Estimation of Time-Varying Velocities of Moving Objects by Time-Frequency Representations

Igor Djurović, Srdjan Stanković

Abstract— **An approach to the estimation of motion parameters of moving objects in a video-sequence, by using the SLIDE (Subspace-based line detection algorithm) algorithm, is considered. The proposed procedure projects video-frames to the coordinate axes, in order to obtain synthetic images containing information about the motion parameters. These synthetic images are mapped to the FM signals by using constant μ-propagation. The problem of velocity estimation is reduced to the instantaneous frequency (IF) estimation. IF estimators, based on time-frequency (TF) representations, are used. Three TF representations: spectrogram (SPEC), Wigner distribution (WD), and S-method (SM), are used and compared to this aim. A tradeoff between concentration of the TF representation (velocity estimation accuracy) and reduction of the cross-terms (possibility for estimation of the multiple objects parameters) is achieved by the SM. A performance analysis of the algorithm is done. Theoretical results are illustrated on several numerical examples.**

I. INTRODUCTION

Time-frequency (TF) representations have a significant practical importance for high-resolution analysis of signals with time-varying parameters. Here we mention several interesting areas where the TF based methods have been used: mechanical and biological systems, musical and speech signals, radar signals, analysis of optical interferograms and multidimensional signals, etc. [1]-[7]. Estimation of time-varying velocities using TF representations is the topic of this paper. The presented approach is based on frame projections and the SLIDE (Subspace-based line detection) algorithm [8], [9]. This algorithm is originally introduced to analyze line parameters. Numerous, very interesting, applications are

presented in [8]. A key step of the SLIDE algorithm is to map a line into a complex sinusoid. By using frame projections on coordinate axes the velocity of an object is mapped into an image, containing lines [9]. In order to transform line parameters to the complex sinusoid parameters, the SLIDE algorithm with constant μ-propagation is applied. Velocity parameters were estimated based on spectral characteristics of complex sinusoids [9]-[11]. Very accurate results are obtained when objects have uniform velocities. However, for nonuniform velocities the SLIDE algorithm produces a frequency modulated (FM) signal. This is the reason why the instantaneous frequency (IF) estimators, based on the TF distributions, are used in this paper. The IF is estimated as a position of the TF representation maxima. Displacement techniques [12]-[14], are applied for improvement of the estimator accuracy.

The primary goal of the proposed approach, as well as of the original method [9], is to determine motion parameters of several relatively small moving objects in video-sequences in an efficient way, reducing the problem from the analysis of a 3D video-sequence to the spectral analysis of 1D FM signals. Aerial traffick monitoring [9] is one of the practical applications areas. There are several other methods for tracking moving objects in video-sequences [9], [15]-[18]. They can be applied in the case of translational motion model or for objects with constant velocities. The proposed method does not assume any specific parametric motion model and it can be applied for tracking objects with varying velocities. IF (velocity) estimators based on the TF representations provide a highly accurate estimates. The proposed estimators are robust with respect

to noise influence. Note that there are several other approaches for velocity estimation in video-sequences where the Fourier transform (FT) is used [19]-[21]. Recently, spectral analysis based methods were employed for a more general case of the optical flow estimation where displacement is calculated for each point in a sequence [22], [23]. Furthermore, FT based techniques are applied for segmentation and analysis of dynamic spatiotemporal image sequences [19], [24], [25]. An overview and comparison of the optical flow estimation techniques can be found in [26]. Almost all optical flow estimation techniques assume a three stage procedure: (a) Smoothing and/or filtering; (b) Determination of the spatiotemporal derivatives; (c) Integration of measurements to produce the 2D flow fields. Our procedure can also be represented by a three stage algorithm: (a) Transformation of the 3D video-sequence to the FM signals; (b) Calculation of a TF representation; (c) Motion parameters extraction by IF estimation. For a video-sequence with large amount of noise a spatiotemporal pre-filtering can be employed in the initial stage of the algorithm [26].

The paper is organized as follows. Determination of signals by the SLIDE algorithm is described in Section II. An algorithm for time-varying velocity estimation is presented in Section III. Application of the various TF representations is considered in Section IV. The performance analysis of proposed algorithm with application of the displacement technique is given in Section V.

II. Theoretical Background

Assume that a video-sequence $i(x, y, t)$ may be considered as a superposition of M relatively small moving objects $s^m(x, y)$, $m = 1, ..., M$, and a stationary background $f(x, y)$, where (x, y) are spatial coordinates and t is a time-instant. Let the moving objects be centered around positions given by the coordinate pairs $(\varphi_x^m(t), \varphi_y^m(t))$, $m = 1, ..., M$, at an instant t. Then the video-sequence can be modeled as:

$$i(x, y, t) = f(x, y) +$$

$$+ \sum_{m=1}^{M} s^m(x - \varphi_x^m(t), y - \varphi_y^m(t)). \quad (1)$$

The initial positions of objects are $(x_0^m, y_0^m) = (\varphi_x^m(0), \varphi_y^m(0))$, $m = 1, ..., M$, while the projections of the velocities are

$$\tilde{\mathbf{v}}^m = (v_x^m(t), v_y^m(t))$$

$$= \left(\frac{d\varphi_x^m(t)}{dt}, \frac{d\varphi_y^m(t)}{dt} \right), m = 1, 2, ..., M. \quad (2)$$

Velocity is represented in pixels per frame.

The main idea of our approach is mapping the 3-D video-sequence (1) into two 1-D FM signals, in order to reduce the problem dimension and to employ highly accurate spectral analysis techniques for motion parameter estimation. These signals will contain information about object motion parameters. The first step in this procedure is determination of two synthetic images by using projections of the video-sequence onto coordinate axes. Consider the projections of frames on x and y axes

$$P_x(x, t) = \sum_y i(x, y, t) \quad P_y(y, t) = \sum_x i(x, y, t). \quad (3)$$

By inserting (1) in (3), the projection $P_x(x, t)$ can be written as

$$P_x(x, t) = \sum_y i(x, y, t) = \sum_y f(x, y) +$$

$$+ \sum_y \sum_{m=1}^{M} s^m(x - \varphi_x^m(t), y - \varphi_y^m(t)) =$$

$$= F_x(x) + \sum_{m=1}^{M} s_x^m(x - \varphi_x^m(t)), \quad (4)$$

where $F_x(x) = \sum_y f(x, y)$, and $s_x^m(x, y) = \sum_y s^m(x, y)$, $m = 1, ..., M$, are projections of the background and objects, respectively. Similar expression holds for the y-axis projection

$$P_y(y, t) = F_y(y) + \sum_{m=1}^{M} s_y^m(y - \varphi_y^m(t)). \quad (5)$$

In the following we will analyze only the estimation of parameters of moving object along

the x-axis. In order to eliminate influence of the background, differentiation of $P_x(x,t)$ with respect to t is done

$$\frac{\partial P_x(x,t)}{\partial t} = P_x(x,t+1) - P_x(x,t) =$$

$$= \sum_{m=1}^{M} [s_x^m(x-\varphi_x^m(t+1)) - s_x^m(x-\varphi_x^m(t))]. \quad (6)$$

Now we want to prove that (6) possesses patterns that correspond to the motion parameters. By expanding $\varphi_x^m(t+1)$ into the Taylor series

$$\varphi_x^m(t+1) = \varphi_x^m(t) + \sum_{k=1}^{\infty} \frac{1}{k!}\frac{d^k\varphi_x^m(t)}{dt^k} \quad (7)$$

from (6) follows

$$\frac{\partial P_x(x,t)}{\partial t} =$$

$$= \sum_{m=1}^{M} \left[s_x^m\left(x - \varphi_x^m(t) - \sum_{k=1}^{\infty}\frac{1}{k!}\frac{d^k\varphi_x^m(t)}{dt^k}\right) - \right.$$

$$\left. - s_x^m(x - \varphi_x^m(t))\right]. \quad (8)$$

The first term on the right hand side of (8) can approximately be written as:

$$s_x^m\left(x - \varphi_x^m(t) - \sum_{k=1}^{\infty}\frac{1}{k!}\frac{d^k\varphi_x^m(t)}{dt^k}\right) \approx$$

$$\approx s_x^m(x - \varphi_x^m(t)) -$$

$$- \sum_{k=1}^{\infty}\frac{1}{k!}\frac{d^k\varphi_x^m(t)}{dt^k}\frac{\partial s_x^m(x - \varphi_x^m(t))}{\partial x}. \quad (9)$$

Then, the projection partial derivative with respect to t is (6)

$$\frac{\partial P_x(x,t)}{\partial t} \approx - \sum_{m=1}^{M} \left(\sum_{k=1}^{\infty}\frac{1}{k!}\frac{d^k\varphi_x^m(t)}{dt^k}\right) \times$$

$$\times \frac{\partial s_x^m(x - \varphi_x^m(t))}{\partial x} =$$

$$= \sum_{m=1}^{M} \left(\sum_{k=1}^{\infty}\frac{1}{k!}\frac{d^k\varphi_x^m(t)}{dt^k}\right) \Pi_x^m(x - \varphi_x^m(t)) \quad (10)$$

where $\Pi_x^m(x) = -\partial s_x^m(x)/\partial x$. Therefore, we obtained the synthetic image $\partial P_x(x,t)/\partial t$ with x and t coordinates that keep information about objects' coordinate $\varphi_x^m(t)$, $m = 1,...,M$. Direct estimation of the motion parameters from $\partial P_x(x,t)/\partial t$ can be inaccurate due to the noise influence.

