On the security of ownership watermarking of digital images based on singular value decomposition

Huo-Chong Ling,[a] **Raphael C.-W. Phan,**[b] **and Swee-Huay Heng**[a]
[a]Multimedia University, Faculty of Engineering, Centre for Multimedia Security and Signal Processing, Research Group of Cryptography and Information Security, Cyberjaya, Selangor 63100 Malaysia
[b]Loughborough University, Department of Electronic and Electrical Engineering, LE11 3TU, United Kingdom
E-mail: hcling@mmu.edu.my

Abstract. *We show that the two countermeasures proposed in a paper on the security of ownership watermarking of digital images based on singular value decomposition by Loukhaoukha and Chouinard do not solve the false-positive detection problem in contrast to designers' claim and therefore should not be used for proof of ownership application. © 2011 SPIE and IS&T.* [DOI: 10.1117/1.3534865]

1 Introduction

Loukhaoukha and Chouinard[1] have recently proposed two countermeasures for the problem of false-positive detections[2,3] of watermarks that exists in Abdallah et al.[4] and Aslantas[5] singular value decomposition (SVD)-based watermarking schemes without making changes to the original algorithm of those schemes. They claimed that the countermeasures could be applied as an add-on to any SVD-based watermarking schemes that suffered from false-positive detections.

The first countermeasure uses a one-way hash function, such as MD5 or SHA-1 during the embedding process to compute digests of the SVD matrices U_{W} and V_{W} of the embedded watermark W. The digests or hash values of U_{W} and V_{W} (denoted as $H_{U_{\mathrm{W}}}$ and $H_{V_{\mathrm{W}}}$) are then kept by the owner so that he can use them in the extraction process if he wants to claim ownership. During the extraction process, the ownership claimant supplies the digests $H_{U_{\mathrm{W}}}$ and $H_{V_{\mathrm{W}}}$ of his own watermark and $H_{U_{\mathrm{W}}}$ and $H_{V_{\mathrm{W}}}$ are verified with the hash values of the received (and possibly altered by an attacker) matrices $\widetilde{U}_{\mathrm{W}}$ and $\widetilde{V}_{\mathrm{W}}$ (denoted as $H_{\widetilde{U}_{\mathrm{W}}}$ and $H_{\widetilde{V}_{\mathrm{W}}}$). If $H_{U_{\mathrm{W}}} \neq H_{\widetilde{U}_{\mathrm{W}}}$ or $H_{V_{\mathrm{W}}} \neq H_{\widetilde{V}_{\mathrm{W}}}$, then the extraction process halts; otherwise, the extraction process continues. In other words, they claimed that only the rightful owner was able to extract the embedded watermark W if $H_{U_{\mathrm{W}}} = H_{\widetilde{U}_{\mathrm{W}}}$ ($U_{\mathrm{W}} = \widetilde{U}_{\mathrm{W}}$) and $H_{V_{\mathrm{W}}} = H_{\widetilde{V}_{\mathrm{W}}}$ ($V_{\mathrm{W}} = \widetilde{V}_{\mathrm{W}}$).

The second countermeasure uses an image-encryption (respectively, decryption) method on the watermark W before the embedding process (respectively, after the extraction process). Before the embedding process, watermark W is encrypted to give W_{E} and then W_{E} is embedded in cover image I to obtain the watermarked image I_{W}. In the extraction process, W_E^* is obtained from the possibly corrupted watermarked image I_W^* and decrypted to get the watermark W^*, which is perceptually similar to the owner's watermark W. Loukhaoukha and Chouinard[1] claimed that a false-positive attack would result in the first extracted image be the attacker's watermark $\widetilde{W}$ because the attacker is using the matrices $U_{\widetilde{\mathrm{W}}}$ and $V_{\widetilde{\mathrm{W}}}$, instead of proper matrices U_{W} and V_{W}. However, because the attacker has to feed the first extracted watermark (i.e., $\widetilde{W}$ to the decryption process), his final watermark will thus be an encrypted image, which does not help in his ownership claim.

