VIDEO WATERMARKING IN 3D DCT DOMAIN

M. Carli, R. Mazzeo, and A. Neri

AE Department
University of Roma TRE, Roma, Italy
phone: + (39) 0655177061, fax: + (39) 0655177026, email: (carli, neri)@uniroma3.it
web: http://www.comlab.uniroma3.it

ABSTRACT

In this work a novel scheme of embedding data in a video sequence is presented. The method is based on a modified Quantization Index Modulation (QIM) embedding algorithm in the 3D-DCT domain. Different zones in the 3D DCT cube have been analyzed in order to obtain better performances in term of robustness and imperceptibility. Analysis of the results shows the effectiveness of the proposed scheme.

1. GENERAL INFORMATION

The increasing presence of multimedia data in every day life is the result of the fusion and spreading of several factors as Internet, third-generation mobile communication, WLAN easy to access and use networks, and so on.

The availability of low cost systems to process digital data allows an easy manipulation of them. Despite of the maliciousness or not of the modification or duplication of digital data, the availability of digital tools to protect the digital information is needed.

A commonly used method to increase the security is the watermarking of digital data. By watermarking the original data before its distribution over the network, a copyright data, or mark, is embedded in the version to be distributed so that no other user (malicious or not) will be able to produce an identical copy of the original data and yet does not contain the embedded message.

Among the several constraints to be observed, every embedding system should not affect the perceptual quality of the video (*invisibility*) producing noticeable distortions in the received data. Besides, the mark should not be significantly altered by malicious (an attempt to alter the mark) or unintentional (compression, transmission or filtering) operations (*robustness*). Furthermore, it should not be possible to remove the mark from the video, even if the embedding scheme is known (*security*).

These constraints are in some sense antagonist. Increasing the robustness, for example, generally increases the visibility of the mark.

Several embedding methods have been proposed in the literature; a first classification of these methods can be done according to the particular domain in which the embedding process is performed. The mark can be inserted either in the spatial domain [1-3] or in an ad hoc transform domain such as DCT, Fourier, or wavelet [4-6]. In a previous work [7], the authors proposed a 3D DCT based embedding system in order to obtain enough capacity to hide the luminance component, down sampled version of a frame, 8 bit / pixel, inside a shot of a video. This capacity allows improving the performances of an error concealment method [9].

The experimental results have shown the effectiveness of the idea and push us to improve the method to be more robust to wanted or unwanted data modification. In that work [7], three different embedding schemes were applied to 3D DCT domain and the results were investigated. In particular, two schemes were obtained by applying the 2D modified Cox's embedding scheme [5], while the third one consisted in a hybrid application of Least Significant Bit substitution method to the 3D volume.

In this work we employed a different scheme to select the key frame of a video shot detection technique and a different embedding scheme in the 3D DCT domain. Furthermore, several schemes have been studied and compared to determine the most effective procedure concerning both visibility and robustness constraints.

The rest of the paper is organized as follows. In Section 2 the 3D DCT transform is briefly introduced. In Section 3 the embedding scheme is presented, while in Section 4 preliminary results are presented. Finally, Section 5 draws the conclusions.

2. 3D DCT TRANSFORM

In literature, 3D transform based algorithms have been recently experimented, especially for video compression as they should perform better than current block-based motion compensated predictive methods. In particular the 3D DCT has been proposed for both image [3, 4] and video compression. Recently this domain has also been considered for data embedding purposes [7].

The 3D Discrete Cosine Transform (DCT) is a real-valued, separable, orthonormal transform whose basis functions are cosine functions.

Let **X** be a 3D signal of size *m* by *n* by *t*, which, in our case, corresponds to a set of consecutive video frames. Let **Y** be the 3D DCT of **X**, of size *m* by *n* by *t*. The matrix **Y** can be computed as follows:

$$\mathbf{Y}_{\mathrm{kij}} = \alpha_k \alpha_i \alpha_j \sum_{t=0}^{N_f-1} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \mathbf{X}_{tmn} \cos\frac{\pi(2t+1)k}{2N_f} \bullet$$

$$\bullet \cos\frac{\pi(2m+1)i}{2M} \cos\frac{\pi(2n+1)j}{2N}$$

where

$$\alpha_k = \begin{cases} \frac{1}{\sqrt{T}}, & k=0 \\ \frac{2}{\sqrt{T}}, & else \end{cases} \quad \alpha_i = \begin{cases} \frac{1}{\sqrt{M}}, & i=0 \\ \frac{2}{\sqrt{M}}, & else \end{cases}$$

$$\alpha_j = \begin{cases} \frac{1}{\sqrt{N}}, & j=0 \\ \frac{2}{\sqrt{N}}, & else \end{cases}$$

Since the 3D DCT is a separable function, the above expression can be evaluated by applying the 1D DCT transform along each of the three dimensions of **X**.

