On the Security of the Schur-Based Watermarking Schemes

On the Security of the Schur-Based Watermarking Schemes

Victor Pomponiu
Information Systems Technology and Design
Singapore University of Technology and Design
Singapore
victor_pomponiu@sutd.edu.sg

Davide Cavagnino, Marco Botta
Dipartimento di Informatica
Universita degli Studi di Torino
Torino
{davide.cavagnino, marco.botta}@unito.it

Abstract—**This paper presents a security analysis of robust watermarking methods based on Schur decomposition in a general scenario. The security is defined as the difficulty to remove the watermark and to estimate the secrets used in the embedding process, supposing that the adversary possesses several watermarked digital contents. The theoretical analysis and extensive experimental results carried out prove that these schemes fail to secure the digital contents against malicious attacks.**

Keywords—ambiguity problem; copyright protection; robust watermarking; Schur decomposition; Singular value decomposition (SVD); watermarking security;

I. INTRODUCTION

The Internet technologies and mobile services have significantly increased in the last decade leading to the problem of rightful ownership and copyright protection. Digital watermarking, which aims at addressing these concerns, is the process of embedding a secret mark into the digital content while preserving its fidelity [1]. For this purpose the hidden mark is called robust. Instead, a fragile mark is used for integrity and authentication assessment. Independently of their type, the marks can be detected (recovered) directly from the watermarked content and in some cases the host media is also necessary.

Watermarking algorithms can be divided into two categories, namely those that work in the pixel domain [2, 3] and those that work in a transform domain, e.g., the Karhunen-Loève transform (KLT) [4], the Singular value decomposition (SVD) [5-12], etc. The Schur decomposition [13-18] is another domain that has been recently used for robust watermarking. These schemes show good performances in terms of robustness, capacity and imperceptibility. However, to our knowledge, the security of these schemes has not been investigated yet.

In the last few years, watermarking security has become a new research field [19-24], which implies the existence of an adversary that wants to circumvent watermarking. The security analysis adopted follows a cryptanalytic approach: all of the parameters of the watermarking scheme are assumed to be public while the security relies only on a secret key which will be used for watermarking several digital contents.

The key technical contributions of this study can be summarized as follows:

- A security analysis of robust data-hiding schemes based on Schur decomposition is carried out.
- An attack is devised which takes into consideration the amount of information available to the attacker. In the first instance, it aims to remove the watermark by assuming that the block size is known, while in the second scenario the attack drops this assumption and tries to estimate it.

The outline of this paper is as follows: in the next section, the Schur decomposition is introduced while in Section 3 we review the most representative Schur-based watermarking algorithms. Section 4 presents several attack scenarios, while experimental results and concluding remarks are given in Sections 5 and 6, respectively.

II. SCHUR DECOMPOSITION

The Schur decomposition is a factorization method applied to square matrixes similar to the singular value decomposition (SVD). Given a real square matrix $A^{n\times n}$ of size $n\times n$ the Schur factorization of A is given by:

$$Schur(A) = U^t S U \tag{1}$$

where $S^{n\times n}$ is an upper triangular matrix with the eigenvalues of A along the diagonal and $U^{n\times n}$ is a unitary matrix, that means $U \cdot U^t = I^{n\times n}$, with its columns represented by the Schur vectors. If A is a positive definite matrix, its Schur decomposition and its singular value decomposition coincide.

Compared to the SVD transform, the use of Schur decomposition for data hiding is limited since a Schur factorization exists only for square matrices. An interesting property of the Schur decomposition that is extensively exploited by the watermarking schemes is the 'stability' of the Schur vectors [13]. In particular, the relationship between the coefficients of the first Schur vector of matrix U is to have closely related values, that are able to survive to common signal processing attacks [13-16]. Furthermore, the coefficients can be perturbed without significantly damaging the content fidelity. The genesis of this idea comes from several observations [8-10] that hold for the SVD transform.

