Hindawi Publishing Corporation
EURASIP Journal on Advances in Signal Processing
Volume 2010, Article ID 579295, 20 pages
doi:10.1155/2010/579295

Review Article

Time-Frequency Analysis and Its Application in Digital Watermarking

Srdjan Stanković

Faculty of Electrical Engineering, University of Montenegro, 20000 Podgorica, Montenegro

Correspondence should be addressed to Srdjan Stanković, srdjan@ac.me

Received 14 February 2010; Accepted 7 June 2010

Academic Editor: Sridhar Krishnan

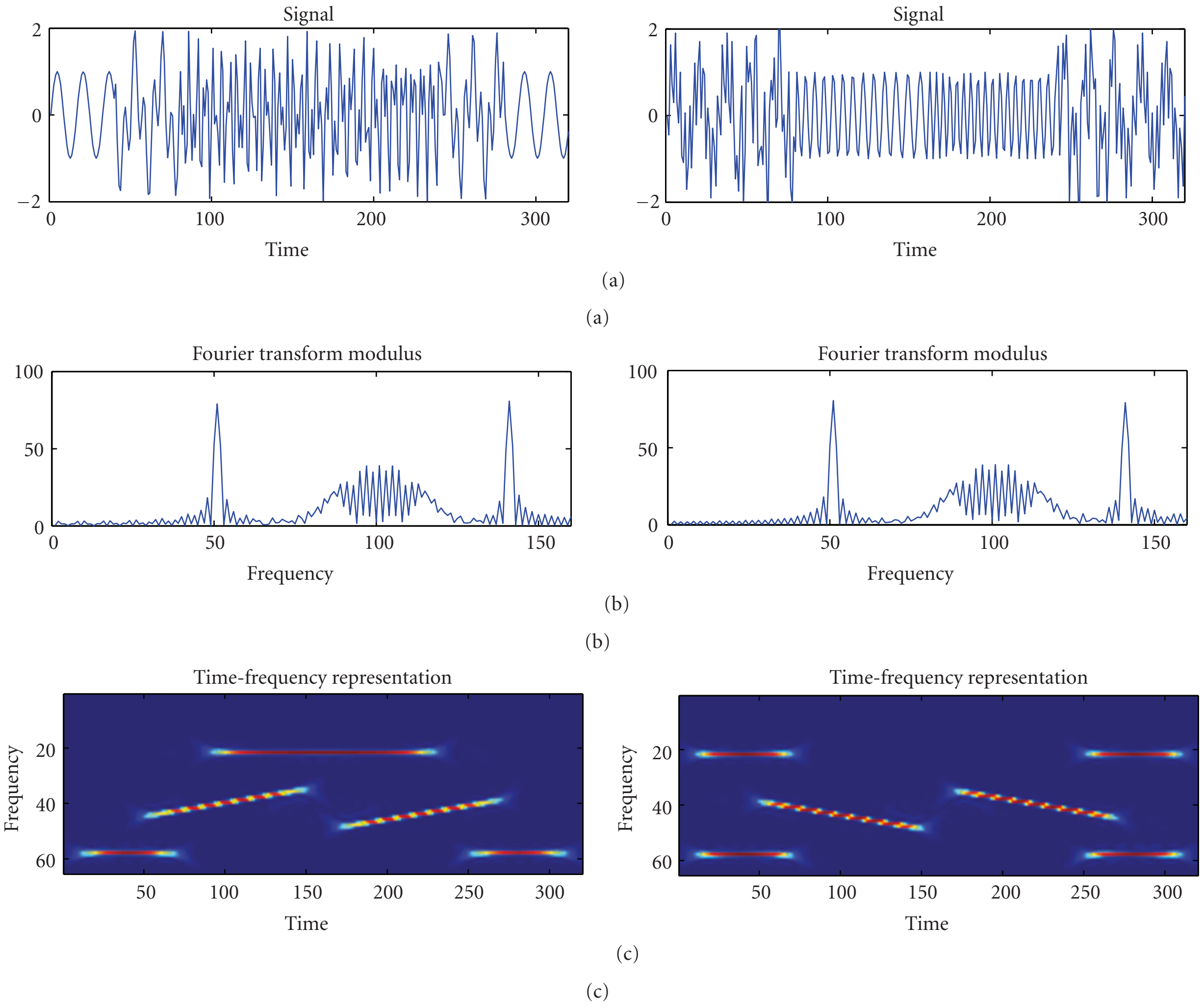
Copyright © 2010 Srdjan Stanković. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

A review of time-frequency analysis and some aspects of its applications in digital watermarking are presented. The main advantages and drawbacks of various time-frequency distributions are first discussed. The aim of this theoretical overview is to facilitate an appropriate distribution selection in a specific application. Different aspects of the time-frequency analysis when applied to digital watermarking are then presented. In particular, the method that maps time-frequency characteristics of a host signal to the pseudo noise watermark sequence is thoroughly discussed. This approach is presented in the multidimensional form and then applied to digital audio, digital image, and digital video watermarking. Finally, the theoretical considerations are illustrated by various numerical and real-life examples.

1. Introduction

Theoretical aspects of time-frequency analysis have been intensively studied over the last two decades [1–27]. In parallel, their various applications have been exploited as well. Namely, for an efficient analysis of nonstationary signals, such are radar, sonar, biomedical, seismic, and multimedia signals, time-frequency representations are required. Time-frequency distributions are most commonly used for this purpose. Many of the researchers have made significant efforts in defining a distribution that is optimal for a wide class of frequency-modulated signals [8–11]. As a result, a number of time-frequency distributions have been proposed. However, the efficiency of each of them is more or less limited to a specific class of signals and, consequently, to a specific application. One of the goals of this paper is to highlight the most important features of some popular time-frequency distributions and to give an idea of how to choose the most appropriate distribution depending on the signal form. The linear, quadratic, higher-order, and multiwindow time-frequency distributions are considered. The short-time Fourier transform, as the most commonly used linear transform, is firstly discussed. Next, the Wigner distribution, as the best known quadratic distribution, is presented. Also, the Cohen class and some specific distributions belonging to this class are considered [1, 5–7]. It is shown that the quadratic distributions are optimal for a linear frequency-modulated signal. However, if the instantaneous frequency variation within the analysis window is faster, multiwindow, or higher order distributions should be used, [14–16, 19–27]. The Hermite functions-based multiwindow approach is also discussed. Finally, highly concentrated distributions with complex-lag argument are presented. To facilitate in better understanding of the presented theoretical considerations, numerous illustrative examples have been provided.

The second part of the paper considers time-frequency-based watermarking techniques. The watermarking of digital audio, digital image, and digital video is discussed [28–37]. A short overview of some existing approaches is first given and they are related to digital audio and digital image [28–50]. Watermarking using time-frequency techniques is usually employed in either of the following two ways. The first one uses the time/space domain of the host signal to embed the watermark with specific time-frequency characteristics. The time-frequency analysis is then used for detection. The second way uses time-frequency distributions to create or embed watermark in the time-frequency domain. A flexible procedure that can be used for different kinds of signals is discussed more extensively. Therein, the watermark is shaped according to the time-frequency characteristics of



(a)

(a)

(b)

(b)

(c)

(c)

Figure 1: Nonstationary signals. (a) Time domain representations. (b) Fourier domain representations. (c) Spectral components' variations along the time axes.

the host signal. The detection is performed in the time-frequency domain. This particular approach is presented in the multidimensional form, and it is applied to digital audio, digital image, and digital video. It provides a high degree of robustness and imperceptibility. Hence, even when the watermark is very weak, a reliably detection can still be achieved. Also, the watermark gets completely hidden by the time-frequency characteristics of the host signal.

2. Time-Frequency Analysis

The Fourier transform provides spectral content of a signal. It has been a valuable tool in various applications. However, for nonstationary signals the Fourier transform cannot give satisfactory results since the information about frequency components variations in time is required.

Furthermore, it can happen that two different signals have the same spectral contents, as illustrated in Figures 1(a) and 1(b). Based on Figure 1(c), however, we can conclude that the time-frequency representations of the two signals are quite different. This example is a simple illustration of the importance of time-frequency analysis for signals whose spectral contents vary with time. Various time-frequency distributions are used for this purpose. The ideal time-frequency representation can be described as [19–21]

$$\mathrm{ITF}(t,\omega) = 2\pi A^2 \delta(\omega - \Phi'(t)), \tag{1}$$

where a signal of the form $x(t) = Ae^{j\Phi(t)}$ is considered. This representation provides the signal local energy distribution, as well.

A question that naturally arises at this point is whether there exist a single representation that would be ideal for any signal at hand. The answer is no, hence a number of time-frequency distributions have been introduced.

2.1. The Short-Time Fourier Transform. The simplest and most commonly used time-frequency representation is

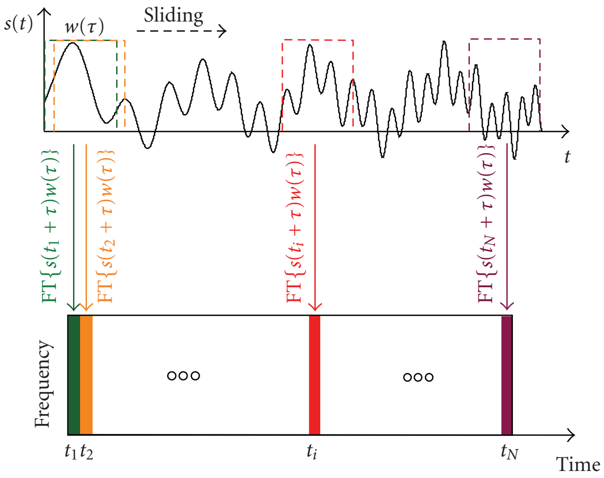

$s(t)$
$w(\tau)$
Sliding
t
$\mathrm{FT}\{s(t_1+\tau)w(\tau)\}$
$\mathrm{FT}\{s(t_2+\tau)w(\tau)\}$
$\mathrm{FT}\{s(t_i+\tau)w(\tau)\}$
$\mathrm{FT}\{s(t_N+\tau)w(\tau)\}$
Frequency
ooo
ooo
$t_1 t_2$
t_i
t_N
Time

Figure 2: Illustration of the STFT calculation.

obtained by using the short-time Fourier transform (STFT) defined as [2]

$$\mathrm{STFT}(t,\omega)=\int_{-\infty}^{\infty}x(t+\tau)w(\tau)e^{-j\omega\tau}d\tau. \tag{2}$$

Thus, it is a windowed version of the Fourier transform. The sliding window function is denoted by $w(\tau)$, where τ is the lag coordinate. An illustration of the STFT calculation is shown in Figure 2.

Note that the spectral content is calculated for each windowed part of the signal. The central point of the sliding window is the time instant for which the spectrum is calculated. The influence of the window size is critical, as it will be discussed later.

The energetic version of the STFT is known as the spectrogram. It can be written as

$$\begin{aligned}\mathrm{SPEC}(t,\omega)&=|\mathrm{STFT}(t,\omega)|^2\\&=2\pi A^2W(\omega-\Phi'(t))*_{\omega}\mathrm{FT}\left\{e^{j(Q(t,\tau))}\right\},\end{aligned} \tag{3}$$

where $W(\omega)$ is the Fourier transform of the time domain window $w(t)$, $\Phi'(t)$ is the first derivative of the phase function, the frequency convolution is denoted by $*\omega$, while $Q(t,\tau)=\Phi^{(2)}(t)(\tau^2/2!)+\Phi^{(3)}(\tau^3/3!)+\cdots+\Phi^{(n)}(\tau^n/n!)$ is the spread factor which depends on the second and higher order phase derivatives. Note that an ideal representation will be obtained if the signal is constant frequency modulated ($\Phi^{(i)}(t)=0$, for $i\geq 2$) and if $W(\omega-\Phi'(t))\rightarrow\delta(\omega-\Phi'(t))$, that is, for a large time domain window. However, if the signal is not constant frequency modulated, a large window size can produce a low time resolution and vice versa. In general, there is a trade-off between the time and frequency resolution, and it is best described with the uncertainty principle, as

$$M_TM_w\geq\frac{1}{2}, \tag{4}$$

where,

$$M_T=\frac{\int_{-\infty}^{\infty}\tau^2|w(\tau)|^2d\tau}{\int_{-\infty}^{\infty}|w(\tau)|^2d\tau},\qquad M_w=\frac{\int_{-\infty}^{\infty}\omega^2|W(\omega)|^2d\omega}{\int_{-\infty}^{\infty}|W(\omega)|^2d\omega} \tag{5}$$

are the measures of duration in time and frequency, respectively. The signal should satisfy $w(t)\sqrt{t}\rightarrow 0$ as $t\rightarrow\pm\infty$.

An illustration of the window size influence on the spectrogram resolution is given in Figure 3. A four-component signal is considered. In order to achieve a good time resolution, two short sinusoidal components should be analyzed by using a narrow window. However, to obtain a good frequency resolution the third component (a sinusoid with a long duration) should be analyzed by using a wide window. In Figure 3(a), a narrow window is used, and it results in a good time resolution, while the frequency resolution is low. However, when a large window size is used, a low time resolution is obtained. Hence, the two short sinusoids have almost merged into one (see Figure 3(b)). Due to the presence of a linear frequency-modulated component (chirp signal), the spread factors are present in both cases.

