Compressive Sensing Reconstruction of Video Data based on DCT and Gradient-Descent Method

Isidora Stanković, *Member IET*, Andjela Draganić, *Member IEEE*

***Abstract* — The reconstruction of video with missing data using a gradient-based compressive sensing algorithm is presented in this paper. The video is assumed to have missing data randomly spread over the frames. Also, it has been assumed that the video frames are sparse in the DCT domain. The value of the gradient is calculated in the sparsity domain and then used to update the pixel values. The efficiency of the video reconstruction using the gradient-descent algorithm is analyzed in terms of the number of available samples and sparsity level, showing better reconstruction performance comparing to the greedy algorithm. Also, it has been shown that the gradient-based algorithm provides high quality reconstruction for real video sequences.**

***Keywords* — compressive sensing, sparse signals, video reconstruction, gradient algorithm, greedy algorithm**

I. Introduction

In the last decade, the reconstruction of data with missing samples is an attractive problem in the field of signal processing. Compressive sensing (CS) is the area dealing with such problems. Theory of missing data reconstruction can be explained within the framework of sparse signals. A sparse signal has a small number of non-zero entries in a transformation domain. Knowing that a signal is sparse, it can be uniquely reconstructed with fewer samples than required by the Shannon-Nyquist sampling theorem [1]-[7].

The application of CS algorithms is widely present in the areas such as signal processing, biomedicine, multimedia, telecommunications, etc. In recent literature, major effects were made to adapt these algorithms within the dynamic framework [6]-[9].

The CS algorithms can be divided into two large groups: the first one is based on the analysis of the transformation coefficients (greedy algorithms) and the second is based on convex relaxation algorithms. The key difference between the two groups of the techniques is the way how the signal sparsity property and the measurements are used. In this paper two reconstruction algorithms will be presented, an orthogonal matching pursuit (OMP) algorithm from the first group and the gradient-based algorithm, a common technique for signal reconstruction from the second group of algorithms.

Application of the CS algorithms in image processing was previously studied in literature [10],[11]. Image processing with CS algorithms is still an attractive research topic in the sense of improving the quality of the images in poorer noisy conditions. Video processing using the CS algorithms is a new developing field with the main challenge being its complex time-varying structure.

The reconstruction of missing pixels randomly spread over the whole frames will be considered in this paper. In the reconstruction process, the information about the video frames sparsity is important but not crucial for reconstruction. The sparsity property of a frame is implicitly assumed.

The paper is organized as follows. After the introduction, the basic definitions of sparse and dynamic signals are represented in Section 2. The reconstruction algorithms are presented in Section 3. The comparison of these algorithms and video reconstruction results are given in Section 4 and the conclusions are in Section 5.

II. Basic Definitions

In most cases, images are transformed and analyzed in the DCT domain. Because of the energy compaction property of an image, usually they have only few non-zero coefficients in the DCT domain. If an image has many small non-zero coefficients spread in the DCT domain, these can be set to zero, as they do not contain significant information about the image. Hence, we may assume that an image $x(n,m)$ is sparse in the DCT domain and that there is a reduced set of available samples at the positions $(n,m) \in \{(n_1,m_1),(n_2,m_2)...,(n_M,m_M)\}$. The image with the available samples (measurements) only that is used for initial reconstruction, is defined by

$$\mathbf{y} = (x(n,m), \text{ for } (n,m) \in \{(n_1,m_1),(n_2,m_2)...,(n_M,m_M)\}).$$

In the initial calculation, it is common to assume that the values of pixels at the positions of missing samples is equal to zero, while at the available positions the pixels are equal to the original ones.

A video can be defined as a sequence of images. If we consider an image to be a two-dimensional signal, the

This work is supported by the Montenegrin Ministry of Science, project grant funded by the World Bank loan: CS-ICT "New ICT Compressive sensing based trends applied to: multimedia, biomedicine and communications".

Authors are with the University of Montenegro, Faculty of Electrical Engineering, Dzordza Vasingtona bb, 81000 Podgorica, Montenegro. Corresponding author is Isidora Stankovic
(e-mail: isidora.lj.stankovic@gmail.com).

video is a dynamic version of these signals. Two steps are used to define a CS reconstruction algorithm for that dynamic signals: reconstruction of two-dimensional DCT signal in the initial frame and then the prediction and update in each next frame. The algorithms will be described after basic formulations for sparse dynamic signals are introduced.

