Semi-Blind Watermarking via NMF Algorithm on Random Blocks of an Image

Sergul Aydore [1] **Serdar Kozat** [2] **Kivanc Mihcak** [3]

Abstract

Watermarking is designed to protect intellectual property of a digital media. We propose a method for semi-blind watermarking by introducing a novel rank reduction technique "modified Non-Negative Matrix Factorization (NMF)" for secure and robust watermarking. The main idea of the study is applying "modified NMF" algorithm to pseudo-randomly selected rectangles of the images. Our experiments show that our approach is able to detect watermark of an attacked image with rotation up to certain degrees invisible to the human's eye.

1. Introduction

Watermarking is a process of embedding information (i.e. watermark) into digital data (audio, video or image) which is not perceivable to the human's eye. One of the common application of watermarking is preventing the copyright of the owner. A secret key is used during the procedure of watermark embedding and this key is also shared with the decoder. However, a watermarked data might face malicious attacks or unintentional modifications which makes watermark detection harder. Therefore, watermark embedding and detection have to be robust against such changes.

Watermarking techniques can be applied in two domains: (i) spatial domain (ii) transform domain (Hartung & Kutter, 1999; Tao et al., 2014). Previous studies utilized transform techniques such as Discrete Cosine Transform (DCT) (Das et al., 2014), Discrete Wavelet Transform (DWT) (Hu et al., 2011) and rank reduction techniques such as Singular Value Decomposition (SVD) (Liu & Tan, 2002). However, SVD-based approaches are not secure enough since including negative values might reduce the image quality. In order to address this issue, many recent watermarking studies focused on hybrid approaches which combine SVD with DCT or DWT (Lai & Tsai, 2010; Gupta & Raval, 2012; Singh et al., 2015). As an alternative to SVD, nonnegative matrix factorization (NMF) has also attracted researchers due to its nonnegative property (Ghaderpanah & Hamza, 2006; Ouhsain & Hamza, 2009).

In this study, we propose to use a modified version of an NMF as a tranform domain. Our approach distinguishes from the standard NMF algorithm by fixing the pseudo-randomly generated base matrix which is shared by encoder and decoder and updating only the coefficient matrix of the decomposition. Hence, the method turns out to be a convex optimization problem. Furthermore, we apply this logic to randomly selected blocks of the image that increases security.

2. Watermarking Game

The Watermarking game, we consider in here, includes three basic parts: an encoder, an attacker/channel and a decoder. The encoder embeds watermark to host image so that the host image and the watermarked image are perceptually the same. We denote encoder as a function that takes the host image, a message and the secret key as its inputs and produces the watermarked image. The channel component represents a malicious attack on the watermarked image. In semi-blind settings, the decoder has an access to the side information in order to make a decision using the key and the noisy image as inputs.

2.1. Encoding

Given a matrix $\mathbf{S}$, representing the host image, the cost function to be minimized is the square Euclidean distance between $\mathbf{S}$ and $\mathbf{WH}$, i.e., $\mathrm{f}(\mathbf{S}, \mathbf{WH}) = ||\mathbf{S} - \mathbf{WH}||^2$, with the constraints $\mathbf{W}$,$\mathbf{H} \geq 0$. For initialization, both $\mathbf{W}$ and $\mathbf{H}$ are generated pseudo-randomly, but only the coefficient matrix $\mathbf{H}$ is updated at every iteration. The additive update rule for $\mathbf{H}$ at k-th iteration to minimize the square Euclidean distance is given by $\mathbf{H}^{(k+1)} = \mathbf{H}^{(k)} - \eta_k \nabla_{\mathbf{H}} f$, where η_k is the learning rate. When all constants η_k are selected to be equal to a small positive number η, the additive update rule reduces to the standard gradient descent update given as

$$\begin{aligned} \mathbf{H}^{(k+1)} &= \mathbf{H}^{(k)} + \eta \nabla_{\mathbf{H}} [(\mathbf{S} - \mathbf{WH}^{(k)})^T (\mathbf{S} - \mathbf{WH}^{(k)})] \\ &= \mathbf{KH}^{(k)} + \eta \mathbf{W}^T \mathbf{S}, \end{aligned} \quad (1)$$

[1] Amazon, NY, USA [2] Bilkent University, Ankara, Turkey [3] Bogazici University, Istanbul, Turkey. Correspondence to: Sergul Aydore <sergulaydore@gmail.com>.

