Compressive sensing in Video Applications

Irena Orović, Seri Park, and Srdjan Stanković

Abstract **— In this paper, a new approach to estimate motion parameters in compressive sensed video sequences is proposed. The proposed procedure combines sparse reconstruction algorithms and time-frequency analysis applied to μ-propagation signal. This concept allows providing precise velocity estimation even under a reduced number of randomly chosen video frames. The theory is applied and illustrated on synthetic and real video sequence.**

Keywords **— Compressive sensing, video signals, motion parameters estimation, velocity, random frames recording, reconstruction algorithms**

I. Introduction

Time-frequency analysis has been widely used in the applications dealing with non-stationary signals characterized by time-varying spectral content [1]-[4]. These applications include radars, sonars, communications, biomedical and multimedia systems [5]-[10]. In the case of 3-dimension (3D) video signals, the time-frequency analysis is usually considered as a powerful tool for object tracking, parameters estimation, optical flow estimation, security issues, surveillance, etc. For instance, time-frequency distributions were combined with the SLIDE (subspace-based line detection) algorithm, [11],[12], to provide a high-precision method for estimation of moving objects velocities in video sequences [13]-[15]. The video frames are firstly projected onto the coordinate axes, and the projections are further used to produce the frequency modulated (FM) signals. The motion parameters can be obtained by estimating time-frequency parameters of these FM signals [13]. In order to provide efficient estimation results, we need to consider a suitable time-frequency distribution that provides high concentration without the cross terms. For that purpose, the S-method provides most of the desirable properties and improves the performance over the spectrogram, without increasing significantly the realization complexity [16]. In this paper we consider the case when we are left with fewer frames than necessary for precise motion parameters estimation. This situation may be caused by discarding some distorted frames, or may be a consequence of compressive sensing/recording with the aim to reduce storage and transmission requirements. Moreover, considering the weather factors (i.e., blocking camera lens due to heavy rain) as well as video camera maintenance aspects, one cannot always assume a continuous video data streaming – thus data loss. Missing frames will produce an incomplete projection vector, which will further affect the time-frequency representation by introducing certain kind of noise [17],[18]. Consequently, it will produce large errors during the velocity estimation. In order to provide the velocity estimation as in the case of full data set, we need to recover the compressive sensed data using reconstruction algorithms. Namely, under the certain conditions, the signals can be reconstructed from a small number of random measurements [19]-[23], whereas the signal must fulfill certain conditions, such as sparsity. The information about sparse signal is contained in the significantly smaller number of coefficients, compared to the total length of the signal. When searching for the best sparse approximation, we will use the l_1 minimization approach which is solved using convex optimization algorithms [20],[21].

The paper is organized as follows. The motion parameters estimation based on the time-frequency analysis is presented in Section II. The concept of velocity estimation based on the Compressive sensed video sequences is proposed in Section III. The experimental results are given in Section IV.

II. Motion parameters Estimation in Video Sequences

High precision motion estimation in video-sequences has been usually done using the techniques based on the spectral analysis methods. Especially in traffic control and safety fields, many studies have been dedicated for an accurate vehicle trajectory (vehicle motion) generation to derive more precise vehicle speeds using video image processing techniques [24]-[28]. However, due to the camera calibration and data loss issues, speed estimation using video image is an on-going research topic. In the case of time-varying velocities, the time-frequency analysis has been combined with other state of the art methods for motion parameters estimation, such as SLIDE algorithm [11],[12]. Namely, the velocity estimation problem can be solved using the SLIDE and μ-propagation (constant or variable). The μ-propagation approach maps the sequence of video frames into FM signals (for constant velocity case) or signals with highly nonlinear phase (time-varying velocity).