Then, transformation of the synthetic image $\partial P_x(x,t)/\partial t$ to an FM signal is performed, in order to produce a signal whose parameters can be analyzed by the spectral analysis techniques. The constant μ-propagation [8], can be used for this purpose. This technique stems from the detection of direction-of-arrival of a signal to a sensor array. Its application to the estimation of line parameters is proposed in [8]. Mapping of the 2-D function $\partial P_x(x,t)/\partial t$ into the frequency domain, by using the constant μ-propagation, can be performed as [8], [9]

$$z_x(t) = \sum_x [\partial P(x,t)/\partial t]e^{j\mu x} =$$

$$= \sum_{m=1}^{M} \left(\sum_{k=1}^{\infty}\frac{1}{k!}\frac{d^k\varphi_x^m(t)}{dt^k}\right)\sum_x \Pi_x^m(x-\varphi_x^m(t))e^{j\mu x}$$

$$= \sum_{m=1}^{M} \left(\sum_{k=1}^{\infty}\frac{1}{k!}\frac{d^k\varphi_x^m(t)}{dt^k}\right)\Phi_x^m(\mu)\exp(j\mu\varphi_x^m(t))$$

$$(11)$$

where $\Phi_x^m(\mu)$ is the FT of signal $\Pi_x^m(x)$

$$\Phi_x^m(\mu) = \sum_x \Pi_x^m(x)\exp(j\mu x). \quad (12)$$

Now the motion parameters $\varphi_x^m(t)$ of moving objects can be extracted from the signal $z_x(t)$, by using some of the spectral analysis tools. For constant velocity of the moving object, the FT based methods are employed [8], [9]. However, for objects with varying velocity more accurate results may be obtained by using the TF representations, as it will be shown later.

As a result of the object motion the value of $\partial P_x(x,t)/\partial t$ will have two components: positive and negative. Since these two components produce the same spectral content, we introduce a modified function

$$\tilde{P}_x(x,t) = \begin{cases} \frac{\partial P_x(x,t)}{\partial t} & \frac{\partial P_x(x,t)}{\partial t} \geq 0 \\ 0 & \frac{\partial P_x(x,t)}{\partial t} < 0 \end{cases} \quad (13)$$

in order to remove one of these components. Therefore, the analysis of the object motion parameters is reduced to the analysis of motion of one of the object edges. We will assume that the function $\Pi_x^m(x)$ can approximately be written as

$$\Pi_x^m(x) \approx A_m \delta(x - \Delta x^m) \qquad (14)$$

where Δx^m contains information about the object dimension. >From (12) and (14) follows

$$\Phi_x^m(\mu) \approx A_m \exp(j\mu\Delta x^m). \qquad (15)$$

This relationship holds for relatively small variations of object position between the consecutive frames.

Consider the case of constant velocities

$$v_x^m(t) = \frac{d\varphi_x^m(t)}{dt} = v_x^m = \text{const}$$

$$\frac{d^k\varphi_x^m(t)}{dt^k} = 0 \text{ for } k \geq 2. \qquad (16)$$

From (11), (14), and (16), follows

$$z_x(t) = \sum_{m=1}^{M} A_m v_x^m \exp(j\mu\Delta x^m + j\mu\varphi_x^m(t)). \qquad (17)$$

The FT of $z_x(t)$ is

$$Z_x(\omega) = \sum_t z_x(t)e^{-j\omega t} =$$

$$= \sum_{m=1}^{M} A_m v_x^m \exp(j\mu\Delta x^m + j\mu x_0^m)$$

$$\times \delta\left(\omega - \mu\frac{d\varphi_x^m(t)}{dt}\right) =$$

$$= \sum_{m=1}^{M} A_m v_x^m \exp(j\mu\Delta x^m + j\mu x_0^m)\delta(\omega - \mu v_x^m). \qquad (18)$$

In the case of single moving object ($M = 1$), we have:

$$Z_x(\omega) = A_1 v_x \exp(j\mu\Delta x + j\mu x_0)\delta(\omega - \mu v_x). \qquad (19)$$

Thus, position of the FT maximum determines the object velocity

$$\hat{\omega} = \arg\max_{\omega} |Z_x(\omega)| = \mu v_x \Longrightarrow v_x = \hat{\omega}/\mu. \qquad (20)$$

Observe that the velocity estimation can be done by using position of the FT maximum. The signal phase contains information about the object initial position x_0. Since there is an ambiguity in the phase function, it cannot be reliable for this purpose. One approach that solves this problem, based on the variable μ-propagation, is proposed in [27].

Note that we have considered an idealized motion model. In a real video sequence, a spatiotemporal noise is present. It can be reduced by filtering in the space domain and/or by employing some more sophisticated derivation technique in (6). Details on this topic can be found in [26].

Example 1: Consider a time-invariant background of the size 256×256. The scene contains two moving objects with initial positions $(\varphi_x^1(0), \varphi_y^1(0)) = (x_0^1, y_0^1) = (2, 8)$ and $(\varphi_x^2(0), \varphi_y^2(0)) = (x_0^2, y_0^2) = (13, 217)$. We have considered 100 frames. The objects' velocities are: $(d\varphi_x^1(t)/dt, d\varphi_y^1(t)/dt) = (v_x^1, v_y^1) = (1.5, 2)$ and $(d\varphi_x^2(t)/dt, d\varphi_y^2(t)/dt) = (v_x^2, v_y^2) = (2, -1.2)$. The projection $P_x(x, t)$ is shown in Fig. 1a, while $\partial P_x(x, t)/\partial t$ is shown in Fig. 1b. Fig. 1b shows the lines which are the basis for application of the SLIDE algorithm to the velocity estimation. The signal $z_x(t)$ is produced by using the constant μ-propagation with $\mu = 1$. It is depicted in Fig. 1c. The FT of $z_x(t)$ is shown in Fig. 1d. Estimated parameters are: $(\hat{v}_x^1, \hat{v}_y^1) = (1.508, 2.011)$ and $(\hat{v}_x^2, \hat{v}_y^2) = (2.011, -1.194)$.

III. SLIDE ALGORITHM AND OBJECTS WITH VARIABLE VELOCITIES

In the previous application a reasonable minimum number of frames for velocity estimation is 20, i.e., time interval of 0.8s for the video sequence with 25 frames/s. Since the velocity can significantly vary within this interval, our intention is to develop an algorithm which will be able to track time-varying velocities in each considered frame. The previous algorithm can be modified by applying the TF representations to the signal $z_x(t)$.

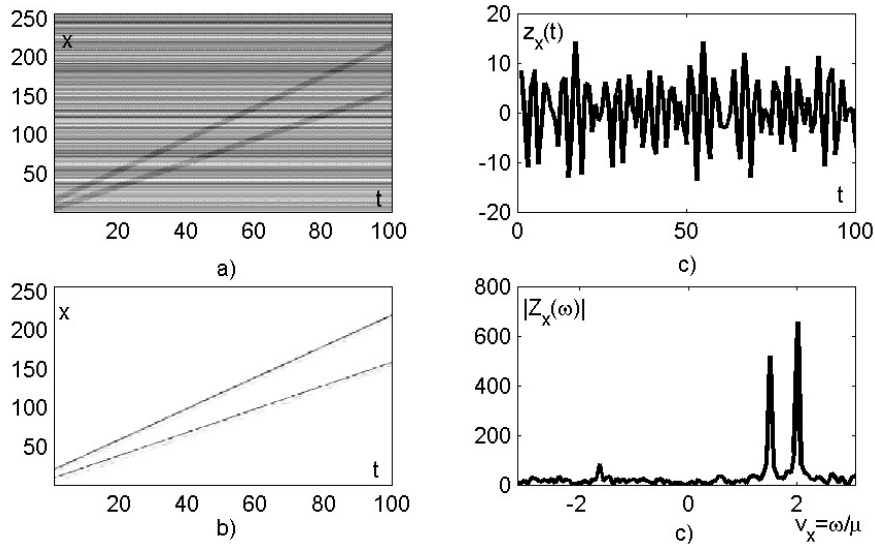Consider a linear velocity variation, with the

Fig. 1.   Velocity estimation by using the SLIDE algorithm: a) $P_x(x,t)$; b) $\partial P_x(x,t)/\partial t$; c) $z_x(t)$; and d) $|Z_x(\omega)|$.

position function:

$$\varphi_x^m(t) = x_0^m + v_{x0}^m t + a_x^m t^2/2, m = 1, ..., M \tag{21}$$

where x_0^m is the initial position, v_{x0}^m is the initial velocity, and a_x^m is the acceleration. The object velocity is

$$v_x^m(t) = \frac{d\varphi_x^m(t)}{dt} = v_{x0}^m + a_x^m t. \tag{22}$$

The signal $z_x(t)$ can be represented as

$$z_x(t) = \sum_{m=1}^{M} (v_{x0}^m + a_x^m t + a_x^m/2)$$

$$\times \Phi_x^m(\mu) \exp(j\mu\varphi_x^m(t)) =$$

$$= \sum_{m=1}^{M} (v_x^m(t) + a_x^m/2) A_m \exp(j\mu\Delta x^m)$$

$$\times \exp(j\mu(x_0^m + v_{x0}^m t + a_x^m t^2/2)). \tag{23}$$

Variable phase function associated with the m-th object is given by

$$\phi_x^m(t) = \mu(\Delta x^m + x_0^m + v_{x0}^m t + a_x^m t^2/2) \tag{24}$$

while the corresponding IF is

$$\omega_x^m(t) = \frac{d\phi_x^m(t)}{dt} = \mu(v_{x0}^m + a_x^m t) = \mu v_x^m(t). \tag{25}$$

Therefore, we will use the IF estimation for analysis of object's velocity. Numerous IF estimators are used in practice. An excellent review of these estimators is given in widely used Boashash paper [28]. For monocomponent signals, the commonly used approach is the IF estimation based on the position of the TF distribution maxima [29]-[31]. Several approaches are also developed for the IF estimation in the case of multicomponent signals [32], [33]. Here, we will use the following procedure:

1) Determination of $z_x(t)$ (or $z_y(t)$);

2) Mapping of the signal $z_x(t)$ into the TF plane by using a TF representation:

$$z_x(t) \xrightarrow{(t,\omega)} \mathrm{TF}_{zx}(t,\omega). \tag{26}$$

3) IF estimation corresponding to the first object velocity. It will be done by using the maximum of the TF distribution:

$$\hat{\omega}_x^1(t) = \arg\max_{\omega} \mathrm{TF}_{zx}(t,\omega). \tag{27}$$

4) Obtaining a new TF representation by taking zero-values in the region around the determined maximum $[\hat{\omega}_x^1(t) - \delta, \hat{\omega}_x^1(t) + \delta]$. This TF representation will be denoted by $\mathrm{TF}_{zx}^{(1)}(t,\omega)$. It will be used for the IF estimation of the next component:

$$\hat{\omega}_x^2(t) = \arg\max_{\omega} \mathrm{TF}_{zx}^{(1)}(t,\omega). \tag{28}$$

Step 4 is repeated for each moving object $m = 3, ..., M$.

