Hindawi Publishing Corporation
EURASIP Journal on Applied Signal Processing
Volume 2006, Article ID 86712, Pages 1–14
DOI 10.1155/ASP/2006/86712

Time-Frequency Signal Synthesis and Its Application in Multimedia Watermark Detection

Lam Le and Sridhar Krishnan

Department of Electrical and Computer Engineering, Ryerson University, Toronto, ON, Canada M5B 2K3

Received 29 March 2005; Revised 28 January 2006; Accepted 5 February 2006

Recommended for Publication by Alex Kot

We propose a novel approach to detect the watermark message embedded in images under the form of a linear frequency modulated chirp. Localization of several time-frequency distributions (TFDs) is studied for different frequency modulated signals under various noise conditions. Smoothed pseudo-Wigner-Ville distribution (SPWVD) is chosen and applied to detect and recover the corrupted image watermark bits at the receiver. The synthesized watermark message is compared with the referenced one at the transmitter as a detection evaluation scheme. The correlation coefficient between the synthesized and the referenced chirps reaches 0.9 or above for a maximum bit error rate of 15% under intentional and nonintentional attacks. The method provides satisfactory result for detection of image watermark messages modulated as chirp signal and could be a potential tool in multimedia security applications.

1. INTRODUCTION

Chirp signals are present ubiquitously in many areas of science and engineering. Chirps are identified in natural signals such as animal sounds (birds, frogs, whales, and bats), whistling sound, as well as in man-made systems such as in radar, sonar, telecommunications, physics, and acoustics. For example, in radar applications, chirp signals are used to analyze the trajectories of moving objects. Due to its inherent ability to reject interference, linear frequency modulated signals or chirp signals are also used widely in spread spectrum communication. Chirps are also involved in biomedicine applications such as in the study of electroencephalogram (EEG) and electromyogram (EMG) data. Recently, the boom in Internet makes it easier for digital contents to be copied and reproduced in large quantities beyond the control of content providers. Digital watermark is the tool created to work against this problem, it can prove the content's origin, protect the copyrights, and prevent illegal use. In watermarking of audio signals and images [1, 2], the chirp message is embedded in the signals and then detected at the receiver based on its frequency change rate. A more detailed discussion on watermarking applications is provided in Section 2 of this paper.

Due to their immense importance, detection and estimation of chirp signals in the presence of high noise level and other signals has attracted much attention in many recent research papers. There are various detection methods for chirps in the time domain, joint time-frequency domain, and the ambiguity domain. Some of the common techniques are the optimal detection [3] based on the square inner product between the observed and referenced chirps, the maximum likelihood which integrates along all possible lines of the time-frequency distribution (TFD), the wavelet transform, and evolutionary algorithm [4, 5]. One of the most common techniques for linear chirp detection is the Hough-Radon transform (HRT) [6–8]. HRT detects the directional elements that satisfy a parametric constraint in the image of the time-frequency (TF) plane by converting the signal's TFD into a parameter space. HRT is an effective method for detecting, error correcting of linear chirp, and it can be applied to small image of the TF plane. However, the complexity of the HRT algorithm increases substantially with the size of the image. The other approach for chirp detection and estimation, which is the main focus of this paper, is based on time-frequency signal synthesis. Signal synthesis was first applied in signal design to generate signal with known, required time-frequency properties such as in the design of time-varying filter and signals separation. A time domain signal can be synthesized from its time-frequency distribution using least square method or polynomial-phase transform. In least square approach [9, 10], the signal is constructed

by minimization of the error function between the signal TFD and the desired TFD. Improved algorithms have been tested for Wigner-Ville distribution as well as its smoothed versions and they yielded satisfactory results. The discrete polynomial-phase transform approach [11–13], on the other hand, models the signal phase as polynomial and uses the higher ambiguity function to estimate the signal parameters.

In this paper, we introduce a new way to detect the image watermark messages modulated as linear chirp signals from the TF plane by signal synthesis method using polynomial-phase transform. The success rate of the method depends considerably on the initial estimation of the instantaneous frequency (IF) from the TF plane and which in turn, depends on the TFD selection. A good TFD candidate would be the one providing high localization and cross-term free for a variety of signals in different noise levels at different frequency modulation rates. The rest of the paper is organized as follows: an analysis on localization of the common TFDs is discussed in Section 3. A review on signal synthesis based on discrete polynomial transform (DPT) is provided in Section 4. Section 5 is for the application of the proposed image watermark detection scheme. And finally, the result and discussion related to the proposed technique are provided in Section 6. But first we will have a brief review on watermark applications in joint time-frequency domain.

2. TIME-FREQUENCY DIGITAL WATERMARKING

Digital watermarking is the process involving integrating a special message into digital contents such as audio, video, and image for copyright protection purposes. The embedded data is then extracted from the multimedia as a proof of ownership. Various digital watermarking methods have been researched by many authors in the past years. The watermark techniques differ depending on their applications and characteristics such as invisibility, robustness, security, and media category. In addition, watermark methods can also be classified by the type of watermark message used as well as the processing domain [14]. The watermark message used can be any noise type, that is, pseudo-noise sequence, Gaussian random sequence, or image type such as ones in the form of binary image, stamp, and logo. The processing domain, where the insertion and extraction of watermark taken place, is usually spatial domain or frequency domain. The techniques based on frequency domain such as DCT, wavelet and Fourier transform have become very popular recently. However very few works have been done to exploit the unique properties and advantages of watermarking in joint time-frequency domain.

In [15], watermark insertion and extraction are both done in time-frequency domain. In the embedding process, watermark message $w(t, f)$, in time-frequency domain, is added to the cells of Wigner-Ville distribution $X(t, f)$ of the signal $x(t)$. The locations of cells are carefully selected so that the message will be invisible when inverting the watermarked Wigner distribution back to spatial domain. In the detection process, the Wigner-Ville distribution of the original message is subtracted from that of the watermarked message to retrieve the watermark.

The fragile watermark approach proposed in [16] does not require the whole original signal to recover the watermark. A quadratic chirp is modulated with a pseudo-random (PN) sequence before being added to the diagonal pixels of the image in the spatial domain. Only the original value of the diagonal pixels is enough for recovering the watermark bits. After removing the PN effect, the watermark pattern can be analyzed using a TFD.

In [1, 2], we introduced the novel watermarking method using a linear chirp based technique and applied it to image and audio signals. The chirp signal $x(t)$ (or $\underline{\mathbf{m}}$) is quantized and has value -1 and 1 as in $\mathbf{m}^q$. $\mathbf{m}^q$ is then embedded into the multimedia files. The detail of the embedding and extracting of watermark is followed.

2.1. Watermark embedding

Each bit m_k^q of $\mathbf{m}^q$ is spread with a cyclic shifted version $\mathbf{p_k}$ of a binary PN sequence called watermark key. The results are summed together and generate the wideband noise vector $\mathbf{w}$:

$$\mathbf{w} = \sum_{k=0}^{N} m_k^q \mathbf{p_k}, \tag{1}$$

where N is the number of watermark message bits in $\mathbf{m}^q$.

The wideband noise $\mathbf{w}$ is then carefully shaped and added to the audio or DCT block of the image so that it will cause imperceptible change in signal quality. In the audio watermarking application as proposed in [2], to make the watermark message imperceptible, the amplitude level of the wideband noise $\mathbf{w}$ is scaled down to be about 0.3 of the product between the dynamic range of the signal and the noise itself and then lowpass filtered before being added to the signal. The fact that audio signals have most of their energy limited from low to middle frequencies will allow embedding the frequency-limited watermark with greater strength. This method is therefore more robust compared to the method in [17] especially to attacks in the high frequency band such as MP3 compression, lowpass filtering, and resampling. In the image watermarking application in [1] and this paper, the length of $\mathbf{w}$ to be embedded depends on the perceptual entropy of the image.

