Reliable SVD based Semi-blind and Invisible Watermarking Schemes

Subhayan Roy Moulick
Indian Institute of Science Education and Research, Kolkata, India

Siddharth Arora
Somerville College
University of Oxford, U.K.

Chirag Jain
Price WaterHouse Coopers, India

Prasanta K. Panigrahi
Indian Institute of Science Education and Research, Kolkata, India

ABSTRACT

A semi-blind watermarking scheme is presented based on Singular Value Decomposition (SVD), which makes essential use of the fact that, the SVD subspace preserves significant amount of information of an image and is a one way decomposition. The principal components are used, along with the corresponding singular vectors of the watermark image to watermark the target image. For further security, the semi-blind scheme is extended to an invisible hash based watermarking scheme. The hash based scheme commits a watermark with a key such that, it is incoherent with the actual watermark, and can only be extracted using the key. Its security is analyzed in the random oracle model and shown to be unforgeable, invisible and satisfying the property of non-repudiation.

Keywords

Singular Value Decomposition (SVD), Principal Components, Semi Blind Watermark, Invisible Watermark, Hash Code

1. INTRODUCTION

The advent of the internet has made it possible to easily store and share digital information and multimedia. While this has largely benefited all, protection of digital multimedia content has become an increasingly important issue for content owners and service providers. A common and well-proposed solution to this problem is the *digital watermark*. It is an important tool for copyright protection that embeds data into multimedia content, which can be later detected or extracted. Because of its key role in security and establishing ownership in multimedia files, watermarking is an area of active interest among a wide spectrum of researchers. In the last two decades considerable amount of research has been devoted to study various methods for efficiently watermarking images, which are both practically viable and theoretically sound.

Taking advantage of the optimal image decomposition property of Singular Value Decomposition (SVD) for embedding a watermark in an image, several SVD based watermarking schemes have been proposed. SVD is a general linear algebraic technique, whereby a given matrix (image in this case), is diagonalized such that most of the signal energy is localized in a few singular values [1]. A digital image A of size $M \times N$ can be represented by its SVD as

$$A = USV^T,$$

where U and V are orthogonal matrices of size $M \times M$ and $N \times N$, respectively. S is a diagonal matrix of size $M \times N$, with the diagonal elements representing the Singular Values (SVs). Columns of matrix U, also known as left singular vectors, are the eigenvectors of AA^T, while columns of matrix V (right singular vectors) are eigenvectors of A^TA. It is worth noting that, the singular vectors of an image specify the image geometry, while the singular values specify the luminance (energy) of the image. For an image of size $M \times N$ (let $M > N$, without any loss of generality), the singular vectors have $O(N^2)$ elements, as compared to just $O(N)$ diagonal elements in the singular value matrix. Hence, this makes the use of singular vectors for information hiding more appropriate than using the singular values [2]-[3].

It is due to the elegant properties of SVD, that it has gained much attention, and has been studied rigorously. [4] used a hybrid technique, whereby they employed wavelet transform and SVD followed by a recursive dither modulation algorithm, to insert signature information and textual data into the cover medical images. In addition, differential evolution was also applied to design the quantization steps optimally for controlling the strength of the watermark. [5] divided the cover image into blocks and applied SVD to each block; the watermark was then embedded in all the non-zero singular values according to the local features of the cover image. [6] used SVD as a medium to embed information and proposed a steganography scheme. [7] presented a SVD-based watermarking technique considering human visual characteristics through block selection with Discrete Cosine Transformation (DCT) followed by SVD. [8] also used DCT and SVD to construct a watermarking scheme. [9] proposed a zero-watermarking scheme combining Discrete Wavelet Transform (DWT) and SVD. They extracted features from the cover image by applying DWT and SVD to each non-overlapping block. The embedding of zero watermarking was realized through the exclusive or (XOR) operation between the singular value of each block and the pixel value of the actual binary character watermark. [10] proposed a DWT and SVD-based scheme for digital watermarking. [11] proposed a blind scheme based on DWT, SVD and support vector regression. Here, the embedding algorithm hides a watermark bit in the low-low (LL) sub-band of the cover image's principal components of the block. Additionally particle swarm optimization has been utilized to optimize the scheme. [12] recently presented a non-blind, adaptive watermarking scheme, based on DWT and SVD, where they used principle components and perceptual tuning. [13] used a family of chaotic maps and SVD. To encrypt the watermark logo and to improve the security of watermark image Jacobian elliptic maps were used, along with quantum maps to determine the location of image's block for the watermark embedding. Yet another commonly used method for watermarking is through image segmentation [14].