In Sec. 2 we show that both countermeasures do not solve the false-positive detection in contrast to what is claimed by Loukhaoukha and Chouinard.[1]

2 Theorectical Analysis and Experiments

In the first countermeasure, the hash values of the watermark W's SVD matrices U_{W} and V_{W} (denoted as $H_{U_{\mathrm{W}}}$ and $H_{V_{\mathrm{W}}}$) provided by the ownership claimant does not bind to the watermarked image I_{W}. Because there is no proof showing that $H_{U_{\mathrm{W}}}$ and $H_{V_{\mathrm{W}}}$ belong to the rightful owner of the watermark W, therefore an attacker A, who repeats the same hashing process on the SVD matrices U_A and V_A of his own watermark W_A to obtain H_{U_A} and H_{V_A}, can claim that the watermarked image I_{W} belongs to him because the extraction process will verify that $H_{U_A} = H_{\widetilde{U}_{\mathrm{W}}}$ ($U_A = \widetilde{U}_{\mathrm{W}}$) and $H_{V_A} = H_{\widetilde{V}_{\mathrm{W}}}$ ($V_A = \widetilde{V}_{\mathrm{W}}$).

An interesting fact is that since A attacks the scheme, he can provide his own kept $H_{\widetilde{U}_{\mathrm{W}}}$ and $H_{\widetilde{V}_{\mathrm{W}}}$ which are similar to H_{U_A} and H_{V_A} during the extraction process. The same case applies to the rightful owner, whereby in order to claim the watermarked image I_{W}, he has to provide his own kept versions of $H_{\widetilde{U}_{\mathrm{W}}}$ and $H_{\widetilde{V}_{\mathrm{W}}}$ that are similar to $H_{U_{\mathrm{W}}}$ and $H_{V_{\mathrm{W}}}$ during the extraction process. Therefore, both have equal rights to the watermarked image I_{W} and no one can prove more than the other. The flaw occurs because the designers[1] view the countermeasure in the owner's perspective, ignoring the fact that the attacker can repeat the same steps as the owner.

In the second countermeasure, an encrypted watermark W_{E} is used in the embedding process. Hence, in the extraction process, after the encrypted watermark is extracted from the watermarked image I_{W}, it has to be decrypted to obtain the watermark W^*, which is perceptually similar to the original watermark W. The owner needs to keep the SVD's $U_{\mathrm{W_E}}$ and $V_{\mathrm{W_E}}$ components of the encrypted watermark W_{E} so that he can supply the components later in the extraction process. If during the extraction process, an attacker A provides encrypted SVD matrices $U_{\widetilde{\mathrm{W}}_{\mathrm{E}}}$ and $V_{\widetilde{\mathrm{W}}_{\mathrm{E}}}$ of his own encrypted watermark $\widetilde{W}_{\mathrm{E}}$, instead of proper matrices $U_{\widetilde{\mathrm{W}}}$ and $V_{\widetilde{\mathrm{W}}}$ as mentioned by the designers,[1] then he can still obtain his own encrypted watermark $\widetilde{W}_E^*$. The encrypted watermark $\widetilde{W}_E^*$ is later decrypted to obtain $\widetilde{W}^*$, which is perceptually similar to $\widetilde{W}$. This countermeasure fails because the designers[1] did not

Paper 10142LR received Aug. 19, 2010; revised manuscript received Oct. 26, 2010; accepted for publication Dec. 13, 2010; published online Jan. 20, 2011.

1017-9909/2011/20(1)/010501/2/$25.00 © 2011 SPIE and IS&T

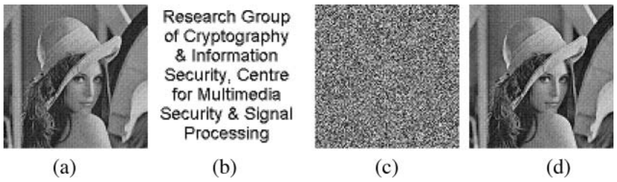
Research Group of Cryptography & Information Security, Centre for Multimedia Security & Signal Processing
(a)
(b)
(c)
(d)

Fig. 1 (a) Cover image, (b) owner's watermark (c) owner's encrypted watermark and (d) watermarked image.

notice that the attacker can simply repeat the same process as the owner in the proof-of-ownership game because they both claim ownership and the extraction process obtains the encrypted SVD matrices from the ownership claimant.