The spectrogram satisfies the marginal properties

$$\begin{gathered}\int_t\mathrm{SPEC}(t,\omega)dt=|X(\omega)|^2,\\\frac{1}{2\pi}\int_\omega\mathrm{SPEC}(t,\omega)d\omega=|x(t)|^2.\end{gathered} \tag{6}$$

The total energy is obtained as

$$Ex=\frac{1}{2\pi}\int_t\int_\omega\mathrm{SPEC}(t,\omega)dt\,d\omega. \tag{7}$$

An important property of the STFT is its linearity. Namely, the STFT of a multicomponent signal $x(t)=\sum_{i=1}^{M}x_i(t)$ is $\mathrm{STFT}_x(t,\omega)=\sum_{i=1}^{M}\mathrm{STFT}_{x_i}(t,\omega)$. Consequently, if the signal components do not intersect in the time-frequency plane the spectrogram will be equal to the sum of spectrograms of each of the signal components. This is evident from Figure 3.

2.2. Quadratic Time-Frequency Distributions.

Quadratic time-frequency distributions have been introduced in order to improve the time-frequency resolution. Namely, they remove the spread factors for linear frequency-modulated signals. Among them, the Wigner distribution is the most commonly used. It has also been used as a base to define several interesting time-frequency distributions.

The windowed version of the Wigner distribution is called the pseudo-Wigner distribution. It is defined as [1, 2]

$$\begin{aligned}&\mathrm{WD}(t,\omega)\\&\quad=\int_{-\infty}^{\infty}x\left(t+\frac{\tau}{2}\right)x^*\left(t-\frac{\tau}{2}\right)w\left(\frac{\tau}{2}\right)w^*\left(-\frac{\tau}{2}\right)e^{-j\omega\tau}d\tau.\end{aligned} \tag{8}$$

The spread factor in this case is $Q(t,\tau)=\Phi^{(3)}(t)\tau^3/(2^23!)+\Phi^{(5)}\tau^5/(2^45!)+\cdots$.

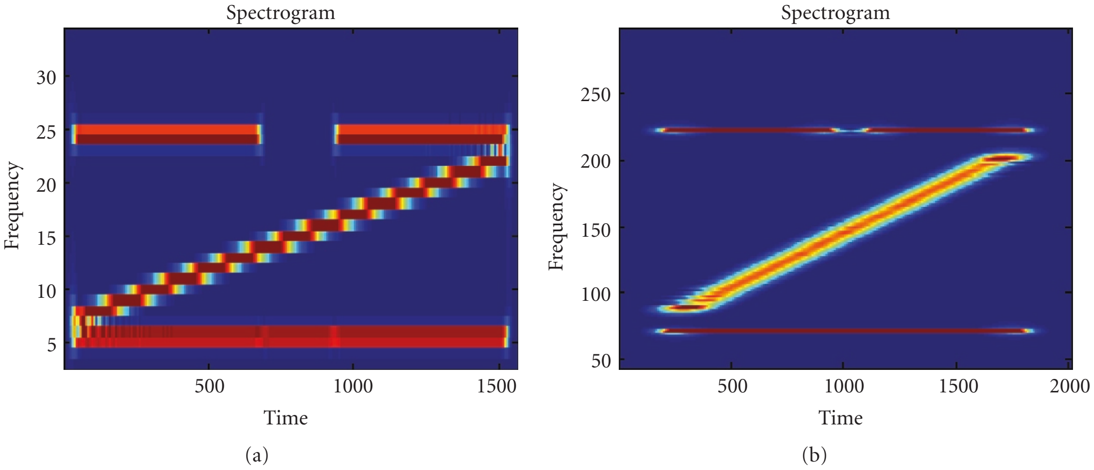


FIGURE 3: STFT of the four-component signal calculated by: (a) narrow window, (b) large window.

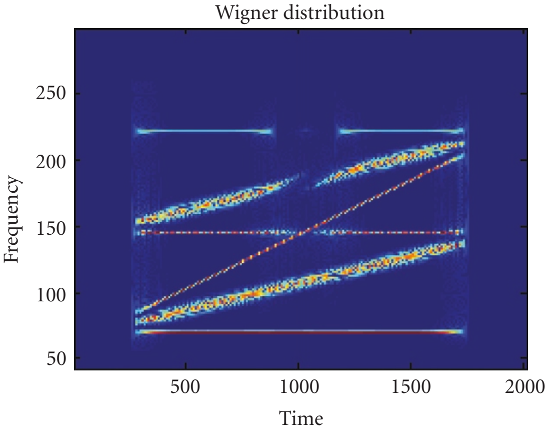


FIGURE 4: Wigner distribution of a four-component signal.

The Wigner distribution satisfies the marginal conditions. Note that it is always real, while the numerical realization requires oversampling with factor 2. However, the Wigner distribution is not linear. Namely, for a multicomponent signal $x(t) = \sum_{i=1}^{M} x_i(t)$, the Wigner distribution is of the form

$$\mathrm{WD}(t,\omega) = \sum_{i=1}^{M} \mathrm{WD}_{x_i x_i}(t,\omega) + \sum_{\substack{i=1 \\ i \neq k}}^{M} \sum_{k=1}^{M} \mathrm{WD}_{x_i x_k}(t,\omega). \tag{9}$$

Thus, beside the autoterms, the interactions between different signal components ($\mathrm{WD}_{x_i x_k}(t,\omega)$), called cross-terms, appear. This is a major drawback of this distribution. The Wigner distribution of the four-component signal, from the previous example, is given in Figure 4.

In this case, the auto components are well concentrated. However, the cross-terms presence is significant. Namely, the time-frequency representation contains frequency components that do not exist within the signal itself. It could lead to a wrong analysis result.

The cross-terms could be reduced by using a filter function in the ambiguity domain. The ambiguity function is the two dimensional Fourier transform of the Wigner distribution, that is,

$$A(\tau,\theta) = \mathrm{FT}_{2\mathrm{D}}\{\mathrm{WD}(t,\omega)\}. \tag{10}$$

In the ambiguity domain, the auto-terms are located around the origin. Thus, a two-dimensional filter, called the kernel function, is used to filter out the cross-terms (that are generally located away from the origin) [1, 5–7]:

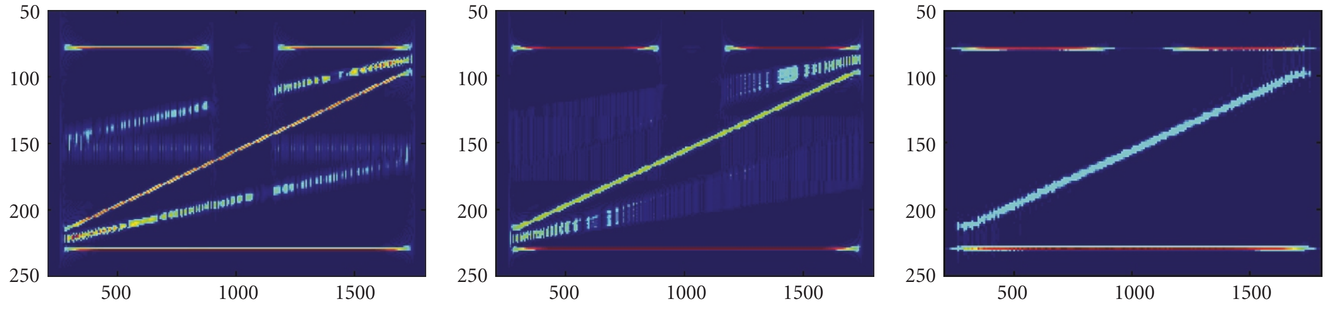
$$A_f(\tau,\theta) = A(\tau,\theta)c(\tau,\theta), \tag{11}$$

where $A(\tau,\theta) = \int_{-\infty}^{\infty} x(t+\tau/2)x^*(t-\tau/2)e^{-j\theta t}dt$ is the ambiguity function, and $c(\tau,\theta)$ is the kernel function. The time-frequency distribution based on the filtered function is obtained as

$$\begin{aligned} \mathrm{CD}(t,\omega) &= \mathrm{IFT}\{A(\tau,\theta)c(\tau,\theta)\} \\ &= \frac{1}{2\pi}\iiint_{-\infty}^{\infty} c(\tau,\theta)x\left(u+\frac{\tau}{2}\right)x^*\left(u-\frac{\tau}{2}\right) \\ &\quad \times e^{-j\theta t}e^{-j\omega\tau}e^{j\theta u}d\theta\, du\, d\tau. \end{aligned} \tag{12}$$

This is the definition of the Cohen class of distributions. By choosing the corresponding kernel functions, some specific distributions belonging to the Cohen class can be obtained. They satisfy the marginal properties if $c(0,\theta) = c(\tau,0) = 1$ holds. For example, the Choi-Williams distribution [5] is obtained for the kernel function $c(\tau,\theta) = e^{-(\tau^2\theta^2/\sigma^2)}$ where σ is a scaling factor to control its attenuation rate. The kernels for some distributions are defined in Table 1, [7].

The four-component signal analyzed by the Choi-Williams and Born-Jordan distribution is shown in Figure 5. The Choi-Williams distribution with two values of its kernel parameter σ is used. The parameter $\sigma = 2\pi$ results in a wider width of the kernel assuring good auto-terms concentration, but a significant amount of cross-terms is still present (Figure 5(a)). The Born-Jordan distribution



(a)

(b)

(c)

Figure 5: Time-frequency representations by: (a) Choi-Williams distribution ($\sigma = 2\pi$), (b) Born-Jordan distribution, and (c) Choi-Williams distribution ($\sigma = 0.5$).

Table 1: Some distributions from the Cohen class.

Distribution	Kernel
Choi-Williams	$e^{-\theta^2\tau^2/\sigma^2}$
Born-Jordan	$(\sin(\theta\tau/2))/\theta\tau/2$
Zao-Atlas-Marks	$\vert\tau\vert(\sin(\theta\tau/2)/\theta\tau/2)w(\tau)$
Sinc distribution	$\mathrm{rect}(\theta\tau/\alpha)$
Wigner distribution	$w^2(\tau/2)$

producing almost the same auto-terms concentration is shown in Figure 5(b). Note that the side-lobes and cross-terms are more emphasized than in the Choi-Williams distribution with $\sigma = 2\pi$ [7]. However, if the Choi-Williams distribution with the attenuation parameter $\sigma = 0.5$ is used (Figure 5(c)), the cross-terms get suppressed, while the auto-term concentration becomes reduced (for the chirp component).

Note that there is a trade-off between the cross-terms reduction and auto-term concentration (it depends on parameters of the kernel function). Obviously, distributions belonging to the Cohen class lie in between the two extreme cases: the spectrogram that eliminates the cross-terms with a low auto-term concentration and the Wigner distribution that provides high resolution, but with emphatic cross-terms. It is possible to obtain an ideal time-frequency concentration only if signal dependent kernels are used [8, 10].

Next, we can ask the following question. Is there a distribution that provides the auto-terms concentration as good as in the Wigner distribution, while eliminating the cross-terms like the spectrogram does? In order to define a distribution with these properties, let us start with the following relationship between the short-time Fourier transform and the Wigner distribution:

$$\mathrm{WD}(t,\omega) = \frac{1}{\pi}\int_{-\infty}^{\infty} \mathrm{STFT}(t,\omega+\theta)\mathrm{STFT}^*(t,\omega-\theta)d\theta. \quad (13)$$

Clearly, the convolution along the frequency axis improves the auto-terms concentration, but it introduces the cross-terms. Thus, the convolution should be performed only over the same auto-terms, avoiding different signal components being convolved. It can be performed by introducing a frequency domain window (see Figure 6).

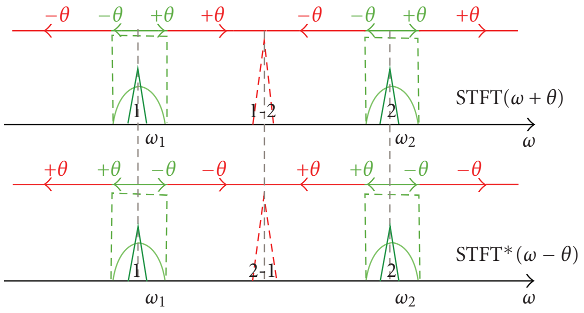


Figure 6: Illustration of the STFTs convolution.