5) The velocities based on the IFs are given by:

$$\hat{v}_x^m(t) = \hat{\omega}_x^m(t)/\mu, \ m = 1, 2, ..., M. \quad (29)$$

6) Accelerations are estimated as:

$$\hat{a}_x^m = \frac{1}{t - t_0}[\hat{v}_x^m(t) - \hat{v}_x^m(t_0)], m = 1, 2, ..., M. \quad (30)$$

The number of moving objects can be estimated based on the comparison of the TF representation $\text{TF}_{zx}^{(m)}(t, \omega)$ with the assumed threshold value. If it is less than a threshold, one may assume that there are no more moving objects in the sequence, and that the procedure can be stopped. An approach for the threshold value determination is presented in [33].

IV. TF REPRESENTATIONS

Since the TF distributions play a key role in our approach, a short review of some commonly used distributions will be done in this section. Comprehensive analysis of these and other important TF distributions can be found in [34]-[36].

A. Spectrogram

The simplest TF representation is the spectrogram (SPEC), defined as a squared modulus of the short-time FT (STFT)

$$\text{SPEC}_{zx}(t, \omega) = |\text{STFT}_{zx}(t, \omega)|^2 =$$

$$\left| \sum_\tau z_x(t + \tau)w(\tau)e^{-j\omega\tau} \right|^2 \quad (31)$$

where $w(\tau)$ is a lag window function. In the case of relatively slow varying velocity, as compared to the signal phase, the following approximation holds [35]:

$$\text{SPEC}_{zx}(t, \omega) \cong \sum_{m=1}^{M} \frac{2\pi A_m^2[v_x^m(t) + a_x^m/2]^2}{\mu a_x^m}$$

$$\times w^2\left(\frac{\omega - \mu a_x^m t - \mu v_{x0}^m}{\mu a_x^m}\right) = \sum_{m=1}^{M} \text{SPEC}_{zx}^m(t, \omega) \quad (32)$$

where:

$$\text{SPEC}_{zx}^m(t, \omega) = \frac{2\pi A_m^2[v_x^m(t) + a_x^m/2]^2}{\mu a_x^m}$$

$$\times w^2\left(\frac{\omega - \mu a_x^m t - \mu v_{x0}^m}{\mu a_x^m}\right). \quad (33)$$

From (32) and (33) one may observe that:

1) Taking the window $w(0) > w(t)$ for $\forall t \neq 0$, the maximum of $\text{SPEC}_{zx}^m(t, \omega)$ is positioned along the IF:

$$\omega = \mu(a_x^m t + v_{xo}^m). \quad (34)$$

2) For constant velocity ($a_x^m = 0$) the SPEC is ideally concentrated along the IF:

$$\text{SPEC}_{zx}^m(t, \omega) \sim \delta(\omega - \mu v_{x0}^m). \quad (35)$$

3) By increasing a_x^m, the resolution and amplitude of the TF components decrease. It may cause error in the IF estimation. This is the reason for introducing other highly concentrated TF distributions, like the Wigner distribution (WD).

B. Wigner Distribution

The WD is defined by:

$$\text{WD}_{zx}^m(t, \omega) = \sum_\tau w(\tau)w(-\tau)$$

$$\times z_x^m(t + \tau)z_x^{m*}(t - \tau)e^{-j2\omega\tau}. \quad (36)$$

In the case of the signal $z_x^m(t)$, (23), it is:

$$\text{WD}_{zx}^m(t, \omega) =$$

$$= \sum_\tau w(\tau)w(-\tau)A_m[v_x^m(t + \tau) + a_x^m/2]$$

$$\times \exp[j\mu(x_0^m + \Delta x_m + v_{x0}^m(t+\tau) + a_x^m(t+\tau)^2/2)]$$

$$\times A_m[v_x^m(t - \tau) + a_m/2]\exp[-j\mu(x_0^m + \Delta x_m$$

$$+ v_{x0}^m(t - \tau) + a_x^m(t - \tau)^2/2)]e^{-j2\omega\tau}. \quad (37)$$

For $v_x^m(t+\tau) \cong v_x^m(t-\tau) \cong v_x^m(t)$, and $a_m \ll v_x^m(t)$, follows:

$$\text{WD}_{zx}^m(t, \omega) \cong A_m^2[v_x^m(t)]^2 \sum_\tau w(\tau)w(-\tau)$$

$$\times \exp[j\mu 2v_{x0}^m\tau + j\mu 2a_x^m t\tau]e^{-j2\omega\tau} =$$

$$= \frac{1}{2} A_m^2 [v_x^m(t)]^2 W(\omega - \mu(a_x^m t + v_{x0}^m)) \quad (38)$$

with $W(\omega) = \text{FT}\{w(\tau/2)w(-\tau/2)\}$. For a wide window width, the following holds:

$$\text{WD}_{zx}^m(t,\omega) \cong \pi A_m^2 [v_x^m(t)]^2 \delta(\omega - \mu(a_x^m t + v_{x0}^m)). \quad (39)$$

Note that, for a linear FM signal, the WD is ideally concentrated along the IF (39). In the considered application this is a very favorable property of the WD.

Unfortunately, the WD is nonlinear, i.e., for multicomponent signal $z_x(t) = \sum_{m=1}^{M} z_x^m(t)$ it exhibits very emphatic cross-terms:

$$\text{WD}_{zx}(t,\omega) = \sum_{m=1}^{M} \text{WD}_{zx}^m(t,\omega) +$$

$$+2\,\text{Re}\left\{ \sum_{\substack{m,n=1 \\ m \neq n}}^{M} \text{WD}_{zx}^{mn}(t,\omega) \right\}, \quad (40)$$

where $\text{WD}_{zx}^{mn}(t,\omega)$, for $m \neq n$, is the cross-term:

$$\text{WD}_{zx}^{mn}(t,\omega) = \sum_{\tau} w(\tau)w(-\tau)$$

$$\times z_x^m(t+\tau) z_x^{n*}(t-\tau) e^{-j2\omega\tau}. \quad (41)$$

This is a serious drawback of the WD since the cross-terms can make the IF estimation impossible. This is the reason for introducing the reduced interference distributions [37], [38]. All of them reduce cross-terms, and at the same time decrease the auto-terms concentration [35]. The distribution, called the S-method (SM) [38], will be used here for multicomponent signals. The SM produces the auto-terms concentration close to those in the WD, with significantly suppressed cross-terms.

C. S-method

The SM can be realized in a very simple way by using the STFT

$$\text{SM}_{zx}(t,\omega) = \sum_{\theta} P(\theta)$$

$$\times \text{STFT}_{zx}(t,\omega+\theta)\text{STFT}_{zx}^*(t,\omega-\theta). \quad (42)$$

Two limit cases of the SM are the SPEC for $P(\theta) = \delta(\theta)$, and the WD for $P(\theta) = 1$. Taking an appropriate frequency window $P(\theta)$

$$P(\theta) = \begin{cases} 1 & |\theta| \leq L\Delta\omega \\ 0 & |\theta| > L\Delta\omega, \end{cases} \quad (43)$$

where $\Delta\omega$ is the frequency resolution of the STFT, it is possible to achieve the concentration of auto-terms as in the WD, and reduce cross-terms as in the SPEC. More details about the SM, frequency window width and reduction of the cross-terms can be found in [38]-[40]. Note that significant improvement in the concentration can be achieved by using relatively narrow frequency window $P(\theta)$, (for example $L = 1, 2, 3$ is appropriate in numerous applications).

D. Modification of TF Representations

It can be seen that the signal $z_x(t)$ has amplitude proportional to the object velocity (23). Thus, the amplitude of the TF distributions is proportional to the squared object velocity (33), (39). By using the modified TF representation

$$\text{TF}'(t,\omega) = \text{TF}(t,\omega)/f(\omega), \quad (44)$$

better IF estimation can be achieved. In the examples we used the function $f(\omega)$ of the form

