Journal of Information Hiding and Multimedia Signal Processing ©2012 ISSN 2073-4212
Ubiquitous International Volume **3**, Number **2**, April **2012**

On the Security of Digital Watermarking Scheme Based on SVD and Tiny-GA

Khaled Loukhaoukha

Laval University, Quebec, QC, Canada, G1K 7P4
khaled.loukhaoukha.1@ulaval.ca

Received February 2011; revised April 2012

Abstract. *In the recent paper entitled A digital watermarking scheme based on singular value decomposition and tiny genetic algorithm by C.-C. Lai, a robust digital image watermarking scheme based on singular value decomposition (SVD) and a tiny genetic algorithm (Tiny-GA) is proposed. In this paper, we demonstrates that this watermarking algorithm is fundamentally flawed and has a very high probability of false positive detection of watermarks.*
Keywords: Digital watermarking, false positive detection, singular value decomposition and genetic algorithm.

1. **Introduction.** The end of 20^{th} century has been marked by an extraordinary technical revolution from analog to digital. However, the advantages of the digital revolution were not achieved without drawbacks such as generating anxiety in terms of duplication and modification of digital documents. Because of this, researchers were motivated more than ever to protect multimedia documents by new techniques. In this context, digital watermarking was introduced in order to guarantee the ownership and the integrity of digital documents by embedding a watermarks into these documents.

According to the domain in which the watermark is embedded, the watermarking algorithms proposed in literature are mainly grouped into two classes: spatial and transform domains. In the spatial domain, the pixel values of the original cover image are modified to embed the watermark. On the other hand, for transform domain, the watermark is embedded by transforming the original cover image into a set of a frequency domain coefficients according to the watermark. Most of the transforms used in watermarking techniques are based on discrete Fourier transform (DFT), discrete cosine transform (DCT) and the discrete wavelet transform (DWT), where the watermark is embedded in the frequency (DFT, DCT or DWT) coefficients. However, decomposition of images in terms of a standard basis set in the frequency domain is not necessarily the optimal representation for an image. Therefore, other transform representations were explored for watermarking using linear algebra methods as such as singular value decomposition (SVD) based methods [1, 2, 3]. Furthermore, it has been shown that SVD-based watermarking algorithms are very robust against a wide range of attacks. Unfortunately, as reported in literature by Zhang et al. [4, 5], Rykaczewski [6] and Xiao et al. [7], Loukhaoukha and Chouinard [8] some SVD-based watermarking algorithms are non-effective and do exhibit prohibitively high probabilities of false positive detections.

In this paper, we show that the watermarking algorithm proposed in [9] suffers from a high probability of false positive detection. After having defined the singular value decomposition in Section 2, the embedding and extracting procedures of this watermarking algorithm [9] is presented in Section 3. In Section 4, an analysis is conducted to show that this watermarking scheme is non-effective in terms of probability of false positive detections of watermarks. Illustrative SVD-based watermarking examples to highlight the false positive detection vulnerability are exposed in Section 5. Concluding remarks are given in Section 6.

2. **Singular value decomposition.** The theory of singular value decomposition (SVD) was established for real square matrices in the 1870s by Beltrami [10] and Jordan [11], for complex matrices by Autonne in 1902 [12] and has been extended to rectangular matrices by Eckart and Young [13] in the 1939. Recently, singular value decomposition has been used in image processing applications, including image compression [14], image hiding [15] and noise reduction [16].

Let I be an image matrix of size N N. It can be represented using singular value decomposition as:

$$I = U \cdot S \cdot V^T = \sum_{k=1}^{N} u_k \cdot s_k \cdot v_k^T \tag{1}$$

with $U = [u_1, u_2, ..., u_N]$, $V = [v_1, v_2, ..., v_N]$, and

$$S = \begin{bmatrix} s_1 & 0 & \cdots & 0 \\ 0 & s_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & s_N \end{bmatrix}$$

Here, U and V are orthogonal matrices of size $N \times N$, whose column vectors are the left-singular and the right-singular vectors, respectively. S is an $N \times N$ diagonal matrix containing nonnegative terms. The diagonal elements $s_1, s_2, ..., s_N$ of matrix S are the singular values of matrix I, satisfying the ordering: $s_1 \geq s_2 ... \geq s_N$.