To embed the watermark into the image, the model based on the *just noticeable difference* (JND) paradigm was utilized. The JND model based on DCT was used to find the perceptual entropy of the image and to determine the perceptually significant regions to embed the watermark. In this method, the image is decomposed into 8×8 blocks. Taking the DCT on the block b results in the matrix $X_{u,v,b}$ of the DCT coefficients. The watermark embedding scheme is based on the model proposed in [18]. The watermark encoder for the DCT scheme is described as

$$X^*_{u,v,b} = \begin{cases} X_{u,v,b} + t^C_{u,v,b} w_{u,v,b}, & \text{if } X_{u,v,b} > t^C_{u,v,b}, \\ X_{u,v,b}, & \text{otherwise,} \end{cases} \tag{2}$$

where $X_{u,v,b}$ refers to the DCT coefficients, $X^*_{u,v,b}$ refers to the watermarked DCT coefficients, $w_{u,v,b}$ is obtained from

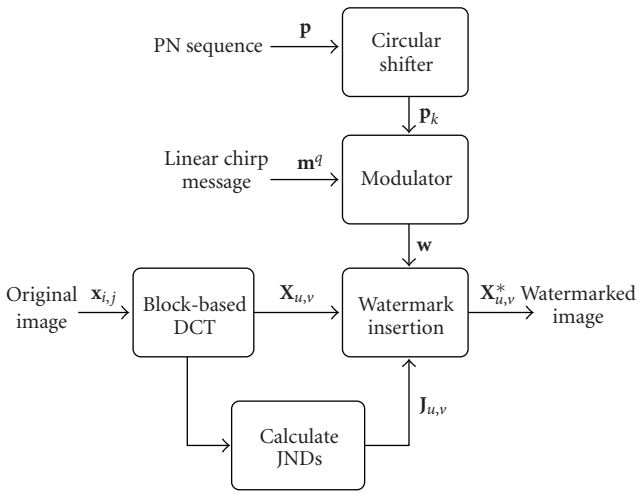


Figure 1: Watermark embedding scheme.

the wideband noise vector $\mathbf{w}$, and the threshold $t^C_{u,v,b}$ is the computed JND determined for various viewing conditions such as minimum viewing distance, luminance sensitivity, and contrast masking. Figure 1 shows the block diagram of the described watermark encoding scheme.

2.2. Watermark detecting

Figure 2 shows the original image, the chirp used as watermark message, and the watermarked image based on our approach. The watermark is well hidden in the image that it is imperceptible and causes no difference in the histogram. The presence of the chirp message is undetectable in the spatial and time-frequency domain thanks to the perceptual shaping processing. Figure 3 shows the block diagram of the described watermark decoding scheme. The detection scheme for the DCT-based watermarking can be expressed as

$$\begin{aligned} \hat{w}_{u,v,b} &= \frac{X_{u,v,b} - \hat{X}^*_{u,v,b}}{t^C_{u,v,b}}, \\ \hat{\mathbf{w}} &= \begin{cases} \hat{w}_{u,v,b} & \text{if } X_{u,v,b} > t^C_{u,v,b}, \\ 0 & \text{otherwise,} \end{cases} \end{aligned} \tag{3}$$

where $\hat{X}^*_{u,v,b}$ are the coefficients of the received watermarked image, and $\hat{\mathbf{w}}$ is the received wideband noise vector. Due to intentional and nonintentional attacks such as lossy compression, shifting, down-sampling the received chirp message $\hat{\mathbf{m}}^q$ will be different from the original message $\mathbf{m}^q$ by a bit error rate BER. We use the watermark key, $\mathbf{p}_k$ to despread $\hat{\mathbf{w}}$, and integrate the resulting sequence to generate a test statistic $\langle \hat{\mathbf{w}}, \mathbf{p}_k \rangle$. The sign of the expected value of the statistic depends only on the embedded watermark bit m^q_k. Hence the watermark bits can be estimated using the decision rule:

$$\hat{m}^q_k = \begin{cases} +1, & \text{if } \langle \hat{\mathbf{w}}, \mathbf{p}_k \rangle > 0, \\ -1, & \text{if } \langle \hat{\mathbf{w}}, \mathbf{p}_k \rangle < 0. \end{cases} \tag{4}$$

The bit estimation process is repeated for all the transmitted bits.

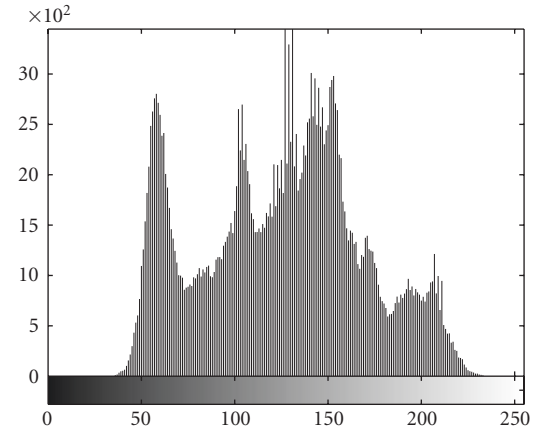
3. SELECTION OF TFD

The frequency change of a signal over time (instantaneous frequency) is an important tool for analysis of nonstationary signals. The instantaneous frequency (IF) is traditionally obtained by taking the first derivative of the phase of the signal with respect to time. This poses some difficulties because the derivative of the phase of the signal may take negative values thus misleading the interpretation of instantaneous frequency. Another way to estimate the IF of a signal is to take the first central moment of its time-frequency distribution. Time-frequency distribution (TFD) has been used widely as an analysis tool for the study of nonstationary signals. It involves mapping a one-dimensional signal $x(t)$ into a two-dimensional function $\mathrm{TFD}_x(t, f)$, which provides the information on spectral characteristics of the signal with respect to time. Time-frequency representations (TFR) are classified into two main groups: linear and quadratic. One example of linear TFR is the short time Fourier transform which has the tradeoff between time and frequency resolution. Quadratic (or bilinear) TFR such as spectrogram and Wigner-Ville uses energy distribution of the signal over time and frequency to represent the temporal and spectral information. There are a large number of possible time-frequency distributions and they are classified based on the desired properties such as cross-term removal and joint time-frequency resolution. There is always a tradeoff between resolution and cross-term suppression. The removal of cross-term (smoothing) also takes away some of the signal energy and reduces the joint time-frequency resolution. When it comes to evaluation of a TFD, besides the factors such as accuracy of IF estimation, high resolution in joint time-frequency domain, ability to suppress cross-terms, one should also consider the effects of noise on the TFD's performance.

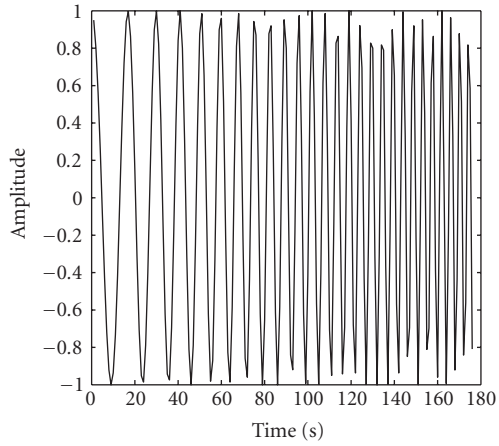
We have done several simulations to compare the properties of different TFDs on various signals, types, and levels of noise. The TFDs involved in the test are spectrogram (SP), Wigner-Ville distribution (WVD), pseudo Wigner-Ville distribution (PWVD), smoothed pseudo Wigner-Ville distribution (SPWVD), Choi-Williams distribution (CWD), chirplet transform (CT), and the matching-pursuit-decomposition-based time-frequency distribution (MPTFD) technique. Our simulation results show that SPWVD, SP, CT, and MPTFD can provide TFDs with better localization than the rest in various conditions.

Among the examined TFRs, only matching pursuit decomposition technique (MPTFD) and the chirplet transform are adaptive in nature. Chirplet transform computation is extensive depending on the number of chirps used. MPTFD has its adaptiveness based on the decomposition algorithm [19, 20] and the choice of the dictionary. Both methods can be adjusted to generate TFD which is clean and cross-term free but at the expense of heavy computation. We prefer to leave them out of the comparison since computational efficiency is also one of the requirements for the TFD applications in multimedia security.

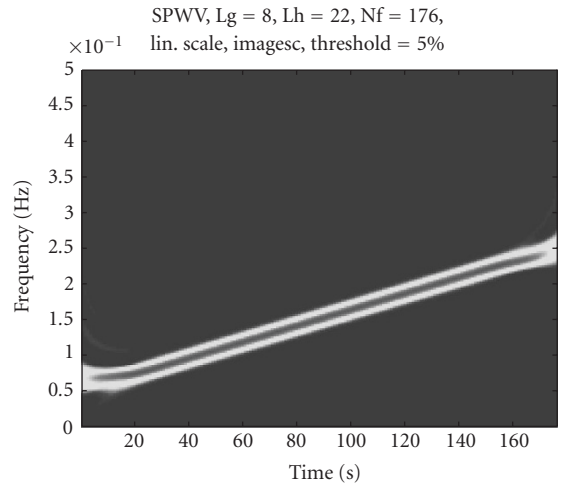
(a)



(b)

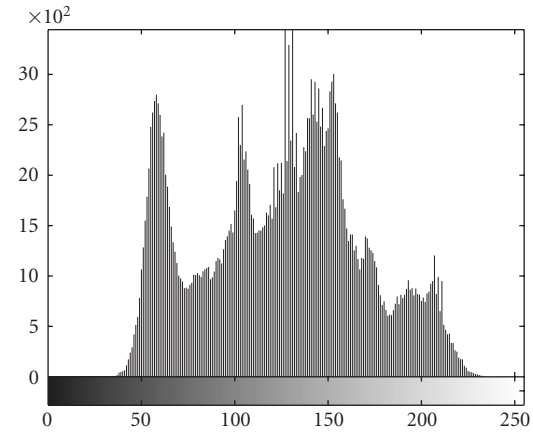


(c)



(d)

(e)



(f)

Figure 2: (a), (b) image with no watermark embedded and its histogram, (c) time domain representation of the linear chirp (watermark), (d) TFD of the linear chirp, (e) the image in (a) with watermark embedded, and (f) its histogram.

