[\alpha, \beta, \mu, \delta]^T$. The α -parameter controls the steepness or pointiness of the density. A larger value of the parameter α implies lighter tails. The rightmost panel in Figure 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 β -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 leftmost panel in Figure 1) around μ , which is a centrality parameter. The rightmost panel in Figure 1 also shows the dependency on the parameter β . It can be noted that the skewness increases as β increases. Lastly, the

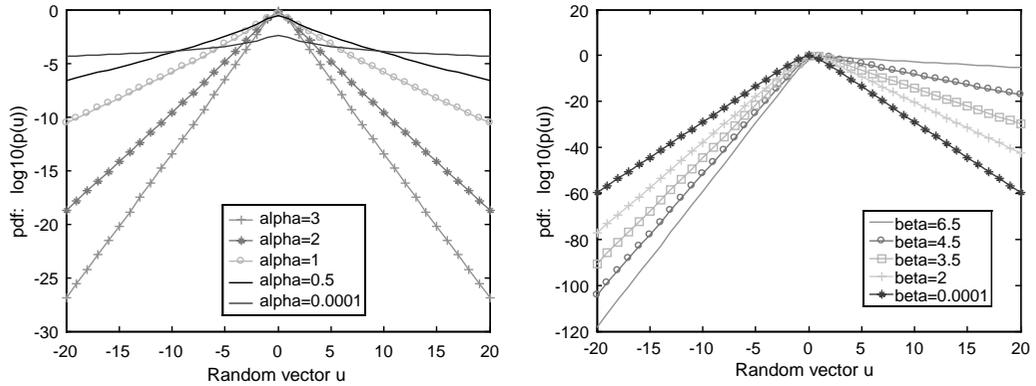


Figure 1 NIG density (logarithmic scale) for different values of α and β . Left: α varies, and $\beta = \mu = 0$, $\delta = 1$. Right: β varies, and α , $\mu = 0$, $\delta = 1$

δ -parameter is a scale-like parameter in the sense that the rescaled parameters $\alpha \rightarrow \alpha\delta$, and $\beta \rightarrow \beta\delta$ are invariant under location-scale changes of μ .

For the four-parameter NIG, Hanssen and Øigård derived a cumulant-based estimator (Bandorff-Nielsen, 1997; Hanssen and Øigård, 2001). 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:

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

Thereafter, the parameter estimators can easily be derived as follows (Bandorff-Nielsen, 1997; Hanssen and Øigård, 2001):

$$\delta = \sqrt{k^{(2)}\zeta(1 - \rho^2)} \quad (6)$$

$$\alpha = \frac{\zeta}{\delta\sqrt{1 - \rho^2}} \quad (7)$$

$$\beta = \alpha\rho \quad (8)$$

$$\mu = k^{(1)} - \rho\sqrt{k^{(2)}\zeta} \quad (9)$$

However, the above estimates require a fairly large set of accurate data samples that can produce statistically consistent estimators (Bandorff-Nielsen, 1997; Hanssen and Øigård, 2001). Although these parameters can be easily estimated from the noise-free data set, if the noise variance σ^2 is known, then by subtracting σ^2 from the estimate of $k^{(2)}$, the parameters can also be easily estimated from noisy observations.

This is due to the fact that zero mean Gaussian noise only contributes to the second-order cumulant and is independent of the input signals.

3. The extended NNSC model and algorithm

3.1 Modelling NNSC of natural images

The basic idea in standard sparse coding is very simple. It can be described as follows: denote $(X = x_1, x_2, \dots, x_n)^T$ as observed n -dimensional random vectors and $(S = s_1, s_2, \dots, s_m)^T$ as hidden m -dimensional components (here only considering the case of $m \leq n$). Then any natural image can be modelled as a linear superposition of some features a_i :

$$X(x, y) = \sum_i^m a_i(x, y)s_i + \varepsilon \quad (10)$$

where (x, y) is the pixel co-ordinate in an image, $X(x, y)$ denotes the input image data, a_i (the i th column of A) are called basis vectors, s_i (the i th row of S) are mutually independent sparse variables and ε is Gaussian noise. The image model of NNSC is the same as that shown in Equation (10). The significant point is here that the input matrix X , basis vectors A and latent sparse coefficients S are non-negative in the 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 behaviour of simple-cells in the receptive fields in the mammalian primary visual cortex in the brain, also known as V1. Furthermore, V1 receives the visual data from the lateral geniculate nucleus (LGN) in the form of a separated ON-channel and OFF-channel, and each channel's input data are positive (Hoyer, 2003). Therefore, the image model of sparse coding is not suitable to model V1 simple-cell behaviour. In order to see how V1 recodes its input data, a model suitable for NNSC has been studied recently.

