Blind Image Watermarking using Normalized STDM robust against Fixed Gain Attack

Makram W. Hatoum
FEMTO-ST Institute
Université of Bourgogne Franche-Comté
UMR 6174 CNRS, France
makram.hatoum@univ-fcomte.fr

Rony Darazi
TICKET Lab
Antonine University
Hadat-Baabda, Lebanon
rony.darazi@ua.edu.lb

Jean-François Couchot
FEMTO-ST Institute
Université of Bourgogne Franche-Comté
UMR 6174 CNRS, France
jean-francois.couchot@univ-fcomte.fr

***Abstract*—Spread Transform Dither Modulation (STDM), as an extension of Quantization Index Modulation (QIM) is a blind watermarking scheme that achieves high robustness against random noise and re-quantization attacks, with a limitation against the Fixed Gain Attack (FGA). In this paper, we improve the STDM watermarking scheme by making the quantization step size dependent on the watermarked content to resist the FGA attack. Simulations on real images show that our approach achieves strong robustness against the FGA attack, the Additive White Gaussian Noise (AWGN) attack, and the JPEG compression attack while preserving a higher level of transparency.**

***Index Terms*—Digital watermarking, QIM, STDM, Robustness, Transparency, FGA, AWGN, JPEG compression.**

I. Introduction

With the explosive growth of the internet and wireless networks, digital watermarking has received a great attention in research, which aims to protect digital contents like images [1], videos [2], audios [3], and PDF documents [4]. Several image watermarking schemes were performed in spatial domain, where the watermark is embedded directly in the pixel intensity values of an image, or in the frequency domain, where the watermark is embedded into the frequency domain of an image transform using the DCT, DFT, DWT or SVD [5], [6]. Watermarking schemes are generally classified as additive class known by Spread Spectrum (SS) schemes introduced by Cox et al. [7]–[9] and substitutive class known by Quantization Index Modulation (QIM) introduced by Chen and Wornell [10]. The basic QIM quantize a sample element to the nearest quantization point according to a message bit m. An important extension of QIM is the Spread Transform Dither Modulation (STDM), which embed a single bit of the message into a host-vector by quantizing the projection of the host-vector onto a random vector p. STDM combines the robustness of Spread Spectrum system and the effectiveness of QIM. Thus, it achieves higher robustness against additive noise and re-quantization attacks, but still largely vulnerable to the FGA attack. In this latter type of attack, the received signal is multiplied by a scaling factor ρ, which scales the watermark vector and shifts it away from its original quantization cell. Therefore, the decoder will be unable to correctly estimate the embedded message. However, Bartolini $et\ al.$ [11] analyzed the performance of STDM against the FGA attack and quantization attack. Based on their analysis, STDM has superior robustness against the quantization attack than the FGA attack, since even for values of a scaling factor ρ close to 1, the error probability becomes excessively high. Perez-Gonzalez $et\ al.$ [12] proposed the Rational Dither Modulation (RDM) to provide invariance to FGA attack. In RDM, the feature signal for quantization is constructed using the ratio of the previously generated watermarked sample and the current host sample. RDM achieves a better performance against the FGA attack, with a limitation against the additive noise. A number of solutions have been proposed using perceptual models [13]–[18] based on Watson's model [19] to improve the fidelity and provide robustness to FGA attack. Watson provides a perceptual model for computing the slack associated with each DCT coefficient within an 8×8 block, and those slacks are used to select the projection vector and/or to determine the quantization step size during the embedding and decoding process. However, this step could introduce a security problem, since that the projection vector could be selected easily by an opponent. Most of the proposed watermarking schemes are applied in the frequency domain and could only be implemented on images to resist the FGA attack, since that they are dependent on the luminance and contrast making of images.The robustness is examined in the DCT domain, $i.e.$ the DCT coefficients are quantized rather than the pixel values. In this paper, we modified the traditional STDM watermarking scheme and applied it on the grayscale images in the spatial domain and frequency domain to compare it with other proposed methods. Our method is more flexible and is not dependent on the perceptual model to achieve the robustness to the FGA attack.