However, this assumption was tested on a small dataset of natural images [13] that is insufficient to draw any statistically relevant conclusion. For instance, in order to test this assumption

TABLE I. THE AVERAGE NCC VALUE BETWEEN THE FOUR COEFFICIENTS OF THE FIRST SCHUR VECTOR FOR THE UCID IMAGE DATABASE. THE VALUES REPORTED IN [13] ARE THE AVERAGE VALUES OF TEN IMAGES

Metric	$NCC_{u_{1,1},u_{2,1}}$	$NCC_{u_{1,1},u_{3,1}}$	$NCC_{u_{1,1},u_{4,1}}$	$NCC_{u_{2,1},u_{3,1}}$	$NCC_{u_{2,1},u_{4,1}}$	$NCC_{u_{3,1},u_{4,1}}$
UCID	0.9868	0.9720	0.9624	0.9996	0.9760	0.9903
[13]	0.8931	0.8771	0.8754	0.9672	0.9466	0.9575

we selected 1338 grayscale images, taken from the UCID collection [25], and split them into non-overlapping blocks of size 4×4 pixels. Then, for each block we computed the normalized cross-correlation (*NCC*) between the coefficients of the first Schur vector. The average results obtained, together with the one reported in [13], are given in Table I. Similar to what has been reported in [13], i.e., "… $u_{2,1}$ and $u_{3,1}$ are the closest elements" of the first Schur vector, our results clearly confirm that the highest correlation obtained was obtained between $u_{2,1}$ and $u_{3,1}$ for the UCID datasets.

III. SCHUR-BASED WATERMARKING

Recently, several Schur-based image watermarking methods for copyright protection have been proposed [13-18]. Depending on the Schur features selected to store the watermark, these methods can be divided into two categories:

- Eigenvalue-based, which insert the watermark into the diagonal elements of the matrix S using an additively (multiplicatively) rule or quantization-based [14-18].
- Vector-based, that hide the watermark bits into the Schur vectors [13].

In general, the former schemes require the original host content in order to detect the watermark, which limit their applicability. Moreover, their security issues are similar to those of the SVD-based schemes which have been extensively studied in [5].

Instead, the latter schemes do not require any additional information to detect the watermark and, to our knowledge, their security has not been investigated yet. For instance, in [13] each color channel (i.e., RGB) of the host image is split into non-overlapping blocks of size 4×4 followed by the Schur decomposition. Then, in order to insert the watermark bits several coefficients of the first Schur vector are modified using the following embedding rule:

$$\begin{cases} u_{2,1}^{w} = \mathcal{S}(u_{2,1})\left(u_{avg} + \frac{T}{2}\right), b_{i,j} = 1 \\ u_{3,1}^{w} = \mathcal{S}(u_{3,1})\left(u_{avg} + \frac{T}{2}\right), b_{i,j} = 0 \\ u_{2,1}^{w} = \mathcal{S}(u_{2,1})\left(u_{avg} - \frac{T}{2}\right), b_{i,j} = 0 \\ u_{3,1}^{w} = \mathcal{S}(u_{3,1})\left(u_{avg} - \frac{T}{2}\right), b_{i,j} = 1 \\ u_{avg} = \frac{|u_{2,1}|+|u_{3,1}|}{2} \end{cases} \quad (2)$$

where $u_{2,1}$, $u_{3,1}$ are the original coefficients and $u_{2,1}^{w}$, $u_{3,1}^{w}$ are the watermarked ones, $\mathcal{S}(\cdot)$ is the sign operator, $|\cdot|$ is the modulus, $b_{i,j}$ represents the bit of the encrypted watermark w, T denotes the embedding threshold and u_{avg} is the average value of $u_{2,1}$ and $u_{3,1}$. Basically, depending on the encrypted watermark bit, the values of $u_{2,1}$, and $u_{3,1}$ are modified such that their difference is exactly T. Higher values of T yield more reliable detection, less security, and potential watermark visibility. The color watermark image encryption is done by shuffling its pixels with the Arnold transform followed by the binarization operation. The number of shuffling rounds is given by a secret key k. Note that this procedure does not increase the security of the embedded bits, i.e., making them more difficult to remove.