A certain video frame that appears at time instant t and contains a moving object can be represented as:

$$I(x, y, t) = b(x, y) + s(x - x_0 - v_x t, y - y_0 - v_y t) \quad (1)$$

where $s(x,y)$ represents the moving object, B is background, while t is the considered frame. The initial object position is (x_0, y_0) and velocity (v_x, v_y). Further, we are observing the

frame projections onto the axes (without loss of generality we can focus on x axis, since the same holds for y axis). Hence, we can write:

$$P(x,t) = \sum_y I(x,y) = \sum_y b(x,y) +$$

$$+ \sum_y s(x-x_0-v_xt, y-y_0-v_yt) = B(x) + S(x-x_0-v_xt) \quad (2)$$

Assuming that the background is constant, its influence is eliminated by calculating the derivative of projection with respect to t:

$$\frac{\partial P(x,t)}{\partial t} = v_x \frac{\partial S(x-x_0-v_xt)}{\partial x} =$$

$$= \hat{P}(x-x_0-v_xt) \approx P(x,t-1) - P(x,t) \quad (3)$$

In order to provide velocity estimation, we define the signal in the form:

$$z(t) = \sum_x \hat{P}(x-x_0-v_xt)e^{j\mu x}, \quad (4)$$

whose instantaneous frequency corresponds to moving object velocity. In order to estimate the instantaneous frequency, the time-frequency analysis should be applied to $z(t)$. Particularly, it has been shown that the Wigner distribution (WD) can provide efficient results. It is a quadratic time-frequency representation that is given by:

$$WD(t,\omega) = \sum_\tau z(t+\tau/2)z(t-\tau/2)e^{-j\omega\tau}. \quad (5)$$

Therefore, the problem is recast as the instantaneous frequency estimation in the time-frequency domain. In practical applications, it is more suitable to use the S-method instead of the Wigner distribution, since it does not produce the cross-terms in the case of multicomponent signals and it is more suitable in noisy case (which is often encountered when dealing with real video sequences). The S-method is defined as follows:

$$SM(t,\omega) = \sum_{i=-L}^{L} STFT(t,\omega+i\theta)STFT^*(t,\omega-i\theta), \quad (6)$$

where $2L+1$ is the frequency window width, (*) is the complex conjugate, while $STFT(t,\omega)$ is the short-time Fourier transform:

$$STFT(t,\omega) = \sum_\tau w(\tau)z(t+\tau)e^{-j\omega\tau}, \quad (7)$$

with $w(t)$ being the window function.

III. VELOCITY ESTIMATION BASED ON COMPRESSIVE SENSED VIDEO SEQUENCES

In most applications, signal acquisition at high sampling rates requires large data storage and transmission capacities. Therefore, it would be very feasible if we can sample at lower rates and reconstruct the signal later for the analysis. It can be especially interesting and useful in different video applications, such as surveillance, where it is needed to store and transmit large number of frames during a long time period. Moreover, considering significant amount of traffic surveillance cameras on the roadway, the proposed method will promote "smart" use of current infrastructure. The concept of compressed sampling/sensing lies in the mathematical foundation that it is possible to reconstruct a sparse (or almost sparse) signal from a small set of randomly chosen samples using the powerful convex optimization algorithms (computationally efficient convex programming [21]). In the sequel we consider the compressive sensing/recording of video sequences. Note that the compressive sensing can be efficiently applied to each video frame in order to reduce significantly the number of acquired pixels (and consequently the number of frames). Since each frame is processed independently, this belongs to the compressive image sampling and reconstruction. Instead, we consider the possibility to acquire just a small random set of frames in time and to assure motion parameters estimation from that incomplete set of frame.

Consider the subset of frames:

$$J(x,y,T) \subset I(x,y,t) \quad (8)$$

for a set of random time instants $T = \{T_1, T_2, ..., T_M\}$, while $card\{t\} = N$, $N > M$. This further means that the projection vector, i.e. μ-propagation vector, contains a small incomplete set of samples. Hence, instead of the whole signal $z(t)$, we actually have a small set of M measurements $z(T)$. Since, we need to calculate the STFT for velocity estimation, for each windowed signal part i.e. for each instant T_i, we might actually observe the measurement vector in the form:

$$\mathbf{y}(T_i) = w(\tau)z(T_i+\tau), \quad for \quad \forall T_i \in T. \quad (9)$$