A distribution that allows for such convolution is called the S-method [17]. It is defined as

$$\mathrm{SM}(t,\omega) = \frac{1}{\pi}\int_{-\infty}^{\infty} P(\theta)\mathrm{STFT}(t,\omega+\theta)\mathrm{STFT}^*(t,\omega-\theta)d\theta. \quad (14)$$

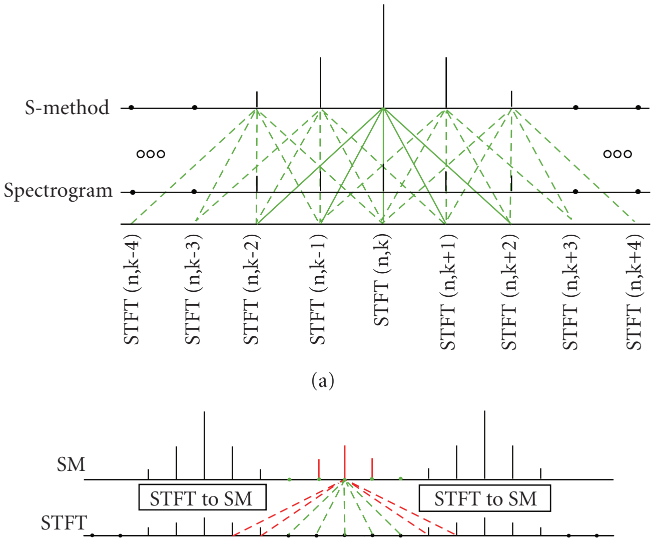
The frequency domain finite window is denoted by $P(\theta)$ ($P(\theta) = 0$ for $|\theta| > L$). Observe that for $P(\theta) = \pi\delta(\theta)$ and $P(\theta) = 1$ the spectrogram and the Wigner distribution are obtained, respectively.

The discrete version of the S-method is

$$\begin{aligned}\mathrm{SM}(n,k) &= \sum_{i=-L}^{L} \mathrm{STFT}(n,k+i)\mathrm{STFT}^*(n,k-i)\\ &= \mathrm{SPEC}(n,k)\\ &\quad + 2\,\mathrm{Re}\left\{\sum_{i=1}^{L}\mathrm{STFT}(n,k+i)\mathrm{STFT}^*(n,k-i)\right\}.\end{aligned} \quad (15)$$

The discrete window width is $2L+1$. It determines the number of summation terms in (15). Note that they improve the spectrogram concentration toward that of the Wigner distribution. An illustration of the S-method calculation is shown in Figure 7.

The calculation of the S-method is illustrated for the central point of an auto-term as well as for the point located in between the two auto components (Figures 7(a) and



(a)

(b)

Figure 7: Illustration of the S-method calculation for a given time instant n.

7(b)). It is important to observe that the summation has to be performed only over the auto-terms. If the other terms are included, the concentration will not be improved. In addition, the noise could be also picked up. The window has to be narrower than the minimal distance between the auto-terms. If this is not the case, the interactions between the auto-terms will produce the cross-terms. Namely, as illustrated in Figure 7(b), the cross-terms appear if the window includes the summation terms marked by red color.

The adaptive S-method with a variable window adjusted to the auto-terms is introduced in [18]. However, in many applications the fixed window size of $L = 3$ has been shown to provide very good results, since the convergence within the window is fast, and it is mostly achieved after a few summation terms.

The S-method of the four-component signal is given in Figure 8.

Note that all signal components (with constant and linear frequency modulations) are well concentrated even for $L = 3$. By increasing L, for $L > 5$ the cross-terms start to appear (the minimal distance between the auto-terms becomes less than $2L + 1$).

The spectrogram and the S-method ($L = 3$) of a real speech signal are shown in Figure 9.

The speech signal time-frequency resolution is improved by using the S-method.

Observe that the spread factor in the quadratic distributions will be present if the instantaneous frequency contains third and higher order phase derivatives. Hence, further concentration improvement can be obtained by using the multiwindow approach or by using higher order time-frequency distributions, as discussed below.

2.3. Multiwindow Time-Frequency Distributions. The concept of multiwindow time-frequency distributions has been developed by using optimally concentrated orthogonal windows [25–27]. The Hermite functions that are localized in both time and frequency domain can be used as orthogonal windows. The multiwindow spectrogram is defined as a weighted sum of the spectrograms:

$$\begin{aligned}\mathrm{SPEC}_{\mathrm{MW}}(t,\omega) &= \sum_{p=0}^{K-1} c_p(t)\mathrm{SPEC}_p(t,\omega)\\ &= \frac{1}{2\pi}\sum_{p=0}^{K-1} c_p(t)\left|\int x(\tau)\Psi_p(\tau-t)e^{-j\omega\tau}d\tau\right|^2.\end{aligned} \tag{16}$$

The total number of the spectrograms and Hermite functions is K, while $c_p(t)$ are the weighting coefficients. The pth order Hermite function is defined as

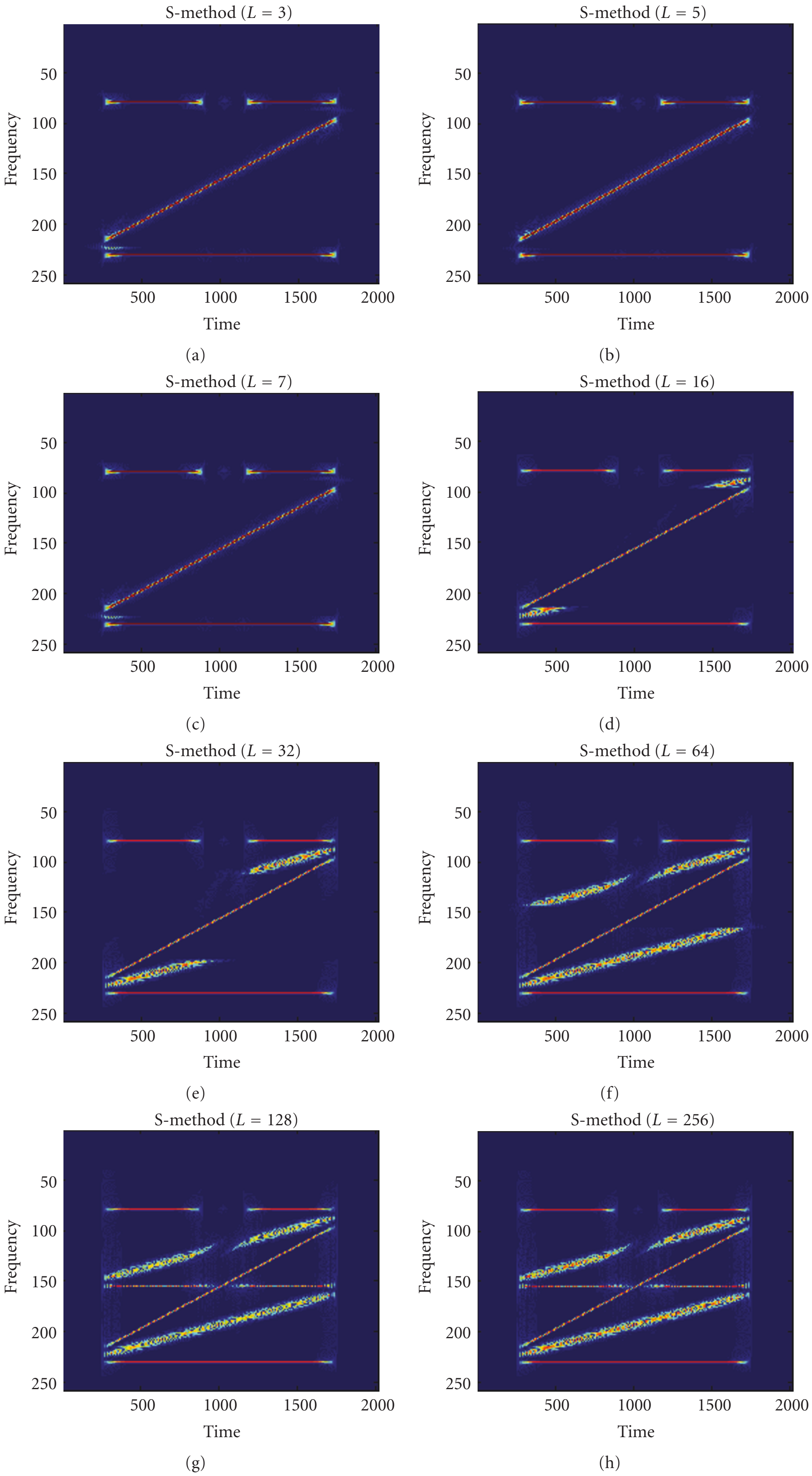
$$\Psi_p(t) = \frac{(-1)^p e^{t^2/2}}{\sqrt{2^p p!\sqrt{\pi}}}\frac{d^p\left(e^{-t^2}\right)}{dt^p}. \tag{17}$$

This function can be obtained by using a recursive realization as follow

$$\begin{aligned}\Psi_p(t) &= t\sqrt{\frac{2}{p}}\Psi_{p-1}(t)\\ &\quad -\sqrt{\frac{p-1}{p}}\Psi_{p-2}(t), \quad \forall p \geq 2,\end{aligned} \tag{18}$$

while:

$$\Psi_0(t) = \frac{1}{\sqrt[4]{\pi}}e^{-(t^2/2)}, \qquad \Psi_1(t) = \frac{\sqrt{2}t}{\sqrt[4]{\pi}}e^{-(t^2/2)}. \tag{19}$$



S-method (L = 3)

(a)

S-method (L = 5)

(b)

S-method (L = 7)

(c)

S-method (L = 16)

(d)

S-method (L = 32)

(e)

S-method (L = 64)

(f)

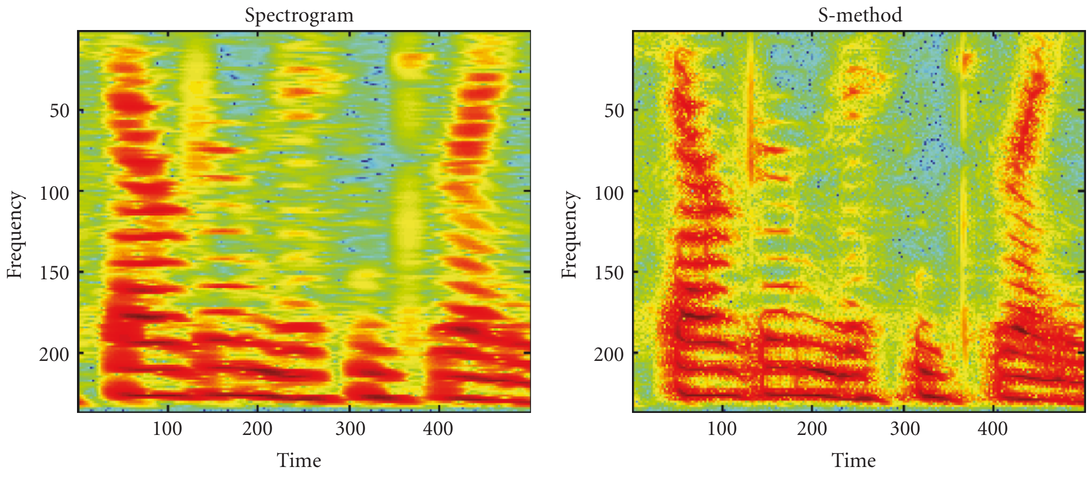
S-method (L = 128)

(g)

S-method (L = 256)

(h)

Figure 8: The S-method of the multicomponent signal (various window sizes L are used).



(a)

(b)

Figure 9: Time-frequency representations of a speech signal: (a) spectrogram, (b) S-method.

The weighting coefficients are calculated by

$$\sum_{p=0}^{K-1} c_p(t)\frac{\int A^2(t+\tau)\Psi_p^2(\tau)\tau^i d\tau}{\int A^2(t+\tau)\Psi_p^2(\tau)d\tau} = \begin{cases}1, & \text{for } i=0\\ 0, & \text{for } i>0,\end{cases} \tag{20}$$

where $A(t)$ is the signal amplitude. The weighting coefficients for a signal with a constant amplitude are given in Table 2.

It is important to emphasize that the spread factor is reduced proportionally to the highest order of Hermite functions used in the multiwindow spectrogram:

$$Q(t,\tau) = \Phi^{(K+1)}(t)\frac{\tau^{K+1}}{(K+1)!} + \Phi^{(K+2)}(t)\frac{\tau^{K+2}}{(K+2)!} + \cdots \tag{21}$$

Thus, the first term in the spread factor comes from the $(K+1)$th phase derivative. An additional concentration improvement can be achieved by introducing the multiwindow S-method [51]. It can be written in the discrete domain as

$$\begin{aligned}&\mathrm{SM}_{\mathrm{MW}}(n,k)\\ &\quad= \sum_{p=0}^{K-1} c_p(n)\mathrm{SPEC}_{\mathrm{MW}}(n,k)\\ &\quad+ \sum_{p=0}^{K-1} 2\,\mathrm{Re}\left\{\sum_{i=1}^{L} c_p(n)\mathrm{STFT}_p(n,k+i)\mathrm{STFT}_p^*(n,k-i)\right\}.\end{aligned} \tag{22}$$

In this case the spread factor is

$$\begin{aligned}Q(t,\tau) &= \Phi^{(K+1)}(t)\frac{\tau^{K+1}}{2^K(K+1)!}\\ &\quad+ \Phi^{(K+3)}(t)\frac{\tau^{K+3}}{2^{K+2}(K+3)!}\\ &\quad+ \cdots, \quad \text{for even } K,\\ Q(t,\tau) &= \Phi^{(K+2)}(t)\frac{\tau^{K+2}}{2^{K+1}(K+2)!}\\ &\quad+ \Phi^{(K+4)}(t)\frac{\tau^{K+4}}{2^{K+3}(K+4)!}\\ &\quad+ \cdots, \quad \text{for odd } K.\end{aligned} \tag{23}$$

An example, where the multiwindow spectrogram and the multiwindow S-method are used is shown in Figure 10 (the components are with constant amplitude and each of them is treated separately).