$$f(\omega) = |\omega|^2 + (k\Delta\omega)^2. \quad (45)$$

Example 2: Here, we consider a moving object with the initial position $(x_0, y_0) = (106, 12)$, initial velocity $(v_{x0}, v_{y0}) = (0.5, 0.5)$, and acceleration $(a_x, a_y) = (0.02, 0.01)$ in a video sequence of 100 frames. The initial frame is shown in Fig. 2a, while the last frame is shown in Fig. 2b. The projection functions $P_x(x,t)$ and $P_y(y,t)$ are shown in Fig. 2c and 2d, respectively, while functions $\partial P_x(x,t)/\partial t$ and $\partial P_y(y,t)/\partial t$ are given in Fig. 2e and 2f. The FTs of signals $z_x(t)$ and $z_y(t)$ are shown in Fig. 3a and 3b. They provide only the information about the spectral content, i.e., the "average" velocity in a sequence. Estimation of the varying velocity can be performed by using the TF distributions: the SPECs (Fig. 3c and 3d), and the WDs (Fig. 3e and 3f). The velocity estimations,
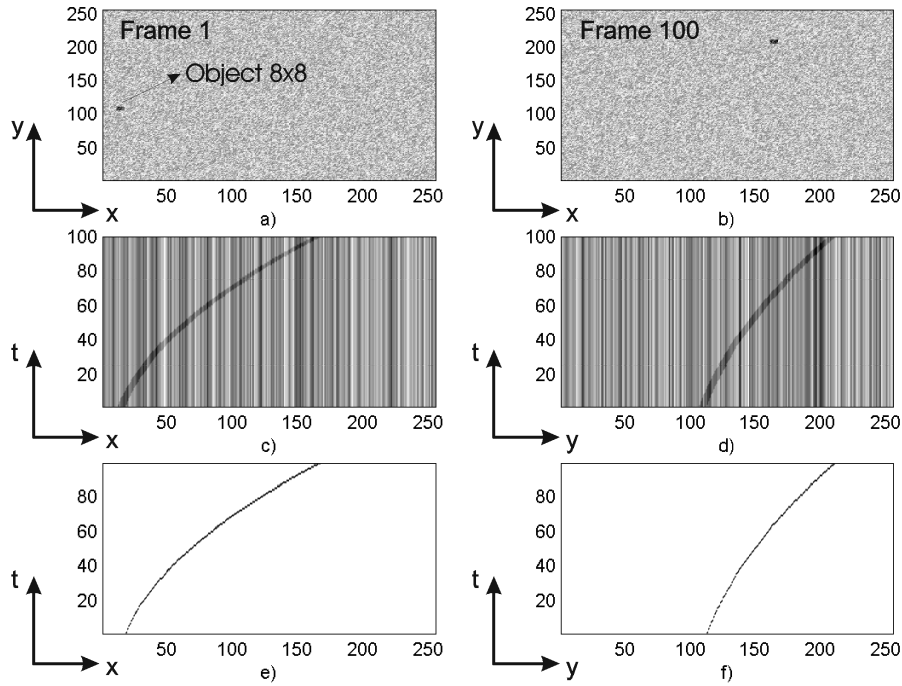
Fig. 2.   Video sequence with single moving object: a) Frame 1; b) Frame 100; c) $P_x(x,t)$; d) $P_y(y,t)$; e) $\partial P_x(x,t)/\partial t$; and f) $\partial P_y(y,t)/\partial t$.

performed by using maxima of the SPECs and the WDs, are shown in Fig. 4. The true velocities are marked by dashed lines while their estimates are depicted with thick lines. The mean squared errors (MSE) are MSE_x=0.01001 and MSE_y=0.01318 for the SPECs, and MSE_x=0.00164 and MSE_y=0.00317 for the WDs.

Example 3: In this example, a sequence with three moving objects is considered. The initial positions of the objects are $(x_0^1, y_0^1) = (106, 12)$, $(x_0^2, y_0^2) = (8, 12)$, and $(x_0^3, y_0^3) = (198, 250)$, while the initial velocities are: $(v_{x0}^1, v_{y0}^1) = (0.5, 0.5)$, $(v_{x0}^2, v_{y0}^2) = (2.8, 4)$, and $(v_{x0}^3, v_{y0}^3) = (-0.93, -0.2)$. Accelerations of the objects are $(a_x^1, a_y^1) = (0.02, 0.01)$, $(a_x^2, a_y^2) = (-0.05, -0.08)$, and $(a_x^3, a_y^3) = (-0.02, -0.045)$. The initial and the 100th frame are shown in Fig. 5a and 5b, respectively. The projections $P_x(x,t)$ and $\partial P_x(x,t)/\partial t$ are shown in Fig. 5c and 5d. The SPEC, the SM for $L = 2$, and the WD are shown in Fig. 6. The WD exhibits significant

amount of cross-terms (Fig. 6c). The velocity estimates are shown in Fig. 7. Fig. 7a and 7b represent estimates obtained by using the SPEC, while the estimation performed by using the SM with $L = 2$ is shown in Fig. 7c and 7d. Procedure (26)-(30) is applied for the velocity estimation. The improvement of estimation accuracy is obtained by using the modified form of the TF distributions (44), Figs. 7b,d. It can be seen that the SPEC forms, Fig. 7a and 7b, cannot track fast velocity variations, due to the small TF resolution. That is the reason why, in the short intervals, the spectrogram based velocity estimation recognizes constant values. The SM overcomes this drawback since it is highly concentrated along the IF, (Fig. 7c and 7d).

Example 4: A real video-sequence is considered in this example. It lasts 12.4s and it contains 187 video-frames (15 frames/s) of the size 320×240 with 256 gray levels. The sequence is recorded in the QVGA Motion-JPEG format. Scene contains the moving object (walker) and the background. Object po-
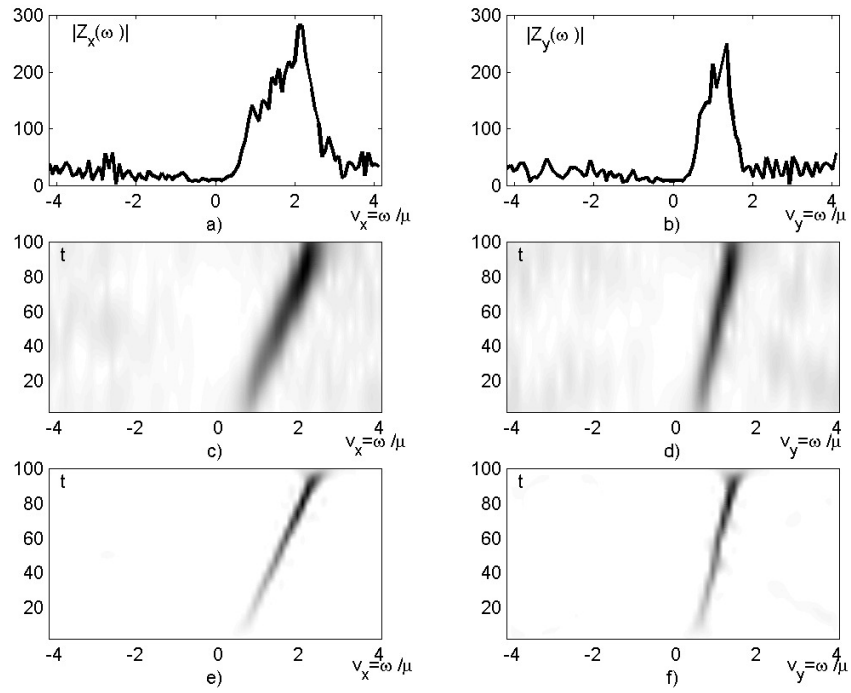
Fig. 3. FTs and TF representations of $z_x(t)$ and $z_y(t)$: a) $Z_x(\omega)$; b) $Z_y(\omega)$; c) SPEC of signal $z_x(t)$; d) SPEC of signal $z_y(t)$; e) WD of signal $z_x(t)$; and f) WD of signal $z_y(t)$.
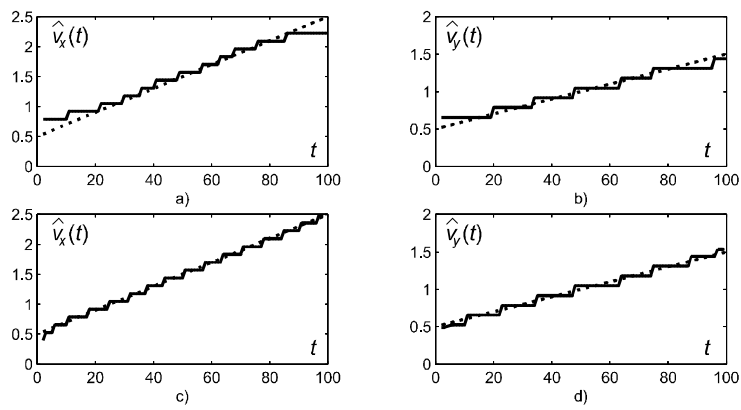

Fig. 4. Velocity estimation based on: a) $\mathrm{SPEC}_{zx}(t,\omega)$; b) $\mathrm{SPEC}_{zy}(t,\omega)$; c) $\mathrm{WD}_{zx}(t,\omega)$; and d) $\mathrm{WD}_{zy}(t,\omega)$. True value - dashed line; Velocity estimate - thick line.
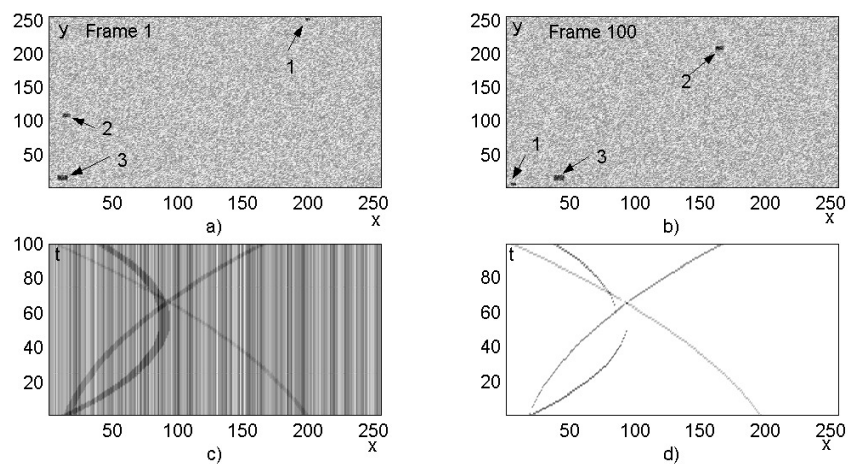
Fig. 5. Video-sequence with three moving objects: a) Frame 1; b) Frame 100; c) $P_x(x,t)$; and d) $\partial P_x(x,t)/\partial t$.
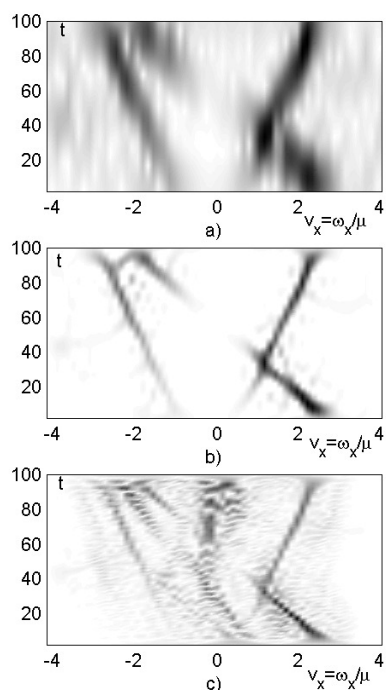