In this paper, we first present a SVD based semi-blind watermarking scheme, whose efficacy and security depends on the fact that the SVD subspace preserves significant amount of information of an image and that SVD is a one way decomposition. The semi-blind watermarking schemes (or semi-private watermarking schemes), do not require the cover image to detect the watermark. To introduce the watermark, the principal components and the singular values of the watermark image are embedded in the cover (original) image, and the detector only needs a complimentary set for watermark extraction from the embedded watermark image.

Subsequently, an invisible hash based watermarking scheme is derived from the semi-blind scheme. The invisible watermarking schemes, unlike visible ones, commit to a key, with embedding the watermark image in the cover (original) image, which remains incoherent with the real watermark image. The watermark can be extracted only when the original key is available. The security of the proposed invisible hash based watermarking scheme is analyzed in the random oracle model and shown to be unforgeable, invisible and satisfying the property of non-repudiation.

The paper is organized as follows. Section 2 presents the semi-blind watermarking scheme and demonstrates its efficacy. In Section 3, a construction of a hash code based invisible watermarking scheme is given, which is proved to be secure in the random oracle model. Finally, the practicality of the proposed schemes for multichannel (color) images are examined in Section 4, followed by conclusion in Section 5.

2. SEMI-BLIND WATERMARKING SCHEME

In consideration of the fact that the singular vectors U and V have majority of image information, and SVD is a one way decomposition, we propose a scheme whereby the principal components of watermark image are embedded into the singular values of original image. Using this scheme, only a set of singular vectors are required to be known at the detector. The security of this emerged from the fact that SVD is a one way decomposition.

Using SVD, a matrix A, representing a mono-channel image can be represented as:

$$A = USV^T, \tag{1}$$

$$W => U_w S_w V_w^T => A_{wa} V_w^T, \tag{2}$$

where *A* is the original (mono-channel) image and *W* is the watermark to be embedded in *A*. $A_{wa} = U_w S_w$ are also known as principal components. Embedding the principal components, A_{wa}, with a corresponding diagonal singular value matrix S of the original image, we get S_1 and the corresponding watermarked image A_w.

$$S_1 = S + \alpha A_{wa}, \tag{3}$$

$$US_1V^T => A_w, \tag{4}$$

Let A_w^* denote the possibly distorted watermarked image at the detector. To recover back the watermark image W^* from A_w^*, following are the steps involved,

$$(A_w^* - A) => A_1, \tag{5}$$

$$(U^{-1} A_1 (V^T)^{-1})/\alpha => A_{wa}^*, \tag{6}$$

$$A_{wa}^* V_w^T => W^*, \tag{7}$$

We now try to search for a reference image, denoted by *P*, in the watermarked image using our scheme. We first find the SVD of the original reference image *P* as,

$$P => U_p S_p V_p^T,$$

To find a possibly distorted reference image in the distorted watermarked image A_{wa}^*, the singular vectors of reference image are used as follows,