This paper recalls some backgrounds on STDM in Section 2. The proposed N-STDM method is presented in Section 3. Experiments on real images are shown in Section 4. Finally, in Section 5 we give our conclusion and future work.

II. Background

STDM [10] is a special case of QIM, where the quantization occurs entirely in the projection of the host signal x onto a random projection vector p. By this way, the embedding-induced distortion is spreading into all groups of samples

instead of one.
The embedded function is given by:

$$\begin{aligned} y &= x + (Q_m(x^T p, \Delta) - x^T p)p \\ &= x + \left(round\left(\frac{x^T p - d_m}{\Delta}\right)\Delta + d_m - x^T p \right) p \end{aligned} \quad (1)$$

where Δ represents the quantization factor, $round()$ is the rounding value to the nearest integer, and d_m denotes the dither level based on the message bit $m \in \{0, 1\}$:

$$d_0 = -\frac{\Delta}{4} \text{ and } d_1 = \frac{\Delta}{4} \quad (2)$$

To extract the embedded message, the detection can be performed with a minimum distance decoder as the form:

$$\hat{m} = arg \min_{m \in \{0,1\}} \mid y^T p - Q_m(y^T p, \Delta) \mid \quad (3)$$

When the FGA attack is applied on the watermarked signal y, it becomes:

$$z = \rho \cdot y \quad (4)$$

and $\hat{m}$ will be decoded as follow:

$$\hat{m} = arg \min_{m \in \{0,1\}} \mid \rho \cdot y^T p - Q_m(\rho \cdot y^T p, \Delta) \mid \quad (5)$$

$$\begin{aligned} Q_m(\rho \cdot y^T p, \Delta) &\neq \rho \cdot Q_m(y^T p, \Delta) \\ Q_m(\rho \cdot y^T p, \Delta) &= round\left(\frac{\rho \cdot y^T p - d_m}{\Delta}\right)\Delta + d_m \\ \rho \cdot Q_m(y^T p, \Delta) &= round\left(\frac{y^T p - d_m}{\Delta}\right)\rho \cdot \Delta + \rho \cdot dm \end{aligned}$$

The decoded message can be thus different from the embedded one, and the robustness against the FGA attack cannot be obtained.

III. Proposed N-STDM Method

The traditional STDM scheme is affected by the FGA attack. Hence, we need to make the quantization step size dependent on the watermarked samples.
We normalized the STDM embedding function as:

$$y = x + \left(\|x\| Q_m\left(\frac{x^T p}{\|x\|}, \Delta\right) - x^T p \right) p \quad m \in \{0, 1\}. \quad (6)$$

$$\|x\| = (|x_1|^{\frac{1}{u}} + |x_2|^{\frac{1}{u}} + ... + |x_n|^{\frac{1}{u}})^u \quad (7)$$

where n is the length of the extracted vector from the cover elements, $|x_n|$ is the absolute value of element x_n, and $\|x\|$ is a norm function that could be expressed as a l^2-norm when $u = 1/2$ and l^1-norm when $u = 1$ etc. Section IV details the influence of $\|x\|$ against the FGA attack while varying u.
Thus, if we perform the FGA attack s.t. $z = \rho y$, then the modification is also applied to the minimum distance decoder as:

$$\begin{aligned} \hat{m} &= arg \min_{m \in \{0,1\}} \mid \tilde{z} - \|z\| Q_m\left(\frac{\tilde{z}}{\|z\|}, \Delta\right) \mid \\ &= arg \min_{m \in \{0,1\}} \mid \rho \cdot \tilde{y} - \rho \cdot \|y\| Q_m\left(\frac{\rho \cdot \tilde{y}}{\rho \cdot \|y\|}, \Delta\right) \mid \\ &= arg \min_{m \in \{0,1\}} \mid \rho \left(\tilde{y} - \|y\| Q_m\left(\frac{\tilde{y}}{\|y\|}, \Delta\right)\right) \mid \end{aligned} \quad (8)$$

