Combining Adaptive Diffusion with Difference for Detecting Nonstationary Signals

TAN Xiao-Gang[1] LI Li-Ping[1] WEI Ping[1]

Abstract Time-frequency analysis aims to construct a density function of time and frequency to reveal the frequency components in signals to be analyzed and the evolution of the frequency of signals with time. The Wigner distribution (WD) is one of the most fundamental and widely used methods for analyzing nonstationary signals in the fields of radar, communication, *etc.* However, the application of the WD is greatly limited by the existence of interference terms. The adaptive diffusion method proposed to remove the interference terms of the WD by Julien Gosme, *et al.* is to be invalid in the presence of interference terms generated by signals, whose distributions are interwoven together in the time-frequency plane of the WD. We combine the diffusion technique with difference method for removing these interference terms to improve the resolution and readability of the time-frequency representation of the Cohen class for detecting nonstationary signals.

Key words Adaptive diffusion, interference terms, signal-dependent kernel function, nonstationary signal

1 Introduction

The nonstationary signals, such as linear modulation, sinusoidal modulation, and polynomial phase signal, have long been of great research interest given their importance in a great variety of applications in radar, communication, *etc.* The methods for analyzing nonstationary signals include linear time-frequency representation, various quadratic time frequency distributions (TFDs) (the Cohen class TFD, the affine class TFD and the reassignment TFD)[1∼3], adaptive kernel function TFD[4,5], and parametric TFD[6∼8]. The Wignen distribution (WD) is one of the most important and widely used tools for analyzing these nonstationary signals, which has good theoretical properties such as covariance with respect to time and frequency shifts, marginal properties, high resolution, and energy conservation. But the presence of interference terms greatly limits its applications. These interference terms can be removed by averaging the WD with appropriate kernels, which yield positive time-frequency densities. However, this reduces the time-frequency resolution. Spectrograms and scalograms are examples of positive quadratic distributions obtained by smoothing the WD, but both of them provide poor resolution because their time-frequency resolution is fixed, regardless of the local geometry of the signal signature. Other various smoothing versions of the WD have been introduced to reduce interference terms and maintain resolution as good as possible. The early smoothing versions, however, suffer from fundamental limitations: they are signal-independent, or their kernel functions are independent of signal to be processed, that is, they do not adapt to signals. These kernels may perform very well for certain signals while very badly for other signals[4]. All this has prompted the development of signal-dependent kernels or adaptive kernels. Diffusion method is a signal-dependent method and was first used to improve the readability of the Cohen class time-frequency representations in [9]. The diffusion process yields signal-dependent TFDs, which are called adaptive-kernel function (or signal-dependent kernel function) TFDs as well. The diffusion processing has emerged as an active area of research in recent years. Reference [10] proposed a method called the forward-and-backward (FAB) diffusion process. The FAB method was further generalized for color processing via the Beltrami flow by adaptively modifying the structure tensor that controls the nonlinear diffusion process. The proposed structure tensor is neither positive definite nor negative, and switches between these states according to image features. This results in a FAB diffusion flow where different regions of the time-frequency distribution are either forward or backward diffused according to the local geometry within a neighborhood. The diffusion process was expanded to fourth-order diffusion equations to optimize the trade-off between noise removal and edge preservation[11]. A new adaptive diffusion process was proposed to locally control both the orientation and the strength of smoothing[12,13], which smoothed the WD by a signal dependent kernel. It provides a uniform framework not only for the Cohen class but for the affine class as well. Without much work, one can easily transpose the diffusion method to the affine class of time-scale representations and get a locally adapted distribution that is covariant about time shifts and scale changes in the signal[12∼15]. So we will restrict ourselves to the diffusion in the TFDs of the Cohen class in this paper.