3. EMBEDDING ALGORITHM

The first step of the proposed system is the partitioning of the video in shots. To this end, if the normalized correlation between two consecutive frames drops below a given threshold, then a shot boundary is detected. Specifically, in the numerical results reported here the threshold has been set to 0.3, after a set of experiments performed on several videos with different content.

The number of frames included in a shot also determines the size of the 3rd dimension of the 3D signal support. To be compliant with a previous work, and to be able to better compare the performance, in the simulations each shot is composed by 8 frames. The embedding algorithm is applied every 8 frames consecutively.

In this work we apply the Scalar Costa Scheme embedding algorithm (SCS)[12, 13] in the 3D-DCT domain. In the following we breifely summarized the method described in the cited papers.

The binary representation of the watermark message **m** is encoded into a sequence of watermark letters **d**. The elements d_n belong to an alphabet $\mathrm{D} = \{0,1,...,D-1\}$. In the following, let us denote with **x** the host signal and with s the watermarked one. For a Costa-type embedding of the watermarks letters **d** into the signal **x**, a sequence $\mathbf{q} = \mathbf{w}/\alpha$, which is nearly orthogonal to **x**, has to be determined. This is reduced to the sample wise operation

$$q_n = Q_\Delta\left\{x_n - \Delta\left(\frac{d_n}{D} + k_n\right)\right\} - \left(x_n - \Delta\left(\frac{d_n}{D} + k_n\right)\right)$$

where $k_n \in [0,1)$ is an element of cryptographically secured pseudorandom sequence **k** and $Q_\Delta\{\bullet\}$ denotes a scalar uniform quantization process with step size Δ. Finally, the transmitted watermark sequence is given by: $\mathbf{w} = \alpha\mathbf{q}$

The watermarked data is given by $\mathbf{s} = \mathbf{x} + \mathbf{w} = \mathbf{x} + \alpha\mathbf{q}$.

At the receiver side, the marked signal is $\mathbf{r} = \mathbf{x} + \mathbf{w} + \mathbf{v}$.

In order to extract the mark, the following rule is applied:

$$y_n = Q_\Delta\{r_n - k_n\Delta\} - \{r_n - k_n\Delta\}$$

For binary SCS $(D=2)$, $|y_n| \le \Delta/2$, where y_n should be close to zero if $d_n = 0$ is transmitted and close to $\pm\Delta/2$ for $d_n = 1$. In our experiment we use $\alpha = 1$.

Some considerations can be carried out about the embedding scheme. The amount of data to be embedded, the robustness to be achieved, and especially the purpose of the application are important factors to be considered. The computational complexity obviously increases if the embedding is performed every 8 frames. On the other side, such a temporal embedding scheme allows to gain in robustness versus frame dropping, synchronization and transmission errors. For instance, when data hiding is finalized to support error concealment [9], the effectiveness of the restoration is directly related to the quality of the recovered mark. In fact, in that case, the mark consists in a thumbnail of the frame itself: by embedding more often the full gray scale version of it, the probability to extract a cleaner version of the mark is higher, resulting in improved visual quality of the processed video. More considerations about this problem are in the conclusions.

The mark we want to embed is the "best" representative of each shot. Then for each shot, a single frame summarizing its content, named "key frame", is extracted. At this end, several algorithms, with different computational complexity are available. Here, to face the computational burden we resorted to a simple algorithm that selects as key frame the central frame of each shot. Since we are dealing with video, the idea is that the frame in middle will present the lower interframe difference from the other frames in the shot.

The binary mark sequence is then computed by sub-sampling the luminance component of the key frame. Specifically, sequentially scanning the 8 bit plans corresponding to the gray-scale image, we obtain a binary sequence consisting of 8 binary frames.