Fig. 1: The original images (first and third columns) and corresponding watermarked images (second and fourth columns) using the N-STDM method for u=2 with a 4096-bit message embedded and PSNR=45 dB.

where:

$$\tilde{y} = y^T p \quad (9)$$

Therefore, the non-linear impact of the ρ factor in the Quantization is now linear and consequently will not affect the process of decoding the message.

IV. Experimentation and Simulation Results

The algorithm is parametrized by two variables, which are the u factor presented in (7) and the elements that are considered to compute the norm. The first experiments aimed at finding optimal values for these parameters with respect to the results against the FGA attack. We compared our proposed method against the FGA attack while varying u using two forms. In the first one (Global form), we compute the norm value $\|x\|$ of the whole pixels of the cover image which will be used to embed all the watermark bits. In the second one (Local form), we compute the norm value $\|x\|$ of each vector, in which we will embed the i^{th} bit of the watermark. Therefore, each bit will have a specific norm value. In the experiments, the 10 grayscale images with size 512×512 presented in Fig. 1 have been used as a host signal, the length of the projection

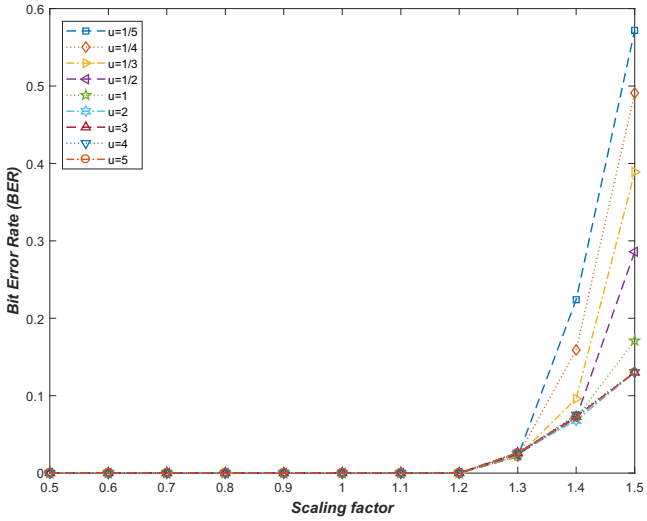


Fig. 2: Robustness of N-STDM (Global form) against Fixed Gain Attack in term of BER while varying u with PSNR=45 dB.

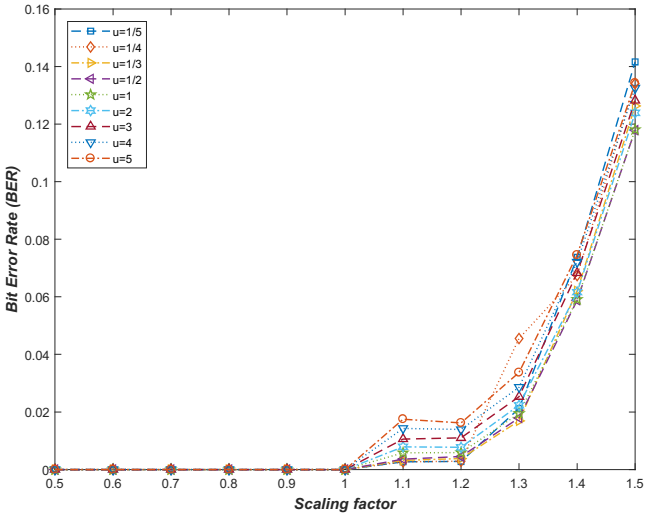


Fig. 3: Robustness of N-STDM (Local form) against Fixed Gain Attack in term of BER while varying u with PSNR=45 dB.

vector is set to 64, which allows a 4096-bit message to be embedded into each image, and all the parameters are adjusted to have watermarked images with the same Peak Signal-to-Noise Ratio (PSNR=45 dB).