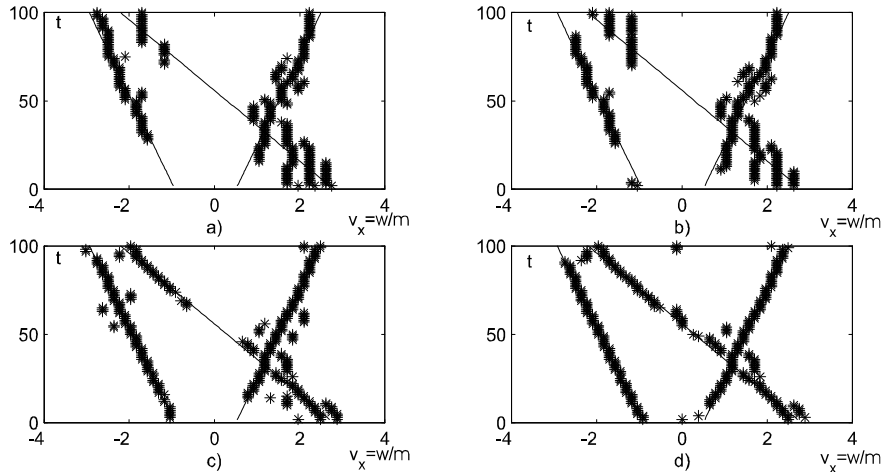Fig. 6. TF representations for video-sequence with three moving objects: a) SPEC; b) SM with $L = 2$; and c) WD.

Fig. 7.   Velocity estimation of three moving objects: a) SPEC; b) Modified SPEC; c) SM; and d) Modified SM.

sition in frames 1, 11, 21, ..., 181, is depicted in Fig. 8a[1]. It can be seen that the object slowly accelerates from frame 1 to 120, with rapidly increasing velocity around the frame 120. After that, velocity decreases around frame 150 (within one walker step), increases again until the final decrease in the frame 180. Since the object has moved along the x-coordinate, we made only velocity estimation of x-component. The synthetic image $\partial P_x(x,t)/\partial t$ is shown in Fig. 8b. From this figure we can observe the pattern that contains motion parameters. Relatively small patterns may be noticed around the basic one, that represent moving of the object's legs and hands. Also, noise-like patterns are observed from $\partial P_x(x,t)/\partial t$. Noise sources are: variations in the background, error caused by the compression method, and variations in the scene illumination. Error caused by variations in the scene illumination produces vertical lines in $\partial P_x(x,t)/\partial t$. This error appears several times in the first 50 frames and three such frames are denoted with arrows. The constant μ-propagation with $\mu = 0.4$ is applied to the function $\partial P_x(x,t)/\partial t$ from Fig. 8b. The spectrogram of the signal $z_x(t)$ is shown in Fig. 8c, while the SM is shown in Fig. 8d. Velocity estimation obtained by using these two TF representations is presented in Fig. 8e. Dot-

ted line represents results obtained by using the SPEC, while thick line depicts estimate obtained by using the SM. It can be easily seen that the SPEC is not precise velocity estimator, while the SM tracks accurately the object velocity within the entire interval. It is important to note that, in this case, an accurate velocity estimation would be very difficult by employing any specific parametric motion model.

V. Performance Analysis

The accuracy of velocity estimation depends on several factors, such as: discretization, stochastic error (caused by: approximations in the algorithm, background influence, variation of the illumination, etc.), and bias in the TF representations. An analysis of the mentioned errors is presented in this section.

A. Discretization Error and Displacement Technique

The maximum velocity value, that can be estimated by using the presented algorithm, depends on parameter μ. Since the sampling interval in the signal $z_x(t)$ is $\Delta t = 1$, then the maximum frequency in the TF representations is

$$\omega_{\max} = \frac{2\pi}{2\Delta t} = \pi. \qquad (46)$$

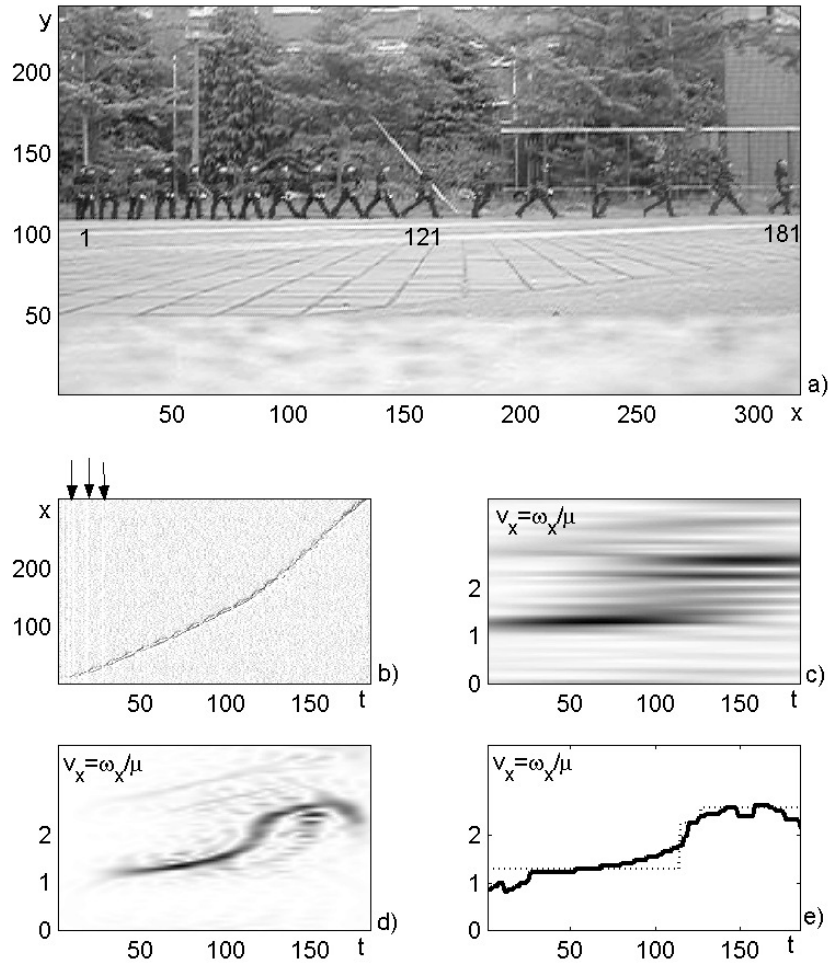[1] The video sequence can be found at: http:// www.tfsa.ac.me/dscn1442.mov

Fig. 8.   Velocity estimation in the real video-sequence: a) Object positions in frames 1, 11, 21, ..., 181; b) $\partial P_x(x,t)/\partial t$ (arrows depict frames with illuminance variation); c) $\mathrm{SPEC}_{zx}(t,\omega)$; d) $\mathrm{SM}_{zx}(t,\omega)$; and e) Velocity estimates: Dotted line - SPEC; Thick line - SM.

Thus, the velocity maximum which can be estimated is $v_{\max} = \omega_{\max}/\mu = \pi/\mu$. Larger velocity would cause the aliasing effect and the estimation could not be done accurately. For example, $\mu = 0.5$ produces $v_{\max} \approx 2\pi$ pixels/frame. This is a reasonable operating range for many practical applications. Resolution of the TF distributions is

$$\Delta v_x = \frac{2v_{\max}}{N_w} = \frac{2\pi}{\mu N_w} \qquad (47)$$

where N_w is the window width in the short-time FT (31). The MSE caused by a discrete nature of the TF distribution is

$$\mathrm{MSE}(\Delta v_x) = \frac{1}{12}\left[\frac{2\pi}{\mu N_w}\right]^2 . \qquad (48)$$

The uniform probability of error in the interval $[-\Delta v_x/2, \Delta v_x/2]$ is assumed. It can be seen that with a decrease of μ the maximum detectable velocity increases, while the accuracy of estimation decreases. The methods for reducing this error are based on zero-padding of signal $z_x(t)$ in the time-domain, or on displacement techniques. For this application we have found that the latter are more appropriate [12], [13]. They are based on interpolation

of the TF distributions around their maxima. Here, we have used the technique described in [12]. The velocity estimation can be obtained as

$$\hat{v}'_x(t) = \hat{v}_x(t) + \delta_x(t)\Delta v_x \qquad (49)$$

where $\hat{v}_x(t)$ is the estimate of velocity performed by using the maxima of the TF representation, while $\delta_x(t)$ is the displacement. For the TF distributions calculated by using the Hanning window, the displacement is given by [12]

$$\delta_x(t) = \frac{1.5[Q_{+1} - Q_{-1}]}{Q_{-1}[1 + Q_{+1}/Q_0] + Q_0 + Q_{+1}} \qquad (50)$$

where $Q_i = |STFT(t, \mu\hat{v}_x(t) + i\Delta\omega)|$, $i = -1, 0, 1$, for the estimation based on the STFT, and $Q_i = |WD(t, \mu\hat{v}_x(t) + i\Delta\omega)|$, $i = -1, 0, 1$, for the estimation based on the WD. If the SM is used, Q_i in (50) is

$$Q_i = |SM(t, \mu\hat{v}_x(t) + i\Delta\omega)|^\alpha, \; i = -1, 0, 1 \qquad (51)$$

with $1/2 \le \alpha \le 1$ (for $L = 0 \Longrightarrow \alpha = 1/2$, and for $L \ge N_w/2 \Longrightarrow \alpha = 1$).

Example 5: Consider a moving object with parameters as in Example 2. Estimation of the velocity component $v_x(t)$, based on the SM (with $L = 2$) maxima, is shown in Fig. 9a. The improvement of estimation precision obtained by using the displacement technique is shown in Fig. 9b. The corresponding MSEs are given in Fig. 9, as well. The MSE is decreased more than 5 times by using the displacement.