From Figure 10 it is obvious that the multiwindow versions of distributions outperform their standard counterparts. The multiwindow approach reduces the noise influence as well [27].

This multiwindow approach can also be interpreted by using the Cohen class of distributions, where it can be written as a two-dimensional convolution of the Wigner distribution and the kernel function: $\mathrm{WD}(t,\omega)_{**}\Phi(t,\omega)$. The kernel function producing the multiwindow Wigner distribution is obtained as

$$\Phi(t,\omega) = \sum_{p=0}^{K-1} c_p(t,\omega)L_p(t,\omega), \tag{24}$$

where L_p is the pth order Laguerre function, and it is the Wigner distribution of the pth order Hermite function.

According to the previous consideration, the spread factor can be gradually reduced by increasing the number

Table 2: The weighting coefficients for $K = 1, 2, \ldots, 11$.

	c_0	c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8	c_9	c_{10}
$K = 1$	1										
$K = 2$	1.5	−0.5									
$K = 3$	1.75	−1	0.25								
$K = 4$	1.875	−1.375	0.625	−0.125							
$K = 5$	1.937	−1.625	1	−0.375	0.062						
$K = 6$	1.968	−1.781	1.312	−0.687	0.219	−0.031					
$K = 7$	1.984	−1.875	1.546	−1	0.453	−0.125	0.016				
$K = 8$	1.992	−1.929	1.710	−1.273	0.727	−0.289	0.070	−0.008			
$K = 9$	1.996	−1.961	1.820	−1.492	1	−0.507	0.179	−0.039	0.003		
$K = 10$	1.998	−1.978	1.890	−1.656	1.246	−0.754	0.344	−0.109	0.021	−0.002	
$K = 11$	1.900	−1.561	0.955	−0.223	−0.357	0.573	−0.460	0.237	−0.079	0.016	−0.001

of Hermite functions, that is, by increasing the number of spectrograms in (16) and (22). Similar resolution improvements can be obtained by using polynomial time-frequency distributions, where each additional order of the distribution results in a removal of one more term within the spread factor [3].

2.4. Distributions with Complex-Lag Argument. A significant components concentration improvement can be achieved by introducing higher order distributions with the complex-lag argument [19, 21]. A general form of these distributions is [22]

$$\begin{aligned}\mathrm{GCD}(t,\omega) &= \int_{-\infty}^{\infty} x\left(t+\frac{\tau}{N}\right)x^{*}\left(t-\frac{\tau}{N}\right) \\ &\quad\times \prod_{i=1}^{N/2-1}\left[x^{(a_i+jb_i)}\left(t+\frac{\tau}{N(a_i+jb_i)}\right)\right. \\ &\quad\left.\times\, x^{-(a_i+jb_i)}\left(t-\frac{\tau}{N(a_i+jb_i)}\right)\right]e^{-j\omega\tau}d\tau. \end{aligned} \tag{25}$$

A special case follows for [19]

$$\begin{aligned} a_i + jb_i &= e^{j2\pi k/N}, \\ \text{for } i &= 1,2,\ldots,\frac{N}{2}-1, \\ k &= 1,2,\ldots,N-1. \end{aligned} \tag{26}$$

The fourth-order distribution ($N = 4$) is obtained for $i = 1$, that is, $\pm(a_i + jb_i) = \pm e^{j\pi/2} = \pm j$. Thus, it has the form

$$\begin{aligned}\mathrm{GCD}_4(t,\omega) &= \int_{-\infty}^{\infty} x\left(t+\frac{\tau}{4}\right)x^{*}\left(t-\frac{\tau}{4}\right) \\ &\quad\times x^{-j}\left(t+j\frac{\tau}{4}\right)x^{j}\left(t-j\frac{\tau}{4}\right)e^{-j\omega\tau}d\tau. \end{aligned} \tag{27}$$

The term with the complex-lag argument is calculated by

$$\begin{aligned} &x^{-j}(t+j\tau)x^{j}(t-j\tau) \\ &\quad= e^{j\ln\left|\int_{-\infty}^{\infty}\mathrm{STFT}(t,\omega)e^{j\omega(t-j\tau)}d\omega/\int_{-\infty}^{\infty}\mathrm{STFT}(t,\omega)e^{j\omega(t+j\tau)}d\omega\right|}. \end{aligned} \tag{28}$$

For a multicomponent signal, it is calculated for each component separately, where the STFT is used to separate them [20]. This procedure can be generalized for an arbitrary distribution order [24].

For $i = 1, 2$ and $N = 6$ we have $\pm(a_i + jb_i) = \pm e^{j\pi/3}, \pm e^{j2\pi/3}$. It defines the sixth-order distribution. Observe that, regarding the spread factor reduction, each distribution order is related to the previous one in the same way the Wigner distribution is related to the spectrogram. Namely, the spread factors for the fourth- and the sixth-order distributions are

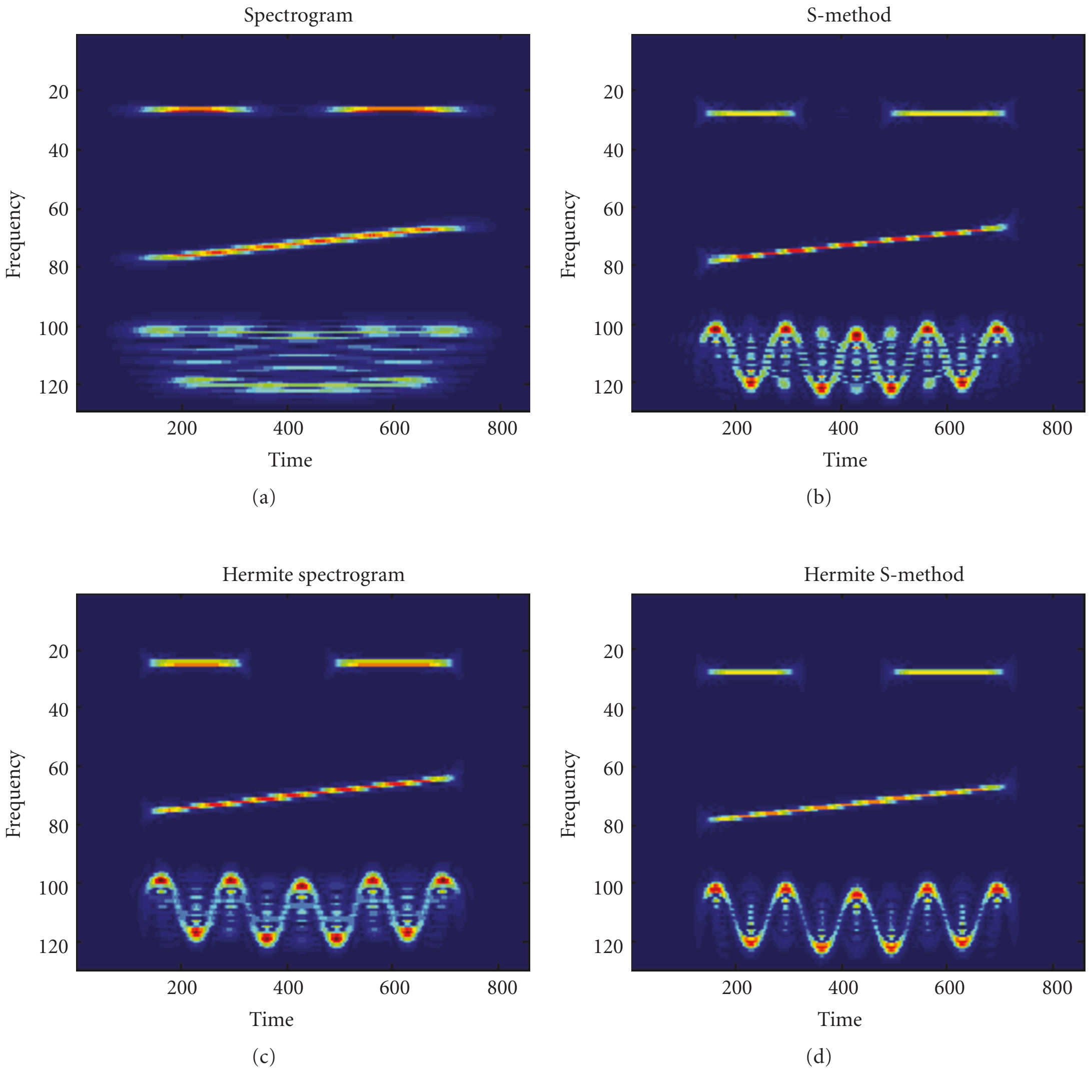
$$\begin{aligned} Q_4(t,\tau) &= \Phi^{(5)}(t)\frac{\tau^5}{5!4^4} + \Phi^{(9)}(t)\frac{\tau^9}{9!4^8} + \cdots, \\ Q_6(t,\tau) &= \Phi^{(7)}(t)\frac{\tau^7}{7!6^6} + \Phi^{(13)}(t)\frac{\tau^{13}}{13!6^{12}} + \cdots, \end{aligned} \tag{29}$$

The complex-lag distributions are in particular useful when the instantaneous frequency variation within the window is very fast. The examples where the distribution order is increased in order to improve time-frequency resolution are shown in Figure 11.

Note that the Wigner distribution produces poor results for both signals, since it cannot follow the instantaneous frequency variations.

3. Digital Watermarking

Digital watermarking has been used to protect multimedia data. Demands in this area increase proportionally with the number of internet applications. Namely, these applications are associated with a need for copyright protection of digital audio, digital image, and digital video. Note that the cryptographic methods could be used for this purpose. However, once the data are decoded they can be unlimitedly copied. This has been one of the primary reasons for


Spectrogram

S-method

Hermite spectrogram

Hermite S-method

Figure 10: Time-frequency representations by using: (a) spectrogram, (b) S-method, (c) multiwindow spectrogram and (d) multiwindow S-method.

developing the watermarking techniques. The watermarking, in general, consists of embedding a secret information that can be reliably detected within the host signal. Obviously, this information should be imperceptible within the host data. Depending on the application type, the watermarking can be robust, fragile, or semifragile. The robust watermark should be resistant to various nonmalicious or malicious attacks. Nonmalicious attacks are commonly used signal processing techniques such as compression algorithms, filtering, and so forth, while the malicious attacks are the signal processing techniques that are intentionally used to remove the watermark. The fragile watermark is used to prove data authenticity. Thus, if the content of a signal has been changed, the watermark should no longer exist. The semifragile watermark should be robust to a slight modification, such as for example a certain degree of compression.

Depending on the type of host signal (speech/audio signals, image, video, etc.) various watermarking approaches are developed. Also, different domains have been used: the time domain (or the space domain), the spectral domains such as DFT, DWT, and DCT domain, and a joint time/space-frequency domain. The existing watermarking techniques are mainly based on either the time or frequency domain. However, in both cases, the time-frequency characteristics of the watermark do not correspond to the time-frequency characteristics of the host signal. It may result in the watermark being not imperceptible, because it is present in the time-frequency regions where the signal components do not exist.

3.1. An Overview of Some Time-Frequency-Based Watermarking Techniques. The time-frequency domain can be very efficient regarding the watermark imperceptibility and robustness. This section presents some key time-frequency-based watermarking procedures with the aim to inspire more contributions on this topic.