Fig. 2 shows the robustness of N-STDM method against FGA attack when varying u between 1/5 and 5 using the global form. The BER decrease when u increases with preferable results when u is higher than 1.

Fig. 3 shows the robustness of N-STDM against FGA attack using the local form. In this situation, the N-STDM has a good performance whenever is the value of u. But in some cases, it still better to use the global form since the same norm value $\|x\|$ is used to embed all the bits of the watermark, and by

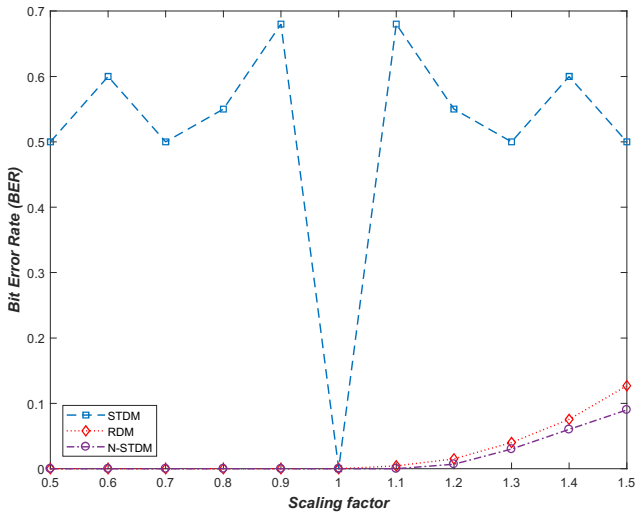


Fig. 4: Robustness against Fixed Gain Attack in term of BER with PSNR=45 dB.

this way, all the watermarked vectors in the image will have an identical imperceptibility. To conclude these experiments, the parameter values that provide the results with the lowest errors are u=2 and a global form for the norm. In the subsequent experiments, these values have been selected.

Next, we have evaluated the correction of the N-STDM watermarking scheme. In other words, we have verified that when a message is embedded into a host signal, and when there is no attack, the message is extracted without any error. To do so, 1000 grayscale images with size 512×512 extracted from Boss image database [20] are used as host signal. The length of the projection vector is set to 64, which allows a 4096-bit message to be embedded into each image. The quantization step size Δ is adjusted in order to have the same PSNR=45 dB for all the watermarked images. For all the images, the Bit-Error-Rate (BER) of the extracted watermark is always equal to 0, which leads us the convincing idea that any embedded message could be retrieved and without errors when the attacks are not applied.

After that, the visual aspect of the presented approach has been studied. The length of the projection vector is set to 64, which allows a 4096-bit message to be embedded into each image. Fig. 1 shows a part of grayscale images. The second and fourth columns display the obtained watermarked images using the N-STDM method with a PSNR=45 dB. These watermarked images look almost the same as the original ones, and it is impossible to distinguish them by human eyes.

A. Comparison in Spatial Domain

In this section, we compared the robustness in the spatial domain of our proposed approach with the traditional STDM watermarking scheme and RDM [12] against the FGA attack and AWGN attack. The experiments are conducted on the grayscale images of size 512×512. The length of the projection vector is set to 64 which allows a 4096-bit message to be