Each encrypted watermark bit is recovered, without the aid of any auxiliary information, using the following extraction rule:

$$\begin{aligned} \tilde{b} &= 0, \ if \ \ u_{2,1}^{\tilde{w}} \leqslant u_{3,1}^{\tilde{w}} \\ \tilde{b} &= 1, \ if \ \ u_{2,1}^{\tilde{w}} > u_{3,1}^{\tilde{w}} \end{aligned} \quad (3)$$

where $\tilde{b}$ denotes the recovered bit of the extracted encrypted watermark $\tilde{w}$. In order to detect the presence of the watermark in an image, the extracted watermark is compared with the original one.

IV. THE PROPOSED ATTACK

In general, the common signal processing attacks aim to remove the watermark by globally perturbing the watermarked content. These attacks do not take into account the details of the watermarking method. Instead, the proposed attack tries to exploit the details of the watermarking method based on Schur decomposition and to use them to perturb only the features conveying the watermarking bits.

To achieve this goal we need to apply a specific perturbation to the coefficients of the Schur vectors. The unknown parameters of [13] are the embedding threshold T, the secret key k, the block size and the selected coefficients of the Schur decomposition.

Algorithm 1. Remove the hidden bits from the watermarked image I_w.

Input: I_w, the block size and the selected coefficients of the first Schur vector in each block.

Output: Attacked image, I_{wa}, which does not contain the watermark.

1) Split I_w of size n×n pixels into non-overlapping blocks of size 4×4 pixels. The total number of blocks is $nb = n/4 \times n/4$.
2) FOR i = 1 to nb
3) Apply Schur decomposition the block $B_w(i)$
4) $Schur[B_w(i)] = U_w^t S_w U_w$

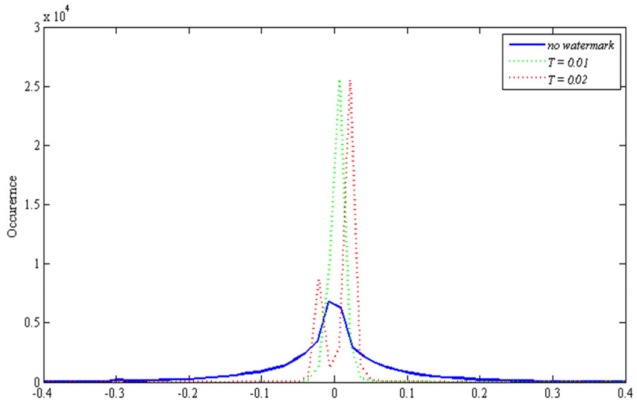

Figure 1. Example of the embedding threshold T can be detected during the searching process. The histogram of the absolute difference, D_w, between the coefficients $u^w_{2,1}$ and $u^w_{3,1}$ for different cases: no watermark (blue solid line), and inserted watermark with T = 0.01 (green dotted line) and T = 0.02 (red dotted line). For a better visualization we restricted the values of D_w to the interval [–0.4, 0.4]. The block size was of 4×4 pixels.

5) Obtain the watermarked Schur coefficients $(u^w_{2,1}, u^w_{3,1})$
6) IF $(u^w_{2,1} > u^w_{3,1})$
7) $u^{wa}_{2,1} = u^w_{3,1}, u^{wa}_{3,1} = u^w_{3,1}$
8) ELSE
9) $u^{wa}_{3,1} = u^w_{2,1}, u^{wa}_{2,1} = u^w_{2,1}$
10) END IF
11) Reconstruct the attacked image block:
12) $B_{wa}(i) = U_{wa} S_{wa} U^t_{wa}$
13) END FOR
14) Reconstruct I_{wa} by combining all attacked blocks.

The attack rule (Steps 6 - 9) aims to perturb one of the coefficients such their difference becomes equal to zero. In this way, the Su et al. [13] extractor will wrongly classify most of the hidden bits, failing to recover the watermark.