Here, we will classify the time-frequency-based watermarking techniques into two categories. The first one is

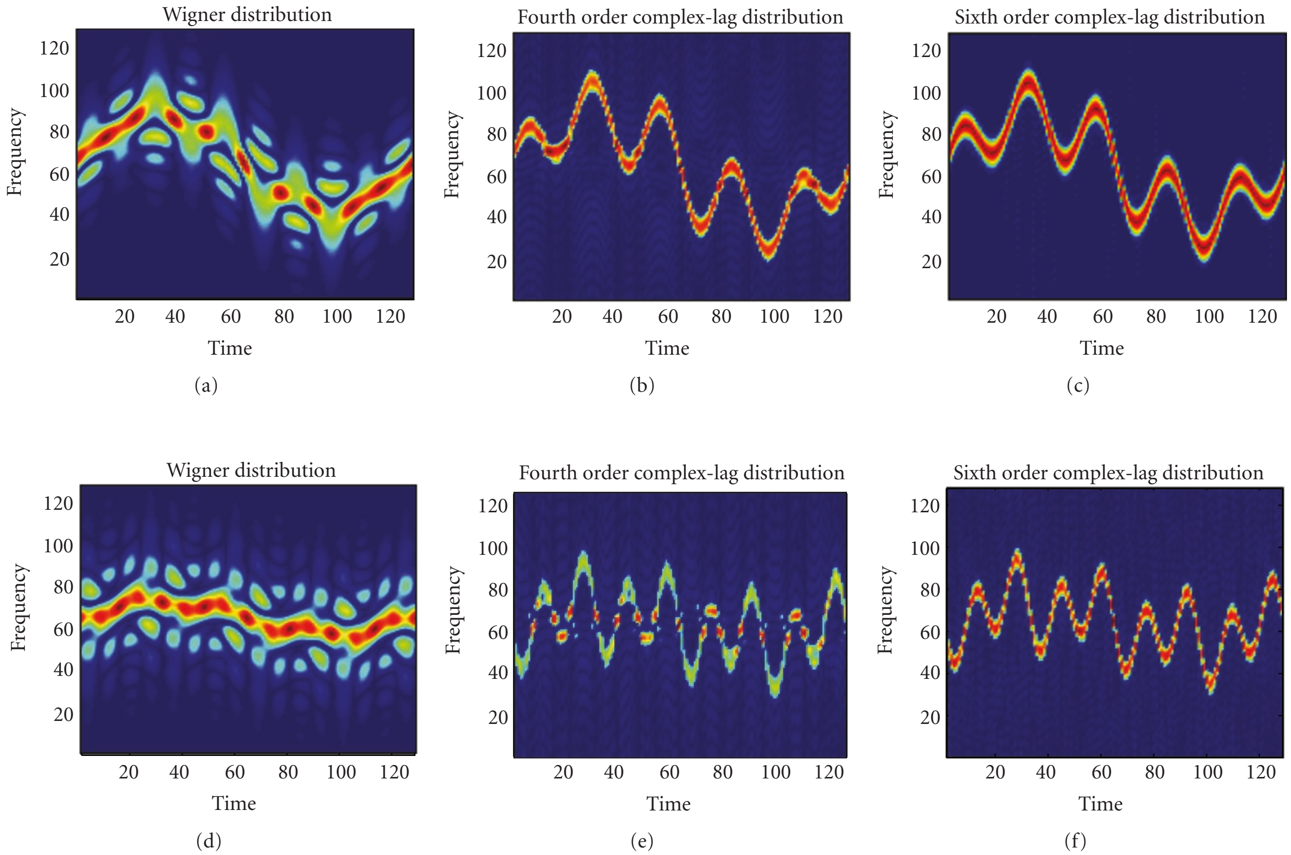


Figure 11: Time-frequency representation of signals with fast instantaneous frequency variation obtained by using: the Wigner distribution, the fourth-order complex-lag distribution, and the sixth-order complex-lag distribution.

the approaches based on watermarks with specific time-frequency characteristics, where the detection procedure is performed within the time-frequency domain. The second one uses the time-frequency domain to embed or to shape the watermark.

(A) Image Watermark with Specific Time-Frequency Characteristics.

(A.1) Among the first time-frequency-based image watermarking procedures is the approach introduced in [38]. Although the watermark is embedded in the space domain it is chosen to have a specific space/spatial-frequency characteristic. Namely, a two-dimensional chirp signal is used as watermark:

$$\begin{aligned} W(x,y) &= 2A\cos(ax^2+by^2) \\ &= A\left(e^{j(ax^2+by^2)}+e^{-j(ax^2+by^2)}\right). \end{aligned} \tag{30}$$

Observe that the Wigner distribution provides an ideal representation for this signal.

The watermark is embedded within the entire image: $I_w(x,y) = I(x,y) + W(x,y)$.

The watermark detection is performed by using

$$\begin{aligned} &P\left(\omega_x,\omega_y;W_v\right) \\ &\quad = \left|\mathrm{FT}_{2\mathrm{D}}\{I_w(x,y)\,W_v(x,y)\}\right|^2 \\ &\quad = \left|\iint_{-\infty}^{\infty} I_w(x,y)\,W_v(x,y)\,e^{-j(x\omega_x+y\omega_y)}dx\,dy\right|^2, \end{aligned} \tag{31}$$

where

$$W_v(x,y) = e^{-j(a_v x^2+b_v y^2+c_v xy)}. \tag{32}$$

The variable parameters a_v, b_v, and c_v are used. Different values of those parameters (a_v, b_v, and c_v) produce a set of projections. The additional term $c_v xy$ can be used to detect some geometrical transformations, as well. Note that the detector has the form of the Radon-Wigner distribution, which ensures that the energy of the watermark is distributed over the hyper plane defined by $(\omega_x,\omega_y) = \nabla\Phi(x,y)$ ($\Phi(x,y)$ is the phase function of the watermark). In order to make a decision weather the watermark exists in the image or not, the maxima of the Radon Wigner distribution

$$M(a_v,b_v,c_v) = \max_{\omega_x,\omega_y} P\left(\omega_x,\omega_y;W_v\right) \tag{33}$$

are compared with an assumed reference threshold. Also, multiple chirp watermarks with small and randomly chosen amplitude are used to increase flexibility of the proposed procedure. The parameters of the chirp signal as well as the random sequence that defines the amplitudes of chirp signals serve as the watermark key. Since the watermark is embedded within the entire image in the space domain, a proper masking that provides imperceptibility should be applied. An analysis of the performances giving an estimation of the detectable watermark amplitude level is provided in [38]. The robustness is tested on various attack, some being a median filter, geometrical transformations (translation, rotation and cropping simultaneously applied), a high-pass filter, local notch filter, and Gaussian noise.

(A.2) Mobasseri et al. [44] have proposed a scheme for robust watermarking based on the polynomial phase. The algorithm combines the approach in [45] (where p bits are embedded in the image) with the 2D chirp-based methods. Here, the image of size $N \times N$ is partitioned into M blocks. A 2D chirp of the form

$$W(x, y) = e^{j\pi(\beta_x x^2+\beta_y y^2)+j2\pi(f_x x+f_y y)} \tag{34}$$

is used, where $\beta_x = \beta_y = \beta$ and $f_x = f_y = f$ are taken. The watermark is embedded in the block located at the pixel (m, n) according to

$$I_w(m, n, x, y) = I_w(m, n, x, y) + k\,\mathrm{Re}[d(m, n)W(x, y)]. \tag{35}$$

The constant that controls the watermark strength is k, and the integer part is denoted by "[]", while $d(m, n)$ are watermark bits taken from $B = \{b_0, b_1, \ldots, b_{p-1}\}$. The knowledge of the pair (β_0, f_0) is required in order to recover B. It can be obtained by using the chirp transform

$$\begin{aligned} &C(m, n, \beta, f) \\ &\quad = \sum_{x=0}^{M-1}\sum_{y=0}^{M-1} I(m, n, x, y)U^*(x, y, \beta, f) \\ &\qquad + \sum_{x=0}^{M-1}\sum_{y=0}^{M-1} k\,\mathrm{Re}[d(m, n)W(x, y)]U^*(x, y, \beta, f), \end{aligned} \tag{36}$$

where:

$$U(x, y, \beta, f) = e^{j\pi\beta(x^2+y^2)+j2\pi f(x+y)}. \tag{37}$$

Finally, the total detection over all blocks can be obtained by

$$C(\beta, f) = \sum_m\sum_n |C(m, n, \beta, f)|. \tag{38}$$

This provides a possibility to use the watermark that cannot be detected by considering a single block only. Thus, in such case it would be necessary to integrate all of them over the entire image. Note that it is also possible to generate different chirps for different blocks instead of using the same chirp for all blocks. It would make the detection even more difficult for unauthorized users.

The embedded bits are recovered by

$$d_r(m, n) = \begin{cases} 1, & \text{for } C(m, n, \beta, f) \geq 0, \\ 0, & \text{for } C(m, n, \beta, f) < 0\ . \end{cases} \tag{39}$$

The proposed method is adapted to be robust to the JPEG compression algorithm. The watermark is embedded within the 8 × 8 blocks by using the quantization matrix Q. Namely, the DCT coefficients and the 2D chirp are quantized by this matrix. However, choosing an appropriate pair (β, f) is necessary to ensure that the watermark survives this quantization. The watermark survival degree can be quantified by

$$e = \frac{1}{M}\sum_{i=0}^{M-1}\sum_{j=0}^{M-1} |I_w^*(i, j) - I^*(i, j)|, \tag{40}$$

where $I^*(i, j)$ is the unmarked compressed block. The watermark is completely removed by compression if $e = 0$ is obtained. The quality of the proposed technique is tested on the image Lena, and it is proven that, for this case, it outperforms the standard spread spectrum technique.

(A.3) The watermarking in the fractional Fourier domain belongs to the time-frequency-based algorithm as well. This approach is defined in [39], and it uses a combination of the space and spatial-frequency domain. Namely, the image is transformed in the fractional Fourier domain for the angles (α_x, α_y):

$$S_{\alpha_x,\alpha_y}\left(u_x, u_y\right) = \mathrm{FRFT}_{\alpha_y}^{y\to u_y}\left\{\mathrm{FRFT}_{\alpha_x}^{x\to u_x}(I(x, y))\right\}, \tag{41}$$

where FRFT denotes the one-dimensional fractional Fourier transform. The FRFT can be treated as a rotation in the time-frequency plane for an angle α, while the inverse transform can be considered as a rotation for the angle $-\alpha$. Thus, the FRFT domain is a combination of the time and frequency domain (the Fourier transform is a special case for $\alpha = \pi/2$). Depending on the angle α, the FRFT assures that the time or the frequency domain is dominant. For α close to $\pi/2$ the frequency domain is dominant, while for small α the FRFT is dominantly in the time domain.The watermark is embedded in the FRFT coefficients reordered into a non-increasing sequence S_i. By analogy with the watermarking in the DCT domain, the first L coefficients are omitted, while the next M coefficients are used. The watermark is embedded as

$$S_i{}^w = S_i + w_i'|\mathrm{Re}\{S_i\}| + jw_i''|\mathrm{Im}\{S_i\}|. \tag{42}$$

A real valued watermark key composed of w_i' and w_i'' is used. The detection is performed by

$$\mathrm{Det} = \sum_{i=L+1}^{L+M} [S_i + w_i'|\mathrm{Re}\{S_i\}| + jw_i''|\mathrm{Im}\{S_i\}|](w_i' - jw_i''). \tag{43}$$

The performance analysis providing the detection threshold is done, the threshold being chosen as

$$T_h \geq \frac{\sigma^2}{2}\sum_{i=L+1}^{L+M} |\mathrm{Re}\{S_i\}| + |\mathrm{Im}\{S_i\}|, \tag{44}$$

where the watermark is a Gaussian white noise with the variance σ^2. The watermark key consists of the watermark sequence and the angles (α_1, α_2). Thus, the algorithm provides two more degrees of freedom, and it offers more possibility to generate watermarks. The watermarking procedure is tested on various images and attacks.

(A.4) Barkat and Sattar have proposed a fragile watermarking procedure for image authentication [43]. The watermark with a particular time-frequency signature is inserted in the image pixels. Although, in general, $N_1 \times N_2$ pixels (according to the image size) exist, a significantly lower number of them is used. The pixels location can be chosen arbitrarily. The authors have used diagonal pixels, modulated by a pseudonoise sequence as a secret key. Various frequency-modulated nonstationary signals can be a watermark, as well. However, the features that could be easily identified should be used. Consequently, different time-frequency distributions should be used for watermark detection. Barkat and Sattar have used a quadratic frequency-modulated signal. It is detected by using the Wigner distribution. The proposed scheme is tested on the following attacks: cropping, translation, JPEG compression, and scaling. Very week and imperceptible attacks were applied (e.g., JPEG with 99% quality is used). It is shown that the watermark cannot be identified after these attacks.

(B) Watermark Created in the Time-Frequency Domain. (B.1) An image watermarking approach is proposed by Al-khassaweneh and Aviyente in [49]. The image rows are used to create a set of one-dimensional signals. Then, the Wigner distribution is calculated for each of them. Also, the watermark sequence is transformed to the time-frequency domain by using the Wigner distribution. Finally, the Wigner distribution of the watermark sequence is embedded in the Wigner distribution of each image row as follows:

$$\mathrm{WD}_{x_w}\left(y,\omega_y\right)=\mathrm{WD}_x\left(y,\omega_y\right)+A\left(y,\omega_y\right)\mathrm{WD}_w\left(y,\omega_y\right),\tag{45}$$

where $A(y,\omega_y)$ is a set of the time-frequency dependent weighting coefficients. The watermarked image is obtained by using the inverse transform. Having in mind that we deal with a real and positive signal, it is defined as

$$\begin{aligned}I_w(x,y)&=\sqrt{\sum_{\omega_y}\mathrm{WD}_{x_w}\left(y,\omega_y\right)}\\&=\sqrt{I^2(x,y)+\left(\sum_{\omega_y}A\left(y,\omega_y\right)\right)w^2(y)}.\end{aligned}\tag{46}$$

However, the previous equation holds only if the two-dimensional function (45) is a valid Wigner distribution. Namely, it is well known that any two-dimensional function cannot be the Wigner distribution. It introduces a very restrictive condition on the function $A(y,\omega_y)$. In the proposed method it is determined by using the time-frequency representation of the corresponding row and taking the middle frequency region.

Al-khassaweneh and Aviyente have suggested a nonblind detection procedure. Namely, the second part of the function in (46) that depends on the watermark is selected. The detection is performed by using the standard correlation detection. A threshold that provides a minimal probability of error is derived. The proposed method is tested on different images and under various attacks. The average probability of error was found to be 0.03.