3.2 The cost function and updating rules

On the basis of the Hoyer's NNSC model (Hoyer, 2002), 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:

$$J(A, S) = \frac{1}{2} \sum_{x,y} \left[X(x, y) - \sum_i a_i(x, y)s_i \right]^2 + \lambda \sum_i f\left(\frac{s_i}{\sigma_i}\right) + \eta \sum_i (a_i^T a_i) \quad (11)$$

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 i th column of A and the i th row of S , λ is the trade-off 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 Equation (1) (see Section 2). According to the estimations of the four-parameter vector $[\alpha, \beta, \mu, \delta]^T$, the function $f(\cdot)$ can be selected as definite function.

In terms of Equation (11), we can obtain the derivatives of \hat{a}_i and \hat{s}_i , shown as follows:

$$\hat{s}_i = a_i^T(x, y) \left[X(x, y) - \sum_{i=1}^n a_i(x, y) s_i \right] - \frac{\lambda}{\sigma_i} f' \left(\frac{s_i}{\sigma_i} \right) = a_i^T e - \frac{\lambda}{\sigma_i} f' \left(\frac{s_i}{\sigma_i} \right) \quad (12)$$

$$\hat{a}_i = \left[X(x, y) - \sum_{i=1}^n a_i(x, y) s_i \right] s_i^T - \gamma a_i(x, y) = e s_i^T - \gamma a_i \quad (13)$$

where $\sigma_i = \sqrt{\langle s_i^2 \rangle}$ and $e = X - AS$ is the residual error between the original image and the reconstructed image of this model. In experiment, we exploited a conjugate gradient algorithm to update basis vectors A and Equation (12) to update S .

4. The denoising algorithm for NNSC shrinkage

4.1 MAP estimator and NIG shrinkage function

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

$$y = s + v \quad (14)$$

where $v \sim N(0, \sigma^2)$, 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 $\hat{s} = g(y)$. Denoting by $p(s)$ the density of s , and by $f(\hat{s}) = -\ln p(\hat{s})$ the negative log-density of \hat{s} . For an unimodal, differentiable posteriori density, \hat{s} can be obtained by solving the following equation:

$$\frac{\hat{s} - y}{\sigma^2} + f'(\hat{s}) = 0 \quad (15)$$

where $f(\hat{s})$ is assumed to be convex and differentiable, and $f'(\hat{s}) = d(f(\hat{s}))/ds$ 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:

$$\hat{s}^* = y - \sigma^2 f'(y) \quad (16)$$

where the problem with this estimator in Equation (16) 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 $f'(\cdot)$ 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 non-Gaussian random variable corrupted by Gaussian noise may be applied:

$$\hat{s} = g(y) = \text{sign}(y) \max(0, |y| - \sigma^2 |f'(y)|) \quad (17)$$

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

$$f'_{NIG}(u) = \frac{\alpha(u - \mu)}{q(u)} \left(\frac{K_0[\alpha q(u)]}{K_1[\alpha q(u)]} + \frac{2}{\alpha q(u)} \right) - \beta \quad (18)$$

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 $[\alpha, \beta, \mu, \delta]^T$ estimated by the sample data. In fact, the function in Equation (17) 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), ie, $S = WX$. Here, X and S have the same definition mentioned in subsection 3.1, ie, respectively denoting the input data matrix and sparse components. 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 the ON-channel minus the 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 $i = 1, 2, \dots, m$, estimate a NIG density model for the non-negative sparse components $s_i = w_i Z$, where w_i is the i th row of W . Determine the four-parameter vector $[\alpha, \beta, \delta, \mu]^T$ in terms of Equations (6)–(9) and find the corresponding NIG shrinkage function g_i according to Equation (17).
- 3) Observing a noisy version \tilde{X} , which has been beforehand centred 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) Applying the shrinkage operator g_i to every component y_i of Y , to obtain $\hat{s}_i = g_i(y_i)$, therefore, $\hat{S} = (\hat{s}_1, \hat{s}_2, \dots, \hat{s}_m)$.
- 5) Do the inverse transformation to obtain estimates \hat{X} of the noise-free data X , ie, $\hat{X} = W^{-1}\hat{S} = W^T\hat{S}$.