4. EXPERIMENTAL RESULTS

All the results here reported, refer to the 'coastguard' sequence (144x176 pixels). Tests performed on other sequences with different content, present similar performances. The DCT cube is 144x176x8 elements and the embedded mark is a cube of 36x88x8 pixels.

The mark has been inserted in three different areas of the DCT cube:

- Starting from coefficient (10, 10, 1), top left area in the DCT cube. *High Zone.*
- Starting from coefficient (54, 44, 1), central part of the cube. *Middle Zone*
- Starting from coefficient (108, 88, 1), bottom right area. *Low Zone*

To test the proposed scheme, we have performed some modifications of the embedding parameters. In the first test, we have varied the quantization step Δ: by increasing it, the mark robustness and also the introduced artifacts increase. In Figure 1 the average PSNR among original and marked frames is depicted.

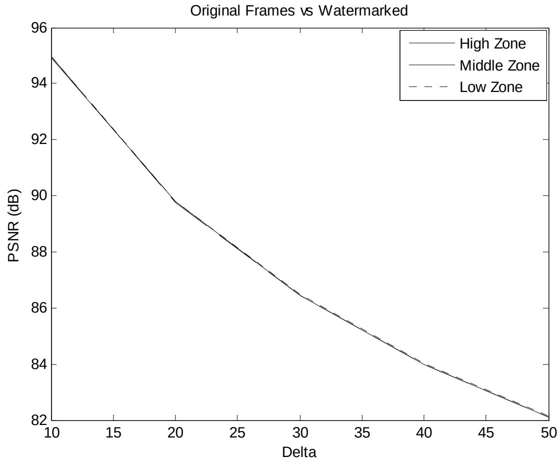

Figure 1: average PSNR between original and watermarked frames for the three different embedding zones.

As expected, Figure 1 shows that by increasing the quantization factor, the average error of the frame decreases. The maximum difference from the three zones is below 0.1 dB. To measure the *perceived quality* of the embedded frames, the Weighted-PSNR (WPSNR) [9] function, based on the Contrast Sensitivity Function, has been computed. In Figure 2, the results are plotted. It can be easily noticed that less artifacts are introduced if the low or the middle areas in the DCT cube are modified. In Figure 3, an example of original and marked frame (Δ=50) is reported. It is useful to mention that, since we are dealing with a video sequence, small localized artifacts in a single frame, are not to likely visible by a user looking at the video.

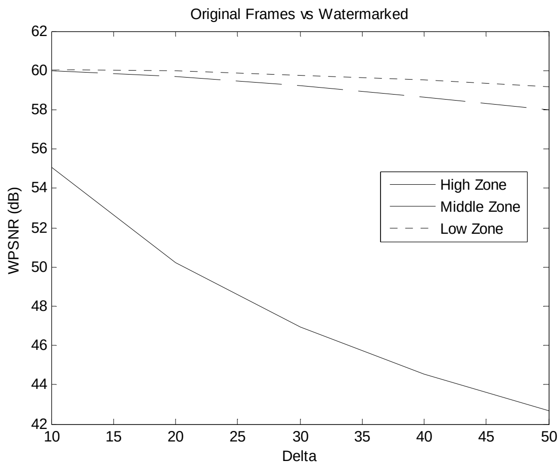

Figure 2: average W-PSNR between original and watermarked frames for the three different embedding zones.

In the second trial the robustness versus some attacks has been tested, in order to select the best embedding scheme and to tune some embedding parameters.

Figure 3: original (left) and marked frame (right). (Δ =50)

In the following the main results obtained are reported.

- JPEG compression of the frames in a shot. The idea is to measure the degradation of the recovered mark, when JPEG compression is performed on the cover signal. In Figures 4-6 is reported the PSNR of the extracted mark versus the original one for 10 compression rates, 5 quantization levels, and for the three embedding zones. It can be noticed that increasing the quantization level, higher robustness is achieved. In Figure 7, the results of comparison among the three methods are reported (Δ=50). As expected, due the JPEG compression algorithm, the quality of the extracted mark is higher in the High Zone scheme . A qualitative comparison can be performed by looking at Figure 8 in which the original and the extracted mark images are reported (Δ=50, Low Zone, JPEG quality Factor = 80). PSNR 68 dB.
- PSNR attack. This is performed by applying the classical Stirmark [10-11]. In Figure 9 the PSNR curves corresponding to High zone, are plotted. In this attack, this zone is more sensitive to the attacks. Nevertheless, the value of the PSNR between original and extracted mark in the worst case is still over 62 dB.