If the block size is not available during the attack we can estimate it through a searching process. It is worth to point out that the embedding rule given in Eq. (2) does not preserve the unitary property of the Schur vector, i.e. $U \cdot U^t = I^{n \times n}$. This is one of the major causes of range overflow and of the inability to perfectly extract the embedded watermark even when the host image has not undergone any perturbations.

Another consequence of Eq. (2) is that it also affects the statistical distribution of the Schur coefficients. For the "clean" coefficients their values are in the range [–1, 1] and the difference between a pair of coefficients can be approximated by a normal distribution, i.e., $N(m, \sigma)$, with m = 0.0049 and σ = 0.1126. Instead, the distribution of the difference between the watermarked coefficients defined as:

$$D_w = |u_{2,1}| - |u_{3,1}| \tag{4}$$

has two peaks, one for each of the watermark possible values, {0, 1}.

Therefore, for each block of I_w, by computing D_w we can figure out the block size and the chosen coefficients used during the watermarking process. We tested this assumption for all the watermarked images of the UCID database, with the block size of 4×4 pixels. In Fig. 1 we report three cases: D_w for no-watermark (i.e., difference between "clean" coefficients), and D_w for the watermark inserted using T = 0.01 and T = 0.02.

It is interesting to note that for low values of T (that increase the invisibility of the watermark) the distribution of D_w does not follow the same pattern and we fail to estimate the block size and the selected coefficients. However, lower values of T do not provide a reliable extraction and robustness against signal processing attacks forcing the content protector to increase the embedding threshold.

The last important observation that we can infer from Fig. 1 is that also T can be accurately estimated from the distribution of D_w: the two peaks are centered on very close to the values -T and T, respectively, depending on the inserted bit. For instance, for the case of the watermark inserted with T = 0.02, the peaks are centered on the values –0.0197 and 0.0197.

V. Simulation Results

This section presents the experimental setup and the results of the proposed attack against the Su et al. [13] watermarking scheme. The 1338 images used in the experiments are all 24bpp color images of size 512×512 pixels taken from the UCID database [25]. For the watermark, we used a logo 24bpp color image of size 128×128 pixels. The quality of the watermarked image is measured using Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity index (SSIM) [26]. The structural index, which is calculated on various windows of an image, better assess the level of the distortion and of the quality perceived by a human. The SSIM index is a real value between –1 and 1, being 1 for two identical images. The similarity between the original watermark, w, and the recovered watermark, w_r, is measured by means of normalized cross-correlation NCC defined as:

$$\begin{aligned} NCC_{w,w_r} &= \frac{\sum_z^3 \sum_i^n \sum_j^n G_w \cdot G_{w_r}}{\sqrt{\sum_z^3 \sum_i^n \sum_j^n G_w^2} \cdot \sqrt{\sum_z^3 \sum_i^n \sum_j^n G_{w_r}^2}} \\ G_w &= w(i,j,z) - \frac{1}{3 \cdot n^2} \cdot \sum_z^3 \sum_i^n \sum_j^n w\,(i,j,z) \\ G_{w_r} &= w_r(i,j,z) - \frac{1}{3 \cdot n^2} \cdot \sum_z^3 \sum_i^n \sum_j^n w_r\,(i,j,z) \end{aligned} \tag{4}$$

The parameters of Su et al. [13] watermarking scheme were set to the following values: T = {0.004, 0.01, 0.02}, block size 4×4 pixels, and k = 3. Increasing the embedding threshold T lowers the quality of the watermarked images but increases the ability to recover the inserted watermark.

It is worth to point out that the scheme is unable to completely recover the watermark. This is mainly due to rounding errors that occurs when reconstructing the watermarked image from the Schur domain to pixel domain [27]. We choose to apply the proposed attack on the watermarked images obtained with the embedding threshold T = 0.01, since it gives the best trade-off between the invisibility, robustness and security of Su et al. [13] scheme. The results for NCC and, the PSNR and SSIM between the host/watermarked image and the attacked ones are presented in Table II.

