On the Security of Robust Image Watermarking Algorithm based on Discrete Wavelet Transform, Discrete Cosine Transform and Singular Value Decomposition

Khaled Loukhaoukha 1,2,∗, Ahmed Refaey 3, Khalil Zebbiche 1,4 and Makram Nabti 1,4

1 Centre de Recherche Développement, Bouchaoui, Algeria
2 Department of Electrical and Computer Engineering, Laval University, Quebec, Canada
3 Department of Electrical and Computer Engineering, University of Western Ontario, London, ON, N6G 5B9, Canada
4 School of Electronics, Electrical Engineering and Computer Science, Queen’s University, Belfast, UK

Received: 8 Jul. 2014, Revised: 9 Oct. 2014, Accepted: 10 Oct. 2014
Published online: 1 May 2015

Abstract: Among emergent applications of digital watermarking, owner identification, proof of ownership and transaction tracking are applications that protect data by embedding the owner’s information (watermark) in it. Recently, a robust image watermarking scheme based on discrete wavelet transform, discrete cosine transform and singular value decomposition was proposed by Hu et al. [1]. However, this scheme has shown some drawbacks. In this paper, we present two ambiguity attacks that clearly demonstrate the ineffectiveness of the above scheme some watermarking applications, such as proof of ownership, transaction tracking and data authentication.

Keywords: Ambiguity attack, image watermarking, discrete wavelet transform, discrete cosine transform, singular value decomposition.

1 Introduction

One of the most important advantages of the numeric era is the widespread use of Internet and computers, which is the result of exchanging digital media. However, illegal reproduction of data has also emerged with this extraordinary revolution and is raising questions and concerns about ownership rights. As a solution to this issue, we found Digital watermarking which consists in embedding digital data into digital contents in order to guarantee the ownership and the integrity. The basic requirements for a secure watermarking scheme are imperceptibility, robustness, capacity and security.

Digital image watermarking algorithms proposed in literature are mainly grouped into two classes: spatial [2, 3] and transform domains [4, 5]. In the literature, watermarking algorithms based on singular value decomposition (SVD) have been proposed [6, 7, 8]. There are several advantages of using SVD in digital watermarking algorithm such as: First, the size of the matrices obtained from SVD transformation is not fixed, they can be square or rectangle matrices. Second, singular values in a digital image exhibit a very good stability, that is, when a small perturbation is added to an image, its singular values do not change significantly. Third, each singular value specifies the luminance of the image while the corresponding pair of left and right singular matrices specifies the geometry of the image [9]. However, beside the above advantages, some SVD-based watermarking algorithms [6, 10, 11, 12, 13] are vulnerable to attacks and do exhibit prohibitively high probabilities of false positive detections as reported in recent literature by Loukhaoukha [14, 15], Rykaczewski [16], Zhang and Li [17], Zhang et al. [18], Loukhaoukha and Chouinard [19] and Xiao et al. [20].

This paper studies the robust image watermarking based on discrete wavelet transform, discrete cosine transform and singular value decomposition proposed by

∗ Corresponding author e-mail: khaled.loukhaoukha.1@ulaval.ca
Hu et al. [1]. We show that this algorithm does not bind an original image to a watermark and cannot be used in some security applications such as owner identification, proof of ownership, and transaction tracking. The rest of this paper is organized as follows. In the section 2, we briefly review the concept of singular value decomposition. The watermarking algorithm proposed by Hu et al. [1] is described in section 3. In section 4, we present two ambiguity attacks on this watermarking algorithm. Experimental results on these attacks are presented in section 5. Conclude are drawn in section 6.