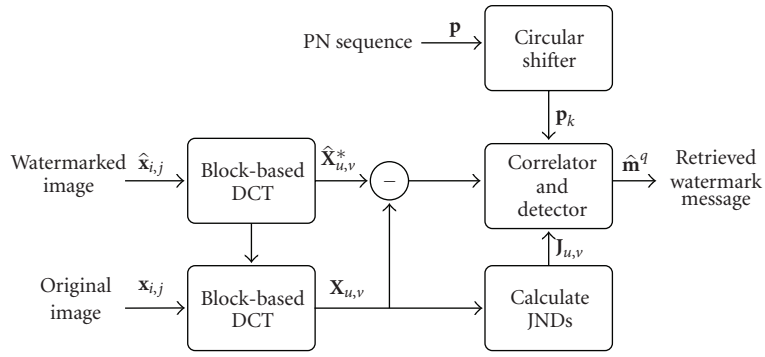


Figure 3: Watermark detection scheme.

Table 1: Multicomponent signal-correlation coefficients between the estimated and referenced IF.

Coef.	WV	PWV	SPWV	CWD	SP
No noise	−0.052	−0.055	0.995	−0.038	0.906
10 dB	0.093	0.100	0.956	0.073	0.893
5 dB	0.105	0.110	0.863	0.087	0.859
1 dB	0.083	0.085	0.697	0.077	0.786
0 dB	0.081	0.082	0.616	0.067	0.732

Table 2: Monocomponent signal-correlation coefficients between the estimated and referenced IF.

Coef.	WV	PWV	SPWV	CWD	SP
No noise	0.961	0.961	0.996	0.992	0.985
10 dB	0.897	0.897	0.996	0.902	0.984
5 dB	0.631	0.633	0.991	0.465	0.981
1 dB	0.222	0.227	0.967	0.209	0.969
0 dB	0.174	0.181	0.956	0.197	0.961

Table 1 gives the result of the correlation coefficients between referenced and estimated instantaneous frequency of a multicomponent signal consists of two linear IF laws intersecting each other under different noise levels. The same simulation was also done on monocomponent FM signal and its results were tabulated in Table 2.

Performance of the TFD estimators varies depending on the input signals' characteristics such as linearity, rate of frequency change, mono- or multicomponent, and the closeness between frequency components in the signal. For monocomponent linear FM signal, almost all estimated IF laws are highly correlated with their corresponding reference. For multicomponent signals, due to the effect of cross-terms, WV and PWV become unreliable tools for estimating IF. SPWVD and SP have high ability to suppress cross-term, their estimated IF is highly correlated with the known IF and less affected by white noise. We prefer SPWV to SP for our image watermark application due to its better joint time-frequency resolution. SPWVD's advanced performance can be contributed to its smoothing kernel design.

All time-frequency distributions which belong to Cohen's class can be represented as a two-dimensional convolution in the equation below [21, 22]:

$$T_x(t,f) = \int_{t'}\int_{f'} \psi_T(t-t', f-f') W_x(t', f')\, dt'\, df', \quad (5)$$

where $W_x(t,f)$ is the Wigner-Ville distribution of the signal $x(t)$ and $\psi_T(t,f)$ is the real value smoothing kernel of the distribution.

The above convolution in time-frequency domain is equivalent to multiplication in the ambiguity domain (τ,ν):

$$T_x(\tau,\nu) = \Psi_T(\tau,\nu) A_x(\tau,\nu), \quad (6)$$

where $\Psi_T(\tau,\nu)$ is calculated as the 2D Fourier transform of the real value kernel $\psi_T(t,f)$:

$$\Psi_T(\tau,\nu) = \int_t\int_f \psi_T(t,f) e^{-j2\pi(\nu t-\tau f)}\, dt\, df, \quad (7)$$

and $A_x(\tau,\nu)$ is the ambiguity function calculated by taking Fourier transform of the Wigner-Ville $W_x(t,f)$:

$$A_x(\tau,\nu) = \int_t x\left(t+\frac{\tau}{2}\right) x^*\left(t-\frac{\tau}{2}\right) e^{-j2\pi\nu t} dt. \quad (8)$$

In the ambiguity domain, the signal auto terms (AT) are centered at the origin while the interference terms (IT) are located away from the origin. The kernel acts as a low-pass filter on the Wigner distribution of the signal, smooths out ITs, and retains the ATs. In order to study the properties of a time-frequency estimator, one has to examine the shape of the corresponding smoothing kernel in the ambiguity domain [21, 22].

Smoothing of interference terms takes away the auto terms and reduces joint TF resolution. Ideally, value of the kernel low-pass filter $\Psi_T(\tau,\nu)$ should be one in the auto term region and zero in the interference term region. If the kernel is too narrow, suppression of IT also takes away some of the AT energy leading to smearing of the TFD. On the other hand, if the kernel shape is too broad, it cannot remove all the

Table 3: Smoothing kernels of the common TFDs.

Distribution	Kernel $\varphi_T(t,\tau)$	Kernel $\Psi_T(\tau,\nu)$
WVD	$\delta(t)$	1
PWVD	$\delta(t)h\left(\frac{\tau}{2}\right)h^*\left(-\frac{\tau}{2}\right)$	$h\left(\frac{\tau}{2}\right)h^*\left(-\frac{\tau}{2}\right)$
SPWVD	$g(t)h\left(\frac{\tau}{2}\right)h^*\left(-\frac{\tau}{2}\right)$	$h\left(\frac{\tau}{2}\right)h^*\left(-\frac{\tau}{2}\right)G(\nu)$
SP	$\gamma\left(-t-\frac{\tau}{2}\right)\gamma^*\left(-t+\frac{\tau}{2}\right)$	$A\gamma(-\tau,-\nu)$
CW	$\sqrt{\frac{\sigma}{4\pi}}\frac{1}{\lvert\tau\rvert}\exp\left[-\frac{\sigma}{4}\left(\frac{t}{4}\right)^2\right]$	$\exp\left[-\frac{(2\pi\tau\nu)^2}{\sigma}\right]$

ITs. This reason explains why a fixed kernel design (not adaptive) cannot work properly for any signal types. High joint time-frequency resolution cannot be achieved at the same time with good interference suppression.

Table 3 lists the smoothing kernels of several estimators in (t,τ) domain and (τ,ν) ambiguity domain [21].

The kernel of the spectrogram,

$$\varphi_T(t,\tau)=\gamma\left(-t-\frac{\tau}{2}\right)\gamma^*\left(-t+\frac{\tau}{2}\right), \tag{9}$$

is the Wigner-Ville distribution of the running window $\gamma(t)$. Its smoothing region is very narrow that it effectively removes all cross-terms at the cost of reduced joint time-frequency resolution. Cross-terms will only be present if the signal terms overlap [21]. In addition, spectrogram suffers from a tradeoff between time and frequency resolution. If a short window is used, smoothing function will be narrow in time and wide in frequency leading to good resolution in time and bad resolution in frequency, and vice versa. The spectrogram is free of cross-terms but it has lower joint time-frequency resolution compared to SPWVD.