B. Stochastic error

Approximations in expressions (10), (14), (15) introduce an error in the algorithm. Additional errors are caused by the background variation, interactions between the objects and the background, camera motion, etc. These errors can be modeled by additive $\nu_a(t)$ and multiplicative $\nu_m(t)$ noises

$$z'_x(t) = z_x(t)(1 + \nu_m(t)) + \nu_a(t) = z_x(t) + \nu_z(t) \qquad (52)$$

where $\nu_z(t)$ represents the noise influenced term:

$$\nu_z(t) = z_x(t)\nu_m(t) + \nu_a(t). \qquad (53)$$

An asymptotic expression for the variance of the WD based IF estimator is derived in [30]. The velocity estimator variance, under assumption that $\nu_z(t)$ is a white Gaussian noise with the variance σ_z^2, is given as

$$\sigma^2(t, N_w) = \text{var}\{\Delta\hat{v}(t)\} =$$

$$\frac{6\sigma_z^2}{\mu^2|A_z(t)|^2}\left(1 + \frac{\sigma_z^2}{2|A_z(t)|^2}\right)\frac{1}{N_w^3} \qquad (54)$$

where $A_z(t)$ is the amplitude of signal $z_x(t)$, (11):

$$A_z(t) = A\left(\sum_{k=1}^{\infty}\frac{1}{k!}\frac{d^k\varphi_x(t)}{dt^k}\right) \qquad (55)$$

while A is the amplitude of motion pattern in $\partial P_x(x,t)/\partial t$ (13). By neglecting the higher order derivatives in (55), the variance $\sigma^2(t, N_w)$ becomes

$$\sigma^2(t, N_w) = \frac{6\sigma_z^2}{\mu^2|Av_x(t)|^2} \times$$

$$\left(1 + \frac{\sigma_z^2}{2|Av_x(t)|^2}\right)\frac{1}{N_w^3}. \qquad (56)$$

Obviously, error caused by noise is an increasing function of the variance σ_z^2, while it is a decreasing function of A, $v_z(t)$, and window length N_w. Large error in the case of small velocity $|v_z(t)| \approx 0$ can be reduced by using modified TF representations (44). It will be illustrated on Example 6.

C. Bias

The bias in the IF estimation comes from the higher order derivatives in the signal's phase, i.e., when velocity can be represented with a higher order polynomial. If we take

$$z_x(t) \approx A_z(t)\exp(j\mu\varphi_x(t)) \qquad (57)$$

the bias of the WD based estimator is [30]

$$\text{bias}(\hat{v}_x(t)) = \sum_{s=1}^{\infty} a_s \frac{d^{2s+1}\varphi_x(t)}{dt^{2s+1}} N_w^{2s} \qquad (58)$$

where a_s are constants determined by the window shape. Neglecting higher order derivatives, the bias can be reduced to [30]

$$\text{bias}(\hat{v}_x(t)) \approx a_1 \frac{d^2 v_x(t)}{dt^2} N_w^2. \qquad (59)$$
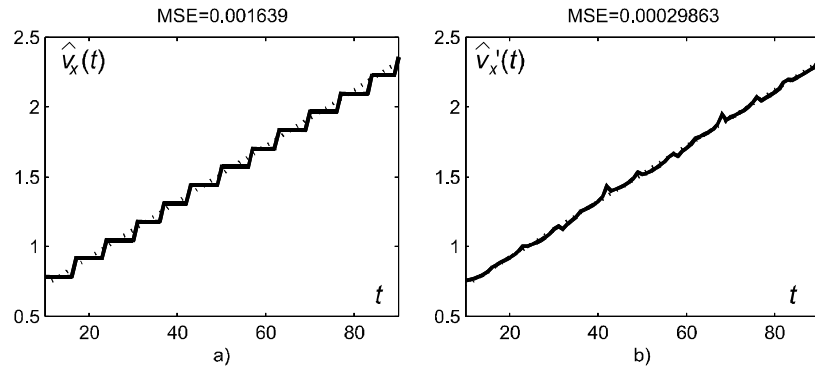
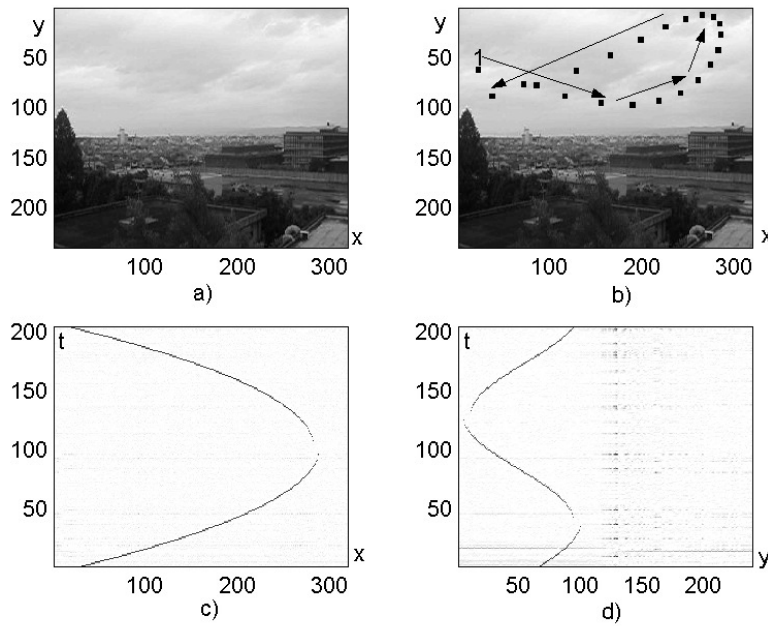Fig. 9. Velocity estimation by using: a) SM; and b) SM with displacement.

Fig. 10. Real video-sequence with artificial object: a) Scene without object: Initial frame; b) Scene with object: Object positions in frames 1, 11, 21, ..., 201, Arrows mark motion direction; c) $\tilde{P}_x(x,t)$; and d) $\tilde{P}_y(y,t)$.

Note that for the rectangular window holds $a_1 = 1/(40\mu)$, see [30, eq. (11)]. The bias is an increasing function with respect to the window width. The expressions for bias have similar forms for other TF representations [30].

D. Optimal Window Width

The estimation MSE can be written as a sum of squared bias and variance

$$\mathrm{MSE}(t, N_w) = \left(\frac{1}{40\mu} \frac{d^2 v_x(t)}{dt^2} N_w^2 \right)^2 +$$

$$\frac{6\sigma_z^2}{\mu^2 |Av_x(t)|^2} \left(1 + \frac{\sigma_z^2}{2|Av_x(t)|^2} \right) \frac{1}{N_w^3}. \quad (60)$$

Minimum of the MSE is achieved for

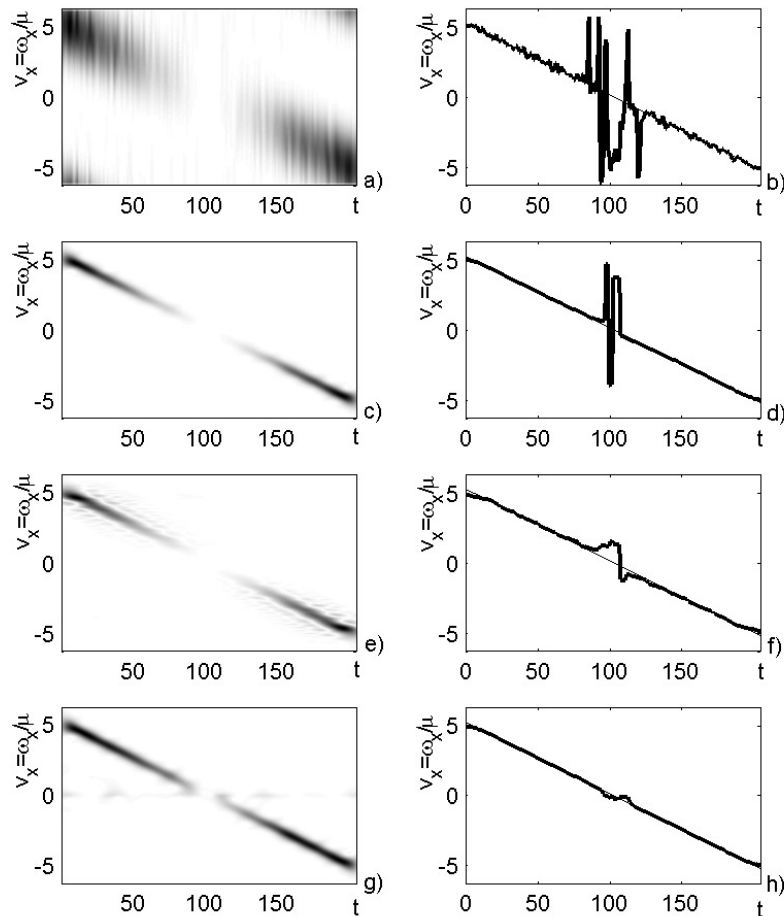$$\frac{\partial \mathrm{MSE}(t, N_w)}{\partial N_w} \Big|_{N_w = N_{opt}(t)} = 0. \quad (61)$$

Fig. 11.   TF representations and velocity estimates along the x-coordinate: Left column: SMs; Right column: $\hat{v}_x(t)$, Thin line: Exact velocity; Thick line: Estimated velocity. First row: $N_w = 8$; Second row: $N_w = 32$; Third row: $N_w = 128$; and Fourth row: Modified SMs $N_w = 32$ and $k = 5$.

Then the optimal window width is

$$N_{opt}(t) = \sqrt[7]{\frac{7200\sigma_z^2 \left(1 + \frac{2\sigma_z^2}{|Av_x(t)|^2}\right)}{|Av_x(t)|^2 \frac{d^2 v_x(t)}{dt^2}}}. \quad (62)$$

Obviously, this formula cannot be used for determination of the optimal window width since it contains unknown motion parameters (velocity and its derivative). Recently, an adaptive algorithm for determination of the suboptimal window width in the IF estimator, based on the TF representations, has been proposed [30]. Application of that adaptive algorithm to the velocity estimation will be the topic of our further research.