The Fourier transform of $\mathbf{y}(T_i)$ will result in low quality STFT, which is not suitable for analysis anymore. Therefore, it is necessary to use the compressive sensing reconstruction algorithms to recover the missing samples starting from the available measurements. For the sake of simplicity, we will omit the notation T_i in the sequel. Nevertheless, the procedure should be identically repeated for each available time instant. By using the compressive sensing notations, we may write [19],[20]:

$$\mathbf{y} = \mathbf{\Phi x}, \quad (10)$$

where $\mathbf{x}=w(\tau)z(t+\tau)$ represents the original (in our case the desirable) windowed μ-propagation vector, while $\mathbf{\Phi}$ describes the random measurement matrix. Furthermore, the signal $\mathbf{x} \in \mathbb{C}^N$ can be represented in Fourier basis $\{\mathbf{\Psi}\}_{k=1}^N = \{e^{j\omega_k\tau}\}_{k=1}^N$, using the weighting coefficients S_k:

$$\mathbf{x} = \sum_{k=1}^N S_k \mathbf{\Psi}_k. \quad (11)$$

The previous relation represents the inversion of (7), where S_k corresponds to the STFT coefficients for certain $T_i \in T$, $i = 1,...,M$. In the vector form, it can be written as:

$$\mathbf{x} = \mathbf{\Psi S}, \quad (12)$$

where $\mathbf{\Psi}$ is a full rank $N \times N$ matrix. From (10) and (12) we can rewrite:

$$\mathbf{y} = \mathbf{\Phi\Psi S} = \mathbf{AS}. \quad (13)$$

The aim is to reconstruct $\mathbf{x}$ or equivalently its spectral representation $\mathbf{S}$ from the incomplete set of measurement $\mathbf{y}$.

For that purpose, we need to solve the underdetermined system of M linear equations with N unknowns. Since, this system may have infinitely many solutions, in compressive sensing applications we are interested in the sparsest one. In that sense, the optimization algorithms based on ℓ_0 - norm minimization should be employed. In practical applications, it is replaced by ℓ_1 - norm, leading to a near-optimal solutions [20]:

$$\min \left\| \tilde{\mathbf{S}} \right\|_{\ell_1} \quad s.t. \quad \mathbf{y} = \mathbf{A}\tilde{\mathbf{S}} . \qquad (14)$$

The above minimization can be solved by using convex optimization algorithms. As a solution, for each considered time instant, we obtain the reconstructed STFT of windowed signal part. The resulting compressive sensing based STFT is used to calculate the S-method according to (6) and then to estimate the instantaneous frequency, i.e. the object velocity.

IV. EXAMPLES

Example 1: Let us observe the simulated video sequence with 100 frames, where the object is moving through the frames sequence having noisy background, Fig. 1. The frame size is 256×256 pixels, while the object size is 8×8. The initial object position is $(x_0, y_0) = (15,9)$. For the first 50 frames velocity is $(v_x, v_y) = (2.3, 2.2)$, while in the next 50 frames $(v_x, v_y) = (1.5, 1.2)$. Assume that we only have 40% of frames, while 60% of frames are missing (due to the compressive sensing/recording). We determine the variable μ-propagation vector, which in the case of compressive sensing (CS) will have only 40% of samples.
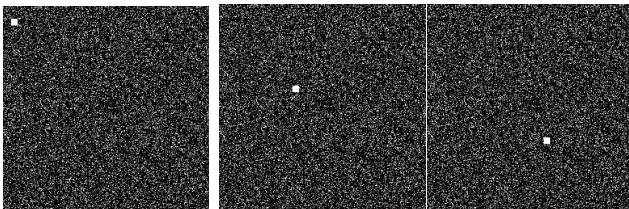

Fig. 1. Synthetic video sequence frames: 5, 35, 80

The S-method is calculated in two ways:

a) Direct calculation using available samples, which is usually known as initial form (Initial S-method) and it is shown in Fig. 2.a.

b) Calculation by applying Compressive sensing based reconstruction in the STFT domain. The corresponding result is Compressive sensing based S-method (CS based S-method), Fig. 2.b.