Figure 1 shows the cover image I, the owner's watermark W, the owner's encrypted watermark W_{E} and the watermarked image I_{W} after being embedded with W_{E}. Figure 2 shows the attacker's watermark $\widetilde{W}$, the encrypted watermark $\widetilde{W}_{\mathrm{E}}$ of the attacker, the extracted encrypted watermark $\widetilde{W}_E^*$ from the watermarked image I_{W}^* using $U_{\widetilde{\mathrm{W}}_{\mathrm{E}}}$ and $V_{\widetilde{\mathrm{W}}_{\mathrm{E}}}$, and the final decrypted watermark $\widetilde{W}^*$. As can be seen from the experimental results, the final watermark $\widetilde{W}^*$ that is decrypted from $\widetilde{W}_E^*$, is perceptually similar to the attacker's watermark $\widetilde{W}$ with the correlation coefficient value of 0.948.

A possible countermeasure against false-positive detections would be to avoid[6,7] using the watermark's SVD matrices U and V in the embedding (and thus extraction) process of the Abdallah et al.[4] and Aslantas[5] schemes. This removes the influence of SVD matrices U and V on the watermark extraction process, which is the main problem causing false-positive detections.

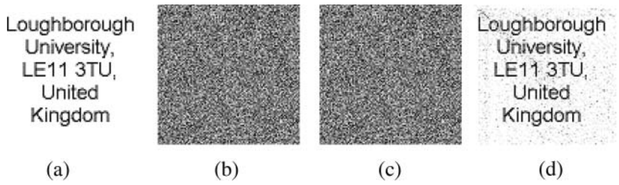
Loughborough University, LE11 3TU, United Kingdom
Loughborough University, LE11 3TU, United Kingdom
(a)
(b)
(c)
(d)

Fig. 2 (a) Attacker's watermark, (b) attacker's encrypted watermark, (c) extracted encrypted watermark, and (d) Final decrypted watermark.

3 Conclusions

We have shown that Loukhaoukha and Chouinard[1] countermeasures are not able to solve the false-positive detection of the attacker's watermark and thus are not suitable for proof-of-ownership application. This is in contrast to the designers' (Ref. 1) claims that the countermeasures are explicitly designed to solve the problem. The shortfall is because the designers only viewed the countermeasures' design from the owner's perspective, instead of also from the attacker's perspective as an ownership claimant. When designing the countermeasures as an add-on to the SVD-based watermarking scheme, the designers did not consider that the attacker can follow the same steps as the owner. This leads to the failure of the countermeasures to solve the false-positive detection problem.

References

1. K. Loukhaoukha and J. Chouinard, "Security of ownership watermarking of digital images based on singular value decomposition," *J. Electron. Imaging* **19**, 013007 (2010).
2. R. Rykaczewski, "Comments on An SVD-based watermarking scheme for protecting rightful ownership," *IEEE Trans. Multimedia.* **9**(2), 421–423 (2007).
3. X. P. Zhang and K. Li, "Comments on An SVD-based watermarking scheme for protecting rightful ownership," *IEEE Trans. Multimedia.* **7**(2), 593–594 (2005).
4. E. Abdallah, A. B. Hamza, and P. Bhattacharya, "Improved image watermarking scheme using fast Hadamard and discrete wavelet transforms," *J. Electron. Imaging* **16**, 033020 (2007).
5. V. Aslantas, "An optimal robust digital image watermarking based on SVD using differential evolution algorithm," *Opt. Commun.* **282**(5), 769–777 (2009).
6. A. A. Mohammad, A. Alhaj, and S. Shaltaf, "An improved SVD-based watermarking scheme for protecting rightful ownership," *Signal Processing* **88**, 2158–2180 (2008).
7. H.-C. Ling, R. C.-W. Phan, and S.-H. Heng, "Analysis on the improved SVD-based watermarking scheme," *Lect. Notes Comput. Sci.* **6059**, 143–149 (2010).