SPWV distributions, on the other hand, have more progressive and independent smoothing control both in time and frequency. SPWVD's advanced performance can be contributed to its smoothing kernel design. The kernel of SPWVD and PWVD in time-frequency domain has the form

$$\psi_T(t,f)=g(t)H(f), \tag{10}$$

where $g(t)$ is the time-smoothing window and $h(t)$ is the running analysis window having frequency-smoothing effect. In the ambiguity domain:

$$\begin{aligned}\Psi_T(\tau,\nu)&=H(\tau)G(\nu)\\&=h\left(\frac{\tau}{2}\right)h^*\left(-\frac{\tau}{2}\right)G(\nu).\end{aligned} \tag{11}$$

In WVD, the kernel is always one, therefore no smoothing is made between the regions of the ambiguity domain. In PWVD, $g(t)=\delta(t)$ leads to $G(\nu)=1$, no smoothing is done to remove IT oscillating in time direction, smoothing is only possible for frequency direction. Since SPWVD smoothing is done in both time and frequency direction, most of its cross-terms are attenuated. Smoothing in time and frequency can be adjusted separably with abundant choices of windows $g(t)$ and $h(t)$. The amount of smoothing in time and frequency increases as the length of window $g(t)$ increases and length of window $h(t)$ decreases, respectively. Although smoothing of interference terms (IT) also takes away the auto terms (AT) and reduces joint TF resolution, SPWVD is still more localized than SP and does not suffer from the time-frequency resolution tradeoff. According to [21, 22], SPWVD separable smoothing kernel has the shape of a Gaussian function and its ability to suppress IT does not depend much on signal types as the Choi-Williams distribution (CWD) kernel. In CWD, independent control of time and frequency smoothing is not possible. This limitation as well as the requirement on marginal property reduce the distribution's ability to remove cross-terms and make it less versatile than SPWVD.

4. DISCRETE POLYNOMIAL-PHASE TRANSFORM AND SIGNAL SYNTHESIS

The discrete polynomial-phase transform (DPT) has been extensively studied in recent years [11–13]. It is a parametric signal analysis approach for estimating the phase parameters of polynomial-phase signals. The phase of many man-made signals such as those used in radar, sonar, communications can be modeled as a polynomial. The discrete version of a polynomial-phase signal can be expressed as

$$x(n)=b_0\exp\left\{j\sum_{m=0}^{M}a_m(n\Delta)^m\right\}, \tag{12}$$

where M is the polynomial order ($M=2$ for chirp signal), $0\le n\le N-1$, N is the signal length, and Δ is the sampling interval.

The principle of DPT is as follows. When DPT is applied to a monocomponent signal with polynomial phase of order M, it produces a spectral line. The position of this spectral line at frequency ω_0 provides an estimate of the coefficient $\hat{a}_M$. After $\hat{a}_M$ is estimated, the order of the polynomial is reduced from M to $M-1$ by multiplying the signal with $\exp\{-j\hat{a}_M(n\Delta)^M\}$. This reduction of order is called *phase unwrapping*. The next coefficient $\hat{a}_{M-1}$ is estimated the same way by taking DPT of the polynomial-phase signal of order $M-1$ above. The procedure is repeated until all the coefficients of the polynomial phase are estimated.

DPT order M of a continuous phase signal $x(n)$ is defined as the Fourier transform of the higher-order $\mathcal{DP}_M[x(n),\tau]$ operator:

$$\begin{aligned}\mathrm{DPT}_M\left[x(n),\omega,\tau\right]&\equiv\mathcal{F}\left\{\mathcal{DP}_M[x(n),\tau]\right\}\\&=\sum_{(M-1)^r}^{N-1}\mathcal{DP}_M[x(n),\tau]\exp^{-j\omega n\Delta},\end{aligned} \tag{13}$$

where τ is a positive number and

$$\begin{aligned}\mathcal{DP}_1[x(n),\tau]&:=x(n),\\\mathcal{DP}_2[x(n),\tau]&:=x(n)x^*(n-\tau),\\\mathcal{DP}_M[x(n),\tau]&:=\mathcal{DP}_2[\mathcal{DP}_{M-1}[x(n),\tau],\tau].\end{aligned} \tag{14}$$

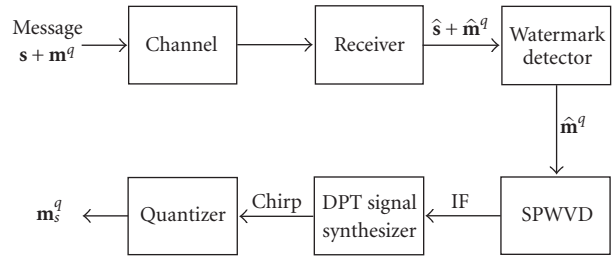


Figure 4: Image watermark detection scheme.

The coefficients a_M (a_1 and a_2) are estimated by applying the following formula:

$$\hat{a}_M = \frac{1}{M!(\tau_M \Delta)^{M-1}} \operatorname{argmax}_\omega \{ | \operatorname{DPT}_M [x(n), \omega, \tau] | \}, \tag{15}$$

where

$$\begin{aligned} \operatorname{DPT}_1 [x(n), \omega, \tau] &= \mathcal{F}\{x(n)\}, \\ \operatorname{DPT}_2 [x(n), \omega, \tau] &= \mathcal{F}\{x(n)x^*(n-)\}, \end{aligned} \tag{16}$$

and

$$\begin{aligned} \hat{a}_0 &= \operatorname{phase}\left\{ \sum_{n=0}^{N-1} x(n) \exp\left\{ -j \sum_{m=1}^{M} a_m (n\Delta)^m \right\} \right\}, \\ \hat{b}_0 &= \frac{1}{N} \sum_{n=0}^{N-1} x(n) \exp\left\{ -j \sum_{m=1}^{M} a_m (n\Delta)^m \right\}. \end{aligned} \tag{17}$$

The estimated coefficients are used to synthesize the polynomial-phase signal:

$$\hat{x}(n) = \hat{b}_0 \exp\left\{ j \sum_{m=0}^{M} \hat{a}_m (n\Delta)^m \right\}. \tag{18}$$