2 Singular value decomposition

Although any given matrix of size $M \times N$ can be decomposed using the singular value decomposition (SVD), our discussion will be limited to square matrices only. The singular value decomposition of an image $I$ of size $N \times N$ is defined as follow:

$$I = U \cdot S \cdot V^T$$

where $U \in \mathbb{R}^{N \times N}$ and $V \in \mathbb{R}^{N \times N}$ are orthogonal matrices called matrices of left and right singular vectors. $S \in \mathbb{R}^{N \times N}$ is a diagonal matrix containing nonnegative terms and also known as singular value matrix. A simple example of the SVD operation is given below :

$$A = \begin{bmatrix} 151 & 140 & 131 \\ 74 & 70 & 72 \\ 135 & 144 & 154 \end{bmatrix} = \begin{bmatrix} -0.6573 & 0.6990 & 0.2816 \\ -0.3365 & 0.0622 & -0.9396 \\ -0.6743 & -0.7124 & 0.1943 \end{bmatrix} \times \begin{bmatrix} 370.6924 & 0 & 0 \\ 0 & 19.5494 & 0 \\ 0 & 0 & 1.9929 \end{bmatrix} \times \begin{bmatrix} -0.5805 & 0.7148 & -0.3900 \\ -0.5737 & -0.0192 & 0.8188 \\ -0.5778 & -0.6991 & -0.4213 \end{bmatrix}^T$$

There is several proprieties of the SVD such as :

– Singular values correspond to the luminance of the image. Suppose that the image represented in figure 1a is decomposed using SVD, then its singular values are first multiplied by 2, and in a second time divided by 2. The obtained images are illustrated in figures 1b and 1c, respectively. It is clear that the luminance of the images changed when their singular values change.

– The singular vectors specify the intrinsic geometry properties of the image. The singular vectors of the same image represented in figure 1a are slightly modified. Firstly, a values of 0.04 is added to matrix $U$, and secondly the same values is subtracted from matrix $V$. The obtained image is represented in figure 2, where one can see that the geometry of image is modified significantly.

– Singular values have a good stability, which means that a slight variation of the singular values do not affect the visual perception of the image. Figure 3 shows an example of the stability of the singular values, where a small perturbation, of value equal to 5, is added in the singular values. One can see that this perturbation do not affect the visual perception of the reconstructed image and the Peak-Signal-to-Noise Ratio (PSNR) between original image and reconstructed image is equal to 52.20 dB

3 Image watermarking scheme based on DWT, DCT and SVD

In this section, we briefly review the watermarking algorithm based on discrete wavelet transform, discrete cosine transform and singular value decomposition.
proposed by Hu et al. [1]. The watermark embedding process comprises the following steps:

1. Transform the original image \( I \) from RGB color space to \( YC_bC_r \) color space to obtain \( I_Y, I_Cb \) and \( I_Cr \) images, using equation (2).

\[
\begin{align*}
Y &= 0.257 \cdot R + 0.504 \cdot G + 0.098 \cdot B + 16 \\
C_b &= -0.148 \cdot R - 0.291 \cdot G + 0.439 \cdot B + 128 \\
C_r &= 0.439 \cdot R - 0.368 \cdot G - 0.071 \cdot B + 128
\end{align*}
\]  

(2)

2. Apply the one-level discrete wavelet transform (DWT) to gray-level image \( I_Y \) to obtain four sub-band images \( I_{Y,LL}, I_{Y,LH}, I_{Y,HL} \) and \( I_{Y,HH} \).

3. Apply the discrete cosine transform (DCT) to the LL sub-band image \( I_{Y,LL} \) to obtain frequency components \( D_Y \).

4. Apply singular value decomposition (SVD) to frequency components \( D_Y \) to obtain matrix \( S_1 \).

\[
D_Y = U_1 \cdot S_1 \cdot V_1^T
\]  

(3)

5. Modify the singular values of \( S_1 \) with the watermark image to obtain matrix \( S_0 \) as defined by equation (4), where \( \alpha \) is the scale factor \((0 < \alpha < 1)\) that controls the strength of the watermark to be inserted.

\[
S' = S_1 + \alpha \cdot W
\]  

(4)

6. Apply SVD to matrix \( S' \) to obtain matrix \( S_2 \).

\[
S' = U_2 \cdot S_2 \cdot V_2^T
\]  

(5)

7. Obtain watermarked image \( D'_Y \) by multiplying matrices \( U_1, S_2 \) and \( V_1^T \).

\[
D'_Y = U_1 \cdot S_2 \cdot V_1^T
\]  

(6)