Based on the Initial S-method and CS based S-method the instantaneous frequency is estimated using *argmax*. The instantaneous frequency estimation is shown in Fig. 3. As shown in Fig. 3, it is clear that the estimation based on the initial S-method calculated from the incomplete set of samples may produce serious errors, while the CS based representation produce precise results. Namely, the MSE between the estimation result obtained from original full data S-method and Initial S-method is 38 dB, while the MSE

between the estimations based on full data S-method and CS based S-method is 0.12 dB.
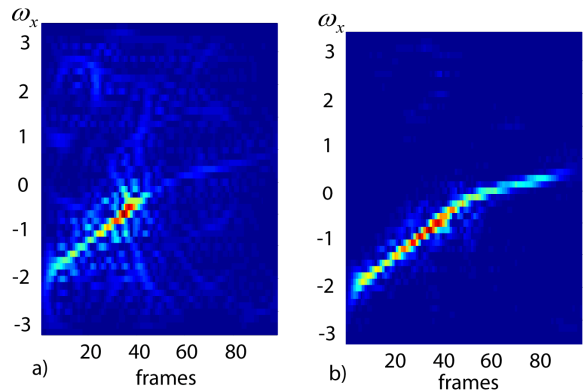

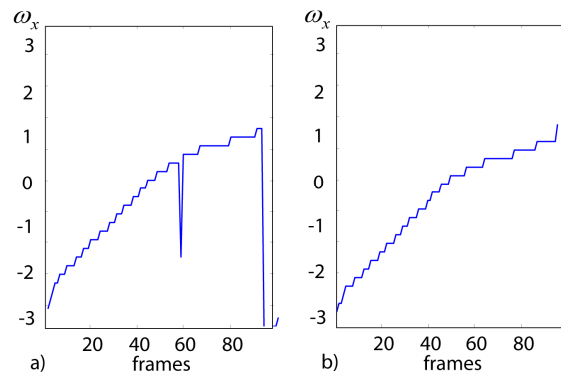**Fig. 2. Time-frequency representations of μ-propagation vector: a) Initial S-method, b) CS based S-method**

Fig. 3. Velocity estimation using: a) Initial S-method, b) CS based S-method

Example 2: In this example we will consider a real world sequence, illustrated by a few frames in Fig 4. The μ-propagation vector is calculated using the same percentage of frames as in the previous example. The Initial S-method and the CS based S-method are shown in Fig. 5. The estimation results are compared in Fig. 6, where we can observe that the CS based S-method significantly improves the estimation results produced from the initial S-method.
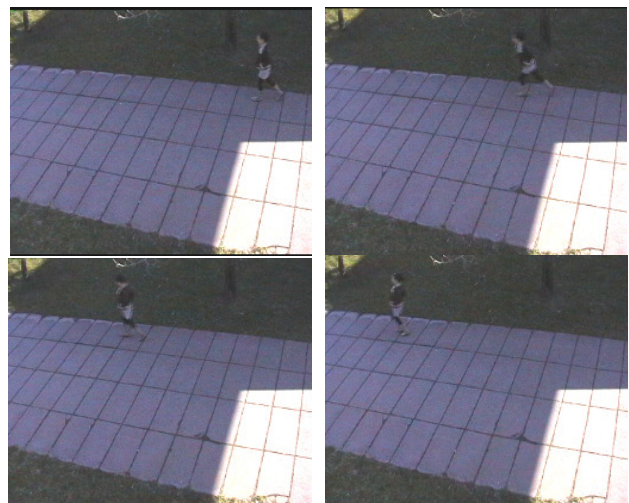

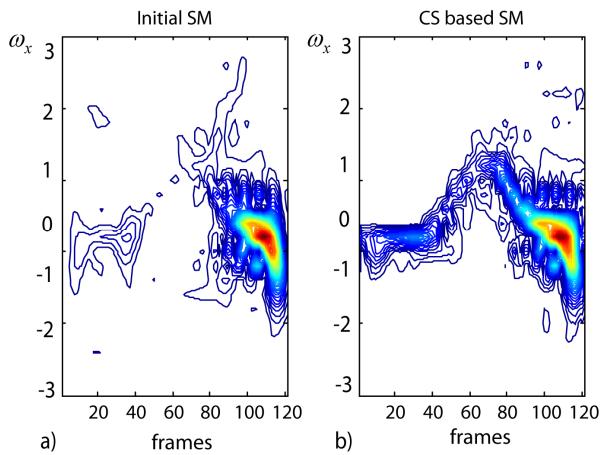**Fig. 4. Frames of real video sequence: 15, 55, 75, 85**

