Compressive Sensing as a Watermarking Attack

Irena Orović, *Member, IEEE,* Andjela Draganić, and Srdjan Stanković, *Senior Member, IEEE*

***Abstract* — The performance of watermark detection under Compressive Sensing (CS) attack is analyzed in the paper. Watermark is created as a pseudorandom sequence and it is embedded into the DCT image coefficients. CS, as method that provides reconstruction of the signals with small number of samples, is used as watermarking attack. Reconstruction procedure assumes certain number of low frequency DCT coefficients, as well as certain number of randomly chosen middle and high frequency DCT coefficients. It is shown that CS can provide good quality image reconstruction with reduced number of samples and, at the same time, to remove the watermark. The theory is supported by experimental results.**

***Keywords* — Digital image watermarking, Watermark detection, Compressive Sensing, Total Variation**

I. INTRODUCTION

The algorithms for digital content protection have been extensively developed in the recent years. One of the methods for multimedia content protection is digital watermarking [1]-[3]. The watermarking procedure consists of watermark embedding and watermark detection [4]-[7]. A signal, called watermark, is embedded into the data coefficients. In order to prove the ownership, it should be detectable within the host data. Watermark could be affected by various attacks (such as filtering, compression, geometrical image transformation, etc.) [8]-[10], that may significantly degrade the watermarking detection performances. In this paper, we analyze the watermark detection under the influence of a new kind of attack, which is based on popular, recently developed, Compressive Sensing (CS) concept.

Compressive Sensing [1], [11]-[14] introduces an alternative way of signal sampling that differs from standard sampling approach, based on the Shannon-Nyquist theorem. According to the CS theory, the signal samples could be acquired randomly, at the rate which is far below Nyquist. CS is based on the powerful mathematical algorithms used for the reconstruction of missing content [1]. To provide high accuracy signal reconstruction with CS reconstruction technique, certain conditions need to be fulfilled. Namely, the signal has to be sparse in a certain transform domain, which means that the information about the signal is concentrated within a small number of coefficients. The second requirement refers to sampling procedure. The signal acquisition/measurement procedure should be incoherent, in order to provide signal reconstruction with small number of available samples. Reconstructed signal can be obtained by using certain optimization algorithm, which can be based on different norms minimization (l_0, l_1, l_2, etc). The optimal solution in large number of applications is provided using l_1-norm minimization [1], [13]. In image processing applications, the commonly used optimization technique is called Total Variation (TV) minimization [1], [15]-[17].

In this paper the CS method is considered as the attack on the watermarked image. CS concept is combined with TV minimization for image reconstruction. The reconstruction using different number of image coefficients for CS measurements is analyzed, as well as the performance of watermark detection under CS attack.

The paper is organized as follows. In Section II, theoretical background on digital watermarking is given, as well as description of the watermarking procedure used in the paper. Basic concepts on the CS and TV method are given in Section III. In this section image reconstruction procedure using CS method is also described. Experimental results and concluding remarks are given in Sections IV and V, respectively.

II. DIGITAL WATERMARKING

Digital watermarking has been introduced for multimedia data protection, copyright protection, tracking of digital copies, etc. Watermark is the secret signal embedded in the multimedia content in a way that does not modify the original content. Depending on the type of host signal, different watermarking techniques were introduced: audio watermarking [1], [2], video [4] and image watermarking approaches, [4]. Watermark could be embedded either directly into the signal domain, or into some of the transform domains [8]. Perceptually, watermark could be classified as perceptible or imperceptible. Perceptible watermark changes the original content and is not very popular nowadays. The commonly used technique is imperceptible watermarking, which will be used in this paper. Further, the watermark should be robust to the various attacks, such as compression, noise, filtering, etc. However, there is usually a trade-off between watermark imperceptibility and robustness.

Beside the watermark embedding procedure, the method for watermark detection should be defined as well. The detection procedure can be blind (without using the original content), or non-blind (when original content is presented).