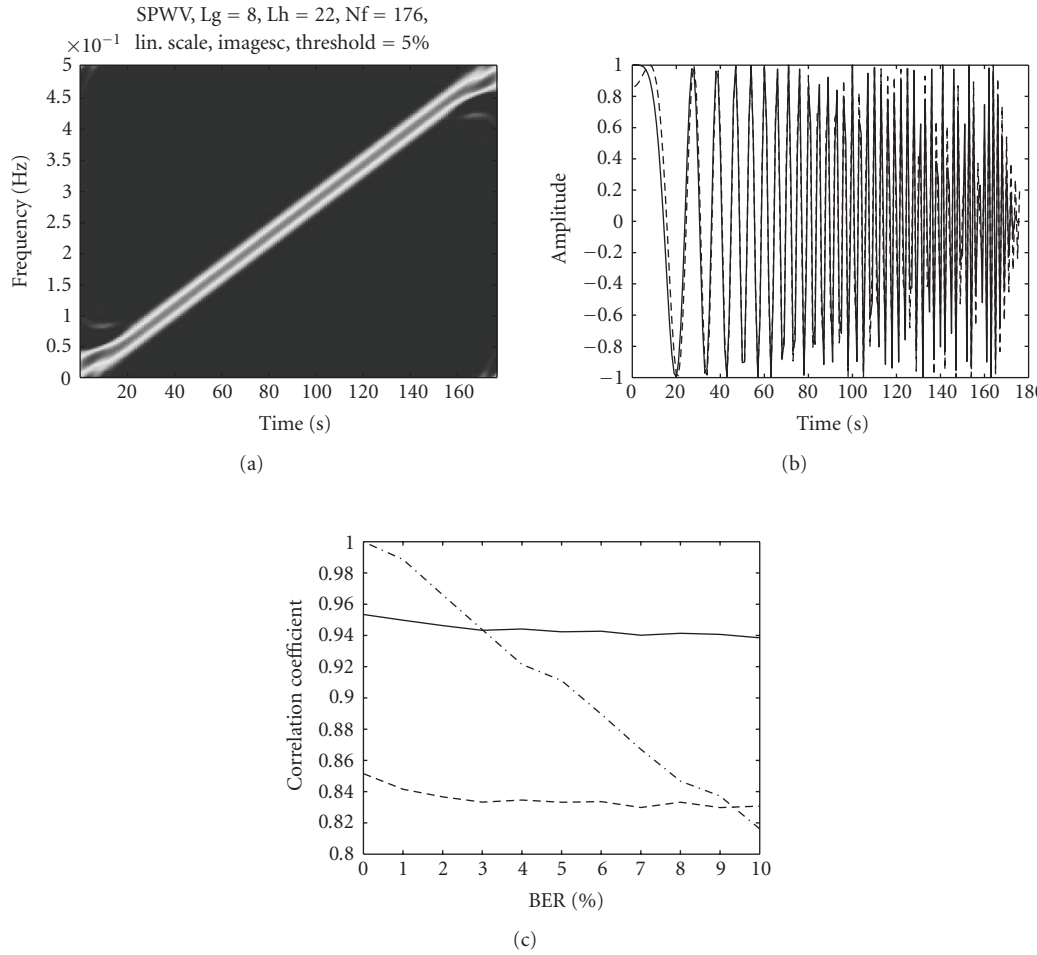
5. APPLICATION: WATERMARK DETECTION IN MULTIMEDIA DATA

The method proposed in this paper synthesizes the polynomial-phase chirp signal using a combination of the time-frequency distribution's property as well as the discrete polynomial-phase transform. This approach and the one in [11] both utilize the fact that the instantaneous frequency equals the derivative of the phase of the signal to estimate the signal phase from the instantaneous frequency. But the method in this paper uses the smoothed pseudo Wigner-Ville distribution as a tool for time-frequency representation of the signal. In addition, instead of using peak tracking algorithm to estimate the instantaneous frequency, the approach proposed in this paper utilizes a very useful property of the TFD theory to generate IF. The IF can be simply obtained by taking the first moment of the TFD. Let $\underline{\mathbf{m}}$ and $\mathbf{m}^q$ be the normalized chirp and its quantized version at the transmitter, respectively. Let $\hat{\mathbf{m}}^q$ be the corrupted quantized chirp at the receiver. To detect the chirp, we apply the time-frequency signal synthesis algorithm described in the previous section. The process involves utilization of phase information which can be obtained from the TFD of the received signal. We use SPWVD to calculate the TFD of $\hat{\mathbf{m}}^q$ instead of using WVD or spectrogram as in our previous works. The detection scheme is illustrated as in Figure 4.

Since the discrete signal that we work on is a quantized version of the chirp signal, its TFD consists of cross-terms in addition to the linear component of the chirp. The cross-terms' energy is smaller than the energy of the linear component, so it can be removed by applying a threshold to the TFD energy. This masking process also removes the noise and unwanted components in the TFD. The current threshold setting is at 0.8 of the maximal energy of the TFD. This value is obtained empirically. A more detailed and systematic analysis of the effect of the environment on the signal can be done so the masking threshold of the TFD can be determined adaptively but this is out of the scope of this paper. The masking process helps to remove unwanted components in the TFD and increase the estimation accuracy of the instantaneous frequency. The monocomponent of interest is extracted from the received signal by *dechirping* with $e^{-j\phi(t)}$, where $\phi(t)$ is obtained by integrating the IF estimated from SPWVD. This extracted monocomponent is then low-pass filtered and translated back into its original location by multiplying with $e^{j\phi(t)}$. The signal at this point can be considered a monocomponent and is subjected to the DPT algorithm as described in the previous section [11, 12].

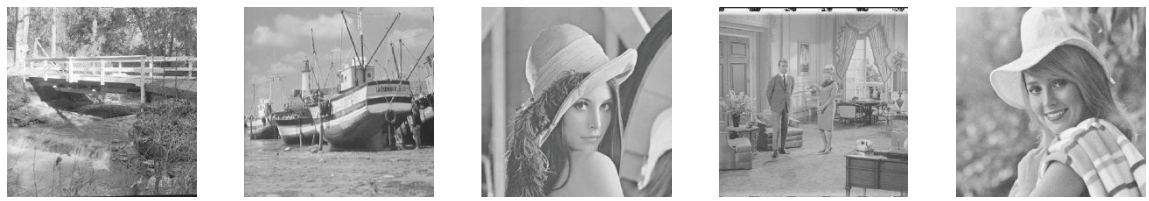
The synthesized version of $\mathbf{m}^q$ is $\mathbf{m}_s^q$ obtained by quantization of $\underline{\hat{\mathbf{m}}}$, where $\underline{\hat{\mathbf{m}}}$ is the chirp estimated from the DPT algorithm. Figure 5(b) shows the original chirp $\underline{\mathbf{m}}$ and its estimated version $\underline{\hat{\mathbf{m}}}$ at BER of 5 percent. Figure 5(c) shows correlation coefficients between the pairs $(\underline{\mathbf{m}}, \underline{\hat{\mathbf{m}}})$, $(\mathbf{m}^q, \hat{\mathbf{m}}^q)$, $(\mathbf{m}^q, \hat{\mathbf{m}}_s^q)$ and they are used as a standard to evaluate the effectiveness of the method.

Figure 6 shows the test images used to evaluate the detection scheme. The size of these images is 512×512. The length of the chirp to be embedded is 176. The sampling frequency f_s is equal to 1 kHz. Therefore, the initial and final frequencies of the chirp to be embedded in the image are constraint to [0–500] Hz. We experimentally found from our previous work that the length of the PN sequence should be at least 10 000 samples for a reliable detection. The number of chirps can be embedded depending on the number of samples in the PN sequence the image can accept. In our watermark



(a) (b) (c)

Figure 5: (a) Time-frequency distribution of the chirp, (b) time domain plot of the original chirp (solid) and synthesized chirp (dashed) corresponding to a correlation coefficient of 0.94 at 5% BER, and (c) correlation coefficients at different BERs between the original and synthesized chirps (solid), between their quantized versions (dashed), and between the quantized original chirp and quantized chirp at the receiver (dash-dotted).



(a) (b) (c) (d) (e)

Figure 6: The test images used in the benchmark.

technique, each image is embedded with only one linear FM chirp. There exists a tradeoff between the data size and robustness of the algorithm. As the length of the PN sequence decreases, the technique will be able to add more bits to the host image but the detection of the hidden bits and resistance to different attacks will be decreased. When the chirp length is increased, the BER resulted from the same attacks compared to the case using the shorter chirp length is decreased. However, as the chirp length increases, the accuracy of the synthesized chirp has a tendency to decrease because any error in the estimated phase coefficients will propagate through the length of the signal. Figure 7 shows the detection result on watermarked image suffered from JPEG compression attack with a quality factor of 20%. Figures 7(a) and 7(b) show the original watermarked and the attacked images with a corresponding BER of 2.84%. The synthesized version of the

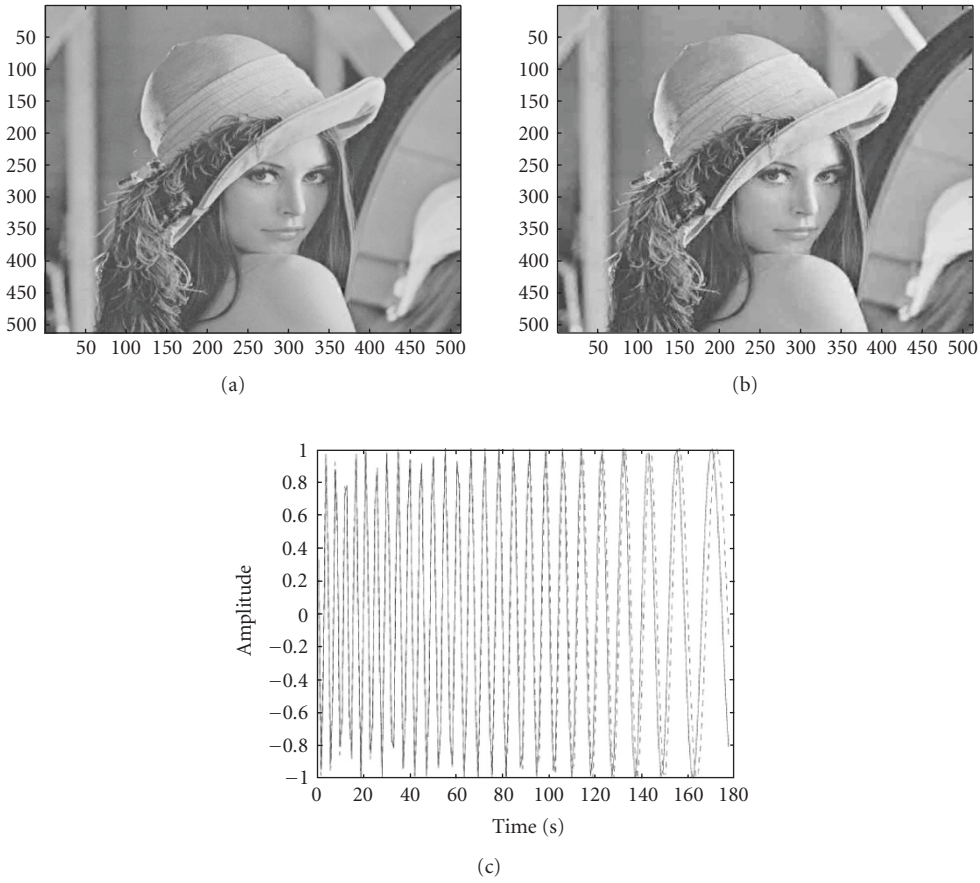


Figure 7: (a) Watermarked image, (b) the same watermarked image after JPEG compression with 20% quality resulting in a BER of 2.84%, and (c) synthesized chirp (solid) and original chirp (dashed) with a correlation coefficient of 0.93.

chirp is highly correlated to the original chirp with a correlation coefficient of 0.93 as shown in Figure 7(c). Our simulation shows that the proposed method successfully detects the watermark under JPEG compression with a quality factor of around 5% or greater. A compression quality factor of less than 5% can result in a BER greater than the detection limit of the proposed method which is about 15%. Figure 8 shows the detection result for the resampling attack case. The watermark image is downsampled and upsampled with corresponding resampling factor of 0.75 and 1.33, respectively. The BER detected in the received chirp is 2.27%. The method successfully detects the chirp with a correlation coefficient of 0.9958 between the original and the synthesized chirps. Similarly, Figure 9 shows the detection result for a watermarked image under wavelet compression attack with a compression factor of 0.3. The corresponding BER and correltion coefficient are 8.5% and 0.9985, respectively.