Fig. 5. The time-frequency representations of variable μ-propagation vector (contour plots): a) Initial S-method, b) CS based S-method
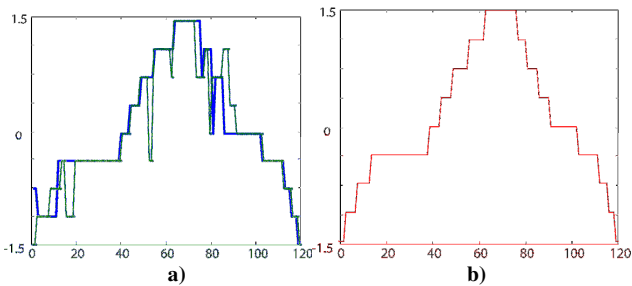

Fig. 6. a) Original velocity estimation, b) velocity estimation using initial S-method (green line) and CS based S-method (blue line)

V. CONCLUSION

An application of reconstruction algorithms to the time-frequency representation of μ-propagation vector is considered. Due to the compressed sensed video sequence the μ-propagation vector is left with small random set of samples, and can be hardly used for estimation of moving object velocity. The missing samples should be therefore recovered using a convex optimization algorithm that returns a reconstructed full data set. The time-frequency representation obtained after recovering missing data is close to the original full data set representation, and thus, can be efficiently used for instantaneous frequency estimation corresponding to the video object velocity. Future research includes a comprehensive multi objects' velocity analyses by applying real-world traffic data stream under congested and non-congested conditions. This will help support further transferability of proposed CS method in other engineering fields.

REFERENCES

[1] B. Boashash, *Time-Frequency Signal Analysis*, in S. Haykin, editor, *Advances in Spectral Estimation and Array Processing*, Prentice Hall, pp. 418-517, 1991.

[2] LJ. Stankovic, S. Stankovic, I. Djurovic, M. Dakovic, "Time-frequency signal analysis," *Research monograph*, Faculty of Electrical Engineering, Podgorica, November, 2011

[3] S. Stanković, I. Orović and C. Ioana, "Effects of Cauchy Integral Formula Discretization on the Precision of IF Estimation: Unified Approach to Complex-Lag Distribution and its Counterpart L-Form," *IEEE Signal Proc. Lett*, vol. 16, no. 4, 2009., pp. 327-330.

[4] S. Stankovic, I. Orovic, E. Sejdic, *Multimedia Signals and Systems*, Springer-Verlag, New York, 2012

[5] I. Orovic, S. Stankovic, M. Amin, "A New Approach for Classification of Human Gait Based on Time-Frequency Feature Representations," *Signal Processing*, Vol. 91, No. 6, pp. 1448-1456, 2011

[6] E. Sejdic and J. Jiang, "'Time-Frequency Analysis of the Heart Sounds," in *Proc. of 2002 ECEGRS*, Ontario, Canada, May 15-17, 2002, pp. 5-9.

[7] S. Stankovic, I. Orovic, N. Zaric, "Robust speech watermarking procedure in the time-frequency domain," *EURASIP Journal on Advances in Signal Processing*, Vol. 2008, No. ID 519206, Page(s) 9 pages, 2008

[8] L. Lee and S. Krishnan, "Time-frequency signal synthesis and its application in multimedia watermark detection," *EURASIP Journal on Applied Signal Processing*, vol. 2006, Article ID 86712

[9] S. Stankovic, I. Djurovic, V. Vukovic, "System architecture for space-frequency image analysis," *Electronics Letters*, 1998, vol.34, no.23, pp. 2224-2245

[10] I. Orovic, M. Orlandic, S. Stankovic, Z. Uskokovic, "A Virtual Instrument for Time-Frequency Analysis of Signals with Highly Non-Stationary Instantaneous Frequency," *IEEE Transactions on Instr. and Measurements*, vol. 60, no. 3, pp. 791 - 803, 2011

[11] H.K. Aghajan, T. Kailath: "SLIDE: Subspacebased line detection", *IEEE Trans. PAMI*, vol.16, no.11, Nov.1994, pp.1057-1073.