A. Dynamic Signal Description

Consider a signal (image) in a matrix form $\mathbf{x}_t$ at an instant t and its transform $\mathbf{X}_t = \mathrm{T}\{\mathbf{x}_t\}$. The signal sparsity is defined as the number of non-zero transform coefficients. The set of these non-zero positions will be denoted by $\mathbf{K}_t$. The relation between the signal transform $\mathbf{X}_t$ at instant t and the signal transform $\mathbf{X}_{t-1}$ in the previous instant can be defined as:

$$\mathbf{X}_t = \mathbf{X}_{t-1} + \mathbf{q}_t \tag{1}$$

where $\mathbf{q}_t$ is the matrix corresponding to the dynamic change of the signal transform $\mathbf{X}_{t-1}$. If the matrix $\mathbf{q}_t$ has non-zero entries at the frequency positions the same as $\mathbf{X}_{t-1}$, then the signal sparsity does not change at the next instant. If the vector has non-zero entries at positions different from $\mathbf{X}_{t-1}$, then new components will be added to the signal transform at that time instant. Adding more components will increase the sparsity of the signal. Some signal components can vanish at an instant t as well. This effect can also be described by matrix $\mathbf{q}_t$.

III. Algorithms for Reconstruction

Reconstruction of two-dimensional signals will be done by converting the signal and transform matrices into a one-dimensional vectors first. Consider a two-dimensional signal $x(n,m)$, $0 \le n \le P-1$, $0 \le m \le Q-1$. To convert it into vector notation, we can rewrite the signal:

$$x(n + P(m-1)) = x(n,m) \ . \tag{2}$$

The available samples at an instant t are denoted in a vector form by $\mathbf{y}_t$. In a similar way, the signal in transformation domain is written as:

$$X(k + Q(l-1)) = X(k,l) \ . \tag{3}$$

for $0 \le k \le P-1$, $0 \le l \le Q-1$. Its vector form at instant t is $\mathbf{X}_t$.

Signal transformation matrix is re-formulated as a two-dimensional matrix as well. Its rows corresponding to the available samples are denoted by $\mathbf{A}_t$.

In the reconstruction of dynamic signals, two major steps are defined: prediction and update. In the prediction step, analysis is done at time t, using the knowledge about sparsity that we have from *t-1*. In the update step, the adaptation of the signal transform coefficients for time t is done.

A. Orthogonal Matching Pursuit Algorithm

To illustrate the dynamic framework, consider from the family of greedy algorithms the iterative Orthogonal Matching Pursuit (OMP). The dynamic case is introduced in [12]. The algorithm reconstructs the signal in the transformation domain. In the conventional OMP, the approach is to use a fixed threshold and eliminate all components that are below that threshold. After the detection of the frequency positions, the signal is reconstructed by straightforward calculations.

Here we use the iterative form of this algorithm. In the first step, we estimate the highest component position in the transformation domain as

$$k_t^{(i)} = \arg\{\max\{\mathbf{A}_t^H \mathbf{y}_t\}\} \ . \tag{4}$$

The signal in the transformation is reconstructed as

$$\mathbf{X}_t^{(i)} = \left(\mathbf{A}_t^{(i)} \mathbf{A}_t^{(i)H}\right)^{-1} \mathbf{A}_t^{(i)H} \mathbf{y}_t \ . \tag{5}$$

with i=0 at the initial iteration. In each next step i the position of highest coefficient is saved in a vector $\mathbf{K}_t$ of non-zero transform coefficients where:

$$\mathbf{K}_t^{(i)} = \mathbf{K}_t^{(i-1)} \cup \{k_t^{(i)}\} \ . \tag{6}$$

After each added component, the error between the original and the reconstructed signals is checked:

$$e^2 = \sum_{(n,m)=(n_1,m_1),\dots,(n_M,m_M)} \left| x_t(n,m) - x_{tR}^{(i)}(n,m) \right|^2 . \tag{7}$$

where $x_t(n,m)$ is the original signal at instant t and $x_{tR}^{(i)}(n,m)$ is the reconstructed signal in the i-th iteration at time instant t. The procedure is repeated using the error signal until the desired error is achieved.