It is important to note that:

- Singular values correspond to the luminance of the image (i.e, image brightness) and the corresponding singular vectors specifies the intrinsic geometry properties of the image [2].
- Many singular values have small values compared to the first singular value s1. If these small singular values are ignored in the reconstruction of the image, the quality of the reconstructed image will degrade only slightly [2].
- A slight variation of the singular values do not affect the visual perception of the image, i.e., singular values do have a good stability.

3. **SVD-based watermarking.** In this section, we present the watermarking algorithm based on singular value decomposition and the tiny genetic algorithm (Tiny-GA) proposed by Lai [9]. The Tiny-GA is used to obtain the optimal scaling factors (SFs) for watermark embedding with respect the two contradictory requirements : imperceptibility and robustness. As will be seen later in Section 4, this algorithm suffer from a high probability of false positive detection of the watermarks.

3.1. **Embedding process.**

1. SVD is applied to the host image I:
$$I = U \cdot S \cdot V^H \tag{2}$$
2. The watermark W is multiplied by a scaling factor a and added to matrix S:
$$D = S + a \cdot W \tag{3}$$
3. SVD is then applied to the new matrix D
$$D = U_W \cdot S_W \cdot V_W^T \tag{4}$$
4. The watermarked image is computed as:
$$I_W = U \cdot S_W \cdot V^T \tag{5}$$

3.2. **Extracting process.** Essentially, the extracting process is the inverse of the embedding process. In watermark extracting process, a *possibly distorted* watermark W^* is extracted from the *possibly distorted* watermarked image I_W^* by essentially reversing the above watermark embedding process.

1. SVD is performed on the (possibly distorted) watermarked image.
$$I_W^* = U^* \cdot S_W^* \cdot V^{*T} \tag{6}$$
2. The (possibly corrupted) matrix D^* is computed as:
$$D^* = U_W \cdot S_W^* \cdot V_W^T \tag{7}$$
3. The (possibly distorted) watermark is extracted as:
$$W^* = \frac{(D^* - S)}{a} = \frac{\left(U_W \cdot S_W^* \cdot V_W^T - S\right)}{a} \tag{8}$$
4. Compute the normalized correlation (NC) between the watermark W and the extracted watermark W^*, given by :
$$NC\,(W, W^*) = \frac{\sum\sum W \cdot W^*}{\sum\sum W^2} \tag{9}$$

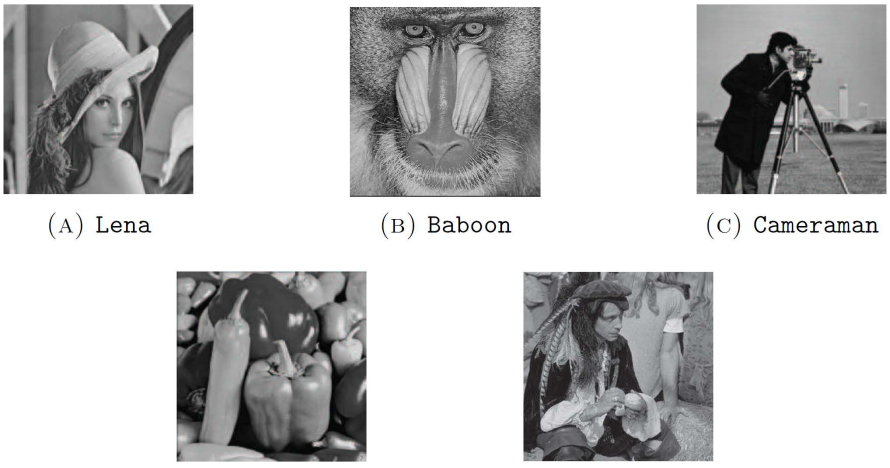
4. **Vulnerability of the false positive detection.** In this section, the high probability of false positive detection vulnerability of the SVD-based watermarking algorithm as proposed by Lai in [9] is demonstrated. Equation (8) indicates that the extracted watermark W^* is mathematically determined by the singular value decomposition matrices U_W, V_W, S_W^* and S.

The matrices U_W and V_W are stored in private key during the embedding process and will be used during the extracting process. The principal operation in extracting consists of computing matrix S_W^* from the eventually attacked watermarked image I_W^*. The singular values of different images with slightly geometric properties will differ relatively little [6]. However, in algorithm [9], if we used instead in equation (8) another matrix S_W^* obtained from the SVD of a different image not related to watermark W, this will result with a very high probability in an estimated watermark W^* visually and geometrically similar to the original watermark W. This is due to the fact that only the singular vectors specifies the geometrical property of the image.