Example 6. A real video-sequence is considered. Its initial frame is shown in Fig. 10a[2]. Sequence has 205 frames, and it lasts 13.6sec. Frame size is 320×240. We transform each frame to a 256-level grayscale image and we calculated synthetic images $\tilde{P}_x(x,t)$ and $\tilde{P}_y(y,t)$. The mean value of $\tilde{P}_x(x,t)$ is 36.74, while for $\tilde{P}_y(y,t)$ it is 56.22. An artificial moving object is added to this sequence. It is a black square of size 5×5 pixels, whose motion can be described as

$$\varphi_x(t) = 14.81 + 5.24t - 0.025t^2,$$

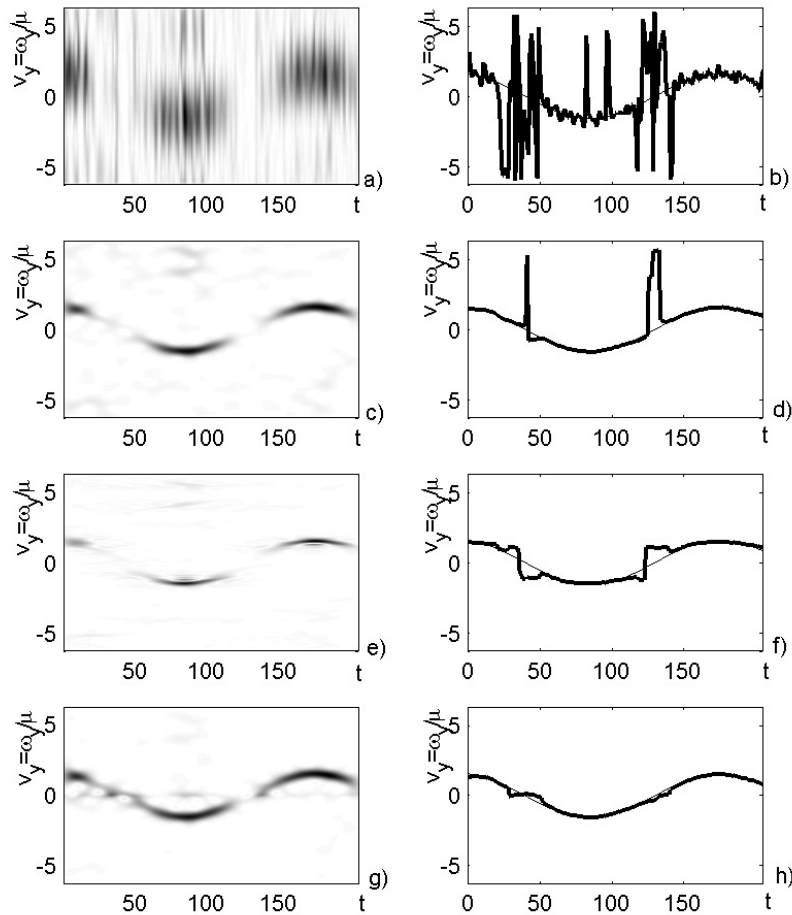[2]The video sequence can be found at: http://www.tfsa.ac.me/dscn1632.mov

Fig. 12. TF representations and velocity estimates along the y-coordinate: Left column: SMs; Right column: $\hat{v}_y(t)$, Thin line: Exact velocity; Thick line: Estimated velocity. First row: $N_w = 8$; Second row: $N_w = 32$; Third row: $N_w = 128$; and Fourth row: Modified SMs $N_w = 32$ and $k = 5$.

$$\varphi_y(t) = 50 + 45\sin(0.035t + 0.185). \qquad (63)$$

We presented positions of the object in frames 1, 11, 21, ..., 201, with arrows denoting motion direction in Fig. 10b. Projections $\tilde{P}_x(x,t)$ and $\tilde{P}_y(y,t)$ (13) are shown in Fig. 10c and 10d. Motion patterns can be observed from these figures. The constant μ-propagation is applied with $\mu = 0.5$. The SMs with $N_w = 8$, $N_w = 32$, and $N_w = 128$, and the modified SM (44) with $N_w = 32$ and $k = 5$, with corresponding velocity estimates, are shown in Figs. 11 and 12. The MSEs are: $\mathrm{MSE}_x = (2.1478, 0.7918, 0.1320, 0.0088)$ and $\mathrm{MSE}_y = (4.8197, 1.1286, 0.1875, 0.0235)$ for these four TF representations, respectively.

It can be seen that the modified SM behaves better than other TF representations, especially around $v_x(t) \approx 0$ and $v_y(t) \approx 0$.

Also, we added artificial noise to the sequence

$$i'(x,y,t) = i(x,y,t) + \sigma\nu(x,y,t), \qquad (64)$$

where $\nu(x,y,t)$ is the white Gaussian noise with unitary variance $E\{\nu(x', y', t')\nu(x'', y'', t'')\} = \delta(x' - x'', y' - y'', t' - t'')$. Two cases are considered: $\sigma = 5$ and $\sigma = 10$. The initial frames with object positions and with projections $\tilde{P}_x(x,t)$ and $\tilde{P}_y(y,t)$ are shown in Fig. 13. The modified SM with $N_w = 32$ and $k = 5$ is applied in these cases, Fig. 14. The MSEs

obtained by the Monte Carlo simulation in 20 trials are: $MSE_x = (0.0119, 0.4741)$ and $MSE_y = (0.0384, 0.1955)$, for these two noise amounts, respectively. In our simulations the larger amount of noise than $\sigma = 12$ produced unreliable estimates. This error can be reduced by using spatiotemporal filters or more sophisticated differentiation function than (6). More details on this topic can be found in [26].

VI. CONCLUSION

TF analysis tools are used for estimation of time-varying velocity of moving objects. The constant μ-propagation is used to transform motion parameters to an FM signal. The object velocity is estimated based on the IF. Different TF representations, employed as the IF estimators, were analyzed and compared. In the case of a single moving object, the WD is a very appropriate tool. The SM is used for velocity estimation of several moving objects. The displacement technique and the modified TF representations are proposed for an improvement of the estimation accuracy.

VII. ACKNOWLEDGMENT

REFERENCES

[1] Special issue on TF analysis, *Proceedings of IEEE*, Vol. 84, No.9, Sept. 1996.

[2] B. Zhang and S. Sato: "A TF distribution of Cohen's class with a compound kernel and its application to speech signal processing," *IEEE Trans. Sig. Proc.*, Vol. 42, No.1, Jan. 1994, pp. 54-65.

[3] G. C. Gaunaurd and H. C. Strifors: "Signal analysis by means of TF (Wigner-type) distributions-Application to sonar and radar echoes," *Proc. IEEE*, Vol. 84, No.98, Sept. 1996, pp. 1231-1248.

[4] N. H. Morgan and A. S. Gevins: "WD of human event-related brain potentials," *IEEE Trans. Bio. Eng.*, Vol. 33, No.1, Jan. 1986, pp. 66-70.

[5] J. Hormigo and G. Cristobal: "High resolution spectral analysis of images using the pseudo-WD," *IEEE Trans. Sig. Proc.*, Vol. 46, No.6, June 1998, pp. 1757-1763.

[6] S. Stanković, I. Djurović and I. Pitas: "Watermarking in the space/spatial-frequency domain using two-dimensional Radon-WD," *IEEE Trans. Im. Proc.*, Vol. 10, No.4, Apr. 2001, pp. 650-658.

[7] F. Cakrak and P. J. Loughlin: "Multiple window time-varying spectral analysis," *IEEE Trans. Sig. Proc.*, Vol. 49, No.2, Feb. 2001, pp. 448-453.

[8] H. K. Aghajan and T. Kailath: "SLIDE: Subspace-based line detection," *IEEE Trans. PAMI*, Vol. 16, No.11, Nov. 1994, pp. 1057-1073.

[9] H. K. Aghajan, B. H. Khalaj and T.Kailath: "Estimation of multiple 2-D uniform motions by SLIDE: Subspace-based line detection," *IEEE Trans. Im. Proc.*, Vol. 8, No.4, Apr. 1999, pp. 517-526.

[10] P. Stoica and R. L. Moses: *Introduction to spectral analysis*, Prentice Hall, 1997.

[11] S. M. Kay, S. Lawrence Marple, Jr.: "Spectrum analysis-A modern perspective," *Proc. IEEE*, Vol. 69, No.11, Nov. 1981, pp. 1380-1419.

[12] G. Andria, M. Savino and A. Trotta: "Windows and interpolation algorithms to improve electrical measurement accuracy," *IEEE Trans. Inst. Meas.*, Vol. 38, No.4, Aug. 1989, pp. 856-863.

[13] E. Dialverogu: "Nonmatrix Cramer-Rao bound expressions for high resolution frequency estimation," *IEEE Trans. Sig. Proc.*, Vol. 46, No.2, Feb. 1998, pp. 463-474.

[14] I. Djurović and LJ. Stanković: "A virtual instrument for TF signal analysis," *IEEE Trans. Inst. Meas.*, Vol. 48, No.6, Dec. 1998, pp. 1086-1092.

[15] T. L. Marzetta: "Fan filters, the 3-D radon transform, and image sequence analysis," *IEEE Trans. Im. Proc.*, Vol. 3, No. 5, May 1994, pp. 253-264.

[16] I. Reed, R. Gadliardi and L. Stotts: "Optical moving target detection with 3-D matched filtering," *IEEE Trans. Aer. El. Syst.*, Vol. 24, No.4, July 1988, pp. 327-336.

[17] J. R. Bergen, P. J. Burt, R. Hingorani and S. Peleg: "A three-frame algorithm for estimation two-component image motion," *IEEE Trans. PAMI*, Vol. 14, No.9, Sept. 1992, pp. 886-896.

[18] G. Adiv: "Inherent ambiguities in recovering 3-D motion and structure from a noisy flow field," *IEEE Trans. PAMI*, Vol. 11, No.5, May 1989, pp. 477-489.