8. Apply the inverse DCT to image \( D'_Y \) to obtain image \( I'_Y,LL \).

9. Apply the inverse DWT to images \( I_{Y,LL}, I_{Y,LH}, I_{Y,HL} \) and \( I_{Y,HH} \) to obtain image \( I_Y \).

10. Use equation (7) on image \( I_Y \) with \( I_{Cb} \) and \( I_{Cr} \) to transform the \( YC_bC_r \) color space into the \( RGB \) color space, and then produce color watermarked image \( I_W \).

\[
\begin{align*}
R &= 1.164 \cdot (Y - 16) + 1.596 \cdot (C_b - 128) \\
G &= 1.164 \cdot (Y - 16) - 0.391 \cdot (C_b - 128) - 0.813 \cdot (C_r - 128) \\
B &= 1.164 \cdot (Y - 16) + 2.018 \cdot (C_b - 128)
\end{align*}
\]  

(7)

Note that, \( S_1, U_2 \) and \( V_2 \) will be kept to be used in the watermark extracting process, which is achieved by the following steps:

1. Transform the watermarked image \( I_W \) from the RGB color space to the \( YC_bC_r \) color space, and then obtain gray-level image \( I_{YW} \).

2. Apply the one-level DWT to image \( I_{YW} \) to obtain four sub-band images \( I_{YW,LL}, I_{YW,LH}, I_{YW,HL} \) and \( I_{YW,HH} \).
3. Apply the DCT to image $I_{W,LL}$ to obtain frequency components $D_{YW}$.
4. Apply the SVD to frequency components $D_{YW}$ to obtain matrix $S_{W1}$.

\[ D_{YW} = U_{W1} \cdot S_{W1} \cdot V_{W1}^T \]  \hspace{1cm} (8)

5. Calculate the difference between $S_{W1}$ and $S_2$ to modify $S_{W1}'$, where $0 < \beta < 1$ and $S_{th}$ is the given threshold.

\[ S_{W1}' = \begin{cases} (1 - \beta) \cdot S_{W1} + \beta \cdot S_2 & \text{if } |S_{W1} - S_2| \geq S_{th} \\ \beta \cdot S_{W1} + (1 - \beta) \cdot S_2 & \text{otherwise} \end{cases} \]  \hspace{1cm} (9)

6. Obtain $S_W''$ by multiplying matrices $U_2$, $S_{W1}'$, and $V_2^T$.

\[ S_W'' = U_2 \cdot S_{W1}' \cdot V_2^T \]  \hspace{1cm} (10)

7. Obtain the extract watermark image using the following equation:

\[ \hat{W} = \frac{S_W'' - S_1}{\alpha} \]  \hspace{1cm} (11)

4 Ambiguity attacks

The basis of the ambiguity attack has first been introduced by Craver et al. [21]. As a consequence of this attack, both the owner and an attacker are able to correctly extract their watermarks from the watermarked image, but neither can prove his ownership of the image, thus, this situation leads to confusion. In this section, we describe two ambiguity attacks on the image watermarking algorithm based on discrete wavelet transform, discrete cosine transform and singular value decomposition proposed by Hu et al. [1], which show that this algorithm cannot be used for owner identification, proof of ownership, and transaction tracking.

4.1 First attack

Suppose that Alice, which is the legitimate owner, has performed the embedding steps (from 1 to 10) to embed her own watermark $W_O$ into the original image $I$ to get watermarked image denoted by $I_{WO}$. To claim the ownership of the image $I$, Alice performs the extracting steps (from 1 to 7). If her watermark is successfully extracted from the watermarked image $I_{WO}$, she is considered as the rightful owner. However, if we suppose that Bob, which considered as an attacker, has done the step 6 of embedding process on his fake watermark, $W_F$, to get the singular vectors $U_{WF}$ and $V_{WF}$. In order to prove his ownership of the original image $I$, he performs the extracting steps as follows:

1. Perform steps from 1 and 5 to obtain $S_{W1}'$.

2. By supplying $U_{WF}$, $V_{WF}$, step 6 can be achieved using the following equation:

\[ S_{WF}' = V_{WF} \cdot S_{W1}' \cdot V_{WF}^T \]  \hspace{1cm} (12)

3.3 The extracted watermark image is obtained using the following equation:

\[ \hat{W}_F = \frac{S_{WF}' - S_1}{\alpha} \]  \hspace{1cm} (13)

Thus, a fake watermark $\hat{W}_F$, which is Bob’s watermark, is successfully extracted from watermarked image $I_{WO}$, which is not supposed to contain it. The block diagram of this attack is represented in Figure 4.