(B.2) Foo et al. in [48] have defined a method for digital audio watermarking based on the time-frequency domain. Here the audio frames are changed, so that the logical value of 1 is assigned. If the original frame is lengthened or shortened, the logical value 1 is assigned, otherwise the "normal frames" correspond to the logical value 0. The watermark is a sequence obtained as a binary code of the alphabet letters, converted to the ASCII code (the example with the binary code 010001100101001101010111 for the letters FSW is used). The crucial part of this method is the selection of frames that will be lengthened or shortened (the frame size of 1024 samples is used). The frames with signal energy level above the masking threshold are selected (the psychoacoustic model is used to determine the masking threshold in each subband). The frames length is changed by adding or removing samples with amplitudes that do not exceed the masking threshold. Four samples are added or removed within the frame of 1024 samples. It ensures that a perceptual distortion will not appear. In order to preserve the total length of the watermarked audio signal, the same number of the lengthened and shortened frames is used. The pair of frames called Diamond frames is used to represent the binary 1, while the logical values 0 are assigned to the unaltered frames.

The detection procedure is nonblind, that is, the original signal is required. A significant difference between the watermarked and the original signals will appear only if a pair of changed frames exists. Thus, it is used for logical values detection. The proposed watermarking scheme has been tested on various musical signals, as well as on a speech signal, and a set of different attacks has been applied (filtering, resampling, noise, cropping, and MP3 compression). Although the results vary for different signals and attacks, in general they are good. The worst results are obtained for the rock and pop music signals with MP3 compression. However, in all cases the owner can be identified.

(B.3) Esmaili et al. have proposed a spread spectrum based watermarking in the time-frequency domain [46]. This technique is used for watermarking of music signals. The watermark is created as

$$w_i(n)=a(n)m_i(n)pn_i(n)\cos(\omega_0(n)n),\tag{47}$$

where $m_i(n)$ is the watermark before spreading, $pn_i(n)$ is the spreading code or the pseudonoise sequence, while ω_0 is the time-varying carrier frequency. The parameter $a(n)$ controls the watermark strength. The masking properties of the human auditory system are used to shape an imperceptible watermark. The pseudonoise sequence is low pass filtered according to the signal characteristics (the Butterwort filter is used). Two different scenarios of masking have been considered. The tone- or noise-like characteristic are determined by using the entropy

$$H(X) = -\sum_{i=1}^{\omega_{\max}} P(x_i)\log_2 P(x_i). \tag{48}$$

The probability of energy for each frequency (within a window used for the spectrogram calculation) is denoted by $P(x_i)$, while $\omega_{\max}$ is the maximum frequency. A half of the maximum entropy $H_{\max}(x) = \log_2\omega_{\max}$ is taken as a threshold between noise-like and tone-like characteristics. If the entropy is lower than $H_{\max}$ it is considered as a tone-like, otherwise it is a noise-like characteristic.

The time-varying carrier frequency is obtained as the instantaneous mean frequency of the host signal, calculated by

$$\omega_i(n) = \frac{\sum_{\omega=0}^{W} \omega \mathrm{TFD}(n,\omega)}{\sum_{\omega=0}^{W} \mathrm{TFD}(n,\omega)}. \tag{49}$$

Finally, after the watermark is modulated and shaped, it is embedded in the time domain as $s_{w_i}(n) = s_i(n) + w_i(n)$.

A simple watermark detection procedure is applied. First, demodulation is performed by using the time-varying carrier, and then the watermark is detected by using the standard correlation procedure with the pseudonoise sequence.

The proposed method has been tested on several music files. It has been shown that, under various attacks, the bit error rates are mostly between 0.02 and 0.08.

(B.4) An interesting audio watermarking approach based on linear chirps has been proposed in [47]. The watermark is created as a chirp signal, which is perceptually shaped according to the host signal samples. Different chirp rates, each representing a unique watermark message, produce different slopes in the time-frequency domain. The efficient time-frequency representation is obtained by using the Wigner distribution. The extracted chirps are postprocessed in the time-frequency plane by an optimal line detection method based on the Hough-Radon transform. It can correctly estimate the slope of the watermark signal despite the broken lines caused by attacks. The simulation results show that the Hough-Radon transform applied to a time-frequency distribution can detect the watermark message correctly at bit error rates up to 20%.

3.2. Watermaking Approach Based on the Time-Frequency-Shaped Watermark.

The approach that will be presented can be used either for audio signals or images [41, 42]. Thus, the embedding and detection procedures for both kinds of signals will be defined and discussed simultaneously, by using the multidimensional notation.

In order to ensure imperceptibility constraints, the watermark should be modeled according to the time-frequency characteristics of the signal components. The concept of nonstationary multidimensional filtering [52] is adapted and used to create a watermark with time-frequency characteristics that correspond to the characteristics of the host signal. The corresponding algorithm consists of the following steps:

(1) selection of the nonstationary parts of signal suitable for watermark embedding;

(2) watermark modeling according to the multidimensional time-frequency characteristics of the host signal;

(3) watermark embedding and watermark detection procedure within the multidimensional time-frequency domain.

Multidimensional time-frequency distributions are employed in order to determine the nonstationary regions. As it will be shown later, the S-method can be efficiently used to analyze dynamics of the regions of speech signals and images. Although the cross-terms are usually undesirable in the time-frequency analysis, they have found to be useful in watermarking. Namely, they may increase performances of a speech watermark detector, and also, increase the efficiency of dynamic regions selection within an image.

The watermark is obtained at the output of a nonstationary filter as follows:

$$w_{\text{key}}(\vec{r}) = \sum_{\vec{\omega}} L_M(\vec{r},\vec{\omega})\mathrm{STFT}_p(\vec{r},\vec{\omega}), \tag{50}$$

where STFT_p is the short-time Fourier transform of a multidimensional random sequence p. The function L_M contains the information about the components within the region D_m^n. It is used to create the watermark that will be adjusted to these components. Thus, we may start with an arbitrary random multidimensional sequence $p(\vec{r})$ and, by using $L_M(\vec{r},\vec{\omega})$, its multidimensional time-frequency characteristic is modeled.

The region D_m^n will be used for watermarking if a time-frequency distribution $\mathrm{TFD}_{D_m^n}(\vec{r},\vec{\omega})$ contains a sufficient number of components whose energy is above a floor value:

$$No\left\{\left|\mathrm{TFD}_{D_m^n}(\vec{r},\vec{\omega})\right| > S\right\} > No_{\mathrm{Re}f}. \tag{51}$$

The function $No\{\}$ returns a number of components that satisfy the condition within the parenthesis, while $No_{\mathrm{Re}f}$ is the reference number of points used to make the decision about the region nonstationary. The parameter S is an energy floor that can be determined as a portion of the TFD maximum:

$$S = \lambda 10^{\lambda\log_{10}(\max(\mathrm{TFD}_{D_m^n}(\vec{r},\vec{\omega})))}. \tag{52}$$

A value of λ between 0 and 1 is taken.

The components' positions within D_m^n are identified by using the support function:

$$L_{H_1}(\vec{r},\vec{\omega}) = \begin{cases} 1, & \text{for } (\vec{r},\vec{\omega}) \in D_m^n, \\ 0, & \text{otherwise.} \end{cases} \tag{53}$$

An additional function is defined in order to consider the significant components only:

$$L_{H_2}(\vec{r},\vec{\omega}) = \begin{cases} 1, & \text{for } (\vec{r},\vec{\omega}) : \left|\mathrm{TFD}_{D_m^n}(\vec{r},\vec{\omega})\right| > \xi, \\ 0, & \text{for } (\vec{r},\vec{\omega}) : \left|\mathrm{TFD}_{D_m^n}(\vec{r},\vec{\omega})\right| \le \xi. \end{cases} \tag{54}$$

The energy threshold is denoted by ξ. Thus, the resulting support function is defined as

$$L_M(\vec{r},\vec{\omega}) = L_{H_1}(\vec{r},\vec{\omega}) \cap L_{H_2}(\vec{r},\vec{\omega}). \tag{55}$$

The watermark embedding is done according to

$$\mathrm{STFT}_{x_w}(\vec{r},\vec{\omega}) = \mathrm{STFT}_x(\vec{r},\vec{\omega}) + \mathrm{STFT}_{w_{\text{key}}}(\vec{r},\vec{\omega}), \tag{56}$$

where $\mathrm{STFT}_{x_w}\mathrm{STFT}_x$ and $\mathrm{STFT}_{w_{\text{key}}}$ are the short-time Fourier transforms of the multidimensional watermarked data, the host data, and the watermark, respectively.

Note that when compared to the signal domain, in the multidimensional time-frequency domain the number of coefficients that contain the information about the watermark is significantly increased. Consequently, the detector response will be enhanced. The standard correlation detector in the multidimensional time-frequency domain is defined as

$$\mathrm{Det} = \sum_{\vec{r}}\sum_{\vec{\omega}} \mathrm{STFT}_{x_w}(\vec{r},\vec{\omega})\mathrm{STFT}_{w_{\text{key}}}(\vec{r},\vec{\omega}). \tag{57}$$

The multidimensional time-frequency domain-based detector provides a low probability of error, even when the number of watermarked samples in the signal domain is small.

3.2.1. Digital Audio Signal.

Let us consider the voiced part of a speech signal. The region

$$D = \{(t,\omega) : t \in (t_1,t_2), \omega \in (\omega_1,\omega_2)\} \tag{58}$$

is determined by the start and the end instances t_1 and t_2 of the voiced speech, as well as by the interval $\omega \in (\omega_1,\omega_2)$ that contains the strongest formants. The S-method is used to define the support function [41, 53]:

$$L_M(t,\omega) = \begin{cases} 1, & \text{for } (t,\omega) \in D,\ \mathrm{SM}(t,\omega) > \xi, \\ 0, & \text{for } (t,\omega) \notin D \text{ or } \mathrm{SM}(t,\omega) < \xi. \end{cases} \tag{59}$$

The region appropriate for watermarking is shown in Figure 12(a). The corresponding support function (Figure 12(b)) is created by using the value $\xi = \lambda 10^{\lambda \log_{10}(\max(\mathrm{SM}(t,\omega)))}$ with $\lambda = 0.7$.

The watermark is embedded in the time domain: $x_w(n) = x(n) + w(n)$. The time-frequency representation of the watermark is shown in Figure 12(c).

As expected, the time-frequency characteristics of the watermark follow those components of the speech signal. Consequently, the watermark is inaudible within the speech signal.

Next, a music signal of the flute is considered, Figure 13(a).

In this case, the fourth-order complex-lag distribution is more appropriate for the region selection than the S-method, because it better follows the frequency variations in the signal (Figures 13(a) and 13(b)). Thus, by using this distribution an inaudible watermark is created.

Note that an important improvement in the watermark detection is obtained if the cross-terms are included [41]. Namely, the watermark is present within the cross-terms, as well. A standard correlation detector in the time-frequency domain that includes the cross-terms can be written in the form

$$D = \sum_{i=1}^{N} \mathrm{SM}^{i}_{w\text{key}}\mathrm{SM}^{i}_{x_w} + \sum_{\substack{i,j=1 \\ i \neq j}}^{N} \mathrm{SM}^{i,j}_{w\text{key}}\mathrm{SM}^{i,j}_{x_w}. \tag{60}$$

The second term in (60) is the result of cross-terms.

Note that this form of detector can be used in other existing detector structures.

The following measure of the detection quality

$$R = \frac{\overline{D}_{wr} - \overline{D}_{ww}}{\sqrt{\sigma_{wr}{}^2 + \sigma_{ww}{}^2}} \tag{61}$$

is used. The mean value and standard deviation of the detector response are denoted by D and σ^2. Indices wr and ww indicate the right and the wrong keys, respectively.

Efficiency of the proposed procedure is demonstrated on various examples. The results for speech signals with maximum frequencies of 4 kHz and 11,025 kHz are presented in [41]. This approach provides a reliable detection for a high SNR (SNR = 32 dB has been used) and under various attacks. The watermark sequence was created by using a pseudorandom Gaussian sequence of 1000 samples.

The probability of error was of order 10^{-7} for: MP3 (constant bit rate 8 kbps and variable bit rate 75–120 kbps are considered), delay monolight echo (180 ms, mixing 20%), echo 200 ms, deep flutter (deep 10, sweeping rate 5 kHz), amplitude (normalize 100%), and additive Gaussian noise (SNR = −35 dB). The worst case is obtained for pitch scaling ±5% and it is of order 10^{-5}. The results for other attacks (time stretch ±15%, wow delay 20%, wow delay 10% and bright flutter, MP3 variable bit rate 40–50 kbps) are of order 10^{-6}.