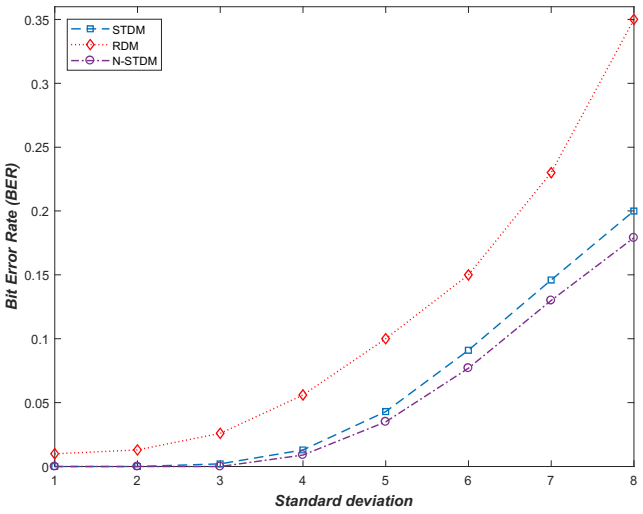


Fig. 5: Robustness against AWGN attack in term of BER with PSNR=45 dB.

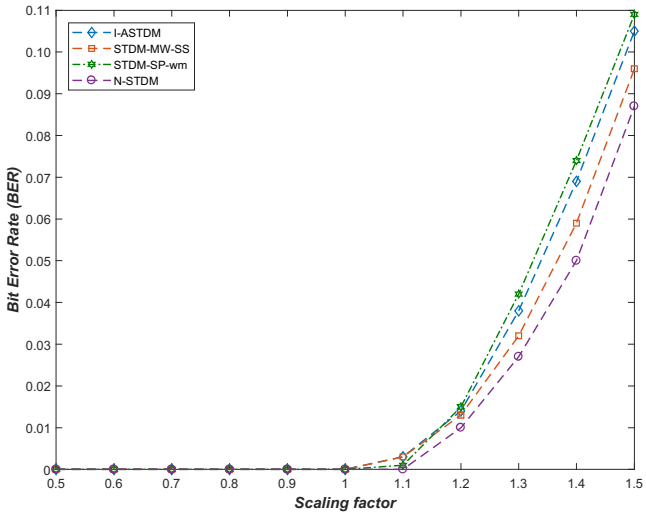
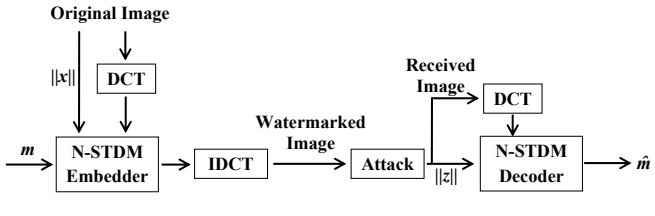
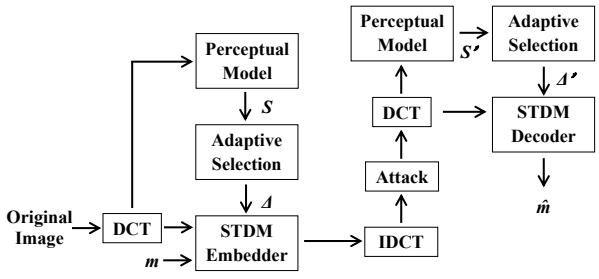


Fig. 7: Robustness against Fixed Gain Attack in term of BER with PSNR=45 dB.

embedded into each image and all the parameters are adjusted to have watermarked images with the same PSNR=45 dB.
As shown in Fig. 4, N-STDM and RDM have good robustness against the FGA attack. STDM has superior robustness against the quantization attack than the FGA attack, since even for values of a gain factor as close to 1, as 1.1 or 0.9, the BER becomes excessively high.
With RDM, the feature signal for quantization is constructed using the ratio of the previously generated watermarked sample and the current host sample. For that, it resists the FGA attack but will be affected by the AWGN attack as shown in Fig. 5. N-STDM make the quantization step size dependent on the watermarked samples, which will scale linearly with the FGA attack. Therefore, this method resists the FGA attack while preserving superior robustness against the AWGN attack.



(a) Block diagram of the N-STDM method.



(b) Block diagram of the family methods based on Watson's model.