![Fig. 4: Block diagram of the first attack.](image)

4.2 Second attack

Consider the scenario where Alice is the owner of the image $I$ which is watermarked by her own watermark $W_O$ to obtain the watermarked image $I_{WO}$. The attacker Bob can directly obtain $I_{WO}$ and performs the embedding steps from 1 to 10 to embed his watermark $W_F$, into the watermarked image $I_{WO}$ to obtain finally the watermarked image $I_{WOF}$. If Alice wishes to claim ownership of the original image $I$, she needs to perform extracting steps
from 1 to 7 to get the extracted watermark $W_O$ from the watermarked image $I_{W_O}$. In the other hand, Bob can claim the ownership of image $I$ by performing the extracting steps from 1 to 7 to his own watermark $W_F$ by claiming $I_{W_F}$ to be original watermarked image instead of $I_{W_O}$. This situation creates ambiguity about who is the legitimate owner of the original image $I$. Finally, no one knows who is telling the truth Alice or Bob. The block diagram of this attack is represented in figure 5.

Fig. 5: Block diagram of the second attack.

In the attacks scenarios, we played the role of Bob (the attacker) and for each attack, two experiments are performed. First, Alice who protects her own images Lena, denoted by $I^1$, and Baboon, denoted by $I^2$, by embedding her own watermark $W_O$, denoted by $I_{W_O}$, using the embedding process described in section 3 to obtain then the watermarked images $I_{W_O}^1$ and $I_{W_O}^2$, respectively.

### 5.1 First attack

In order to extract the embedded watermark from the watermarked images $I_{W_O}^1$ and $I_{W_O}^2$, the extracting process described in section 3 is performed. If the singular vectors $U_{W_F}$, $V_{W_F}$, $U_{W_O}$, and $V_{W_O}$ obtained from the attacker watermark $W_F$, and are used instead of $U_W$ and $V_W$ obtained from the owner watermark $W_O$. The extracted watermarks $\hat{W}_1$ and $\hat{W}_2$ from watermarked images $I_{W_O}^1$ and $I_{W_O}^2$, respectively, are fake watermarks, which are visually and geometrically similar to the attacker watermark. The normalized correlation values (NC) between extracted watermarks, $\hat{W}_1$ and $\hat{W}_2$, and attacker watermark $W_F$ are 0.9715 and 0.9714, respectively. The embedded and extracted watermarks of these experiments are illustrated in figures 7 and 8. One can see that the extracted watermarks are visually and geometrically similar to the attacker watermark although the embedded one in original images is the owner watermark.

5 Experimental results

In this section, experiments were carried out to prove the feasibility of proposed attacks described in section 4. Color original images Lena and Baboon of size $512 \times 512$, and owner and attacker watermarks of size $256 \times 256$, illustrated in figure 6, were used in the experiments.
5.2 Second attack

Using the watermark embedding process described in section 3, the attacker watermark $W_F$ is embedded in the watermarked images $I_{W_O}^1$ and $I_{W_O}^2$, to obtain the watermarked images $I_{W_OF}^1$ and $I_{W_OF}^2$, which have PSNR values with their corresponding original images equal to 44.05 dB and 41.90 dB, respectively. The watermarked images are presented in figure 9b and 9d.

According to the scenario presented in section 4.2, Alice obtains the extracted watermarks $W_{O}^1$ and $W_{O}^2$ from the watermarked images $I_{W_O}^1$ and $I_{W_O}^2$, which are presented in figures 9a and 9c, respectively. The extracted watermarks are presented in figures 10a and 10c, where it is clear that they are visually and geometrically similar to the owner watermark $W_O$. The normalized correlation (NC) values between the owner watermark $W_O$ and the extracted watermarks $W_{O}^1$ and $W_{O}^2$ are equal to 0.999 and 0.999, respectively.