Table 4 shows the watermark detection on all images as shown in Figure 6 under the geometric attacks according to the benchmarking scheme proposed in [23]. A total of 235 attacks are performed on the five images (47 for each image). The proposed technique can detect the watermark for 197 attacks corresponding to a detection rate of 83.82%. Compare to 84% and 90% of the nonblind algorithm proposed by Xia et al. [24] and Cox et al. [25], respectively, the detection result obtained by the proposed method is very satisfactory considering the fact that it can embed multiple-bit chirp message into the image, successfully detect and synthesize the chirp from its corrupted version.

Table 5 shows the detection result of the method proposed by Pereira et al. [26] with a detection rate of 61%. The method can embed 56 bits into the image but it does not need the original image at the receiver to recover the watermark.

The accuracy of the detection algorithm depends on how precise the synthesized signal is compared to the referenced signal. The estimation of instantaneous frequency contributes significantly to the accuracy of the synthesized signal. If the watermark message involved is a monocomponent signal, the step that uses SPWVD to separate and estimate the monocomponent IF can be dropped and DPT can be applied directly to the signal. Since the IF estimation step can be skipped, the contribution of the error it can possibly create is removed in the final synthesis output. The correlation between the synthesized and referenced chirp signals is, therefore, improved. Table 6 shows the result of the chirp detection on the same signal with and without the IF estimation process through SPWVD. The comparison is done for the continuous and quantized versions of the chirps.

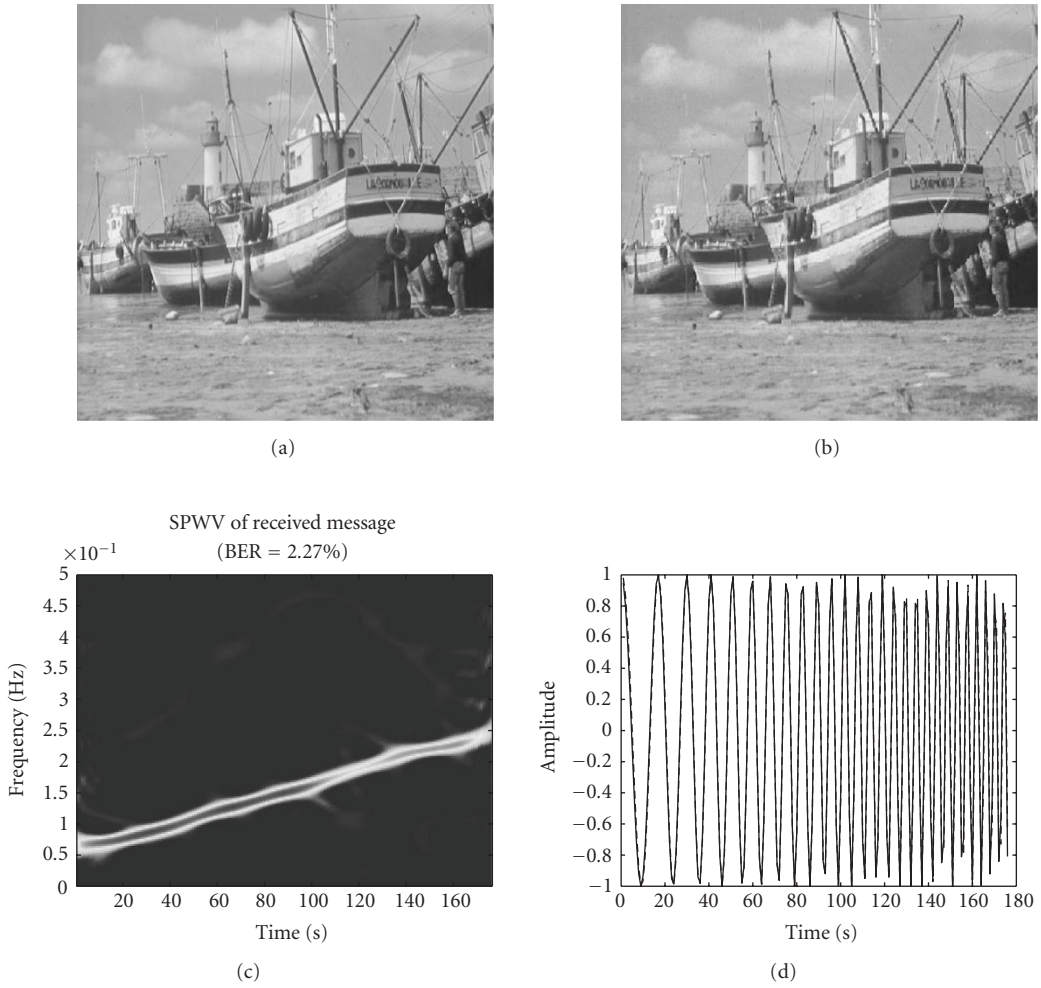


FIGURE 8: (a) Original image, (b) the same image after resampling attack resulted in a BER of 2.27%, (c) TFD of the received chirp, and (d) original chirp (solid) and synthesized chirp (dashed) with a correlation coefficient of 0.9958.

6. DISCUSSION AND CONCLUSION

The success of the estimated polynomial coefficients depends considerably on the initial estimation of the instantaneous frequency. The simulation we performed on different types of signals and noise levels proves that SPWVD is a good choice for determining IF. SPWVD has more versatility to adapt to different types of signals. It can suppress interference terms with least joint time-frequency resolution smearing. We should note that any TFD which is highly localized and cross-term free would be a good choice for the estimation of IF.

The proposed technique, like the HRT method, has the ability to detect the chirp message embedded in image and audio signals and subjected to different BERs due to attacks on the image watermark. The simulations show its robustness for corrupted signal with BER of up to 15%. Since the watermark message is a linear frequency modulated signal, it is easily modeled using polynomial-phase transform. Therefore, the parameters of the chirp such as slope and initial phase, and frequency can be recovered easily and precisely. The proposed technique not only can detect the chirp message but also has the ability of error correction and reconstruction of the original chirp. It can detect and synthesize the chirp signal from distorted TFD having discontinuity in its IF trajectory. Figure 10 shows the TFD of a signal with discontinuity in its IF law and the corresponding synthesized chirp. Both the referenced and synthesized chirps are highly correlated despite the corruption in the instantaneous frequency.