In the next time instant (frame), the iteration does not start from the beginning. That is, in the prediction step we define the vector of non-zero positions as

$$\mathbf{K}_t = \mathbf{K}_{t-1} \ . \tag{8}$$

For the update step, we check the error of the original signal and the reconstructed signal at time instant t:

$$e^2 = \sum_{(n,m)=(n_1,m_1),\dots,(n_M,m_M)} \left| x_t(n,m) - x_{t|t-1R}^{(i)}(n,m) \right|^2 . \tag{9}$$

where $x_{t|t-1R}^{(i)}(n,m)$ represents the signal at time instant t reconstructed using the knowledge that we have from the previous instant *t-1*. If the error is below an acceptable level, then the non-zero positions of the signal transform did not change from the previous time instant. If that is not the case, the iterations from (4) to (7) are done until the error is bellow an acceptable level. At the end of iteration process for each frame, possible zero values of some coefficients are excluded from the set $\mathbf{K}_t$.

B. Gradient-Based Algorithm

Gradient-based algorithm, unlike the greedy algorithms, is based on the estimation of the missing samples in the spatial domain. The key advantage of this algorithm is that it does not change the available samples. For the reconstruction with CS algorithms, we assume that the signal is sparse in the transformation domain. In this gradient reconstruction algorithm, the sparsity is important but not crucial information.

The algorithm can be described as a modified form of the direct search L_1-minimization. The direct search method is based on finding the minimum of the L_1-norm by checking the range of all possible values. This method gives very accurate results but it is computationally insufficient. The gradient-based algorithm is used to find the minimum of the L_1-norm in an efficient way. The illustration of the L_1-norm gradient moving to the solution is shown in Fig. 1.

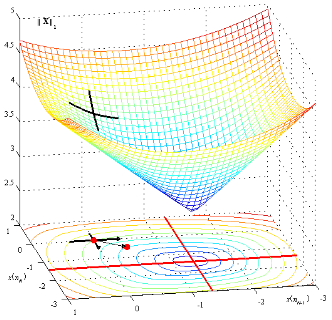

Fig. 1. Illustration of the L_1-norm and its gradient with two missing samples

Because of the dynamics property of a video, the algorithm will be divided into two parts: reconstruction of the signal at the first time instant (i.e. *t=0*), and the reconstruction of signals at other time instances (i.e. *t>0*).

For the initial time instant (frame) *t=0*, the gradient-based algorithm is performed in a usual manner. The algorithm for one-dimensional signals is presented in [13] and the algorithm for images is presented in [10]. The algorithm is as follows:

Let assume a two-dimensional $P \times Q$ signal $x(n,m)$ with M available samples where $M << PQ$. In the initial step, the signal with available measurements $x_t^{(0)}(n,m)$ (for t=0) is defined as:

$$x_0^{(0)}(n,m) = \begin{cases} x(n,m), \text{ for } (n,m) \in \{(n_1,m_1),(n_2,m_2),...,(n_M,m_M)\}, \\ 0, \text{ elsewhere.} \end{cases}$$

For the unavailable pixels we add a constant value Δ. The new signals with available samples and the parameter Δ are created as

$$\begin{aligned} x_{t1}(n,m) &= x_t^{(i)}(n,m) + \Delta\delta(n-k,m-l) \\ x_{t2}(n,m) &= x_t^{(i)}(n,m) - \Delta\delta(n-k,m-l) \end{aligned} \quad (10)$$

After calculating the DCT transform of the new signals, the gradient is calculated by:

$$g_i(k,l) = \frac{\|\mathbf{X}_{t1}\|_1 - \|\mathbf{X}_{t2}\|_1}{2\Delta} \quad (11)$$

where $\mathbf{X}_{t1} = \text{DCT2}\{x_{t1}(n,m)\}$ and $\mathbf{X}_{t2} = \text{DCT2}\{x_{t2}(n,m)\}$.

This procedure is repeated for each k and l corresponding to the positions of missing samples $(k,l) \notin \{(n_1,m_1),(n_2,m_2),...,(n_M,m_M)\}$.

Note that (k,l) in this case are the positions of the missing samples, and not the positions of the transform coefficients, as in equation (3). The values of the gradient for the positions of the available samples $g_i(k,l) = 0$.

The adaptation of each missing pixel is done by:

$$x^{(i+1)}(k,l) = x^{(i)}(k,l) - \mu g_i(k,l) \quad (12)$$

where μ is a constant gradient parameter. Because of the gradient form (Fig. 1.) when the algorithm is close to the solution, it will oscillate around it. This is the reason why the values of μ and Δ are then decreased. In this case, the values are decreased by a factor of 10, i.e. $\mu = \mu/10$ and $\Delta = \Delta/10$.