TABLE II. THE VALUES OF SSIM, PSNR AND *NCC* AFTER EMPLOYING THE PROPOSED ATTACK ON "LENA" AND "BABOON" IMAGES, AND ON UCID DATABASE. FOR THE UCID DATABASE THE AVERAGE VALUES, TOGETHER WITH THEIR ASSOCIATED STANDARD DEVIATIONS, ARE REPORTED. WATERMARKED IMAGES ARE OBTAINED WITH THE EMBEDDING THRESHOLD $T = 0.01$. I_{WA} IS THE ATTACKED IMAGE.

Metric	***Lena***	***Baboon***	***UCID***
$SSIM_{Io,Iw}$	0.982	0.940	0.978±0.086
$SSIM_{Io,Iwa}$	0.989	0.946	0.097±0.093
$SSIM_{Iw,Iwa}$	0.991	0996	0.098±0.096
$PSNR_{Io,Iw}$ *(dB)*	37.66	27.57	33.15±3.2
$PSNR_{Io,Iwa}$ *(dB)*	37.32	27.58	33.06±3.3
$PSNR_{Iw,Iwa}$*(dB)*	41.68	40.94	35.97±3.2
NCC_{w,w_r}	-0.018	-0.007	-0.002±0.005

Although the host image is unavailable during the attack, we use it to better assess the invisibility of the attacked image. For the "Lena" and "Baboon" images the attack generated images with high quality, i.e., PNSR above 40 dB and SSIM above 0.990, while completely removing the inserted watermark, i.e., *NCC* below 0.02.

Similar results have been obtained for the images belonging to UCID database. The average PSNR and SSIM values for the attacked images were lower since it was mounted on the watermarked images with a lower quality. This is due to the fact that Su et al. [13] embedding threshold is fixed and the embedding algorithm does not take into consideration the human visual system [1]. For the UCID database, the maximum values of SSIM, PSNR and *NCC* obtained were 0.997, 42.68 dB and 0.0245, respectively.

The PSNR between the watermarked image and the attacked image is 41.68 dB for the "Lena" image and 40.94 dB for the "Baboon" image. It means that the quality of the attacked image is very good leaving no traces of artifacts visible to human visual inspection.

VI. CONCLUSION

In this paper an attack to the watermarking schemes based on Schur decomposition is designed. In particular, the attack has been applied on a recently proposed algorithm [13]. We successfully removed the watermark by exploiting the traces left by the embedding rule.

Specifically, these traces enable us to identify, from several watermarked contents, the secret parameters of the watermarking scheme, i.e., the embedding threshold, the block size and the selected features of the Schur decomposition. The extensive tests carried out on the UCID database prove the efficacy of the attack to generate high quality unmarked images. Therefore, the attack proves that the watermarking algorithms based on the Schur decomposition, cannot be used to protect the copyright of digital contents.

REFERENCES

[1] I. J. Cox, M. L. Miller, J. A. Bloom, J. Fridrich, T. Kalker. Digital Watermarking and Steganography, 2nd edition. Morgan Kaufmann Publisher, San Francisco, 2008.

[2] P.-Y. Lin, J.-S. Lee, C.-C. Chang. Protecting the content integrity of digital imagery with fidelity preservation. ACM Tran. on Mult, Comp. Com. and App., 7 (3) (2011) 1.

[3] H. He, F. Chen, H.-M. Tai, T. Kalker, J. Zhang. Performance Analysis of a Block-Neighborhood-Based Self-Recovery Fragile Watermarking Scheme. IEEE Tran. on Inf. For. and Sec. 7 (1) (2012) 185.

[4] M. Botta, D. Cavagnino, V. Pomponiu. KL-F: Karhunen-Loève Based Fragile Watermarking. In 5th NSS, 2011, p. 65.

[5] V. Pomponiu, D. Cavagnino, A. Basso, A. Vernone. Data hiding schemes based on Singular Value Decomposition. Advanced Techniques in Multimedia Watermarking: Image, Video and Audio Applications, 2010, p. 254.

[6] R. Gunjan, P. Mitra, M. S. Gaur. Contourlet Based Image Watermarking Scheme Using Schur Factorization and SVD. Advances in Communication, Network, and Computing, LNCS 108, 2012, p. 337.