Proceedings of the 35^{th} International Conference on Machine Learning, Stockholm, Sweden, PMLR 80, 2018. Copyright 2018 by the author(s).

where $\mathbf{K} \triangleq \mathbf{I} - \eta \mathbf{W}^T\mathbf{W}$ and $\mathbf{I} \in \mathbb{R}^{r \times r}$ is an identity matrix. Hence, the iteration rule for $\mathbf{H}$, given the initial non-negative matrices $\mathbf{H}^{(0)}$ and $\mathbf{W}$ as

$$\mathbf{H}^{(n)} \quad = \quad \mathbf{K}^n\mathbf{H}^{(0)} + \eta \left(\sum_{i=0}^{n-1} \mathbf{K}^i \right) \mathbf{W}^T\mathbf{S}. \qquad (2)$$

where $\mathbf{K}^n = \mathbf{K} \dots \mathbf{K}$. Let $\Delta\mathbf{H}$ and $\Delta\mathbf{S}$ be the watermark desired to be embedded to the host image in the modified NMF domain and spatial domain, respectively. Then according to (2), the watermarked image $\mathbf{S} + \Delta\mathbf{S}$ should satisfy:

$$\mathbf{H}^{(n)} + \Delta\mathbf{H} = \mathbf{K}^n\mathbf{H}^{(0)} + \eta(\sum_{i=0}^{n-1} \mathbf{K}^i)\mathbf{W}^T(\mathbf{S} + \Delta\mathbf{S}), \quad (3)$$

that yields $\Delta\mathbf{H} = \underbrace{\eta[(\sum_{i=0}^{n-1} \mathbf{K}^i)]\mathbf{W}^T}_{\triangleq \mathbf{Q}} \, \Delta\mathbf{S}$ Hence, the watermarking problem reduces to solving $\mathbf{S}$ for a given $\Delta\mathbf{H}$ and $\mathbf{Q} \in \mathbb{R}^{r \times m}$. Since the system in is under-determined, the solution is not unique. Hence we seek for a minimum norm solution for $\Delta\mathbf{S}$ which can be written as $(\Delta\mathbf{S})_{\mathrm{MNLS}} = \mathbf{V}\mathbf{\Lambda}^{-1}\mathbf{U}^T\mathbf{Z}$ where $\mathbf{Q} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^T$.

We employ this approach to the randomly selected blocks of the host image. Those regions are selected pseudo-randomly so that attackers could not know the regions without a secret key. Hence, embedding watermark to those regions is more secure than embedding to the full image. We use L for the total number of rectangles and t for the rank of an each rectangle. $\mathbf{S}_l \in \mathbb{R}^{z \times z}$ and $\Delta\mathbf{W}_l \in \mathbb{R}^{z \times t}$ denote l-th rectangle of the host image and basis matrix of the l-th rectangle after modified NMF decomposition, respectively where $1 \leq l \leq L$. Also, $\Delta\mathbf{S}_l \in \mathbb{R}^{z \times z}$ and $\Delta\mathbf{H}_l \in \mathbb{R}^{t \times z}$ denote the watermark desired to be embedded to the corresponding rectangle of the host image and in the modified NMF domain, respectively. Solving the problem becomes more complicated if the selected regions overlap as they bring constraints to the optimization problem. We define a mapping $\mathcal{M} : \mathbb{R}^{a \times b} \to \mathbb{R}^{ab}$, such that $\mathcal{M}(\mathbf{A}) = [\mathbf{A}_{11}\mathbf{A}_{21} \cdots \mathbf{A}_{a1} \cdots \mathbf{A}_{b1}\mathbf{A}_{b2} \cdots \mathbf{A}_{ba}]^T$. Let $\Delta\mathbf{s} \triangleq \mathcal{M}(\Delta\mathbf{S})$, $\Delta\mathbf{H} = [\Delta\mathbf{H}_1, \Delta\mathbf{H}_2, \cdots, \Delta\mathbf{H}_L]$ and $\Delta\mathbf{h} \triangleq \mathcal{M}(\Delta\mathbf{H})$. Furthermore, let define a sparse matrix $\mathbf{C}$ such that $\mathbf{C}\Delta\mathbf{s} = \Delta\mathbf{h}$. For a given $v \in \{1, \cdots, n\}$ and $i \triangleq (v-1)n + 1$ (recall $\mathbf{S} \in \mathbb{R}^{m \times n}$), the sparse matrix is formed as follows; $\mathbf{C}(v : v - 1 + t, i : i - 1 + z) = \mathbf{Q}_l$ if $\Delta\mathbf{s}(i : i - 1 + z) \in \Delta\mathbf{S}_l$. Then the problem reduces to finding $\min \Delta\mathbf{s}$ such that $\mathbf{C}\Delta\mathbf{s} = \Delta\mathbf{h}$ which has the solution $\Delta\mathbf{s} = \mathbf{V}\mathbf{\Lambda}^{-1}\mathbf{U}^T\Delta\mathbf{h}$ where $\mathrm{SVD}(\mathbf{C}) \triangleq \mathbf{U}\mathbf{\Lambda}\mathbf{V}$. Since $\mathcal{M}$ is a one to one mapping we can find $\Delta\mathbf{S}$ from $\Delta\mathbf{s}$.