In the presence of signals, whose distributions are closely spaced or interwoven together in the time-frequency plane of the WD, the interference terms are so strong that they are very stubborn to be smoothed out by the diffusion process. We might very likely meet these signals that produce very strong interference terms in the applications of radar. So we concentrate on attenuating or removing these very strong interference terms to improve the readability and the resolution of the TFDs. This goal can be reached by incorporating the difference method into the diffusion method since the differences of the interference terms are much stronger than those of the signal components.

We have made a brief introduction to the recent development of diffusion processing and the shortage of this method in time frequency analysis. In the next section, we will introduce diffusion models and present an explicitly discrete scheme for the diffusion method. Then, we will make a brief explanation to the problem involved with diffusion method and propose our solution. In the end, we draw a brief conclusion about the diffusion method.

Received September 4, 2006; in revised form February 7, 2007
Supported by Program for New Century Excellent Talents in University
1. School of Electronic Engineering, University of Electronic Science and Technology of China (UESTC), Chengdu 610054, P. R. China
DOI: 10.1360/aas-007-0931

2 Diffusion models

The general form of a diffusion model is given as follows[12].

$$\begin{cases} \frac{\partial C_x(t,f;\tau)}{\partial \tau} = \mathrm{div}(\eta_x(t,f)\nabla C_x(t,f;\tau)) \\ C_x(t,f;\tau=0) = W_x(t,f) \end{cases} \tag{1}$$

where the sign ∇ denotes the gradient operator.

This is an isotropic diffusion process generated originally from the heat diffusion equation. It has been proved that the distribution $C_x(t,f;\tau)$ decided by (1) is covariant with respect to time-frequency shifts for all positive τ. $\eta_x(t,f;\tau)$ is a conductance function depending both on the signal to be analyzed and the time-frequency location (t,f) in the TFD plane. It can be a decreasing function of the distribution magnitude of $C_x(t,f;\tau)$, or a decreasing function of spectrogram (for Cohen class), or other forms. In this paper, a decreasing function of the distribution magnitude of $C_x(t,f;\tau)$ applied for Cohen class is defined as follows.

$$\eta_x(C(t,f;\tau)) = \left[1 + \left(\frac{C_x(t,f;\tau)}{\beta}\right)^{\alpha}\right]^{-1} \tag{2}$$

We can derive the discrete framework from the first formula in (1) by sampling the distribution in the time-frequency grid $(n\cdot\delta t, m\cdot\delta f), (n,m) \in [0, N-1]\times[0, M-1]$. Here, we present an explicitly discrete scheme for the convenience of simulation as follows. The sampling period of τ is $\delta\tau = (4\pi)^{-1}$ in this paper.

$$\frac{\partial C_x(t,f;\tau)}{\partial \tau} = \nabla\left[\eta_x(t,f)\nabla(C_x(t,f;\tau)\right] \tag{3}$$

$$\frac{\partial C_x(t,f;\tau)}{\partial \tau} = \left[\nabla\eta_x(t,f)\right]\left[\nabla C_x(t,f;\tau)\right] + \eta_x(t,f)\left[\nabla^2 C_x(t,f;\tau)\right] \tag{4}$$

$$C_x(n,m;\tau_{k+1}) = C_x(n,m;\tau_k) + \delta\tau\sum_{i=1}^{4} S_i \tag{5}$$

$$\begin{cases} S_1 = D_t(\eta_x(n,m))D_t(C_x(n,m)) \\ S_2 = D_t(\eta_x(n,m))D_f(C_x(n,m)) \\ S_3 = \eta_x(n,m)D_{tt}(C_x(n,m)) \\ S_4 = \eta_x(n,m)D_{ff}(C_x(n,m)) \end{cases} \tag{6}$$

where $D_t(\cdot)$ denotes the partial differential operator with respect to time t, while $D_{tt}(\cdot)$ denotes the second-order partial differential operator with respect to time t, they are defined as

$$\begin{cases} D_t(C_x(n,m)) = C_x(n+1,m) - C_x(n,m) \\ D_{tt}(C_x(n,m)) = D_t(C_x(n+1,m)) - D_t(C_x(n,m)) \end{cases} \tag{7}$$