Moreover, according to watermark embedding process as indicated in (3), one can see that the actual singular vectors of watermark W are not inserted at all. This is a serious limitation of the SVD-based watermark insertion algorithm due to the embedding watermark insertion method. Furthermore, our analysis corroborates the SVD-based watermarking algorithm analysis published by Tian *et al.* [17], who show that all useful

information for *facial recognition* is contained in singular vectors obtained from the singular value decomposition of an image and not in the singular values themselves.

5. **Examples of the false positive detection.** To support the above analysis, we give four examples using Lena and Baboon as original cover images I of size 512×512 and three watermarks of size 128×128, which are *Cameraman*, *Peppers* and *Pirate*, denoted by $W_{Cameraman}$, $W_{Peppers}$ and W_{Pirate} respectively. Note that, the watermark *Pirate* (i.e. W_{Pirate}) is considered as attackers watermark. Figure 1 illustrates, the chosen original and watermarks images.



(A) Lena (B) Baboon (C) Cameraman

(D) Peppers (E) Pirate

FIGURE 1. Cover and watermark images.

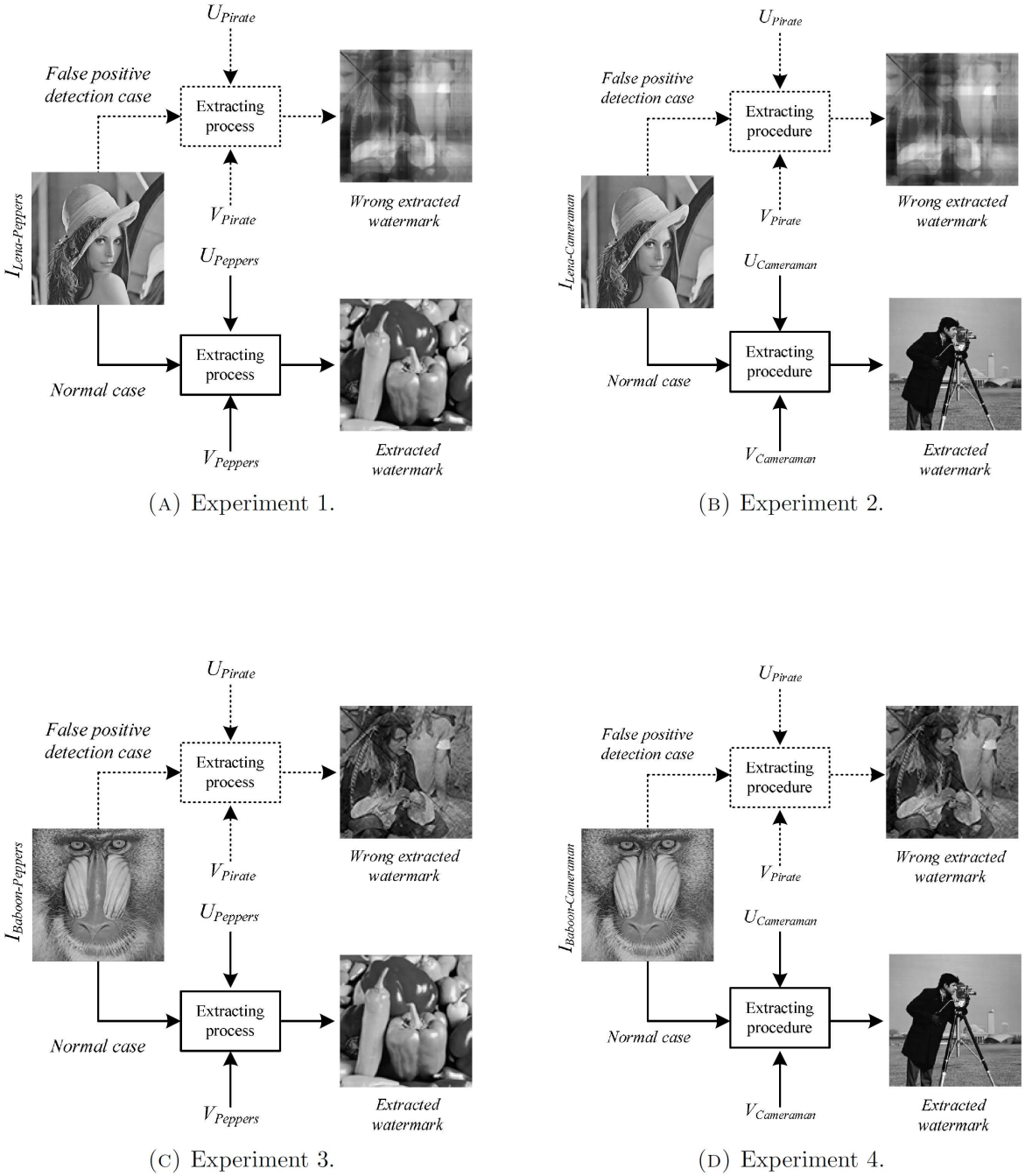
The watermarks $W_{Peppers}$ and $W_{Cameraman}$ are applied to the cover images Lena and Baboon resulting the watermarked images $I_{Lena-Peppers}$, $I_{Lena-Cameraman}$, $I_{Baboon-Peppers}$ and $I_{Baboon-Cameraman}$, respectively. Normally, in the extracting process, matrices $U_{Peppers}$ and $V_{Peppers}$ would be used to extract the watermark embedded in the received watermarked images $I_{Peppers}$ and $I_{Baboon-Peppers}$. Similarly, $U_{Cameraman}$ and $V_{Cameraman}$, are used to extract the watermark from watermarked images $I_{Lena-Cameraman}$ and $I_{Baboon-Cameraman}$. The extracted watermarks should have high correlation (correlation coefficient close to one) with original watermark $W_{Peppers}$ and $W_{Cameraman}$.