3.2.2. Digital Image.

The space-spatial-frequency analysis (two-dimensional time-frequency analysis) is used to select pixels that belong to the image nonstationary regions [42]. The two-dimensional S-method is used as a space-spatial frequency distribution [54]:

$$\begin{aligned} &\mathrm{SM}(n_1,n_2,k_1,k_2) \\ &\quad = \mathrm{SPEC}(n_1,n_2,k_1,k_2) \\ &\qquad + 2\,\mathrm{Re}\left\{\sum_{i_1=0}^{L}\sum_{i_2=1}^{L}\mathrm{STFT}(n_1,n_2,k_1+i_1,k_2+i_2)\right. \\ &\qquad\qquad \left.\times\mathrm{STFT}^*(n_1,n_2,k_1-i_1,k_2-i_2)\right\} \\ &\qquad + 2\,\mathrm{Re}\left\{\sum_{i_1=1}^{L}\sum_{i_2=0}^{L}\mathrm{STFT}(n_1,n_2,k_1+i_1,k_2+i_2)\right. \\ &\qquad\qquad \left.\times\mathrm{STFT}^*(n_1,n_2,k_1-i_1,k_2-i_2)\right\}. \end{aligned} \tag{62}$$

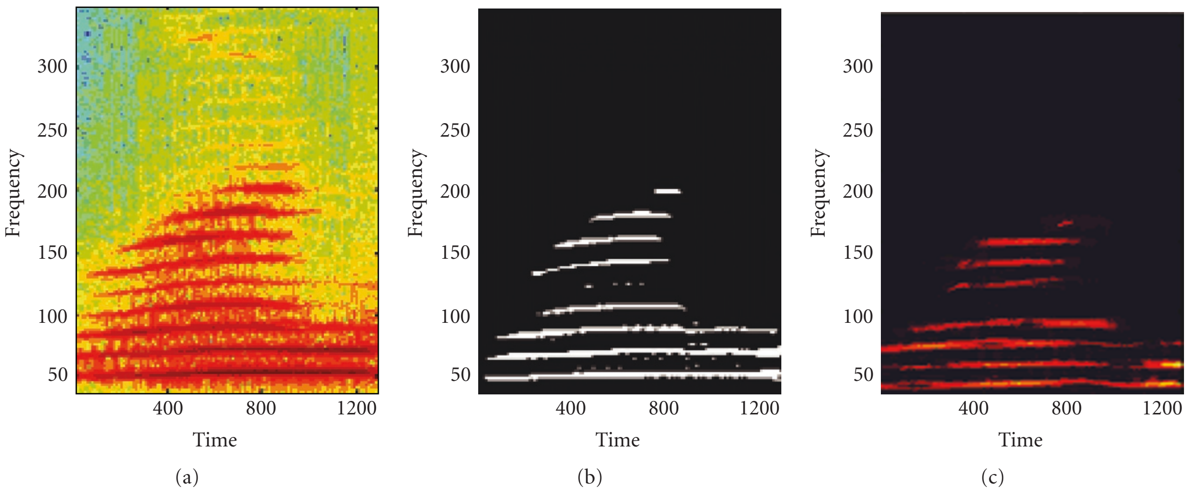


FIGURE 12: (a) Region selected for watermarking. (b) Support function. (c) Time-frequency representation of the watermark. The S-method is used.

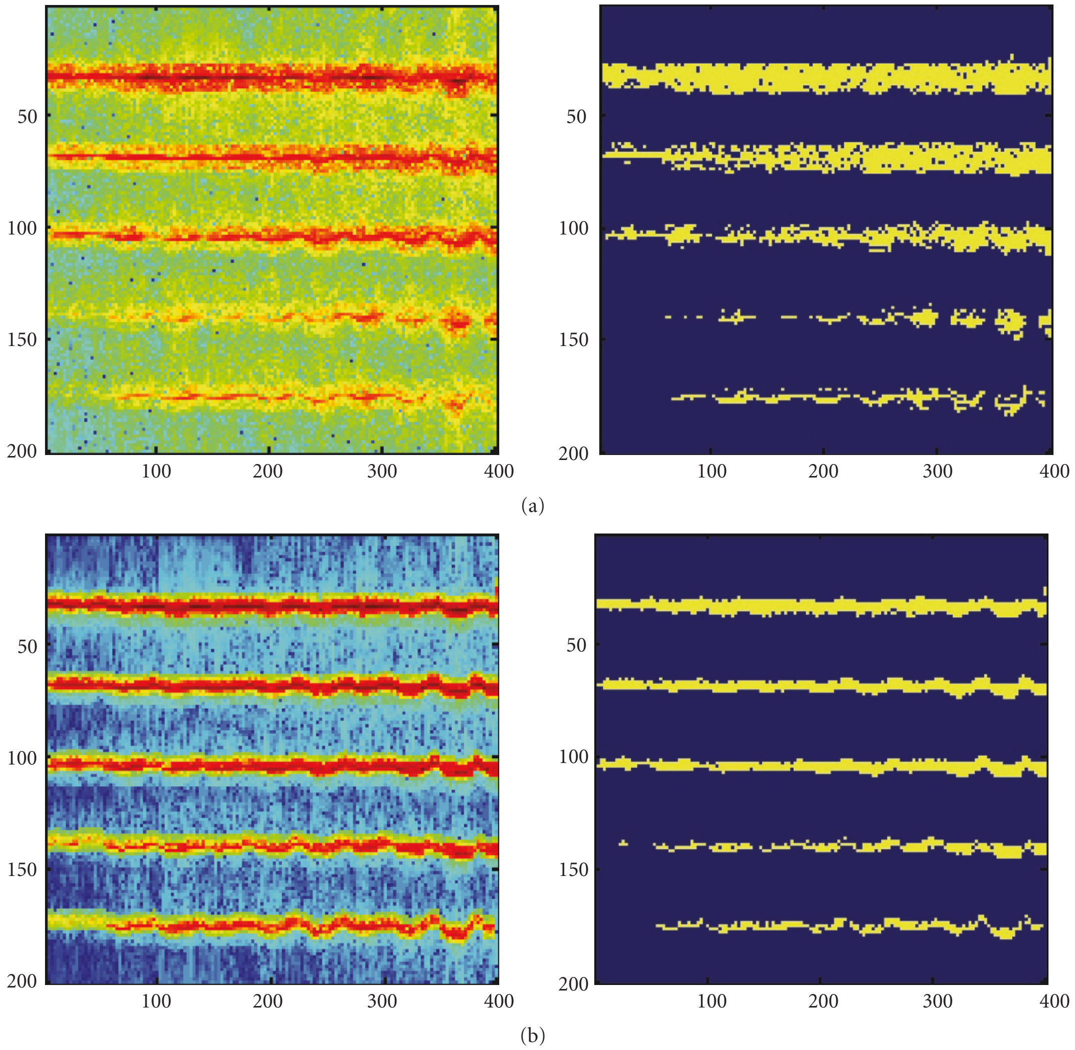
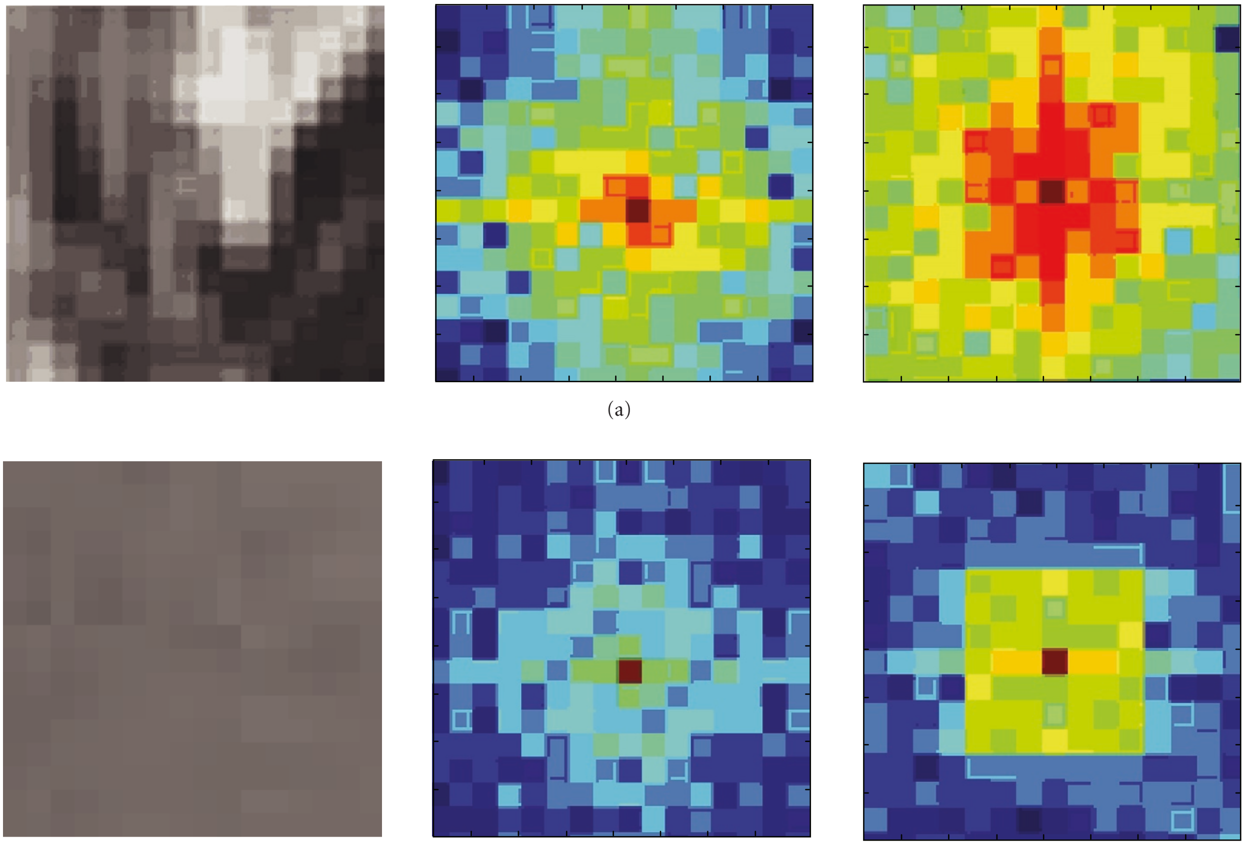


FIGURE 13: (a) S-method and the corresponding support function. (b) Fourth-order complex-lag distribution and corresponding support function.



(a)

(b)

Figure 14: (a) Dynamic region. (b) Stationary region (the spectrogram—based characterization-the second column; the S-method-based characterization—the third column).

By increasing the size L of a two-dimensional window, the cross-terms start to appear. Thus, when compared to the spectrogram, the number of frequency components increases, hence making the region characterization easier. A pixel that belongs to the dynamic region can be selected by using the following procedure.

(1) The S-method is calculated for a $N \times N$ window (windows of size 9×9 up to 16×16 are used). The middle frequency range $D_m^n = \{(\omega_1, \omega_2) : r_1 < \omega_1, \omega_2 < r_2\}$ is used.

(2) The energy floor S is obtained by using the experimentally determined $\lambda = 0.7$.

(3) The region is considered as nonstationary if:

$$No\left\{\left|\mathrm{SM}_{D_m^n}(n_1, n_2, \omega_1, \omega_2)\right| > S\right\} > No_{\mathrm{Re}\,f}, \qquad (63)$$

where $No_{\mathrm{Re}\,f} = \lambda n$, while n is the total number of the points within the region D_m^n.

The examples where the pixels belong to the dynamic and stationary regions, respectively, are shown in Figures 14(a) and 14(b).

The procedure for watermark embedding is just a two-dimensional case of the presented multidimensional approach. Namely, a two-dimensional support function is used:

$$L(n_1, n_2, \omega_1, \omega_2) = \begin{cases} 1, & \text{for } (\omega_1, \omega_2) : |\mathrm{SM}(n_1, n_2, \omega_1, \omega_2)| > S, \\ 0, & \text{for } (\omega_1, \omega_2) : |\mathrm{SM}(n_1, n_2, \omega_1, \omega_2)| \le S. \end{cases} \qquad (64)$$

The S-method as a two-dimensional time-frequency distribution is applied, while the energy threshold is $\xi = S$. The watermark is shaped by the space-spatial frequency characteristic of the image components:

$$w_{\mathrm{key}}(n_1, n_2) = \sum_{\omega_1} \sum_{\omega_2} \mathrm{STFT}_p(n_1, n_2, \omega_1, \omega_2) L(n_1, n_2, \omega_1, \omega_2). \qquad (65)$$

A two-dimensional pseudorandom sequence STFT_p is used.

The watermark embedding and detection are performed in the space-spatial frequency domain:

$$\begin{aligned} I_w(n_1,n_2) &= \sum_{\omega_1}\sum_{\omega_2}\mathrm{STFT}_I(n_1,n_2,\omega_1,\omega_2) \\ &\qquad + \mathrm{STFT}_{w_{\mathrm{key}}}(n_1,n_2,\omega_1,\omega_2), \\ \mathrm{Det} &= \sum_{\omega_1}\sum_{\omega_2}\mathrm{STFT}_{I_w}(n_1,n_2,\omega_1,\omega_2) \\ &\qquad \times \mathrm{STFT}_{w_{\mathrm{key}}}(n_1,n_2,\omega_1,\omega_2). \end{aligned} \tag{66}$$

This procedure is tested on several images (Lena, Peppers, Boat, F16, and Barbara), under various attacks (JPEG80-JPEG40, Median 3 × 3, Median 5 × 5, Average 3 × 3, Impulse noise, Gaussian noise, Lightening, and Darkening). The PSNR was around 50 dB. The number of the selected pixels varied from 3304 for F16 to 7833 for Barbara. The probability of error was compared with the standard DCT-based procedure (with different detector forms), where 22050 coefficients are used. It was shown that the proposed procedure significantly outperforms the standard DCT procedures.