Since the diffusion process is an iterative process, it is important to select an appropriate stopping criterion. Many information criteria are eligible for this goal. In [16], the Shannon entropy was defined as

$$H_C(\tau_k) = -\sum_n\sum_m |C_x(n,m;\tau_k)|\log|C_x(n,m;\tau_k)| \tag{8}$$

And the Rényi entropy was defined in [16] as

$$R_C(\tau_k) = \log_2\delta_t\delta_f + \frac{1}{1-\alpha}\times \log_2\sum_n\sum_m\left(\frac{C_x(n,m;\tau_k)}{\sum_{n'}\sum_{m'}C_x(n',m';\tau_k)}\right)^{\alpha} \tag{9}$$

where δ_t is the sampling period in time and δ_f is the sampling period in frequency.

In this paper, a third-order Rényi entropy is used as our criterion for optimal stopping point. Without a constraint on the diffusion time, the entropy of $C_x(t,f;\tau)$ will rapidly decrease toward the minimum and then slowly increase, while C_x will converge to a uniform distribution over the time-frequency domain, regardless of the analyzed signal. The minimum point of the Shannon entropy was used as the optimal stopping point in [12]. In our work, the difference of the Rényi entropy $R_c(\tau_k) - R_c(\tau_{k-1})$ is used for detecting the optimal stopping point. The difference is negative before the minimum point and always positive after the minimum point. Once $R_c(\tau_k) - R_c(\tau_{k-1})$ is not negative, the iteration will be stopped. So the constant term $\log_2(\delta_t\delta_f)$ in (9) can be neglected.

The diffusion process allows the well-structured signal components to get a high concentration. However, the interference terms and noises will be smoothed out in the diffusion process. The performance of this method is shown in Fig. 1 (see next page).

The isotropic diffusion is controlled by the scalar-valued conductance function for locally tuning diffusion strength. But it is unable to control the diffusion strength and orientation simultaneously. In [12], a tensor T, which is a gradient-square tensor, was incorporated into the general diffusion model in (1) and the so-called adaptive diffusion was derived. The adaptive diffusion can tune both diffusion strength and orientation with the tensor T. The tensor T is defined as

$$T = \begin{bmatrix} L_\sigma * \left(\frac{\partial C}{\partial t}\right)^2 & L_\sigma * \left(\frac{\partial C}{\partial t}\frac{\partial C}{\partial f}\right) \\ L_\sigma * \left(\frac{\partial C}{\partial t}\frac{\partial C}{\partial f}\right) & L_\sigma * \left(\frac{\partial C}{\partial f}\right)^2 \end{bmatrix} \tag{10}$$

where L_σ is a one-dimensional low pass smoothing filter and the $*$ denotes the usual one-dimensional convolution. The length of the filter is decided by computational efficiency (with respect to filtering computation) and filtering effect (and thus the resulting diffusion effect and iteration number). From the point of computational efficiency, the length should be short. But the filtering effect will degrade if the length is too short. Computational simulation has revealed that the resulting diffusion effect improves very little when the length is greater than 21, while the effect is acceptable when the length is 7 (the data length is 256). In our study, the length of the filter is 13. The smoothing filter adjusts the diffusion towards the average orientation, thus results in the attenuation of the noise and interference terms. One can filter the partial differences along rows (time domain filtering) or columns (frequency domain filtering). We first filter the partial differences from rows, then get a stronger smoothing effect from columns. Note that if the FIR filter is used, one should compensate for the time delay, which is caused by the linear phase of the FIR filter. Obviously, T is of rank 1 without this filter. In addition, T is non-negative definite. The kernel of the adaptive diffusion process is right the tensor T.