In this paper, an additive watermarking procedure in the

This work is supported by the Montenegrin Ministry of Science. The authors are with the Faculty of Electrical Engineering, University of Montenegro, 20000 Podgorica, Montenegro. The corresponding author is Irena Orović, phone:+38267516795, fax:+382245873, Email: irenao@ac.me

DCT domain is considered, since the watermarking in the transform domains provides higher robustness to attacks. The block-based DCT watermark embedding procedure is considered. Image is first divided into 8x8 DCT blocks. Watermark is embedded into the DCT coefficients from each block, by using additive procedure. Since the embedding into the strongest DCT coefficients will produce image degradation, these coefficients have been avoided. Watermark embedding in the high frequency part is not suitable from the robustness standpoint, since in this case, the watermark could be easily removed. Hence, the middle frequency coefficients of each block are used for watermarking embedding, which can be defined as:

$$DCT_w = DCT + \alpha w, \tag{1}$$

where DCT are the original middle-frequency DCT coefficients (from 8x8 block), DCT_w are the watermarked DCT coefficients, α is the watermark strength and w is the watermark.

The detection procedure is blind and it is performed by using the standard correlation detector [8], [10]. The Gaussian pdf of the DCT coefficients is assumed, and it is defined as:

$$D = \sum_{i=L+1}^{L+K} w_i DCT_{w_i}. \tag{2}$$

Note that, for different choice of the watermarking coefficients, the optimal detector forms could be used [6]. The detector response for watermark (right key) should be larger than for any wrong key generated in the same way as the watermark:

$$D_{ww} = \sum_{i=L+1}^{L+K} w_i DCT_{w_i} > \sum_{i=L+1}^{L+K} wrong_i DCT_{w_i} = D_{wr}. \tag{3}$$

Indices ww and wr denote the watermark and the wrong keys, respectively. As a measure of a detection quality the ratio R is used:

$$R = \frac{\bar{D}_{ww} - \bar{D}_{wr}}{\sqrt{\sigma^2_{ww} + \sigma^2_{wr}}}, \tag{4}$$

where $\bar{D}$ and σ are mean values and the standard deviation of the detector responses for the right (ww) and the wrong (wr) keys. Based on (4), probability of detection error is calculated as:

$$P_e(R) = \frac{1}{2} erfc(\frac{R}{\sqrt{2}}). \tag{5}$$

III. Compressive Sensing

A. Compressive Sensing concept

CS allows signal reconstruction using small set of randomly chosen samples. Signal acquisition rate could be much smaller than it is required by Shannon-Nyquist theorem. In order to provide reliable reconstruction, the signal has to be sparse in a certain transform domain. Sparse signal condense information of interest into few non-zero samples in the transform domain. An N-dimensional vector could be represented in the transform domain, using the following relation [1], [13]:

$$x = \sum_{i=1}^{N} b_i \psi_i = \psi b, \tag{6}$$

where b_i is weighting coefficient, ψ_i is basis vector, ψ denotes NxN transform matrix whose columns are basis vectors and b is the equivalent of the signal in ψ domain. If x has S non-zero samples in the transform domain (where $M>S$ and $M<<N$ holds), it is said that x is S-sparse. Successful reconstruction requires incoherent measurement procedure, i.e. measurement matrix ϕ should be incoherent with the transform matrix ψ [13]. It means that the correlation between two matrices should be low, as lower correlation leads to a smaller number of measurements required to recover the entire signal. Random matrices satisfy low coherence condition and are often used in measurement process.

Acquired measurements are stored in vector v. Hence, we can write:

$$\begin{aligned} v_{M\text{x}1} &= \phi_{M\text{x}N} x_{N\text{x}1}, \\ v &= \phi x = \phi\psi b = Ab, \end{aligned} \tag{7}$$

where ϕ is measurement matrix and A is CS matrix. System (7) consists of M equations with N unknowns. Therefore, system is undetermined ($M < N$) and has infinite number of solutions. Optimal solution is obtained by finding the sparsest solution, among infinite number of them [13]. For that purpose, optimization algorithms are used. Commonly used optimization technique is based on l_1 – minimization. The optimization problem is defined as:

$$\hat{b} = \min \|b\|_{l_1} \; subject\ to \;\; v = Ab, \tag{8}$$

where $\hat{x}$ is solution of the minimization problem and $\|b\|_{l_1} = \sum_{i=1}^{N} |b_i|$ is l_1 – norm of vector b_1.