5. Experimental results

5.1 Applied 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>. First, selecting randomly 10 noise-free

natural images with 256×512 pixels, 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 $64 \times 100\,000$ is acquired. Considering the non-negativity, we separate X into the ON-channel and OFF-channel, denoted respectively by Y and Z . So, the non-negative matrix $I = (Y; Z)$ with the size of $2 \times 64 \times 100\,000$ is obtained. Then, using the updating rules of A and S in turn, we minimized the objective function given in Equation (11).

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} ($j = 1, 2, \dots, T$; 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 modelling 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 Equation (1) was:

$$p(s) = 0.0476 \frac{\exp(F1 - F2)}{F3} \quad (19)$$

where $F1$, $F2$ and $F3$ are respectively calculated as: $F1 = 0.1757 + 0.05(s - 0.022)$, $F2 = 2.17[(s - 0.022)^2 + 0.0066]^{1/2}$ and $F3 = [(s - 0.022)^2 + 0.0066]^{3/4}$. The resulting NIG density of the shrunked sparse components of s_{1j} was shown in the left of Figure 2 in a log-plot (solid line). It has a negligible skewness and is centred close to the origin. For

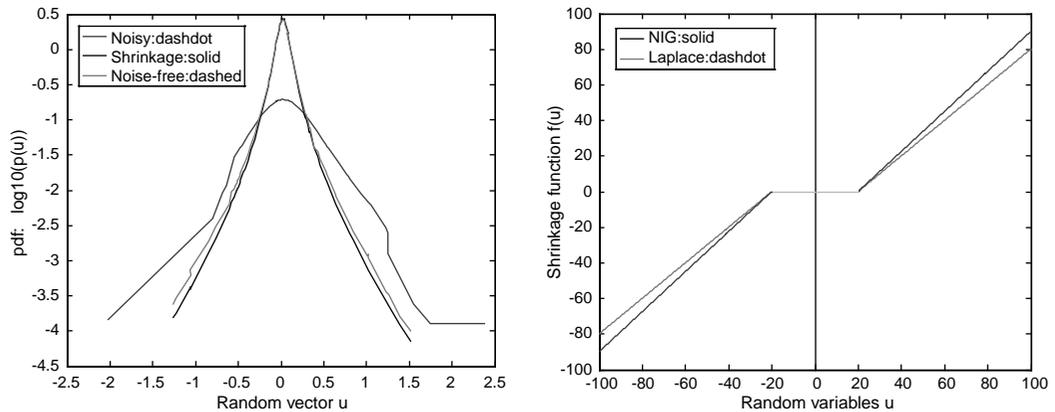


Figure 2 Results of denoising noisy sparse components of s_{1j} by the NIG-based NNSC shrinkage function. Left: NIG density corresponding to the given sparse vector s_{1j} . Dashed: the noise-free s_{1j} . Dash-dotted: the noisy s_{1j} . Solid: the shrunked s_{1j} . Right: the in-out property of the shrinkage function. Solid: NIG shrinkage function. Dashed: Laplace shrinkage function

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 Figure 2. It can be readily seen that the NIG density of the shrunked s_{1j} approaches highly to that of the noise-free s_{1j} . The noise has been reduced effectively, and the shrunked components are concentrated around zero to a much higher degree, compared with 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 .