The adaptive anisotropic diffusion model derived from (1) is defined as follows[12].

$$\begin{cases} \frac{\partial C_x(t,f;\tau)}{\partial \tau} = \nabla(\eta_x(t,f)G_{tf}\nabla C_x(t,f;\tau)) \\ C_x(t,f;\tau=0) = W_x(t,f) \end{cases} \tag{11}$$

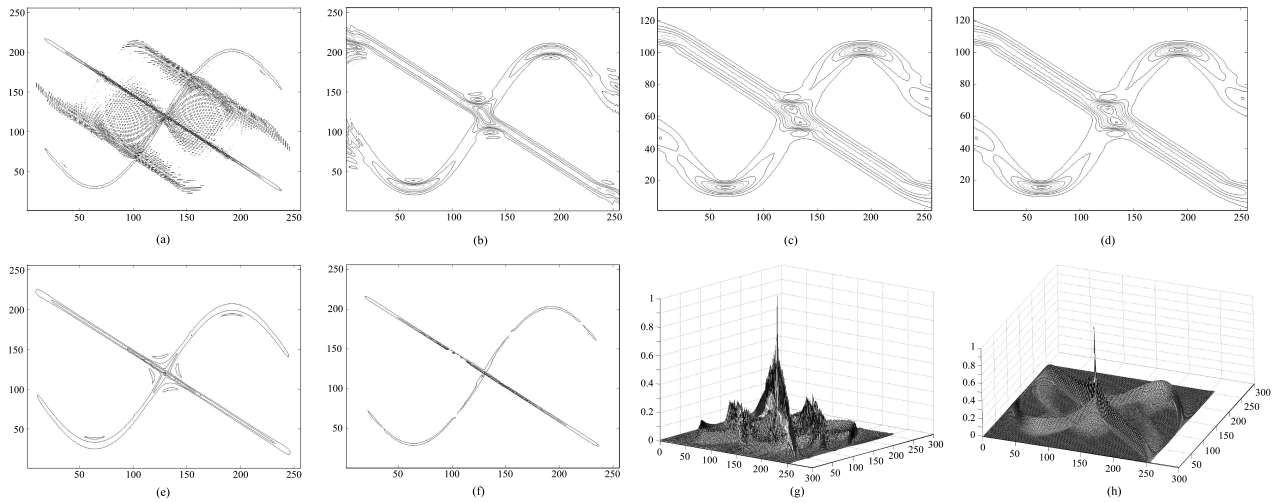

Fig. 1 Comparison of the adaptive diffusion process with well-known TFRs ((a) WD; (b) Smoothing pseudo WD; (c) Spectrogram; (d) Spectrogram modified by the reassignment method; (e) Isodiffusion; (f) Anisodiffusion; (g) 3-D plotting of WD; (h) 3-D plotting of Aisodiffusion. These 256-by-256 TFRs are those of a 256-sample signal embedded in noise, which consist of two frequency-modulated waves with linear and sinusoidal modulations.)

$$G_{tf} = (\boldsymbol{v}_1, \boldsymbol{v}_2) \begin{bmatrix} 1 & 0 \\ 0 & \lambda \end{bmatrix} (\boldsymbol{v}_1, \boldsymbol{v}_2)^{'} \tag{12}$$

where the sign $(\cdot)^{'}$ denotes conjugate transposition, $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ are the orthonormal eigenvectors of the gradient tensor T, and λ a constant denoting the diffusion strength. $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ are the estimation of the directions of ∇C and $\nabla C^{\perp}$, respectively. λ, which is greater than 1, has to be chosen in order to produce more aggressive smoothing along signal components than across them. In our paper, λ is 1.35.

It is easy to verify in the line of [12] that the distributions decided by (11) are covariant with respect to time-frequency shifts and conserve energy.

The anisotropic (or isotropic) process diffuses from strong areas to relatively weak areas, so it is a natural forward diffusion process. The result is a smoothed version of the WD, which has a signal-dependent kernel. Their performances are shown in Fig. 1, comparing with other well-known time frequency representations.