Fig. 6: Block diagrams of the N-STDM method (a) and the family methods based on Watson's model (b).

B. Comparison in Frequency Domain

To test the performance of N-STDM in the frequency domain, we implement the DCT on the grayscale images of size 512×512 as shown in Fig. 6a. First of all, we compute the norm value $\|x\|$ of the original image to be used during the embedding process. After that, we divide the image into 8×8 blocks of pixels, through which we will perform the DCT transform to get the DCT coefficients. A part of these coefficients will be used as a host vector of length L, in which we will embed the i^{th} bit of the watermark message m. Then we perform the inverse DCT transform at each block to get the watermarked image. Our proposed scheme is compared with I-ASTDM [14], STDM-MW-SS [15] and STDM-SP-wm [17]; family methods with small variation related to the perceptual model. The block diagram of those family methods is shown in Fig. 6b, where the quantization step size Δ is modified based on the slacks vectors S computed using Watson's model.
Meanwhile, the FGA attack, AWGN attack, and JPEG compression are used to verify the performance of our proposed scheme. For all the algorithms we used the 2nd-21st DCT coefficients in zig-zag-scanned order of each 8×8 block, in which we embed 1 bit of the watermark. The embedding rate is 1/64, $i.e.$ one bit in each 8×8 block, which allows the embedding of a 4096-bit message into each image. The parameters of all methods are adjusted so that the PSNR value of all the watermarked images is equal to 45 dB. As expected, according to Fig. 7, all the proposed schemes have good robustness against FGA attack in the frequency domain. With the I-ASTDM watermarking scheme, Δ is modified based on

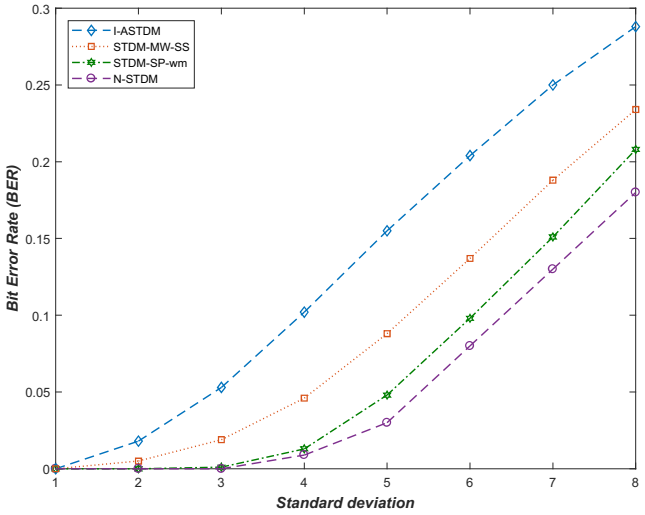


Fig. 8: Robustness against AWGN in term of BER with PSNR=45 dB.

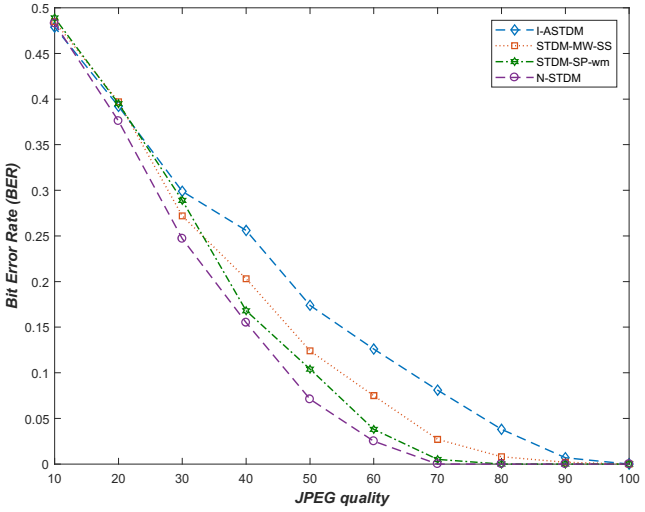


Fig. 9: Robustness against JPEG compression in term of BER with PSNR=45 dB.