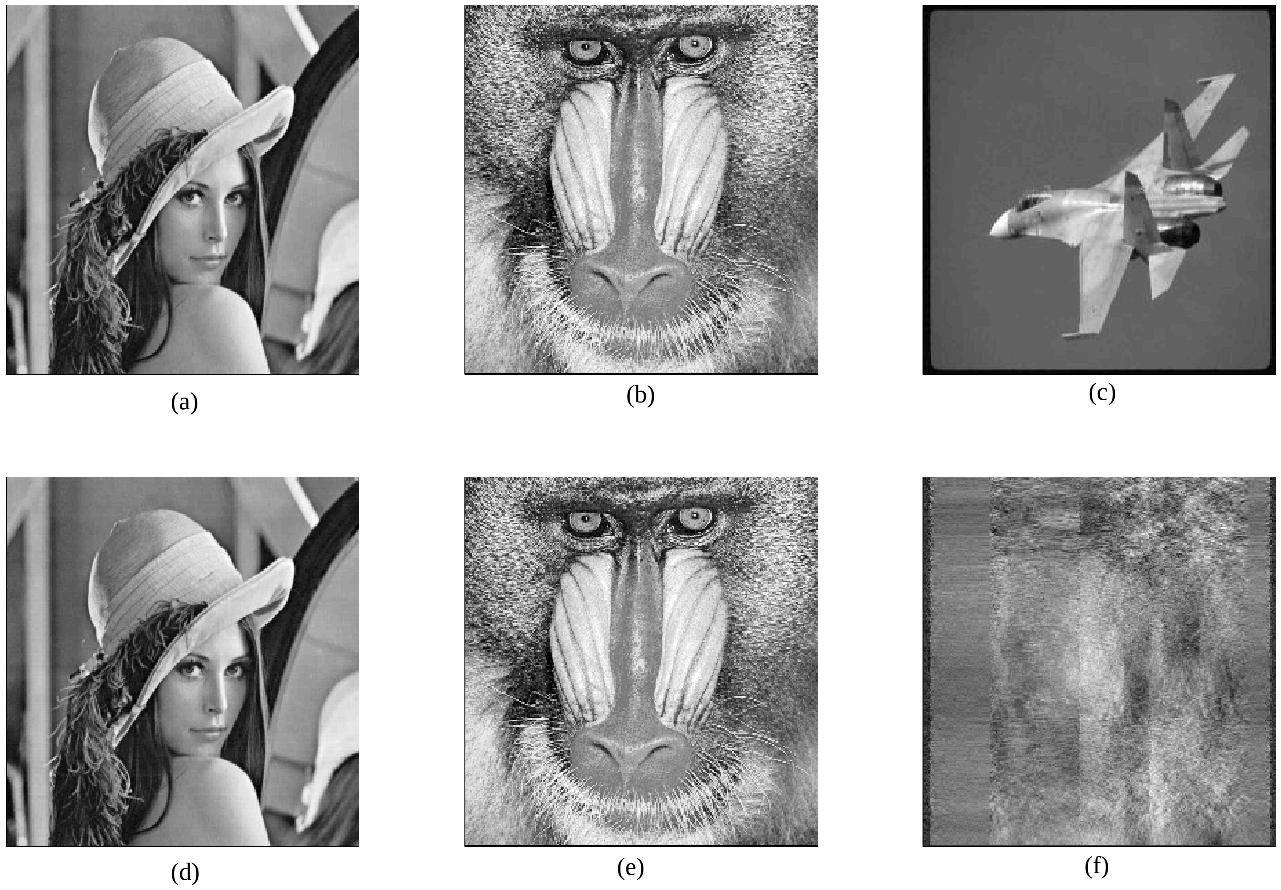
$$P^* => A_{wa}^* V_p^T, \tag{8}$$

Since, one of the singular vectors of watermark is embedded in the original image; watermark extraction without knowing the original principal components is not possible. Thus, no reference image can be extracted from any arbitrary image using the proposed scheme, as demonstrated in figure 1.

3. A HASH CODE BASED INVISIBLE WATERMARKING SCHEME

Following the semi-blind watermarking scheme described in the preceding section, we give an invisible watermarking scheme derived from the same. An invisible watermarking scheme essentially contains a *scrambled watermark* encoded in the original image that can be made visible only with a key. Ideally, an invisible watermarking scheme should have the following properties:

> *Un-forgeability*: No adversary should be able to forge a watermark with an key, id', that the owner (of the original image) has not watermarked, such that,
>
> $$Pr[\mathcal{A}\big(A_w, id\big) = (A'_w, id')] = negligible$$
>
> *Non-repudiation*: Once a signer (owner of the original image) has signed or watermarked an image with some key id, the signer cannot repudiate the associate id, such that,
>
> $$Pr[\mathcal{A}\big(A_w, id) = (A_w, id')] = negligible$$
>
> *Invisibility*: The watermarked image, without the knowledge of the key, id it was watermarked with, should be indistinguishable from an image containing a watermark of white noise from an uniform distribution, such that,
>
> $$Pr[\mathcal{D}\big(A_w\big) = 1] - Pr[\mathcal{D}\big(A_U\big) = 1] = negligible$$