3 Improvement on adaptive method

Though the anisotropic diffusion is powerful to remove interference terms and noise, it is likely to be invalid in the presence of interference terms generated by signals, whose distributions are closely spaced or interwoven in the time-frequency plane of the WD. The oscillating interferences are centered at the middle of auto terms′ time-frequency locations. The frequency of the oscillating interferences is proportional to the Euclidean distance of the two auto terms which produce the interference. The direction of these oscillating interferences is perpendicular to the line that joins the two auto terms[17,18]. This can be shown clearly in Fig. 2. The signal to be processed by the new method is as follows.

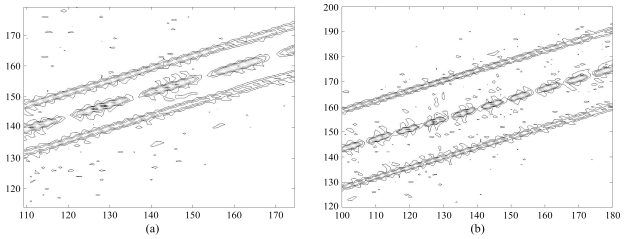

Fig. 2 WD of two linear modulations ((a) Signal consisting of two linear frequency modulations: one is from 0.17 Hz to 0.37 Hz, the other is from 0.2 Hz to 0.4 Hz; (b) Signal consisting of two linear frequency modulations: one is from 0.17 Hz to 0.37 Hz, the other is from 0.23 Hz to 0.43 Hz. These 256-by-256 TFRs are those of a 256-sample signal embedded in noise.)

$$x(t) = s(t) + v(t)$$

where $s(t)$ consists of two parabolic frequency modulations, $v(t)$ is an additive white noise, and the signal-to-noise ratio (SNR) is 15 dB.

In our simulation, the interference terms and noise in the time-frequency plane of the WD reduce its readability seriously. The diffusion method removes the noise completely, but still leaves interference terms stronger than signal components as shown in Figs. 3 (a) and (b). We propose a method to remove these very strong interference terms and thus can enhance the time-frequency resolution of the distributions significantly.

As we have mentioned earlier, the signal components are well structured and, moreover, are the result of the quadratic integration in a finite area. So their second order partial differences are very small and near to zero. And unsurprisingly, the points, whose magnitudes are much greater than zero and their second order partial differences are near to zero, form continuous curves which almost right cover the signal components in the time-frequency plane.

While the interference terms are oscillating swiftly in the time-frequency plane, their second order partial differences are far above zero and locate in small areas. And the points of the interference terms can not form continuous curves. The noise is the same case as the interference terms. This is shown in Fig. 3 (d).

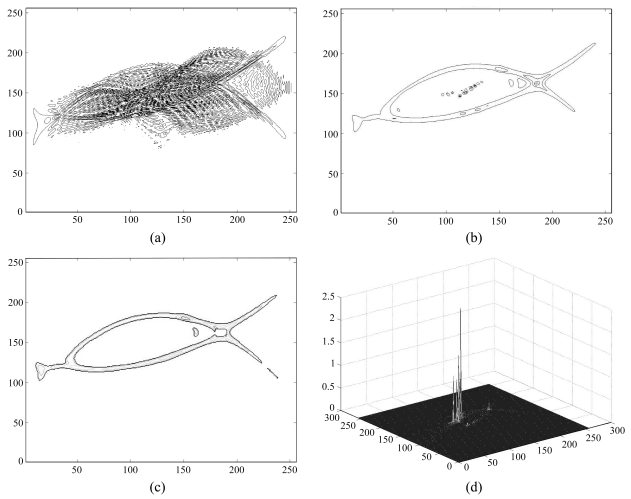

Fig. 3 Performance displaying of difference method ((a) WD; (b) Anisodiffusion; (c) Difference method; (d) 3-D plotting of the second order partial difference of anisodiffusion. These 256-by-256 TFRs are those of a 256-sample signal embedded in noise.)