To illustrate the problem of the false positive detection of watermark, suppose that to extract the watermark embedded in watermarked images $I_{Lena-Peppers}$, $I_{Lena-Cameraman}$, $I_{Baboon-Peppers}$ and $I_{Baboon-Cameraman}$, employ instead matrices U_{Pirate} and V_{Pirate} extracted from watermark W_{Pirate}. The extracted watermark W^* should in principle have no, or very little, correlation with the watermark W_{Pirate} since the embedded watermark in $I_{Lena-Peppers}$, $I_{Baboon-Peppers}$, $I_{Lena-Cameraman}$ and $I_{Baboon-Cameraman}$ are $W_{Peppers}$ for the first two images and $W_{Cameraman}$ for the two last images.

However, as stated in Section 4, because of the slight differences in the geometric properties in the watermarks, the resulting watermark W^* will erroneously be extracted as the wrong watermark W_{Pirate}, although there are slight differences in the singular values.

Figure 2 illustrates the watermark extracting process for two possibles scenario: normal case and false alarm detection case, which are presented respectively with solid lines and dotted lines. One can see that, for the extracting process with both $U_{Peppers}$ and $V_{Peppers}$ matrices on received watermarked images $I_{Lena-Peppers}$ and $I_{Baboon-Peppers}$, as

well as $U_{Cameraman}$ and $V_{Cameraman}$ matrices for watermarked images $I_{Lena-Cameraman}$ and $I_{Baboon-Cameraman}$, the extracted watermarks are visually and geometrically similar to $W_{Peppers}$ and $W_{Cameraman}$, respectively. Furthermore, in the extracting process if the matrices U_{Pirate} and V_{Pirate} are used instead to $U_{Peppers}$ and $V_{Peppers}$, and $U_{Cameraman}$ and $V_{Cameraman}$, this will lead in the false positive detection of the watermark W_{Pirate}.



(A) Experiment 1. (B) Experiment 2.

(C) Experiment 3. (D) Experiment 4.

FIGURE 2. Vulnerability test.

The illustrated examples in figure 2 show that Lais algorithm has robustness limitation against attacks: *objective detection outcome* is not guaranteed and its affected with high probability of false positive watermark detections. The normalized correlation values (NC) between extracted watermark W^*_{Pirate} and attackers watermark $W_{Peppers}$ are given in Table

1 for the cases when cover image Lena is watermarked respectively by the watermarks $W_{Peppers}$ and $W_{Cameraman}$.

TABLE 1. Normalized correlation value (NC) obtained for the example presented in Figure 2 for the cases of false positive detections.