(a) (b) (c)

(d) (e) (f)

Fig. 1: (a) Original Lena image; (b) Original Baboon image; (c) Original Plane image; (d) Watermarked image, baboon embedded in Lena image; (e) Watermark extracted from (d) using the proposed scheme; (f) Distorted reference Plane image extracted from (d).

where, $\mathcal{A}$ and $\mathcal{D}$ are polynomial time adversaries and distinguishers respectively. A_w is an image watermarked with key, id, A_U is an image containing a watermark of white noise from an uniform distribution, $id' \neq id$ and $A'_w \neq A_w$.

The key idea we used here, is to commit a $M \times N$ dimensional watermark image W to a unique id (e.g. derived from the customer's name). The id, for security reasons, must have some random nonce in it.

To do so, we use a collision resistant cryptographic hash function with polynomial span, H, for generating an integer string $h \in \{0, \dots, 255\}^{M*N} \leftarrow H(id)$, and then convert h to a $M \times N$ matrix h_{id} using a row-major format. Informally, our hash function (of polynomial span) is defined as $H : \{0,1\}^* \rightarrow \{0, \dots, 255\}^{M*N}$, such that the following properties hold:

(1) It is computationally hard to find a pre-image, i.e., given $y = H(x)$, to compute x.

(2) It is computationally hard to find a collision, i.e., given $H(x)$, to compute x' such that $H(x') = H(x)$

(3) It is indistinguishable from a random number, i.e., given $H(x)$ and $y \overset{R}{\leftarrow} \{0, \dots, 255\}^{|H(x)|}$, hard to distinguish w.p. $> 1/2$

Finally, an element-wise XOR of the matrix A_{wa}, as in eq. 11 and h_{id}, to obtain the $A^{'}_{wa}$ is performed. Formally, given original image A, a watermark W, and a identity id, we introduce a watermark to the image by first decomposing:

$$A = USV^T,$$

$$W = U_w S_w V_w^T = A_{wa} V_w^T,$$

Subsequently, we encode the watermark on to the image as:

$$S_1 = S + \alpha\Big(A_{wa} \oplus_e h_{id}\Big) \tag{9}$$

Here α is the scaling factor, and $\oplus_e$ is the element-wise XOR, defined as $A \oplus_e B = A_{i,j} \oplus B_{i,j}; \forall i, j$.

$$US_1V^T = A_w, \tag{10}$$

After executing the above computations, we get A_w which contains a unique watermark embedded in A. To extract and verify the invisible watermark, from the image A_w, we essentially make use of the unique id (used earlier to introduce the watermark), and to extract the watermark W from A_w as:

$$A_w - A = A_1, \tag{11}$$

$$\big(U^{-1}A_1V^{T^{-1}}\big)/\alpha = \big(A_{wa} \oplus_e h_{id}\big), \tag{12}$$

$$\big(A_{wa} \oplus_e h_{id}\big) \oplus_e h_{id} = A_{wa}, \tag{13}$$

$$A_{wa}V_w^T = W, \tag{14}$$

3.1 Security Analysis

THEOREM: If H is a collision resistant secure hash function and the semi-blind watermarking scheme (from Section 2) is secure, then the hash code watermarking scheme is secure.

LEMMA 1: If H is a collision resistant secure hash function and the semi-blind watermarking scheme (from Section 2) is secure, then the hash code watermarking scheme is unforgeable.

Proof Sketch: (by contradiction) Suppose $\exists adversary, \mathcal{A}$ which can forge a watermark to produce A'_w using some id' that the owner did not watermark. Then we can either use $\mathcal{A}$ as a subroutine of $\mathcal{B}$ that finds a collision such that $H(id) = H(id')$ or we can use $\mathcal{A}$ as a subroutine of $\mathcal{C}$ that can find a pre-image of the hash function. It is straightforward to construct $\mathcal{B}$. $\mathcal{B}$ queries $q_{i \in poly(n)}$ to the hash oracle $\mathcal{H}$, that returns $H(q_i)$. $\mathcal{B}$ simply uses q_i as key, to watermark different images and give them to $\mathcal{A}$. If $\mathcal{A}$ can produce an key $id \notin q_i$, such that $H(id) = H(q_i)$ for some i, then $\mathcal{B}$ simply returns the $H(id), id'$ to win the game.