B. Watermarking attack using CS Total Variation minimization

One of the commonly used methods for reconstruction of CS images is the Total Variation (TV) minimization. TV is popular not only in image reconstruction [15], [17], but also in restoration and denoising, due to its ability to preserve image edges. On the other side, images are usually not sparse in the transform domain, but their gradient is. Since TV in fact, involves minimization of the gradient, it founds applicability in image processing.

Consider a set of image measurements v. Measurements are taken from the low frequency (v_1) and the middle to high frequency image DCT coefficients (v_2), i.e.:

$$v = v_1 + v_2. \tag{9}$$

Note that v are, in fact, zig-zag reordered DCT coefficients: v_1 are first K_1 coefficients and v_2 are chosen randomly from the rest of the DCT. As most of the image energy is contained in low frequencies, K_1 low-frequency coefficients are necessary to provide good quality of reconstructed image. The TV minimization problem for measurement vector v and transform domain vector b, is defined as:

$$\min_b TV(b) \;\; subject \;\; to \;\; v = Ab. \tag{10}$$

The TV of the signal b represents a sum of the magnitudes of discrete gradient at each point, and can be defined as [1]:

$$TV(b) = \sum_{i,j} \left\| D_{i,j} b \right\|_{l_2}. \tag{11}$$

Gradient approximation for the pixel ij is denoted as $D_{i,j}$ and is described by using relation:

$$D_{i,j} b = \begin{bmatrix} b(i+1, j) - b(i, j) \\ b(i, j+1) - b(i, j) \end{bmatrix}. \tag{12}$$

Now, the discrete form of the Total Variation could be defined using the following relation:

$$TV(b) = \sum_{i,j} \sqrt{(b_{i+1,j} - b_{i,j})^2 + (b_{i,j+1} - b_{i,j})^2}. \tag{13}$$

The minimization problem can be actually formulated as a second-order cone programming and solved by using log-barrier method:

$$\min_{t,w} t_{ij} \;\; subject \;\; to \;\; \left\| D_{i,j} b \right\|_{l_2} \le t_{ij}, \quad v = Ab, \tag{14}$$

where $t_{ij}^2 \ge (b_{i+1,j} - b_{i,j})^2 + (b_{i,j+1} - b_{i,j})^2$.

Note that the quality of the CS reconstructed image is shown to be similar to the watermarked one. Reconstructed image contains far less information about watermark, and thus the watermark detection procedure fails, as it will be shown in the next Section.

IV. Experimental results

In this section, the watermark detection after CS based image reconstruction (under CS attack) is analyzed. Consider image in high resolution (256x256, for example). Image is completely described with 256x256=65536 coefficients. The watermark is embedded using the procedure described in Section II. Image samples are acquired from the DCT domain, using CS procedure described in the previous section. The measurements consist of K_1= 4000 low frequency DCT coefficients (6% of the total number of DCT coefficients), while the rest of the coefficients (that corresponds to the image details) are acquired in the pseudo-random manner (K_2).

The number of the middle and high frequency coefficients is variable, starting from the 20% in total number of measurements, and finished with 50%. Watermarked image is reconstructed from the collected samples, by using CS procedure and TV minimization procedure.