The algorithm is repeated until the error between two iterations is less than some acceptable error level ε:

$$\max\left|x^{(i+1)}(n,m) - x^{(i)}(n,m)\right| < \varepsilon. \quad (13)$$

For each next frame, instead of initializing the missing samples to zero, we assume the values from the previous time instant at the missing samples positions:

$$y_t^{(0)}(n,m) = \begin{cases} x_t(n,m), \text{ for available samples} \\ x_{t-1}(n,m), \text{ for missing samples} \end{cases} \quad (14)$$

where the available samples are at the positions $(n,m) \in \{(n_1,m_1),(n_2,m_2),...,(n_M,m_M)\}$, $x_t(n,m)$ are the available samples at time instant t and $x_{t-1}(n,m)$ are the samples found at the previous instant *t-1*, assumed at the positions of missing samples [12]. The adaptation iterative process does not start from the beginning as in the first instant. After defining the signal with the available samples, other algorithm steps are the same as in the first frame.

IV. Comparison and Numerical Results

A synthetic signal is formed for the comparison of the presented methods. A sparse signal is defined as:

$$x(n,m) = \sum_{i=1}^{K} A_{p_i q_i} \cos\left(\frac{2\pi k_{p_i}(2n-1)}{4P}\right)\cos\left(\frac{2\pi k_{q_i}(2m-1)}{4P}\right) \quad (15)$$

with $P = 64$, and $A_{p_i q_i}$ represent random amplitudes at the positions (k_{p_i}, k_{q_i}) which are also random. The sparsity of this signal is *K*. For the reconstruction it has been assumed that only M randomly positioned signal samples are available. Table 1 represents calculation time (in seconds) needed for the OMP and gradient-based algorithm to reconstruct the signal (15) for different number of available samples M and sparsity K. The red values represent the time of a more efficient algorithm for given M and K. We will denote by "X" the cases where the reconstruction is not achieved due to insufficient number of available samples.

Table 1: Time in seconds needed for OMP and Gradient

K \ M	32^2 OMP	32^2 Gradient	48^2 OMP	48^2 Gradient	60^2 OMP	60^2 Gradient
16	2.61	28.72	2.81	7.97	3.05	3.64
64	X	29.75	2.84	8.79	3.08	3.80
144	X	26.03	2.87	8.63	3.11	3.66
256	X	85.98	X	8.48	3.38	3.68
576	X	X	X	12.55	5.01	3.64
784	X	X	X	68.85	6.91	3.74
1024	X	X	X	X	10.36	3.93
1296	X	X	X	X	15.31	4.48

For relatively low sparsity level, the OMP reconstructs the signal faster than the gradient-based algorithm. For high signal sparsity, the reconstruction using the gradient-based algorithm is more efficient. The calculations time for the gradient-based needs for signal reconstruction mostly depends on the number of available samples and not on the signal sparsity. Since the sparsity of real images and video frames is significant in the application on real videos we will present reconstruction results with gradient-based algorithm.

Frames of a video with missing data are presented in Fig.2. (left) where the missing data are represented by the black and white pixels. Some of the frames from the reconstructed video are shown in Fig. 2. (right). The video is reconstructed using the gradient-based algorithm. [1]

V. CONCLUSIONS

The application of the iterative OMP and the gradient-based algorithms in the reconstruction of dynamic signals is presented. In difference to the case of reconstruction of stationary signals, for the dynamic signals we do not start from the initial value (which, in most cases, is equal to zero), but with the values calculated in the previous time-instant. Because of the property that the calculation complexity is independent from the sparsity level K, the gradient-based algorithm is used in the reconstruction of two-dimensional dynamic (video) signals.

REFERENCES

[1] E. J. Candes, and M. Wakin, “An Introduction to Compressive Sampling”, *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21-30, March 2008.

[2] R. Baraniuk, “Compressive Sensing”, *IEEE Signal Processing Magazine*, vol. 24, no. 4, pp. 118-121, July 2007.

[3] S. Stankovic, I. Orovic, LJ. Stankovic, "An Automated Signal Reconstruction Method based on Analysis of Compressive Sensed Signals in Noisy Environment," Signal Processing, vol. 104, Nov 2014, pp. 43 - 50, 2014.