One can construct the pre-image finding game as follows: $\mathcal{C}$ queries $q_{i \in poly(n)}$ to the hash oracle $\mathcal{H}$, that returns $H(q_i)$. $\mathcal{C}$ simply hash codes all watermarks with $id = q_i$ and gives them to $\mathcal{A}$. Finally when $\mathcal{A}$ returns a different $id' \neq \{q_i\}_{\forall i}$ and possibly different A_w forgery $\mathcal{C}$ computes:
$A'_w - A = A_1; U^{-1}A_1V^{T^{-1}} = A_{wa} \oplus H(id')$
Following that, $\mathcal{C}$ computes $A_{wa} \oplus H(id') \oplus A_{wa} = H(id')$ and returns $\big(H(id'), id'\big)$ to the hash oracle to win the game. However, since we assumed our hash function is pre-image resistant, this is a contradiction.

LEMMA 2: If H is a collision resistant secure hash function and the semi-blind watermarking scheme (from Section 2) is secure, then the hash code watermarking scheme has the property of non-repudiation.

Proof Sketch: (by contradiction) Suppose $\exists adversary, \mathcal{A}$ which can repudiate a watermark A_w signed by some key id with a different key id', then we can use $\mathcal{A}$ as a subroutine of $\mathcal{B}$ to find a collision in the hash function.

The construction is straightforward (as in Lemma 1) where $\mathcal{B}$ queries the hash oracle $\{q_i\}_{i \in poly(n)}$ to obtain $H(q_i), \forall i$. $\mathcal{B}$ watermarks all images with $id = q_i$ and gives them to $\mathcal{A}$. If $\mathcal{A}$ can successfully return any pair $\big(id, id'\big)$, such that $H(id) = H(id')$, $\mathcal{B}$ returns the $\big(id, id'\big)$ to the hash oracle to win the collision finding game. However, since we assumed our hash function is collision resistant, this is a contradiction.

LEMMA 3: If H is a collision resistant secure hash function and the semi-blind watermarking scheme (from Section 2) is secure, then the hash code watermarking scheme is invisible.
Proof Sketch: (by contradiction) Suppose $\exists Distinguisher, \mathcal{D}$ which can distinguish a watermark A_w from A_U (A_U is an image contain a watermark of white noise from an uniform distribution) in polynomial time, then we can use $\mathcal{D}$ as a subroutine of $\mathcal{B}$ to win the distinguishability game.

We construct the distinguishability game as follows: $\mathcal{B}$ queries $q_i \in poly(n)$ to the hash oracle $\mathcal{H}$, that returns $H(q_i)$. $\mathcal{B}$ then hash codes all watermark with $id = q_i$ and gives them to $\mathcal{D}$. Finally as challenge, the hash oracles gives $\big(H(id), U_r \overset{R}{\leftarrow} \{0, \ldots, 255\}^{M*N}\big)$ to $\mathcal{B}$. $\mathcal{B}$ then creates two watermarks $A_w^{(1)}$ and $A_w^{(2)}$ with $H(id)$ and U_r respectively and gives it to $\mathcal{D}$. $\mathcal{D}$ can distinguish the image containing the real watermark with probability more than $1/2$, and returns the correct watermark $A_w^{(b)}$ to $\mathcal{B}$. $\mathcal{B}$ can simply return the hashed string corresponding to $A_w^{(b)}$ to the hash oracle to win the game. However, since we assumed the output of our hash function is indistinguishable from a random sequence, this is a contradiction.

Proof of Theorem: From *Lemma 1*, *Lemma 2* and *Lemma 3*, we can assert that the hash code watermarking scheme is secure for any arbitrary image.

4. EXTENSION TO COLOR IMAGES

In the previous sections, we described two constructions that introduces a watermark or fingerprint to greyscale digital images, which are $M \times N$ matrices. This construction can be easily extended to support color images with $M \times N \times 3$ as well. One common used approach is marking only the luminance component of the image, where the greyscale leveling techniques described in the previous sections is straightforward to apply.