Relation (4) is used as a measure of detection quality, and probability of error detection is calculated according to relation (5). Cameraman image of 256x256 pixels is used for watermarking. Watermark strength is chosen to be 2. Image is reconstructed with different number of measurements. Figure 1 shows original and watermarked image. In the Table 1, PSNR beetwen watermarked and reconstructed image, is calculated. As it can be seen, PSNR of the reconstructed image (with 50% of measurements) is about 37 dB, which gives very good quality of the reconstructed image, similar to the watermarked one.

a) b)

Fig. 1: a) Original, b) watermarked Cameraman image

Fig. 2: Image reconstructed using 50% coefficients, in total

TABLE 1: NUMBER OF MEASUREMENTS USED FOR RECONSTRUCTION AND MEASURES OF IMAGE (PSNR) AND DETECTION (R_1 AND R_2) QUALITY

Number of measurements (%)	K1	K2	PSNR (dB)	R_1 (no attack)	R_2 (CS attack)
20	4000	9000	28.93	7.2	1.23 detection failed
30	4000	16000	31.85	6.36	2.22 detection failed
40	4000	22000	34.10	8.1372	2.8598 detection failed
50	4000	29000	37.2	6.34	2.98 detection failed

Fig. 2 shows image reconstructed using 50% of the coefficients - 4000 low frequency coefficients and 29000 od the middle and high ones. The reconstructed image shows no degardation due to reduction of the coefficients number used for reconstruction. Measure of detection quality for image without attack is denoted as R_1, and for image reconstructed using CS, is denoted with R_2. Figure 3 shows detector responses in the cases of no attack (Fig. 3a) and for CS reconstruction with different number of coefficients (Fig. 3 b-e). Watermark detection fails if the number of measurements is 50% (or below) of the total number of image coefficients. The same conclusion could be made by observing Table 1. For the value of measure R_2 below 3, the detection procedure fails.

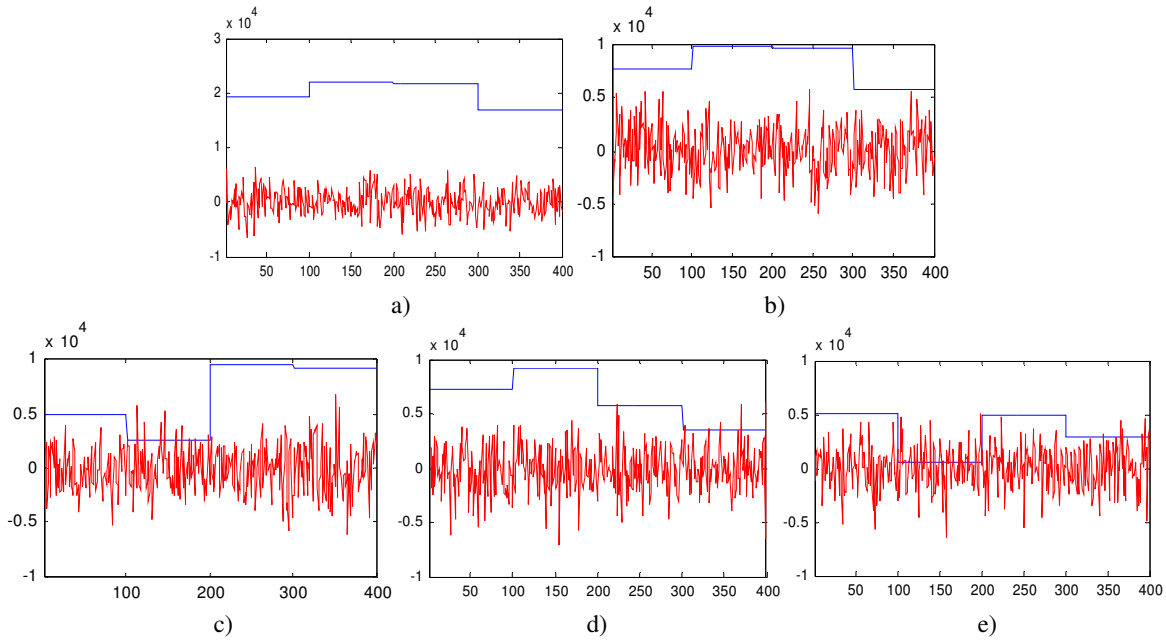

Fig. 3: Detection results for a set of right keys - watermarks (blue) and wrong trials (red): a) without attack, b) CS reconstruction with 50% measurements, c) CS reconstruction with 40% measurements, d) CS reconstruction with 30% measurements, e) CS reconstruction with 20% measurements