[12] H.K. Aghajan, B.H. Khalaj, T.Kailath: "Estimation of multiple 2-D uniform motions by SLIDE: Subspace-based line detection," *IEEE Transactions on Image Processing*, vol.8, No.4, Apr.1998, pp.517-526.

[13] S. Stankovic, I. Djurovic, "Motion parameter estimation by using time frequency representations," *Electronics Letters*, vol.37, no.24, pp.1446-1448, 2001

[14] I. Djurovic, S. Stankovic, A. Ohsumi, H. Ijima, "Motion parameters estimation by new propagation approach and time-frequency representations," *Signal Processing Image Communications*, vol. 19, No. 8 pp. 755-770, 2004

[15] I. Djurovic, S. Stankovic, "Estimation of time-varying velocities of moving objects in video-sequences by using time-frequency representations," *IEEE Trans. on Image Processing*, Vol.12, No.5, pp.550-562, 2003

[16] LJ. Stankovic, "A method for time-frequency signal analysis," *IEEE Trans. on Signal Processing*, vol-42, no.1, Jan.1994.pp.225-229

[17] LJ. Stankovic, S. Stankovic, M. Amin, "Missing Samples Analysis in Signals for Applications to L-estimation and Compressive Sensing," *Signal Processing*, vol. 94, 2014, pp. 401-408

[18] S. Stankovic, I. Orovic, M. Amin, "L-statistics based Modification of Reconstruction Algorithms for Compressive Sensing in the Presence of Impulse Noise," *Signal Processing*, vol.93, no.11, 2013, pp. 2927-2931 , 2013

[19] Baraniuk, R.: 'Compressive sensing' *IEEE SP Magazine*, 2007, **24**, (4), pp. 118-121

[20] E. Candes, and J. Romberg, L1-magic : Recovery of Sparse Signals via Convex Programming, http://www.acm.caltech.edu/l1magic/\#code, 2005, pp. 1-19

[21] S. P. Boyd, L. Vandenberghe, "*Convex Optimization*", Cambridge University Press, Mar 8, 2004 - Business & Economics - 716 pages.

[22] J. A. Tropp, A. C. Gilbert, "Signal Recovery From Random Measurements Via Orthogonal Matching Pursuit," *IEEE Transaction on Information Theory*, vol. 53, no. 12, pp. 4655 – 4666, 2007

[23] M. Elad, *Sparse and Redudant Representations: From Theory to Applications in Signal and Image Processing*, Springer 2010.

[24] D. J. Dailey, F. W. Cathey, and s. Pumrin, "An Algorithm to Estimate Mean Traffic Speed Using Uncalibrated Cameras," *IEEE Trans. on Intelligent Transportation Systems*, vol. 1, no. 2, pp. 98,107, Jun 2000

[25] A. G. Rad, A. Dehghani, and M. R. Karim, "Vehicle speed detection in video image sequences using CVS method", *International Journal of the Physical Sciences*, vol. 5(17), pp. 2555-2563, 2010

[26] J.N . Melo, A. Bernardino, and A.J. Santos-Victor, "Detection and classification of highway lanes using vehicle motion trajectories", *IEEE Transaction on Intelligent Transportation Systems*, vol. 7, no. 2, pp.188, 200, June 2006.

[27] T. Sayed, K. ISmali, M. H. Zaki, and J. Autey, "Feasibility of Computer Vision-based Safety Evaluations", *Transportation Research Records: Journal of Transportation Research Board*, no. 2280, pp. 18-27, 2012

[28] N. Saunier and T. A. Sayed, "Automated Road Safety Analysis with Video Data", *Transportation Research Records: Journal of Transportation Research Board*, no. 2019, pp. 57-64, 2007