[7] C.-C. Lai, An improved SVD-based watermarking scheme using human visual characteristics. Optics Communications 284 (4) (2011) 938.

[8]] C.-C. Chang, P. Tsai, C.-C. Lin. SVD-based Digital Image Watermarking scheme. Pattern Recognition Letters 26 (10) (2005) 1577.

[9] K.-L. Chung, W.-N. Yang, Y.-H. Huang, S.-T. Wu, Y.-C. Hsu. On SVD-based watermarking algorithm. Applied Mathematics and Computation 188 (1) (2007) 54-57.

[10] J. C. Patra, W. Soh, E. L. Ang, P. K. Meher. An Improved SVD-Based Watermarking Technique for Image and Document Authentication. IEEE Conference on Circuits and Systems, 2006, p. 1984.

[11] M.-Q. Fan, H.-X. Wan, S.-K. Li. Restudy on SVD-based watermarking scheme. Applied Mathematics and Computation 203 (2) (2008) 926.

[12] J. Wang, R. Healy, J. Timoney. A robust audio watermarking scheme based on reduced singular value decomposition and distortion removal. Signal Processing 91(8) (2011) 1693.

[13] Q. Su, Y. Niu, X. Liu, Y. Zhu. Embedding color watermarks in color images based on Schur decomposition. Optics Com 285 (7) (2012) 1792.

[14] A. A. Mohammad. A new digital image watermarking scheme based on Schur decomposition. Mult. Tools and App. 59 (3) (2012) 851.

[15] B. C. Mohan, K. V. Swamy. On the use of Schur Decomposition for Copyright Protection of Digital Images. Intl. J. of Comp. and Elec. Eng. 2 (4) (2010) 1793.

[16] A. Choudhary, S. P. S. Chauhan, M. A. Alam, S. Tanveer. Schur decomposition and dither modulation: an efficient and robust audio watermarking technique. CUBE, 2012, p. 744.

[17] C. B. Mohan, V.K. Swamy, S. S. A. Kumar. Comparative performance evaluation of SVD and Schur Decompositions for Image Watermarking. Int. Conf. on VLSI, Comm. and Instr., 2011, 14, p. 25.

[18] P. Liu, J. Yang, J. Wei, F. Chen. A Novel Watermarking Scheme in Contourlet Domain Based on Schur Factorization. ICIEC, 2010, p. 1.

[19] C.-C. Chang, Y.-H. Fan, W.-L. Tai. Four-scanning attack on hierarchical digital watermarking method for image tamper detection and recovery. Pattern Recognition, 41 (2) (2008) 654.

[20] F. Cayre, C. Fontaine, T. Furon. Watermarking security: Theory and practice. IEEE Transactions on Signal Processing 53 (10) (2005) 3976.

[21] L. P. Freire, P. Comesaña, J. R. T. Pastoriza, F. González. Watermarking security: a survey. Transactions on Data Hiding and Multimedia Security I LNCS 4300 (2006) 41.

[22] M. Barni, F. Bartolini, T. Furon. A general framework for robust watermarking security. Signal Processing 83 (10) (2003) 2069-2084.

[23] V. Pomponiu, D. Cavagnino. Security analysis of SVD-based watermarking techniques. I. J. of Mult. Int. Sec. 2 (2) (2011) 120-145.

[24] W. Chen, C. Quan, C. J. Tay. Optical color image encryption based on Arnold transform and interference method. Optics Communications 282 (18) (2009) 3680.

[25] G. Schaefer, M. Stich. UCID - An Uncompressed Color Image Database. SPIE, SRMAM, 2004, p. 472.

[26] Z. Wang, A. C. Bovik. Mean squared error: love it or leave it? A new look at signal fidelity measures. IEEE Sig. Proc. Mag. 26 (1) (2009) 98.

[27] F. Y. Shih, Y.-T. Wu. Robust watermarking and compression for medical images based on genetic algorithms. Information Sciences 175 (3) (2005) 200.