V. Conclusion

In this paper, the ability of watermark detection is analyzed for the case when watermarked image is reconstructed from its small set of samples using CS procedure. The additive block/based DCT watermark embedding procedure is observed. Image is reconstructed by using different number of DCT coefficients as CS measurements. It is shown that under the CS attack the watermark will not be reliably detected, although the reconstructed image has a high quality, visually very close to the quality of original image.

References

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

[2] Y. Nakashima, B. R. Tachibana, "Watermarked Movie Soundtrack Finds the Position of the Camcorder in a Theater," *IEEE Transactions on Multimedia* , vol.11, no.3, pp.443,454, April 2009.

[3] S. Stankovic, I. Orovic, N. Zaric, "An Application of Multidimensional Time-Frequency Analysis as a base for the Unified Watermarking Approach," *IEEE Transactions on Image Processing*, vol. 1, no. 3, pp.736-745, 2010.

[4] Gopika V Mane, G. G. Chiddarwar - Review Paper on Video Watermarking Techniques - published at: "International Journal of Scientific and Research Publications (IJSRP), Volume 3, Issue 4, April 2013 Edition ".

[5] L. Le, S. Krishnan, "Time-Frequency Signal Synthesis and Its Application in Multimedia Watermark Detection," EURASIP Journal on Advances in Signal Processing, 2006.

[6] S. Stankovic, I. Orovic, N. Zaric,"Robust watermarking procedure based on JPEG-DCT image compression," *Journal of Electronic Imaging*, vol. 17, no. 4, Page(s) 043001, 2008.

[7] M. Barni, F. Bartolini, A. De Rosa, A. Piva, "A new decoder for the optimum recovery of nonadditive watermarks," *IEEE Transactions on Image Processing*. 10, (2001), pp.755-766.

[8] 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.

[9] I. Orovic, P. Zogovic, N. Zaric, S. Stankovic, "Speech Signals Protection via Logo Watermarking based on the Time-Frequency Analysis," Annals of Telecommunication, vol. 63, No. 5-6, pp. 276-284, 2008.

[10] S. Stankovic, I. Djurovic, R. Herpers, LJ. Stankovic, "An approach to the optimal watermark detection," *AEUE International Journal of Electronics and Communications*, vol. 57, no. 5, pp. 355-357, 2003.

[11] J. Romberg, "Imaging via Compressive Sampling," Signal Processing Magazine, IEEE , vol. 25, no. 2, pp. 14, 20, March 2008.

[12] S. Stankovic, I. Orovic, "An Ideal OMP based Complex-Time Distribution," *2nd Mediterranean Conference on Embedded Computing MECO* - 2013, pp. 109-112, June 2013, Budva, Montenegro , 2013.

[13] E. J. Candes, M. B. Wakin, "An Introduction To Compressive Sampling," Signal Processing Magazine, IEEE , vol. 25, no. 2, pp. 21, 30, March 2008.

[14] 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, 2013.

[15] L. Rudin, S. Osher, E. Fatemi, "Nonlinear total variation based noise removal algorithms", *Physica D*, 60:259-268, 1992.

[16] J. M. Bioucas-Dias, M. A. T. Figueiredo, "Multiplicative Noise Removal Using Variable Splitting and Constrained Optimization", *IEEE Transactions on Image Processing*, vol. 19, no. 7, July 2010, Pages 1720-1730.

[17] A. Chambolle, "Total variation minimization and a class of binary MRF models," *Proceedings of the 5th international conference on Energy Minimization Methods in Computer Vision and Pattern Recognition,* Lecture Notes in Computer Science Volume 3757, 2005, pp. 136-152.