[19] W. Chen, G.B. Giannakis and N. Nandhakamur: "Spatiotemporal approach for time-varying global image motion estimation," *IEEE Trans. Im. Proc.*, Vol. 5, No.10, Oct. 1996, pp. 1448-1461.

[20] B. Porat and, B. Friedlander: "A frequency domain algorithm for multiframe detection and estimation of dim targets," *IEEE Trans. PAMI*, Vol. 12, No.4, Apr. 1990, pp. 398-401.

[21] P. Milanfar: "A model of the effects of the image motion in the Radon transformation domain," *IEEE Trans. Im. Proc.*, Vol. 8, No.3, Mar. 1999, pp. 438-444.

[22] C. E. Erdem, G. Z. Karabulut, E. Yanmaz and E. Anarim: "Motion estimation in the frequency domain using fuzzy c-planes clustering," *IEEE Trans. Im. Proc.*, Vol. 10, No.12, Dec. 2001, pp. 1873-1879.

[23] P. Milanfar: "Two-dimensional matched filtering for motion estimation," *IEEE Trans. Im. Proc.*, Vol. 8, No.9, Sept. 1999, pp. 1276-1281.

[24] D. Vernon: *Machine Vision*, Prentice Hall, 1991.
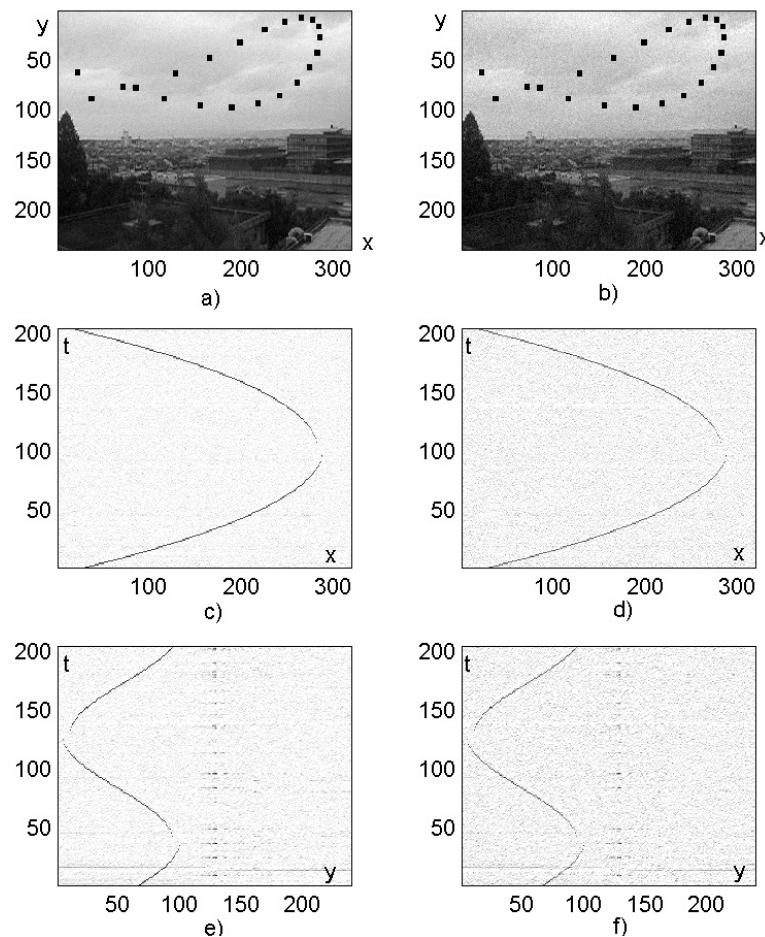
[25] D. Vernon and M. Tistarelli: "Using camera mo-

Fig. 13. Noisy versions of the sequence: Left column: $\sigma = 5$; Right column: $\sigma = 10$; First row: Object positions in frames 1, 11, 21, ..., 201; Second row: $\tilde{P}_x(x,t)$; and Third row: $\tilde{P}_y(y,t)$.

tion to estimate range for robotic parts manipulation", *IEEE Trans. Rob. Aut.*, Vol. 6, No.5, Oct. 1990, pp. 509-521.

[26] J. L. Barron, D. J. Fleet and S. Beauchemin: "Performance of optical flow techniques," *Int. Jour. Comp. Vis.*, Vol. 12, No.1, 1994, pp. 43-77.

[27] S. Stanković and I. Djurović: "Motion parameters estimation by using TF representations," *El. Let.*, Vol.37, No.24, Nov. 2001, pp. 1446-1448.

[28] B. Boashash: "Estimating and interpreting the IF of a signal - Part I: Fundamentals," *Proc. IEEE*, Vol. 80, No.4, Apr. 1992, pp. 520-538.

[29] P. Rao and F.J. Taylor: "Estimation of the IF using discrete WD," *El. Let.*, Vol. 26, 1990, pp. 246-248.

[30] LJ. Stanković and V. Katkovnik: "IF estimation using higher order distributions with adaptive order and window length," *IEEE Trans. Inf. Th.*, Vol. 46, No.1, Jan. 2000, pp. 302-311.

[31] B. Barkat, B. Boashash and LJ. Stanković: "Adaptive window in the polynomial Wigner-

Ville distributions for IF estimation of FM signal in additive Gaussian noise," in *Proc. IEEE ICASSP'1999*, pp. 1317-1320.

[32] S. Barbarossa and O. Lemoine: "Analysis of non-linear FM signals by pattern recognition of their TF representations," *IEEE Sig. Proc. Let.*, Vol. 3, No.4, Apr. 1996, pp. 112-115.

[33] A. Francos and M. Porat: "Analysis and synthesis of multicomponent signals using positive TF distributions," *IEEE Trans. Sig. Proc.*, Vol. 47, No.2, Feb. 1999, pp. 493-504.

[34] L. Cohen: *TF analysis*, Prentice-Hall, 1995.

[35] LJ. Stanković: "The auto-term representation by the reduced interference distributions; The procedure for a kernel design," *IEEE Trans. Sig. Proc.*, Vol. 44, No.6, June 1996, pp. 1557-1564.

[36] F. Hlawatsch and G.F. Boudreaux-Bartels: "Linear and quadratic TF signal representations," *IEEE Sig. Proc. Mag.*, Apr. 1992, pp. 21-67.

[37] J. Jeong and W.J. Williams: "Kernel design for reduced interference distributions," *IEEE Trans.*
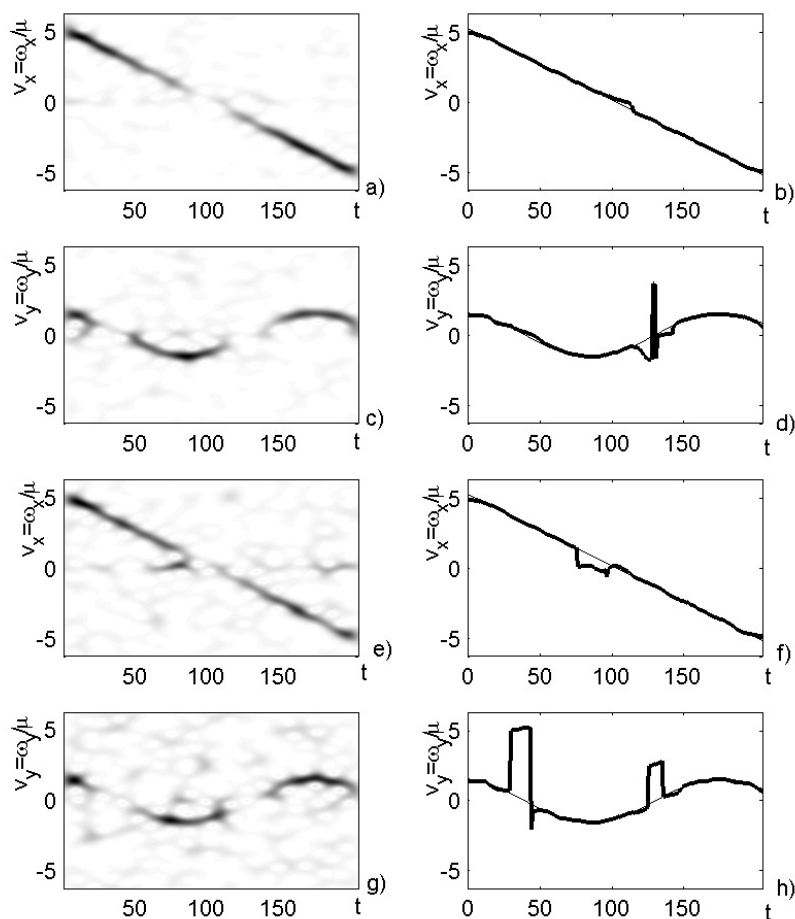
Fig. 14.   TF representations and velocity estimates for noisy sequences: Left column: Modified SMs with $N_w = 32$ and $k = 5$; Right column: Velocity estimates, Thin line: Exact velocity; Thick line: Estimated velocity; First row: $\sigma = 5$, x-component; Second row: $\sigma = 5$, y-component; Third row: $\sigma = 10$, x-component; and Fourth row: $\sigma = 10$, y-component.

Sig. Proc., Vol. 40, No.2, Feb. 1992, pp. 402-412.

[38] LJ. Stanković: "A method for TF signal analysis," *IEEE Trans. Sig. Proc.*, Vol. 42, No.1, Jan. 1994, pp. 225-229.

[39] LJ. Stanković: "Method for improved energy concentration in the TF analysis of multicomponent signals using L-WD," *IEEE Trans. Sig. Proc.*, Vol. 43, No.5, May 1995, pp. 1262-1269.

[40] L. L. Scharf and B. Friedlander: "Toeplitz and Hankel kernels for estimating time-varying spectra of discrete-time random process," *IEEE Trans. Sig. Proc.*, Vol. 49, No.1, Jan. 2001, pp. 179-189.