Watson's model using the slacks of the contrast masking. By this way, Δ will scales linearly with the FGA attack. As for STDM-MW-SS watermarking scheme, the slack values of the images which are computed based on Watson's model are used as a projection vector, and they are used to modify the values of Δ. Therefore, Δ scales linearly with the FGA attack. In the STDM-SP-wm watermarking scheme, the slack vectors are computed based on Watson's model and projected onto a projection vector p to select Δ, which will scale linearly with the FGA attack. Therefore, those methods could only be applied to images and in the frequency domain based on DCT transform. In contrast, N-STDM makes the quantization step size dependent on the watermarked samples to resists the FGA attack and could be applied on images or any other element in the spatial and frequency domain.

The robustness against AWGN attack has been tested in term of BER while varying the standard deviation between 1 and 8. As shown in Fig. 8, our proposed N-STDM watermarking scheme achieve better performance, and the BER always tend to 0 when the standard deviation is lower than 5.

Fig. 9 illustrates the robustness to JPEG compression in term of BER while varying the JPEG quality between 10 and 100, which denotes the compressibility of the JPEG compressor; lower numbers mean lower quality. N-STDM have a better performance against the JPEG compression, and the BER tend to 0 when the JPEG quality is higher than 50.

V. Conclusion and Future Work

In this paper, we have presented a blind image watermarking using normalized STDM robust against FGA attack. Our proposed N-STDM scheme make the quantization step size dependent on the watermarked samples. Therefore, Δ will scale linearly with the FGA attack. We have applied our approach on the grayscale images in the spatial domain and frequency domain based on the DCT transform.

The modified STDM family methods could only be applied on images in the frequency domain to resist the FGA attack, since that the quantization step size has been adjusted based on the luminance and contrast masking of images. But our approach is more flexible, could be applied in the spatial or frequency domain, and any element can be used as support to contain the watermark.

The experimental results confirm that the N-STDM approach achieves the robustness against the FGA attack and also manifest good performance facing the AWGN attack and JPEG compression while preserving a higher level of transparency.

As for future enhancements, we plan to execute the theoretical proof of the practice evidence presented in Section IV, to include further improvement of the N-STDM watermarking scheme and apply it to other types of digital content such as PDF documents.

Acknowledgment

This work is partially funded with support from the National Council for Scientific Research in Lebanon CNRS-L, the Hubert Curien CEDRE programme, the Agence Universitaire de la Francophonie AUF-PCSI programme, and the Labex ACTION program (contract ANR-11-LABX-01-01).

References

[1] R. Darazi, P. Callau, and B. Macq, "Secure and hvs-adaptive exhibition spread transform dither modulation watermarking for digital cinema," in *2009 First IEEE International Workshop on Information Forensics and Security (WIFS)*, Dec 2009, pp. 1–5.

[2] A. Mansouri, A. M. Aznaveh, F. Torkamani-Azar, and F. Kurugollu, "A low complexity video watermarking in h. 264 compressed domain," *IEEE Transactions on Information Forensics and Security*, vol. 5, no. 4, pp. 649–657, 2010.

[3] V. Bhat, I. Sengupta, and A. Das, "An adaptive audio watermarking based on the singular value decomposition in the wavelet domain," *Digital Signal Processing*, vol. 20, no. 6, pp. 1547–1558, 2010.

[4] A. W. Bitar, R. Darazi, J.-F. Couchot, and R. Couturier, "Blind digital watermarking in pdf documents using spread transform dither modulation," *Multimedia Tools and Applications*, vol. 76, no. 1, pp. 143–161, 2017.

[5] C.-C. Lai and C.-C. Tsai, "Digital image watermarking using discrete wavelet transform and singular value decomposition," *IEEE Transactions on instrumentation and measurement*, vol. 59, no. 11, pp. 3060–3063, 2010.