On the ground of the above analysis, we design a detector to detect the points in the plane, whose second order partial differences are near to zero, and of which the magnitudes are greater than a given threshold (Since our processing follows the diffusion process, the magnitude constraint is reasonable).

$$C(t,f) = \begin{cases} C(t,f;\tau), (t,f) \in U(P), & P \in X \\ 0, & \text{otherwise} \end{cases} \tag{13}$$

$$X = \left\{ \begin{array}{l} (t,f) : |\frac{\partial^2 C_x(t,f;\tau)}{\partial t \partial f}| < \lambda_1 \\ \text{and} \quad C(t,f;\tau) > \lambda_0 / \underbrace{\max}_{(t,f)}(C(t,f;\tau)) \end{array} \right\}$$

where $U(P)$ denotes the neighborhood around the point P. According to the uncertainty principle: $\Delta t \Delta f = 1/(4\pi)$, we have $mn \geq 40$, where Δt is the time resolution, Δf the frequency resolution, $m = \Delta t/\delta t, n = \Delta f/\delta f$ and in our paper $\delta t = 1$, $\delta f = 0.5/256$. The radius of the neighborhood around P decides the joint time-frequency resolution. If the radius is zero, the time-frequency curve will be discontinuous and we get the best joint time-frequency resolution. If it is chosen to be greater than zero and not greater than a constant r, the curve will be continuous and smooth and the best joint resolution property remains. For convenience of calculation, we use square neighborhood; so r should be 2 (the area of the neighborhood around P is $(2 \times r + 1) \times (2 \times r + 1) = 25 < 40$). The joint resolution will degrade if r is greater than 2. λ_0 is used to remove interference terms and noise, which have small second order partial differences due to their small magnitudes. It indicates the magnitude constraint as mentioned above and is empirically chosen to be between $0.1 \sim 0.3$. λ_1 indicates the second order partial difference constraint. It is related to the data length and can be empirically chosen to be $4\pi/256 \sim 8\pi/256\,(0.05 \sim 0.1)$. The validity of values of λ_0 and λ_1 is verified by computational simulations. The performance of our method is shown in Fig. 3 (c).

4 Conclusion

The adaptive diffusion produces adaptive optimal kernel function TFDs and is available for processing a large kind of nonstationary signals. Of all the existing adaptive methods, such as adaptive diffusion, radially Gaussian kernel[4,5] and optimal phase kernel[19], the adaptive diffusion has the best adaptability due to its capability of local controlling[12]. The distributions derived from adaptive diffusion method are signal-dependent TFDs. They have the advantage of being adaptive to a wide range of nonstationary signals[12], unlike those bilinear distributions which can not adapt to signals. In this paper, we propose an improvement on the diffusion method for time-frequency representation of the Cohen class, thus making it capable of removing the very strong interference terms generated by signals whose distributions are interwoven in the time-frequency plane of the WD. We efficiently improve the readability and resolution of time-frequency representation of the Cohen class for detecting nonstationary signals.

Acknowledgement

The authors would like to acknowledge the anonymous reviewers for their helpful comments and suggestions that have resulted in many improvements in the presentation of this paper.

References

1 Cohen L. Time-frequency distributions - a review. *Proceedings of the IEEE*, 1989, **77**(7): 941∼981

2 Hlawatsch F, Boudreaux-Bartels G F. Linear and quadratic time-frequency representations. *IEEE Signal Processing Magazine*, 1992, **9**(2): 21∼67

3 Auger F, Flandrin P. Improving the readability of time-frequency and time-scale representations by the reassignment method. *IEEE Transactions on Signal Processing*, 1995, **45**(5): 1069∼1089

4 Baraniuk R G, Jones D L. A signal-dependent time-frequency representation: optimal kernel design. *IEEE Transactions on Signal Processing*, 1993, **40**(11): 1589∼1602