Cover image	Embedded watermark	Extracted watermark	NC(W_{Pirate},W^*_{Pirate})
Lena	$W_{Peppers}$	W^*_{Pirate}	0.9588
Lena	$W_{Cameraman}$	W^*_{Pirate}	0.9594
Baboon	$W_{Peppers}$	W^*_{Pirate}	0.9815
Baboon	$W_{Cameraman}$	W^*_{Pirate}	0.9811

6. **Conclusion.** In this paper, it is shown that the watermarking algorithm proposed by Lai [9] does not guarantee an objective detection outcome and has a very high probability of a false positive detection of the watermarks. We also note that such type of problems has been have been independently addressed by Zhang *et al.* [5] and Rykaczewski [6] for another SVD-based watermarking algorithm proposed by Liu and Tan [1].

REFERENCES

[1] R. Liu and T. Tan, A SVD-based watermarking scheme for protecting rightful ownership, *IEEE Trans. Multimedia*, vol. 4, no. 1, pp. 121-128, 2002.

[2] P. Bao and X. Ma, Image adaptive watermarking using wavelet domain singular value decomposition, *IEEE Trans. Circuits and Systems for Video Technology*, vol. 15, no. 1, pp. 96-102, 2005.

[3] E. Ganic and A. M. Eskicioglu, Robust DWT-SVD domain image watermarking: Embedding data in all frequencies, *Proc. of the Workshop on Multimedia and Security*, vol. 1, pp. 166-174, 2004.

[4] T. X. Zhang, W. M. Zheng, Z. M. Lu and B. B. Liu, Comments on A semi-blind digital watermarking scheme based on singular value decomposition, *Proc. of the 8th International Conference on Intelligent Systems Design and Applications*, vol. 2, pp. 123-126, 2008.

[5] X. P. Zhang and K. Li, Comments on an SVD-based watermarking scheme for protecting rightful ownership , *IEEE Trans. Multimedia*, vol. 7, no. 3, pp. 593-594, 2005.

[6] R. Rykaczewski, Comments on an SVD-based watermarking scheme for protecting rightful ownership, *IEEE Trans. Multimedia*, vol. 9, no. 2, pp. 421-423, 2007.

[7] L. Xiao, Z. Wei, and J. Ye, Comments on Robust embedding of visual watermarks using discrete wavelet transform and singular value decomposition and theoretical analysis, *Journal of Electronic Imaging*, vol. 17, pp. 040501-1 - 040501-3, 2008.

[8] K. Loukhaoukha and J. Y. Chouinard, On the security of ownership watermarking of digital images based on SVD decomposition, *Journal of Electronic Imaging*, vol. 19, no. 1, pp. 013007-1 - 013007-9, 2010.

[9] C. C. Lai, A digital watermarking scheme based on singular value decomposition and tiny genetic algorithm, *Journal of Digital Signal Processing*, vol. 21, no. 4, pp. 522-527, 2011.

[10] E. Beltrami, Sulle funzioni bilineari, *Proc. of Giornale di Mathematiche*, vol. 11, pp. 98-106, 1873.

[11] C. Jordan, Mmoire sur les formes trilinaires, *Journal de Mathmatiques Pures et Appliques*, vol. 19, pp. 35-54, 1874.

[12] L. Autonne, Sur les groupes linaires, rels et orthogonaux, *Bulletin de la socit mathmatique de France*, pp. 121-143, 1902.

[13] C. Eckart and G. Young, A principal axis transformation for non-hermitian matrices, *Bulletin of American Mathematical Society*, vol. 45, no. 2, pp. 118-121, 1939.

[14] P. Waldemar and T. Ramstad, Image compression using singular value decomposition with bit allocation and scalar quantization, *Proc. of Nordic Signal Processing Symposium*, pp. 83-86, 1996.

[15] K. Chung, C. Shen, and L. Chang, A novel SVD and VQ-based image hiding scheme, *Pattern Recognition Letters*, vol. 22, no. 9, pp. 1051-1058, 2001.

[16] K. Konstantinides, B. Natarajan, and G. Yovanof, Noise estimation and filtering using block-based singular value decomposition, *IEEE Trans. Image Processing*, vol. 6, no. 3, pp. 479-483, 1997.

[17] Y. Tian, T. Tan, Y. Wang, and Y. Fang, Do singular values contain adequate information for face recognition?, *Journal Pattern Recognition*, vol. 36, no. 3, pp. 649-655, 2003.