[6] A. K. Singh, M. Dave, and A. Mohan, "Hybrid technique for robust and imperceptible image watermarking in dwt–dct–svd domain," *National Academy Science Letters*, vol. 37, no. 4, pp. 351–358, 2014.

[7] I. J. Cox, J. Kilian, T. Leighton, and T. Shamoon, "Secure spread spectrum watermarking for images, audio and video," in *Image Processing, 1996. Proceedings., International Conference on*, vol. 3. IEEE, 1996, pp. 243–246.

[8] I. J. Cox, J. Kilian, F. T. Leighton, and T. Shamoon, "Secure spread spectrum watermarking for multimedia," *IEEE transactions on image processing*, vol. 6, no. 12, pp. 1673–1687, 1997.

[9] H. S. Malvar and D. A. Florêncio, "Improved spread spectrum: A new modulation technique for robust watermarking," *IEEE transactions on signal processing*, vol. 51, no. 4, pp. 898–905, 2003.

[10] B. Chen and G. W. Wornell, "Quantization index modulation: a class of provably good methods for digital watermarking and information embedding," *IEEE Transactions on Information Theory*, vol. 47, no. 4, pp. 1423–1443, May 2001.

[11] F. Bartolini, M. Barni, and A. Piva, "Performance analysis of st-dm watermarking in presence of nonadditive attacks," *IEEE Transactions on Signal Processing*, vol. 52, no. 10, pp. 2965–2974, 2004.

[12] F. Pérez-González, C. Mosquera, M. Barni, and A. Abrardo, "Rational dither modulation: A high-rate data-hiding method invariant to gain attacks," *IEEE Transactions on Signal Processing*, vol. 53, no. 10, pp. 3960–3975, 2005.

[13] Q. Li and I. J. Cox, "Using perceptual models to improve fidelity and provide invariance to valumetric scaling for quantization index modulation watermarking," in *Acoustics, Speech, and Signal Processing, 2005. Proceedings.(ICASSP'05). IEEE International Conference on*, vol. 2. IEEE, 2005, pp. ii–1.

[14] X. Zhu, "Image-adaptive spread transform dither modulation using human visual model," in *Computational Intelligence and Security, 2006 International Conference on*, vol. 2. IEEE, 2006, pp. 1571–1574.

[15] Q. Li and I. J. Cox, "Improved spread transform dither modulation using a perceptual model: robustness to amplitude scaling and jpeg compression," in *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on*, vol. 2. IEEE, 2007, pp. II–185.

[16] D. Yu, L. Ma, G. Wang, and H. Lu, "Adaptive spread-transform dither modulation using an improved luminance-masked threshold," in *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on*. IEEE, 2008, pp. 449–452.

[17] X. Li, J. Liu, J. Sun, X. Yang, and W. Liu, "Step-projection-based spread transform dither modulation," *IET information security*, vol. 5, no. 3, pp. 170–180, 2011.

[18] W. Wan, J. Liu, J. Sun, X. Yang, X. Nie, and F. Wang, "Logarithmic spread-transform dither modulation watermarking based on perceptual model," in *Image Processing (ICIP), 2013 20th IEEE International Conference on*. IEEE, 2013, pp. 4522–4526.

[19] A. B. Watson, "Dct quantization matrices visually optimized for individual images," in *Human vision, visual processing, and digital display IV*, vol. 1913. International Society for Optics and Photonics, 1993, pp. 202–217.

[20] T. Filler, T. Pevný, S. Craver, and A. D. Ker, Eds., *Information Hiding - 13th International Conference, IH 2011, Prague, Czech Republic, May 18-20, 2011, Revised Selected Papers*, ser. Lecture Notes in Computer Science, vol. 6958. Springer, 2011. [Online]. Available: https://doi.org/10.1007/978-3-642-24178-9