Based on the NIG pdf calculated in Equation (19), the shrinkage function for the noisy sparse components of s_{1j} can be calculated. It was shown in the right panel in Figure 2 as the solid non-linearity. For comparison, we assumed that the classical Laplacian density has modelled the given sparse data. This shrinkage function is given by

$$g(s) = \text{sign}(s) \max(0, |s| - \frac{\sqrt{2}\sigma^2}{d}) \quad (20)$$

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 shrunked 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 Denoising results

Here, the quality of denoised images is evaluated by the values of normalized SNR, which is defined as follows (Grgić *et al.*, 2004):

$$\text{SNR}_n = 10 \log_{10} \left(\frac{\sum_{i=1}^N \sum_{j=1}^M (X_{ij} - \bar{X}_{ij})^2}{\sum_{i=1}^N \sum_{j=1}^M (X_{ij} - \hat{X}_{ij})^2} \right) \quad (21)$$

where M and N denote the size of the image data, X denotes the input image data set, \bar{X} denotes the mean value of X and \hat{X} denotes the denoised image data. The calculated



Figure 3 Denoising experiment on the grasshopper 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



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

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 grasshopper image, which were obtained by our algorithm, were shown in the rightmost panel in Figure 3.

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 Figure 4 and Table 1. According to the experimental results, it can be concluded that our NIG-based NNSC shrinkage method is the best denoiser of the 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 the naked eye 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 moulding 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 fit to the NNSC transformed data. In denoising process, we performed the NIG-based NNSC shrinkage technique on the 'grasshopper' image contaminated by additive Gaussian noise. The results

Table 1 Values of normalized SNR obtained by different denoising algorithms. The noise level is $\sigma=0.5$

Algorithm	SNR_n (Denoised images)	SNR_n (Noise images)
Wavelet-based soft shrinkage	5.6536	
Sparse coding shrinkage	11.3024	1.1701
NIG-based NNSC shrinkage	14.9864	
Wiener filter	4.7728	

showed that this technique is highly efficient in reducing noise. 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

- Akkarakaran, S.** and **Vaidyanathan P.P.** 1999: The role of principal component filter banks in noise reduction. In *Wavelet Applications in Signal and Image Processing VII*. Bellingham, SPIE, The International Society for Optical Engineering Press, 3813, 346–57.
- Bandorff-Nielsen, O.E.** 1997: Normal inverse gaussian distributions and stochastic volatility modeling. *Scandinavian Journal of Statistics* 24, 1–13.
- Bovik, A.C.** 2000: *Handbook of image and video processing*. Academic Press.
- Grgić, S., Grgić, M.** and **Mrak, M.** 2004: Reliability of objective picture quality measures. *Journal of Electrical Engineering*, 55, 3–10.
- Hanssen, A.** and **Øigård, T.A.** 2001: The normal inverse Gaussian distributions as a Flexible model for heavy tailed processes. *Proceedings of the IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing*, Baltimore.
- Hoyer, P.O.** 2002: Non-negative sparse coding. *Neural Networks for Signal Processing XII (Proceedings of the IEEE Workshop on Neural Networks for Signal Processing)*, Martigny, 557–65.
- Hoyer, P.O.** 2003: Modeling receptive fields with non-negative sparse coding. *Neurocomputing* 52–54, 547–52.
- Hyvärinen, A.** 1997: Sparse coding shrinkage: denoising of nongaussian data by maximum likelihood estimation. *Neural Computation* 11, 1739–68.
- Hyvärinen, A.** 1998: Independent component analysis in the presence of Gaussian noise by maximizing joint likelihood. *Neurocomputing* 22, 49–67.
- Hyvärinen, A.** 1999a: Fast independent component analysis with noisy data using Gaussian moments. In *Proceedings of the International Symposium on Circuits and Systems*. Orlando, FL.
- Hyvärinen, A.** 1999b: Gaussian moments for noisy independent component analysis. *IEEE Signal Processing Letters* 6 145–47.
- Hyvärinen, A.** and **Oja, E.** 2001: Independent component analysis. John Wiley.
- Mika, S., Schölkopf, B., Smola, A.J., Müller, K.-R., Scholz, M.** and **Rätsch, G.** 1999: Kernel PCA and de-noising in feature spaces. *Advances in Neural Information Processing Systems* 11. MIT Press, 536–42.
- Oja, E.** 1992: Principal components, minor components, and linear neural networks. *Neural Networks* 5, 927–35.
- Weeks, A.R.** 1996: *Fundamentals of electronic image processing*. SPIE/IEEE Series on Imaging Science and Engineering. Bellingham, SPIE Optical Engineering Press and IEEE Press.