In the following, typical manipulations of watermarked video have been considered. In all the performed tests, for the quantization parameter the value Δ=50 has been used.

- Contrast change: the DCT coefficients present considerable variations especially in the low frequencies. Hence, the High zone is more sensitive to this attach. In Figure 10 the results corresponding to PSNR vs. Contrast change are reported. Note that the value of contrast change=±50% correspond to causing the content of the frame to be unintelligible. The value of PSNR for the extracted mark is 68 dB.
- Intensity value attack: also in this case the modification of intensity values of the video, affects the DCT coefficients resulting a decreased performances for the High zone embedding scheme. Nevertheless, the PSNR of the extracted mark when a Middle zone scheme is adopted, and a variation of +30% for the intensity is used, is still over 70 dB.
- Video compression: the watermarked video has been compressed by using MPEG1, MPEG2, and other video coders freely available on Internet.

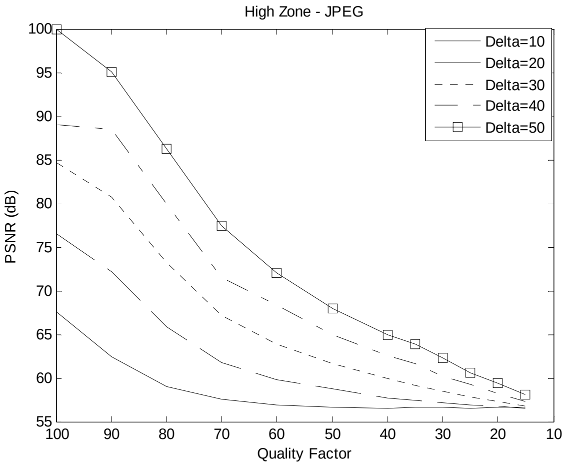

Figure 4: JPEG compression- High Zone scheme

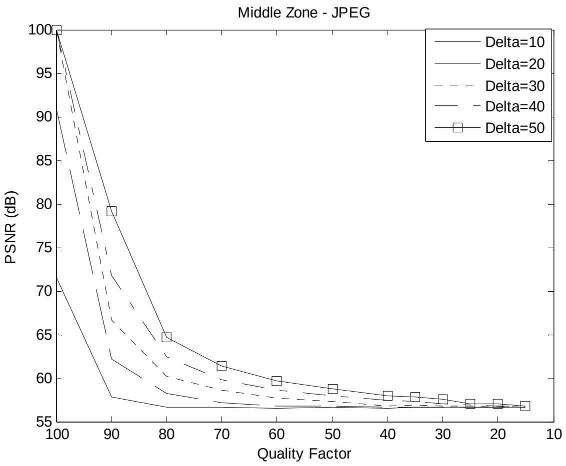

Figure 5: JPEG compression- Middle Zone scheme

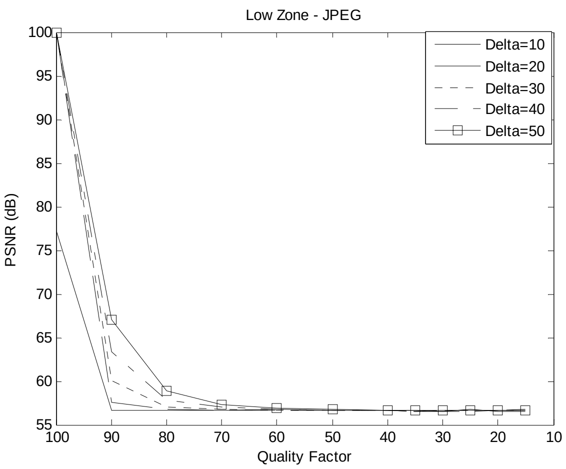

Figure 6: JPEG compression- High Zone scheme

In Figure 11 is reported the PSNR between original and extracted mark corresponding to several compression rates for a MPEG2 coder. The quality factor 110 indicates a non compressed video. The High Zone scheme has always better performances, also at low compression rates. The mark is almost perfectly reconstructed (both MPEG1 and MPEG2 coders) in High and Middle Zone. Little degradations appear instead in Low Zone (87,27 dB). We obtain comparable results also with the commercial video coders.