The luminance component of a 3-dimensional matrix with R, G, B channels is computed as:

$$L = M + m \tag{15}$$

where, $M = max(R, G, B)$ and $m = min(R, G, B)$
Yet another technique, put forward by [15] proposes embedding the watermark on the blue channel, since the human eye is sensitive to this band. Also another straightforward technique is to watermark each channel individually. This can be naturally used to watermark either mono-channel or multichannel images on multichannel cover images.

5. CONCLUSION

We have presented two watermarking schemes based on SVD, a semi-blind watermarking scheme and an invisible hash code based watermarking scheme. The security of the first scheme relies on the fact that the information about the entire watermark is not available without a prior knowledge of the original watermark, and that the principal components along with their singular values are embedded in the watermark. These two ideas help avoid common pitfalls for the case of the semi-blind watermarking scheme. A construction of new invisible hash code based watermarking scheme derived from the proposed semi-blind watermarking scheme has also been proposed here, whose security is proved in the random oracle model and shown to be unforgeable, invisible and satisfying the property of non-repudiation. Lastly, straightforward extensions of the aforementioned schemes for multichannel (color) images are discussed. Given that the SVD is a one way decomposition, and based on the constructions and security proofs described in the paper, we conclude that our methods help ensure rightful ownership of the digitally watermarked image.

6. REFERENCES

[1]. Golub, G.H., Reinsch, C., 1970. Singular value decomposition and least squares solutions. Numerische Mathematik, 14, 403-420.

[2]. Chang, C.C., Tsai, P., Lin, C.C., 2005. SVD-based digital image watermarking scheme. Pattern Recognition Letters, 26, 1577-1586.

[3]. Agarwal, R., Santhanam, M.S., 2008. Digital watermarking in the singular vector domain. International Journal of Image and Graphics, 8, 351-368.

[4]. Lei, B., Tan, E.L., Chen, S., Ni, D., Wang, T., Lei, H., 2014. Reversible watermarking scheme for medical image based on differential evolution. Expert Systems with Applications, 41, 3178-3188.

[5]. Calagna, M., Guo, H., Mancini, L.V., Jajodia, S., 2006. A robust watermarking system based on SVD compression. Proceedings of the 2006 ACM symposium on Applied computing, 1341-1347.

[6]. Bergman, C., Davidson, J., 2005. Unitary embedding for data hiding with the SVD. Electronic Imaging 2005, International Society for Optics and Photonics, 619-630.

[7]. Lai, C.C., 2011. An improved SVD-based watermarking scheme using human visual characteristics. Optics Communications, 284, 938-944.

[8]. Quan, L., Qingsong, A.I., 2004. A combination of DCT-based and SVD-based watermarking scheme. International Conference on Signal Processing, 873-876.

[9]. Zhou, Y., Jin, W., 2011. A novel image zero-watermarking scheme based on DWT-SVD. IEEE International Conference on Multimedia Technology, 2873-2876.

[10]. Potfode, A., Kourav, D., 2016. Digital color image watermarking using DWT and SVD for data security. International Journal of Computer Applications, 141, 17-20.

[11]. Tsai, H.H., Jhuang, Y.J., Lai, Y.S., 2012. An SVD-based image watermarking in wavelet domain using SVR and PSO. Applied Soft Computing, 12, 2442-2453.

[12]. Pandey, P., Kumar, S., Singh, S.K., 2014. Rightful ownership through image adaptive DWT-SVD watermarking algorithm and perceptual tweaking. Multimedia Tools and Applications, 72, 723-748.

[13]. Khorrami, N., Ayubi, P., Behnia, S., Ayubi, J., 2014. A svd-chaos digital image watermarking scheme based on multiple chaotic system. Signal Processing and Information Technology, 9-18.

[14]. Arora, S., Acharya, J., Verma, A., Panigrahi, P.K., 2008. Multilevel thresholding for image segmentation through a fast statistical recursive algorithm. Pattern Recognition Letters, 29, 119-125.

[15]. Kutter, M., Jordan, F.D., Bossen, F., 1997. Digital signature of color images using amplitude modulation. Electronic Imaging, International Society for Optics and Photonics, 518-526.