The novelty of the new method is in the fact that it is very efficient in terms of computational complexity (CC). The computational complexity is determined based on the number of multiplications needed to detect a linear chirp having length N. HRT-based method involves the calculation of WVD [27] and taking the standard HRT [28] on the resulted WVD:

$$\begin{aligned} \mathrm{CC(WVD)} &= O(N^2 \log_2 N), \\ \mathrm{CC(HRT)} &= O(N^2 t), \end{aligned} \tag{19}$$

where t is the number of bins used for the quantization of

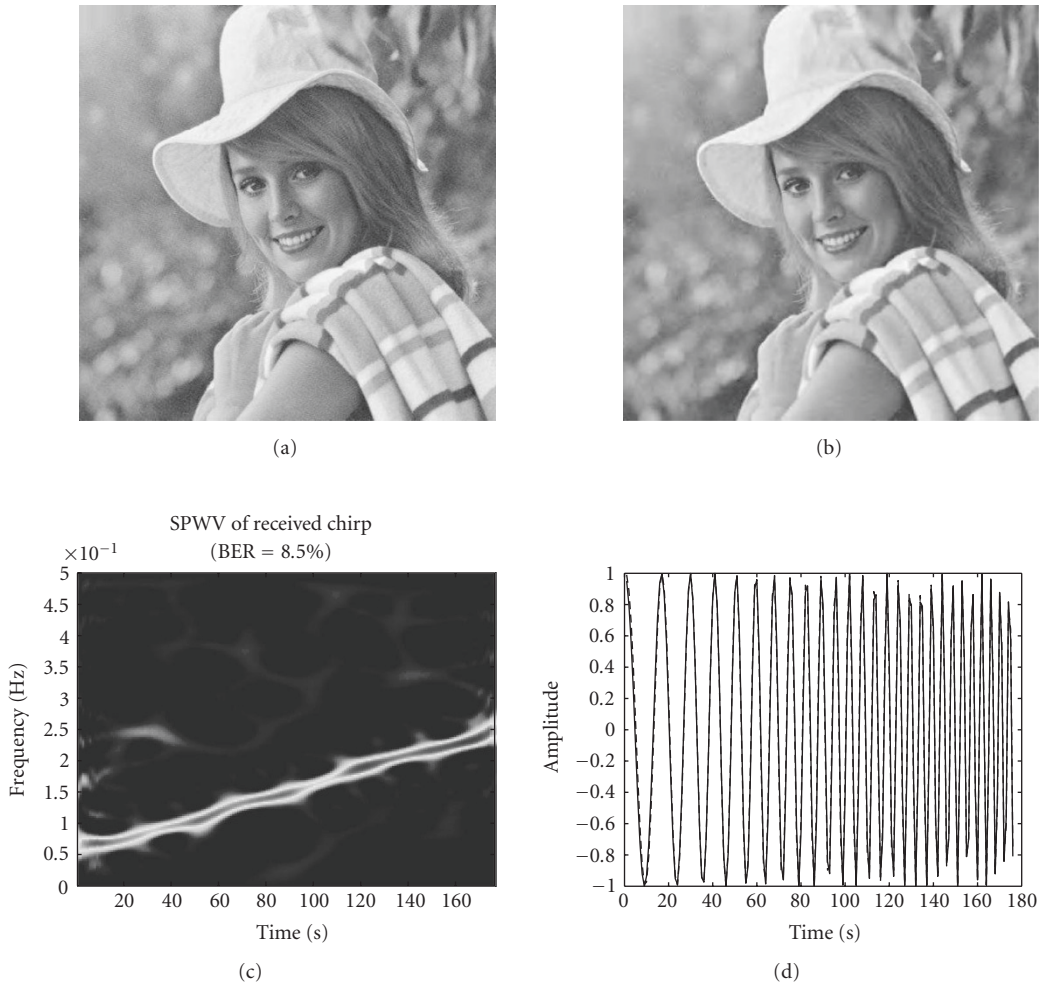


Figure 9: (a) Original image, (b) the same image after wavelet compression attack resulted in a BER of 8.5%, (c) TFD of the received chirp, and (d) original chirp (solid) and synthesized chirp (dashed) with a correlation coefficient of 0.9965.

each axis in the Hough space ($t < \sqrt{2}N$). The total complexity is

$$O(N^2 \log_2 N) + O(N^2 t). \tag{20}$$

The DPT-based method estimates the signal parameter from the TFD (WVD, e.g.). It involves WVD calculation of the chirp signal and taking DPT on the estimated chirp. DPT is calculated by first taking the ambiguity function and then taking the Fourier transform of the ambiguity function. The computational complexity of the ambiguity function calculation and the fast Fourier transform are $O(N)$ and $O(N \log_2 N)$, respectively. Therefore,

$$\mathrm{CC}\,(\,\mathrm{DPT}\,) = O(N) + O(N \log_2 N) \longrightarrow O(N \log_2 N). \tag{21}$$

The total complexity of the DPT-based chirp detection approach is

$$O(N^2 \log_2 N) + O(N \log_2 N). \tag{22}$$

Obviously, computational complexity of the method based on DPT is lower compared to the HRT-based method. The calculation involves DPT-based method for the case WVD is used. If DPT is applied directly to the signal without using the TFD to estimate IF, the total complexity is only $O(N \log_2 N)$. Our simulation shows that on a Pentium 3.0 GHz computer, after the watermark chirp is extracted from the attacked image, the mean time to detect the monocomponent chirp of the proposed DPT-based and HRT-based techniques is 0.0229 and 9.1725 seconds, respectively. This means that the DPT-based detection method is about 400 times faster than the HRT-based method. The similarity between the HRT-based method and DPT-based method proposed in this paper is that both transforms are applied to the TFD. SPWVD is known to have high joint time-frequency resolution and the ability to suppress cross-term due to the smoothing operation involved. Unfortunately, the smoothing operation also increases the total amount of computation needed. Depending on the application at hand, if the chirp to be detected is a monocomponent signal, DPT can be applied directly to it to reduce the computational

Table 4: Result of watermark detection for checkmark benchmark for a chirp length of 176.

Watermark attack	Images					
	1	2	3	4	5	Detection average (%)
Remodulation (4)	4	0	2	3	3	60.00
MAP (6)	6	6	6	6	6	100.00
Copy (1)	1	1	1	1	1	100.00
Wavelet (10)	8	7	9	9	9	84.00
JPEG (12)	12	11	12	12	12	98.33
ML (7)	4	2	6	3	6	60.00
Filtering (3)	3	3	3	3	3	100.00
Resampling (2)	2	1	2	1	2	80.00
Color reduce (2)	2	0	1	1	1	50.00

Table 5: Result of watermark detection for the blind detection scheme proposed by Pereira.

Watermark attack	Images					
	1	2	3	4	5	Detection average (%)
Remodulation (4)	0	0	0	2	0	10
MAP (6)	1	2	2	5	1	37
Copy (1)	1	1	0	1	0	60
Wavelet (10)	3	6	8	7	7	62
JPEG (12)	12	12	12	12	12	100
ML (7)	0	3	3	3	0	26
Filtering (3)	3	3	3	3	3	100
Resampling (1)	1	1	1	1	1	100
Color reduce (2)	0	1	1	2	0	40

complexity. If the signal involved is a multicomponent signal, then the individual components have to be extracted first before DPT can be used. Because of this reason, SPWVD should be used in place of WVD.

So with the new proposed method, tradeoff problem between speed and accuracy in HRT-based method is solved. Faster detection is thus allowed together with accuracy. This is also the motivation for the proposed approach for chirp detection in a real-time application such as image watermarking of multimedia data.

We performed a novel technique to detect watermark message on images with different BERs. Our simulation showed high correlation (> 0.9) between $\underline{\mathbf{m}}$ and $\underline{\hat{\mathbf{m}}}$. As a result of the quantization process, the correlation between $\mathbf{m}^q$ and $\hat{\mathbf{m}}_s^q$ is lower than the previous case but it is still larger than 0.8 at different BERs. These correlation coefficients can help us to confidently make decision on the transmitted chirp's properties. In our simulation, the watermark message embedded in the images is monocomponent linear FM chirp.

Table 6: Correlation coefficient between the synthesized and referenced chirps in continuous and quantized forms with and without the IF estimation process.

BER	With IF estimation		Without IF estimation	
	Continuous	Quantized	Continuous	Quantized
0	0.9534	0.8515	0.9889	0.8861
1	0.9497	0.8416	0.9889	0.8898
2	0.9463	0.8367	0.9884	0.8925
3	0.9433	0.8333	0.9859	0.8901
4	0.9441	0.8346	0.9854	0.8886
5	0.9423	0.8332	0.9838	0.8879
6	0.9427	0.8336	0.9835	0.8889
7	0.9402	0.8298	0.9811	0.8836
8	0.9414	0.8333	0.9814	0.8876
9	0.9407	0.8298	0.9813	0.8856

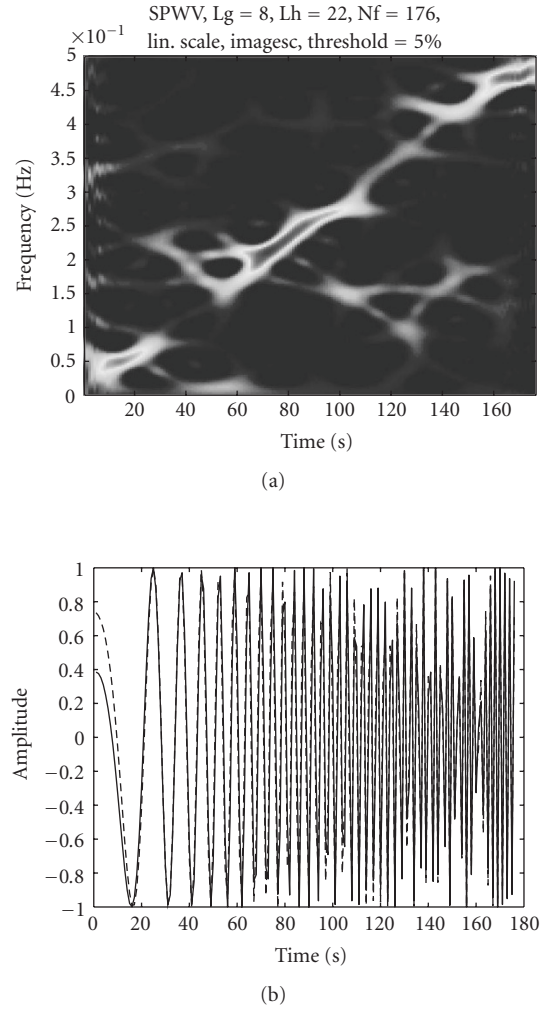


Figure 10: (a) Time-frequency distribution of the chirp with discontinuity in its IF law corresponding to a BER of 17%, and (b) time domain plot of the original chirp (solid) and synthesized chirp (dashed) corresponding to a correlation coefficient of 0.92.