3.2.3. Digital Video. Observe that the proposed approach can be also used for video signal watermarking. The two-dimensional and one-dimensional time-frequency distributions are combined in this case. Namely, the stationary pixels and stationary regions around them are selected by using the two-dimensional analysis, as it was described in the previous subsection. Then, the time dependent sequence $I_t(x,y) = [I_1(x,y), I_2(x,y), \ldots, I_K(x,y)]$ is produced by taking the stationary pixels at the position (x,y), along K consecutive frames. Based on $I_t(x,y)$ the frequency modulated signal is created as

$$z(t) = e^{j\mu(I_t(x,y)-\overline{I}_t(x,y))}, \tag{67}$$

where $\overline{I}_t(x,y) = \mathrm{mean}(I_t(x,y))$, while μ is a constant. The stationarity of the selected pixels, along the time axis, is examined by using the one-dimensional S-method. The experiments show that the minimal number of pixels for reliable watermark detection is about 600. This can be easily achieved, even for a very short video sequence (note that more than 2500 stationary pixels are obtained for a signal of duration of 2 s in the example provided in [42]). This approach was tested under the presence of MPEG4 compression. The obtained probabilities of errors were found to be within the range 10^{-4}–10^{-5}.

4. Conclusion

An overview of most important time-frequency analysis techniques is presented. An appropriate distribution selection procedure for a specific type of signal is discussed. Time-frequency-based watermarking algorithms for digital audio, digital image, and video are reviewed, as well. The watermark is either a signal with specific time-frequency characteristics or a pseudonoise sequence shaped according to the time-frequency characteristics of the host signal. The main advantages of the time-frequency domain over the Fourier, DCT, and signal domain are emphasized. Finally, the presented theory could be used to generalize the existing watermarking approaches defined in either the Fourier or the DCT domain.

Acknowledgments

The author is very thankful to Dr. Irena Orović and Professor Victor Sučić for their help and useful suggestions during the work on this paper.

References

[1] L. Cohen, "Time-frequency distributions—a review," *Proceedings of the IEEE*, vol. 77, no. 7, pp. 941–981, 1989.

[2] F. Hlawatsch and G. F. Boudreaux-Bartels, "Linear and quadratic time-frequency signal representations," *IEEE Signal Processing Magazine*, vol. 9, no. 2, pp. 21–67, 1992.

[3] B. Boashash and B. Ristić, "Polynomial time-frequency distributions and time-varying higher order spectra: application to the analysis of multicomponent FM signals and to the treatment of multiplicative noise," *Signal Processing*, vol. 67, no. 1, pp. 1–23, 1998.

[4] B. Boashash, *Time-Frequency Analysis and Processing*, Elsevier, Amsterdam, The Netherlands, 2003.

[5] H. Choi and W. J. Williams, "Improved time-frequency representation of multicomponent signals using exponential kernels," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 6, pp. 862–871, 1989.

[6] Y. Zhao, L. E. Atlas, and R. J. Marks, "The use of cone-shaped kernels for generalized time-frequency representations of nonstationary signals," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 38, no. 7, pp. 1084–1091, 1990.

[7] L. Stanković, "Auto-term representation by the reduced interference distributions: a procedure for kernel design," *IEEE Transactions on Signal Processing*, vol. 44, no. 6, pp. 1557–1563, 1996.

[8] R. G. Baraniuk and D. L. Jones, "A signal-dependent time-frequency representation. Optimal kernel design," *IEEE Transactions on Signal Processing*, vol. 41, no. 4, pp. 1589–1602, 1993.

[9] F. Hlawatsch and R. L. Urbanke, "Bilinear time-frequency representations of signals: the shift-scale invariant class," *IEEE Transactions on Signal Processing*, vol. 42, no. 2, pp. 357–366, 1994.

[10] R. G. Baraniuk and D. L. Jones, "Signal-dependent time-frequency analysis using a radially Gaussian kernel," *Signal Processing*, vol. 32, no. 3, pp. 263–284, 1993.

[11] M. G. Amin and W. J. Williams, "High spectral resolution time-frequency distribution kernels," *IEEE Transactions on Signal Processing*, vol. 46, no. 10, pp. 2796–2804, 1998.

[12] M. J. Bastiaans, T. Alieva, and L. Stanković, "On rotated time-frequency kernels," *IEEE Signal Processing Letters*, vol. 9, no. 11, pp. 378–381, 2002.

[13] B. Boashash, "Estimating and interpreting the instantaneous frequency of a signal—part 1: fundamentals," *Proceedings of the IEEE*, vol. 80, no. 4, pp. 520–538, 1992.

[14] B. Barkat and B. Boashash, "Design of higher order polynomial Wigner-Ville distributions," *IEEE Transactions on Signal Processing*, vol. 47, no. 9, pp. 2608–2611, 1999.

[15] G. Viswanath and T. V. Sreenivas, "IF estimation using higher order TFRs," *Signal Processing*, vol. 82, no. 2, pp. 127–132, 2002.

[16] L. Stanković, "A multitime definition of the Wigner higher order distribution: L-Wigner distribution," *IEEE Signal Processing Letters*, vol. 1, no. 7, pp. 106–109, 1994.

[17] L. Stanković, "A method for time-frequency analysis," *IEEE Transactions on Signal Processing*, vol. 42, no. 1, pp. 225–229, 1994.

[18] S. Stanković and L. Stanković, "An architecture for the realization of a system for time-frequency signal analysis," *IEEE Transactions on Circuits and Systems II*, vol. 44, no. 7, pp. 600–604, 1997.

[19] S. Stanković and L. Stanković, "Introducing time-frequency distribution with a "complex-time" argument," *Electronics Letters*, vol. 32, no. 14, pp. 1265–1267, 1996.

[20] L. Stanković, "Time-frequency distributions with complex argument," *IEEE Transactions on Signal Processing*, vol. 50, no. 3, pp. 475–486, 2002.

[21] C. Cornu, S. Stanković, C. Ioana, A. Quinquis, and L. Stanković, "Generalized representation of phase derivatives for regular signals," *IEEE Transactions on Signal Processing*, vol. 55, no. 10, pp. 4831–4838, 2007.

[22] S. Stanković, I. Orovic, and C. Ioana, "Effects of Cauchy integral formula on the precision of the IF estimation," *IEEE Signal Processing Letters*, vol. 16, no. 4, pp. 327–330, 2009.

[23] M. Morelande, B. Senadji, and B. Boashash, "Complex-lag polynomial Wigner-Ville distribution," in *Proceedings of IEEE Speech and Image Technologies for Computing and Telecommunications (TENCON '97)*, vol. 1, pp. 43–46, Brisbane, Australia, December 1997.

[24] S. Stanković, N. Žarić, I. Orović, and C. Ioana, "General form of time-frequency distribution with complex-lag argument," *Electronics Letters*, vol. 44, no. 11, pp. 699–701, 2008.

[25] G. Frazer and B. Boashash, "Multiple window spectrogram and time-frequency distributions," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '94)*, vol. 4, pp. 193–296, 1994.

[26] F. Çakrak and P. J. Loughlin, "Multiwindow time-varying spectrum with instantaneous bandwidth and frequency constraints," *IEEE Transactions on Signal Processing*, vol. 49, no. 8, pp. 1656–1666, 2001.

[27] M. Bayram and R. G. Baraniuk, "Multiple window time-frequency analysis," in *Proceedings of the IEEE-SP International Symposium on Time-Frequency and Time-Scale Analysis*, pp. 173–176, June 1996.

[28] I. J. Cox, M. L. Miller, and J. A. Bloom, *Digital Watermarking*, Academic Press, New York, NY, USA, 2002.

[29] M. Barni and F. Bartolini, *Watermarking Systems Engineering*, Marcel Dekker, New York, NY, USA, 2004.

[30] "Special issue on "Identification and protection of multimedia information"," *Proceedings of the IEEE*, vol. 87, no. 7, 1999.

[31] E. Muharemagic and B. Furht, "Survey of watermarking techniques and applications," in *Multimedia Watermarking Techniques and Applications*, B. Furht and D. Kirovski, Eds., chapter 3, pp. 91–130, Auerbach Publication, 2006.

[32] D. Kirovski and H. S. Malvar, "Spread-spectrum watermarking of audio signals," *IEEE Transactions on Signal Processing*, vol. 51, no. 4, pp. 1020–1033, 2003.

[33] M. Steinebach and J. Dittmann, "Watermarking-based digital audio data authentication," *EURASIP Journal on Applied Signal Processing*, vol. 2003, no. 10, pp. 1001–1015, 2003.

[34] A. Nikolaidis and I. Pitas, "Asymptotically optimal detection for additive watermarking in the DCT and DWT domains," *IEEE Transactions on Image Processing*, vol. 12, no. 5, pp. 563–571, 2003.

[35] J. R. Hernández, M. Amado, and F. Pérez-González, "DCT-domain watermarking techniques for still images: detector performance analysis and a new structure," *IEEE Transactions on Image Processing*, vol. 9, no. 1, pp. 55–68, 2000.

[36] A. Briassouli and M. G. Strintzis, "Locally optimum nonlinearities for DCT watermark detection," *IEEE Transactions on Image Processing*, vol. 13, no. 12, pp. 1604–1617, 2004.

[37] Q. Cheng and T. S. Huang, "An additive approach to transform-domain information hiding and optimum detection structure," *IEEE Transactions on Multimedia*, vol. 3, no. 3, pp. 273–284, 2001.

[38] S. Stanković, I. Djurović, and L. Pitas, "Watermarking in the space/spatial-frequency domain using two-dimensional Radon-Wigner distribution," *IEEE Transactions on Image Processing*, vol. 10, no. 4, pp. 650–658, 2001.

[39] I. Djurović, S. Stanković, and I. Pitas, "Digital watermarking in the fractional Fourier transformation domain," *Journal of Network and Computer Applications*, vol. 24, no. 2, pp. 167–173, 2001.

[40] T. D. Wickens, *Elementary Signal Detection Theory*, Oxford University Press, Oxford, UK, 2002.

[41] S. Stanković, I. Orović, and N. Žarić, "Robust speech watermarking procedure in the time-frequency domain," *EURASIP Journal on Advances in Signal Processing*, vol. 2008, Article ID 519206, 9 pages, 2008.

[42] S. Stanković, I. Orović, and N. Žarić, "An application of multidimensional time-frequency analysis as a base for the unified watermarking approach," *IEEE Transactions on Image Processing*, vol. 19, no. 3, pp. 736–745, 2010.

[43] B. Barkat and F. Sattar, "A new time-frequency based private fragile watermarking scheme for image authentication," in *Proceedings of the 7th International Symposium on Signal Processing and Its Applications*, vol. 2, pp. 363–366, July 2003.

[44] B. G. Mobasseri, Y. Zhang, M. G. Amin, and B. M. Dogahe, "Designing robust watermarks using polynomial phase exponentials," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '05)*, pp. 833–836, March 2005.

[45] M. Kutter and S. Winkler, "A vision-based masking model for spread-spectrum image watermarking," *IEEE Transactions on Image Processing*, vol. 11, no. 1, pp. 16–25, 2002.

[46] S. Esmaili, S. Krishnan, and K. Raahemifar, "Audio watermarking using time-frequency characteristics," *Canadian Journal of Electrical and Computer Engineering*, vol. 28, no. 2, pp. 57–61, 2003.

[47] S. Erküçük, S. Krishnan, and M. Zeytinoğlu, "A robust audio watermark representation based on linear chirps," *IEEE Transactions on Multimedia*, vol. 8, no. 5, pp. 925–926, 2006.

[48] S. W. Foo, S. M. Ho, and L. M. Ng, "Audio watermarking using time-frequency compression expansion," in *Proceedings of IEEE International Symposium on Cirquits and Systems (ISCAS '04)*, pp. 201–204, May 2004.

[49] M. Al-khassaweneh and S. Aviyente, "A time-frequency based perceptual and robust watermarking scheme," in *Proceedings of 13th European Signal Processing Conference (EUSIPCO '05)*, Antalya, Turkey, September 2005.

[50] S. Erküçük, *Time-frequency analysis of spread spectrum based communication and audio watermarking systems*, Ph.D. thesis, 2003.

[51] I. Orović, S. Stanković, T. Thayaparan, and L. Stanković, "Multiwindow S-method for instantaneousfrequency estimation and its application in radar signal analysis," *IET Signal Processing*, vol. 90, no. 5, 7 pages, 2010.

[52] L. Stanković, S. Stanković, and I. Djurović, "Space/spatial-frequency analysis based filtering," *IEEE Transactions on Signal Processing*, vol. 48, no. 8, pp. 2343–2352, 2000.

[53] S. Stanković, "About time-variant filtering of speech signals with time-frequency distributions for hands-free telephone systems," *Signal Processing*, vol. 80, no. 9, pp. 1777–1785, 2000.

[54] S. Stanković, L. Stanković, and Z. Uskokovic, "On the local frequency, group shift, and cross-terms in some multidimensional time-frequency distributions: a method for multidimensional time-frequency analysis," *IEEE Transactions on Signal Processing*, vol. 43, no. 7, pp. 1719–1724, 1995.