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

[5] D.L. Donoho, “Compressed sensing”, *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289-1306, April 2006.

[6] D. Angelosante, G. Giannakis, and E. Grossi, “Compressed sensing of time-varying signals”, *16th International Conference on Digital Signal Processing*, pp. 1-8, July 2009.

[7] LJ. Stankovic, S. Stankovic, I. Orovic, M. Amin, "Robust Time-Frequency Analysis based on the L-estimation and Compressive Sensing," IEEE Signal Processing Letters, vol. 20, No. 5, pp. 499-502, 2013.

[8] N. Vaswani, “Kalman Filtered Compressed Sensing”, *Proc. IEEE Int. Conf. Image (ICIP)*, pp. 893-896, 2008.

[9] J. Ziniel, P. Schniter, “Dynamic Compressive Sensing of Time-Varying Signals Via Approximate Message Passing”, (arXiv:1205.4080), 2012.

[10] I. Stankovic, I. Orovic, S. Stankovic, “Image Reconstruction from a Reduced Set of Pixels using a Simplified Gradient Algorithm”, 22nd TELFOR, November 2014.

[11] J. Romberg, “Imaging via Compressive Sampling”, *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 14-20, 2008.

[12] I. Stankovic, “Track Global Ozone Density with Missing Data”, MSc Thesis (supervisor: dr W. Dai), Department of Electrical and Electronic Engineering, Imperial College London, London, United Kingdom, 2015.

[13] LJ. Stankovic, M. Dakovic, S. Vujovic, “Adaptive Variable Step Algorithm for Missing Samples Recovery in Sparse Signals”. *IET Signal Processing*, vol. 8, no. 3, May 2014, pp. 246-256, (arxiv.org/abs/1309.5749)

[1] The video shown is a sequence from the music movie *“Good morning Montenegro”* from 1990. authored by Miodrag Bole Boskovic.

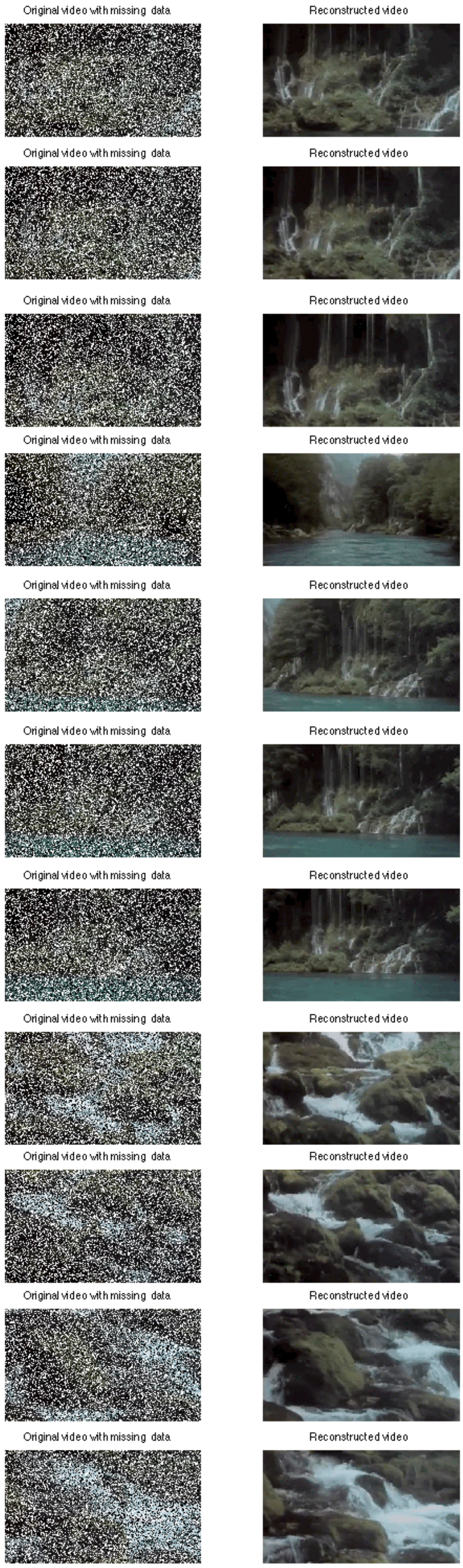

Fig. 2. Noisy video (left) and reconstructed video (right)