In the other hand, we can also extract the watermarks $W_{OF}^1$ and $W_{OF}^2$ from the watermarked images $I_{W_{OF}}^1$ and $I_{W_{OF}}^2$, respectively. These extracted watermarks are presented in figures 10b and 10d, where it is clear that they are visually and geometrically similar to the attacker watermark $W_F$. The normalized correlation (NC) obtained between the attacker watermark $W_F$ and the extracted watermarks $W_{OF}^1$ and $W_{OF}^2$ are 0.999 and 0.999, respectively.
According to the above results, the extracted watermarks make an ambiguity situation about who is the owner of original image $I$ because no one knows which is telling the truth Alice or Bob. Consequently and as mentioned above, the image watermarking algorithm based on discrete wavelet transform, discrete cosine transform and singular value decomposition should not be used for owner identification, proof of ownership, and transaction tracking.

6 Conclusion

In this paper, we have demonstrated that the image watermarking algorithm based on discrete wavelet transform, discrete cosine transform and singular value decomposition proposed by Hu et al. [1] fails to resist two ambiguity attacks. In the first one, using the singular vectors of any fake watermark in the extracting process, the attacker can always claim that this watermark is the embedded one, hence, proves his ownership of the watermarked image. In the second attack, any watermarked image is publicly available which can be re-watermarked by an attacker’s watermark. Later, this attacker one can claim that the embedded watermark is his one. Experimental results prove that this algorithm should not be used for proof of ownership, transaction tracking and data authentication.

References


Khaled Loukhaoukha received the Doctorate (Ph.D) degree in electrical engineering from Laval University, Quebec, Canada in 2010, the Electronics Engineer degree and Master’s degree in electrical engineering from Saad Dahlab University, Blida, Algeria in 1999 and 2002, respectively. He has pursuing his Ph.D degree research project in digital watermarking and image encryption field at the Department of Electrical and Computer Engineering of Laval University. He has pursuing his master’s degree research project in speech coding field at the Division of Architecture and Multimedia Systems of the Center for Development of Advanced Technologies (CDTA), Algiers. He is the author or co-author of more than 20 journal and conference papers. His research interests include information security, digital watermarking, cryptography, image processing, information theory, telecommunication and optimization methods.

Ahmed Refaey Hussein is a postdoctoral fellow at the University of Western Ontario, Canada. Previously, he worked as a professional researcher at the Radiocommunications and Signal Processing Laboratory (LRTS), Laval University, Canada, in the field of wireless communications since 2007-2011. Prior to joining Laval University, he was a System/Core Network Engineer leading a team of junior engineers and technicians in the telecoms field in three large companies which are Fujitsu, Vodafone and Alcatel-Lucent. He received his B.Sc. and M.Sc. degrees from Alexandria University, Egypt in 2003 and 2005, and Ph.D. degree from Laval University, Quebec, Canada in 2011, respectively. His current research is focused on the systems and network security aspects by developing a new lower layer approach that supports important security objectives: authentication, confidentiality and integrity. He is the author and co-author of more than 17 technical papers and 2 patents applications. He is currently serves as a reviewer for IEEE transactions on broadcasting, IEEE transactions on vehicular technology, and springer for signal image and video processing.

Khalil Zebbiche received the "Ingénieur d’états“ degree in computer science from the Ecole Polytechnique, Algiers, Algeria, in 2003. In September 2005, he joined the Queens University of Belfast, Belfast, UK, as a research student and received the PhD degree from the School of Computer Science in 2008. From 2003 to 2005, he was a research scientist at the Algerian National Centre for Research and Development. He is currently holding a senior research position at the Algerian National Centre for Research and Development, Algeria. His research interests include biometrics, security, image watermarking, digital signal and image processing and information theory.

Makram Nabti received the "Ingénieur d’états“ degree in computer science from Ecole Polytechnique of Algiers, Algeria, in 2003. From 2003 to 2005, he worked in the Algerian National Centre for Research and Development as a research developer in information systems and forensic/security applications. He received his Ph.D degree in 2009 from Queen’s university of Belfast, UK. He is currently working in biometrics for forensic and security, and he is investigating iris recognition system based on a multi-scale approach, he is teaching courses in computer science and image processing. His main research interests include biometrics, systems security, image analysis, he was published several papers in scientific journals and international conferences.