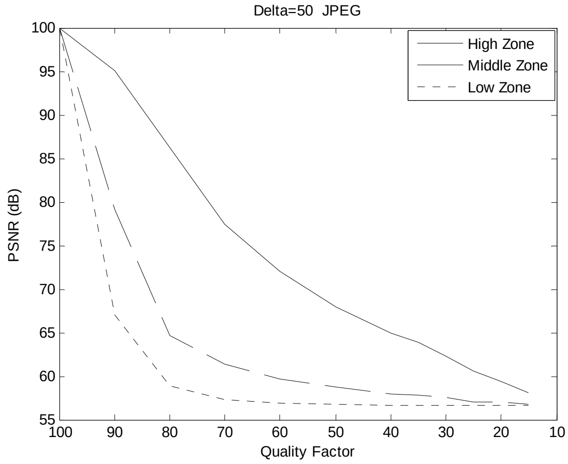

Figure 7: JPEG compression – PSNR comparison at constant quantization parameter

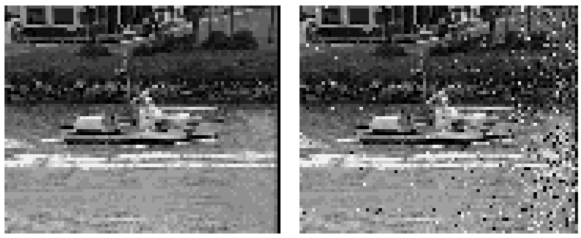

Figure 8: JPEG compression. Original and extracted mark. PSNR=68dB

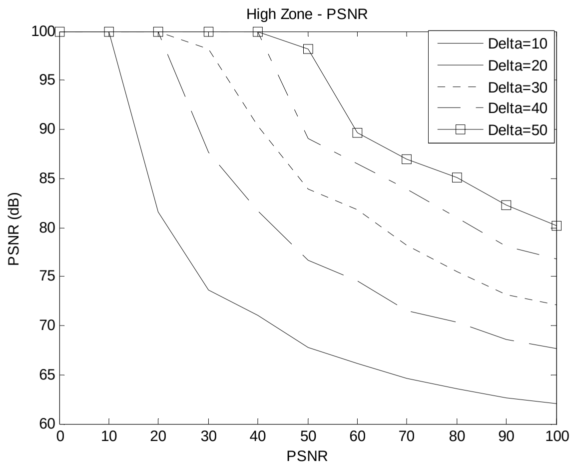

Figure 9: PSNR attack- High Zone scheme

- Frame loss: this situation often occurs during video transmission over packet networks. We tried to simulate this attack by dropping one frame starting to the first one of the 3D DCT block to the last one, verifying the effects on the extracted mark. In the detection phase, in order to rebuilt the sequence of eight frames needed to compute the 3D DCT cube, we simply added as eight frame, the ninth one. The results show the importance of the position of the mark in the 3D cube. It is also useful to shuffle the data to be embedded to increase the robustness.