However, the method can be extended to detect image watermark messages consisting of multicomponent linear chirps.

REFERENCES

[1] A. Ramalingam and S. Krishnan, "A novel robust image watermarking using a chirp based technique," in *Proceedings of the Canadian Conference on Electrical and Computer Engineering*, vol. 4, pp. 1889–1892, Niagara Falls, Ontario, Canada, May 2004.

[2] S. Erkucuk, S. Krishnan, and M. Zeytinoglu, "Robust audio watermarking using a chirp based technique," in *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME '03)*, vol. 2, pp. 513–516, Baltimore, Md, USA, July 2003.

[3] S. Kay and G. F. Boudreaux-Bartels, "On the optimality of the Wigner distribution for detection," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '85)*, vol. 10, pp. 1017–1020, Tampa, Fla, USA, April 1985.

[4] J. S. Dhanoa, E. J. Hughes, and R. F. Ormondroyd, "Simultaneous detection and parameter estimation of multiple linear chirps," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '03)*, vol. 6, pp. 129–132, Hong Kong, April 2003.

[5] J. Torres, P. Cabiscol, and J. Grau, "Radar chirp detection through wavelet transform," in *Proceedings of the 5th Biannual World Automation Congress*, vol. 13, pp. 227–232, Orlando, Fla, USA, June 2002.

[6] R. M. Rangayyan and S. Krishnan, "Feature identification in the time-frequency plane by using the Hough-Radon transform," *Pattern Recognition*, vol. 34, no. 6, pp. 1147–1158, 2001.

[7] Y. Sun and P. Willett, "Hough transform for long chirp detection," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 38, no. 2, pp. 553–569, 2002.

[8] B. D. Carlson, E. D. Evans, and S. L. Wilson, "Search radar detection and track with the Hough transform. I: system concept," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 30, no. 1, pp. 102–108, 1994.

[9] G. F. Boudreaux-Bartels and T. Parks, "Time-varying filtering and signal estimation using Wigner distribution synthesis techniques," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 34, no. 3, pp. 442–451, 1986.

[10] W. Krattenthaler and F. Hlawatsch, "Time-frequency design and processing of signals via smoothed Wigner distributions," *IEEE Transactions on Signal Processing*, vol. 41, no. 1, pp. 278–287, 1993.

[11] A. Francos and M. Porat, "Analysis and synthesis of multicomponent signals using positive time-frequency distributions," *IEEE Transactions on Signal Processing*, vol. 47, no. 2, pp. 493–504, 1999.

[12] S. Peleg and B. Friedlander, "Multicomponent signal analysis using the polynomial-phase transform," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 32, no. 1, pp. 378–387, 1996.

[13] S. Peleg and B. Friedlander, "The discrete polynomial-phase transform," *IEEE Transactions on Signal Processing*, vol. 43, no. 8, pp. 1901–1914, 1995.

[14] S.-J. Lee and S.-H. Jung, "A survey of watermarking techniques applied to multimedia," in *Proceedings of the IEEE International Symposium on Industrial Electronics (ISIE '01)*, vol. 1, pp. 272–277, Pusan, South Korea, June 2001.

[15] B. G. Mobasseri, "Digital watermarking in joint time-frequency domain," in *Proceedings of the International Conference on Image Processing (ICIP '02)*, vol. 3, pp. 481–484, Rochester, NY, USA, September 2002.

[16] F. Sattar and B. Barkat, "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 (ISSPA '03)*, vol. 2, pp. 363–366, Paris, France, July 2003.

[17] W. Bender, D. Gruhl, N. Morimoto, and A. Lu, "Techniques for data hiding," *IBM Systems Journal*, vol. 35, no. 3-4, pp. 313–336, 1996.

[18] C. I. Podilchuk and W. Zeng, "Image-adaptive watermarking using visual models," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 4, pp. 525–539, 1998.

[19] S. Krishnan, "Instantaneous mean frequency estimation using adaptive time-frequency distributions," in *Proceedings of the Canadian Conference on Electrical and Computer Engineering*, vol. 1, pp. 141–146, Toronto, Ontario, Canada, May 2001.

[20] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3397–3415, 1993.

[21] 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.

[22] F. Hlawatsch, T. G. Manickam, R. L. Urbanke, and W. Jones, "Smoothed pseudo-Wigner distribution, Choi-Williams distribution, and cone-kernel representation: ambiguity-domain analysis and experimental comparison," *Signal Processing*, vol. 43, no. 2, pp. 149–168, 1995.

[23] S. Pereira, S. Voloshynovskiy, M. Madueno, S. Marchand-Maillet, and T. Pun, "Second generation benchmarking and application oriented evaluation," in *Proceedings of the Information Hiding Workshop III*, Pittsburgh, Pa, USA, April 2001.

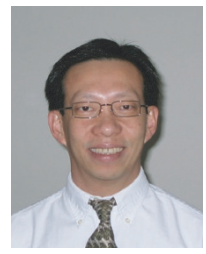
[24] X. Xia, C. G. Boncelet, and G. R. Arce, "A multiresolution watermark for digital images," in *Proceedings of the IEEE International Conference on Image Processing*, vol. 1, pp. 548–551, Santa Barbara, Calif, USA, October 1997.

[25] I. Cox, F. Leighton, and T. Shamoon, "Secure spread spectrum watermarking for multimedia," *IEEE Transactions on Image Processing*, vol. 6, no. 12, pp. 1673–1687, 1997.

[26] S. Pereira, S. Voloshynovskiy, and T. Pun, "Optimal transform domain watermark embedding via linear programming," *Signal Processing*, vol. 81, no. 6, pp. 1251–1260, 2001, special issue: information theoretic issues in digital watermarking.

[27] F. Peyrin and R. Prost, "A unified definition for the discrete-time, discrete-frequency, and discrete-time/frequency Wigner distributions," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 34, no. 4, pp. 858–867, 1986.

[28] S. Pavel and S. G. Akl, "Efficient algorithms for the Hough transform on arrays with reconfigurable optical buses," in *Proceedings of 10th International Parallel Processing Symposium (IPPS '96)*, pp. 697–701, Honolulu, Hawaii, USA, April 1996.



Lam Le received his B.S. degree in chemistry from Saigon, Vietnam, in 1991. He received his B.Eng. and M.A.S. degree in electrical engineering from Ryerson University, Toronto, Canada, in 2003 and 2005, respectively. He has been working in the Signal Analysis Research (SAR) Group led by Dr. Sri Krishnan since May 2002 and his research works are supported by Natural Sciences and Engineering Research Council

of Canada (NSERC), CITO, and Ryerson University. His research interests are in the areas of biomedical system engineering, biomedical signal analysis, discrete polynomial-phase transform, instantaneous frequency extraction, and joint time-frequency distribution of nonstationary signals.

Sridhar Krishnan received the B.E. degree in electronics and communication engineering from Anna University, Madras, India, in 1993, and the M.S. and Ph.D. degrees in electrical and computer engineering from the University of Calgary, Calgary, Alberta, Canada, in 1996 and 1999, respectively. He joined the Department of Electrical and Computer Engineering, Ryerson University, Toronto, Ontario, Canada, in July 1999, and currently he is an Associate Professor and Chairman of the department. His research interests include adaptive signal processing, biomedical signal/image analysis, and multimedia processing and communications.