5 Jones D L, Baraniuk R G. An adaptive optimal-kernel time-frequency representation. *IEEE Transactions on Signal Processing*, 1995, **43**(10): 2361∼2371

6 Friedman J H, Stuetzle W. Projection pursuit regression. *Journal of the American Statistical Association*, 1981, **76**: 817∼823

7 Hube P J. Projection pursuit. *Annals of Statistics*, 1985, **13**(2): 435∼475

8 Goodwin M M, Vetterli M. Matching pursuit and atomic signal models based on recursive filter banks. *IEEE Transactions on Signal Processing*, 1999, **47**(3): 731∼745

9 Perona P, Malik J. Scale space and edge detection using anisotropic diffusion. In: Proceedings of IEEE Computer Society Workshop on Computer Vision. IEEE, 1987. 16∼27

10 Gilboa G, Sochen N, Zeevi Y Y. Forward-and-backward diffusion processes for adaptive image enhancement and denoising. *IEEE Transactions on Image Processing*, 2002, **11**(7): 689∼703

11 Jia D Y, Huang F G, Wen X F. Fourth order anisotropic diffusion equations for noise removal. In: Proceedings of IEEE 7th International Conference on Signal Processing. IEEE, 2004. 994∼997

12 Gosme J, Richard C, Goncalvues P. Adaptive diffusion as a versatile tool for time-frequency and time-scale representations processing a review. *IEEE Transactions on Signal Processing*, 2005, **53**(11): 4136∼4146

13 Gosme J, Goncalvues P. Diffusion equations for adaptive affine distributions. In: Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing. IEEE, 2004. 17∼21

14 Rioul O, Flandrin P. Time-scale energy distributions: a general class extending wavelet transform. *IEEE Transactions on Signal Processing*, 1992, **40**(11): 1746∼1757

15 Bertrand J, Bertrand P. A class of affine Wigner functions with extended covariance properties. *Journal of Mathematical Physics*, 1992, **33**(7): 2515∼2527

16 Baraniuk R, Flandrin P, Jansen M, Michel O. Measuring timefrequency information content using the Renyi entropies. *IEEE Transactions on Information Theory*, 2001, **47**(4): 1391∼1409

17 Mallat S. *A Wavelet Tour of Signal Processing*. Beijing: China Machine Press, 2003. 112∼113

18 Arce G R, Hasan S R. Elimination of interference terms of the discrete Wigner distribution using nonlinear filtering. *IEEE Transactions on Signal Processing*, 2000, **48**(8): 2321∼2331

19 Wisur-Olsen L F, Baraniuk R G. Optimal phase kernels for time-frequency analysis. In: Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing. IEEE, 1996. **3**: 1419∼1422

TAN Xiao-Gang Lecturer at Nanchang Institute of Technology, and in the meanwhile, Ph. D. candidate in signal and information processing at University of Electronic Science and Technology of China (UESTC). He received his master degree in detection technology and automatic equipment from UESTC in 2003. His research interest covers time-frequency analysis and statistical estimation. Corresponding author of this paper. E-mail: xgtan1971@163.com

LI Li-Ping Professor at School of Electronic Engineering, UESTC. She received her bachelor degree in 1984 from UESTC, and received her master degree in signal and information processing and her Ph. D. degree in circuits and systems from UESTC in 1995 and 2000, respectively. Her research interest covers array signal processing, high-speed signal processing, weak signal detection, and parameter estimation. E-mail: liliping@uestc.edu.cn

WEI Ping Professor at School of Electronic Engineering, UESTC. He received his bachelor and master degrees both in the electronic engineering from Beijing Institute of Technology in 1986 and 1989, respectively. He received his Ph. D. degree in communication and electronic system from UESTC in 1996. He is a senior member of CIE and a member of IEEE. His research interest covers spectral analysis, array signal processing, electronic surveillance, and communication signal processing. E-mail: pwei@uestc.edu.cn