2.2. Decoding

Let the attacked image after watermark embedding is denoted as $\mathbf{Y}$ and we assume that $\mathrm{NMF}_{\mathbf{W}}(\mathbf{S}) = \mathbf{H}$ is known for the semi-blind setting. The decoder should verify whether the attacked image $\mathbf{Y}$ is watermarked or not. In order to accomplish this task, $\mathbf{Y}$ is transformed into the transformation (modified NMF) domain via $\mathrm{NMF}_{\mathbf{W}}(\mathbf{Y}) = \tilde{\mathbf{H}}$. Then, decoder decides whether or not the image is watermarked according to the correlation coefficient between $\Delta\mathbf{H}$ (watermark in transform domain) and $(\tilde{\mathbf{H}} - \mathbf{H})$. A threshold τ is determined according to experimental results. Thus, the decoder concludes that the attacked image is watermarked if $\mathrm{corr}(\Delta\mathbf{H}, (\tilde{\mathbf{H}} - \mathbf{H})) > \tau$.

3. Results

We applied the proposed approach to the well-known image Lena of size 512×512 in gray scale. We applied discrete wavelet transform (DWT) with wavelet family Daubechies of length 8. The DC subband of 2-level DWT is used for the watermarking game. Note that we used DWT for the purpose of dimensionality reduction rather than enconding watermark. Hence the final image we used is of size 133×133. We set rectangle size to 20×20 and number of rectangles to 5 for watermark embedding while preserving 30 dB PSNR between the weatermarked and unwatermarked image. We applied rotation attack of various angles. The performance of the approach is evaluated based on the ability to distinguish an attacked watermarked image from an attacked un-watermarked image. For this purpose, we repeated the experiments for 1000 seeds. Figure 1 shows the ROC curves for an increasing amount of rotation angles. Unsurprisingly, the performance decreases as the rotation angle increases. The diagonal dashed curve in navy demontrates the random decision.

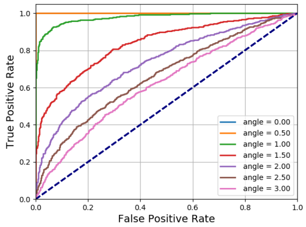

Figure 1. ROC curves for an increasing amount of rotation attacks to Lena image.

4. Conclusion

We discussed the theory of our approach and illustrated the performance of our approach against a geometric attack applied to Lena image. This work can be extended in various ways: (i) implement the approach with frequency domain transforms (ii) computational complexity analysis of the approach (iii) robustness against other types of attacks.

References

Das, Chinmayee, Panigrahi, Swetalina, Sharma, Vijay K, and Mahapatra, KK. A novel blind robust image watermarking in dct domain using inter-block coefficient correlation. *AEU-International Journal of Electronics and Communications*, 68(3):244–253, 2014.

Ghaderpanah, Mohammadreza and Hamza, A Ben. A non-negative matrix factorization scheme for digital image watermarking. In *Multimedia and Expo, 2006 IEEE International Conference on*, pp. 1573–1576. IEEE, 2006.

Gupta, Akshya Kumar and Raval, Mehul S. A robust and secure watermarking scheme based on singular values replacement. *Sadhana*, 37(4):425–440, 2012.

Hartung, Frank and Kutter, Martin. Multimedia watermarking techniques. *Proceedings of the IEEE*, 87(7):1079–1107, 1999.

Hu, Yuping, Wang, Zhijian, Liu, Hui, and Guo, Guangjun. A geometric distortion resilient image watermark algorithm based on dwt-dft. *JSW*, 6(9):1805–1812, 2011.

Lai, Chih-Chin and Tsai, Cheng-Chih. Digital image watermarking using discrete wavelet transform and singular value decomposition. *IEEE Transactions on instrumentation and measurement*, 59(11):3060–3063, 2010.

Lee, Daniel D and Seung, H Sebastian. Algorithms for non-negative matrix factorization. In *Advances in neural information processing systems*, pp. 556–562, 2001.

Liu, Ruizhen and Tan, Tieniu. An svd-based watermarking scheme for protecting rightful ownership. *IEEE transactions on multimedia*, 4(1):121–128, 2002.

Ouhsain, Mohamed and Hamza, A Ben. Image watermarking scheme using nonnegative matrix factorization and wavelet transform. *Expert Systems with Applications*, 36 (2):2123–2129, 2009.

Reddy, A Adhipathi and Chatterji, Biswanath N. A new wavelet based logo-watermarking scheme. *Pattern Recognition Letters*, 26(7):1019–1027, 2005.

Singh, Amit Kumar, Kumar, Basant, Dave, Mayank, and Mohan, Anand. Robust and imperceptible dual watermarking for telemedicine applications. *Wireless Personal Communications*, 80(4):1415–1433, 2015.

Tao, Hai, Chongmin, Li, Zain, Jasni Mohamad, and Abdalla, Ahmed N. Robust image watermarking theories and techniques: A review. *Journal of applied research and technology*, 12(1):122–138, 2014.