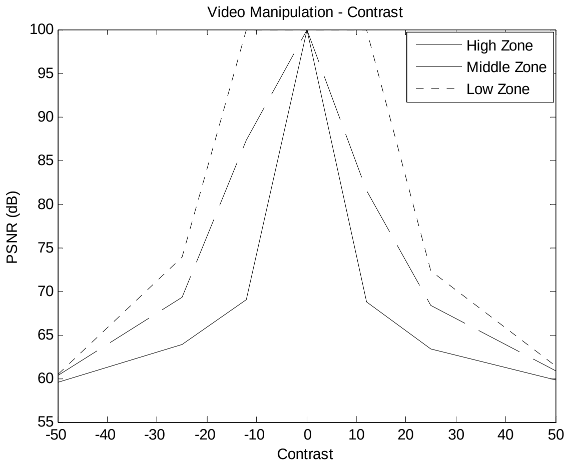

Figure 10: Contrast attack

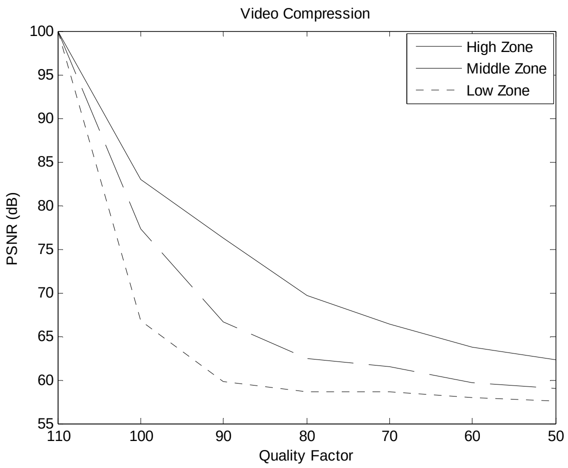

Figure 11: Video Compression attack. High zone scheme outperforms the other two schemes at high and middle levels of compression rates.

5. CONCLUSIONS

In this paper, an embedding scheme, based on Scalar Costa Scheme in 3D DCT domain, for video watermarking has been presented. The performed tests confirm the effectiveness to use a volume in order to hide a huge amount of data with little or no visible distortion to the host video. The quality of the extracted mark remains high even after many attacks. The performed simulations allow selecting the best scheme depending on application. It is important to underline that digital watermarking is not only data hiding. The requirements and the constraints depend on the final application. The variations of the embedding scheme are proposed to understand possible real applications of the general method.

In our future study we are going to apply other watermarking schemes in 3D domain both for authentication and for concealment of the video.

REFERENCES

[1] N. Nikolaidis and I. Pitas, “Robust image watermarking in the spatial domain,” Signal Processing, vol. 66, no. 3, pp. 385–403, May 1998.

[2] R. Wolfgang and E. Delp, “A watermark for digital images,” in IEEE Proc. Int. Conf. Image Proc. ICIP 1996, 1996, pp. 219–222.

[3] D. Gruhl W. Bender and N. Morimoto, “Techniques for data hiding,” Tech. rep., MIT Media Lab, Cambridge, MA, 1994.

[4] M. Barni, F. Bartolini, V. Cappellini, and A. Piva, “A dct-domain system for robust image watermarking,” Signal Processing, vol. 66, no. 3, pp. 357–372, May 1998.

[5] I. Cox, J. Kilian, F. Leighton, and T. Shamoon, “Secure spread spectrum watermarking for multimedia,” IEEE Transaction on Image Processing, vol. 6, no. 12, December 1997.

[6] H. Inoue, A. Miyazaki, and T. Katsura, “An image watermarking method based on the wavelet transform,” in Proceedings of IEEE International Conference on Image Processing, Kobe, Japan, October 1999, pp. 296–300.

[7] M. Carli, C. Adsumilli, V. Razzauti, A. Neri – “Video watermarking in 3D DCT domain,” – Proc. IEEE 2005 International Workshop on Spectral Methods and Multirate Signal Processing, SMMSP 2005, Riga, Latvia, June 20-22, 2005.

[8] S. Pereira, S. Voloshynovskiy, M. Madueño, S. Marchand-Maillet and T. Pun, “Second generation benchmarking and application oriented evaluation”, In Information Hiding Workshop III, Pittsburgh, PA, USA, April 2001

[9] C.B. Adsumilli, M.C.Q. Farias, S.K. Mitra, and M. Carli, “A robust error concealment technique using data hiding for image and video transmission over lossy channels,” IEEE Transactions on Circuits and Systems for Video Technology, vol.15, no 11, Nov. 2005 Page(s):1394 - 1406

[10] F. A. P. Petitcolas, R. J. Anderson, M. G. Kuhn. “Attacks on copyright marking systems”, in David Aucsmith (Ed), Information Hiding, Second International Workshop, IH’98, Portland, Oregon, U.S.A., April 15-17, 1998, Proceedings, LNCS 1525, Springer-Verlag, ISBN 3-540-65386-4, pp. 219-239.

[11] F. A. P. Petitcolas, “Watermarking schemes evaluation.” I.E.E.E. Signal Processing, vol. 17, no. 5, pp. 58–64, September 2000.

[12] J. J. Eggers, R. Bauml, R. Tzschoppe, and B. Girod, “Scalar Costa Scheme for information embedding,” IEEE Transactions on Signal Processing, vol. 51, no. 4, pp. 1003–1019, April 2003.

[13] M. H. M. Costa, “Writing on dirty paper,” IEEE Transactions on Information Theory, vol. 29, no. 3, pp. 439–441. May 1983.