A NOVEL DIGITAL IMAGE WATERMARKING SCHEME USING BLIND SOURCE SEPARATION

Thang V. Nguyen, Jagdish C. Patra and Narendra S. Chaudhari

School of Computer Engineering, Nanyang Technological University, Singapore 639798
thangnguyen@pmail.ntu.edu.sg, ASpatra@ntu.edu.sg, ASNarendra@ntu.edu.sg

ABSTRACT

It is not very convenient to bring along the original image as well as the watermark whenever one need to run the watermark extraction for ownership verification or image authentication. We introduce a novel watermarking scheme by taking the advantage of Blind Source Separation technique. Using only a single 'key image' which can be issued to public, the new watermarking algorithm is able to extract the watermark without requiring the original image and any information of the watermark. The new method, undergoing a variety of experiments, has shown a good performance against many salient attacks.

1. INTRODUCTION

Digital Watermarking, in which a watermark is embedded directly and imperceptibly into digital data to form watermarked data, is one of the most effective techniques to protect digital works from piracy [1]. To estimate the watermark, many algorithms require the original data and the watermark. It is quite inconvenient since the original data should be kept secret and the mark is varied from work to work. In this paper, we introduce a novel watermarking method using Blind Source Separation (BSS) technique that does not require the original data as well as the watermark.

The BSS is an important technique for blindly estimating unknown signals from their observed mixtures [2]. When applied to watermarking, BSS presumes the watermarked data as mixtures of original data and watermark, and does blind separation on it to estimate the watermark. In [3], the authors propose a method that partitions off the original image and the watermark into independent components (ICs), and then combining these ICs to produce watermarked image. This technique, however, requires a lot of computation and usually fails in brutal attacks. Another approach considers the watermarked data as a mixture of the whole host data and the watermark, and manages to generate the other mixtures from the available data. This approach is simple to implement but usually need additional knowledge about the original data or the watermark. For example, in [4], the algorithm needs both secret key and the original image. In [5], the proposed technique requires one more additional watermark and has to process on wavelet domain which is somewhat more complex.

Our novel watermarking method follows the advantage second approach, but further exploit the two-dimensional characteristic of an image to overcome the requirement of additional information while keeping the algorithm simple. The idea is, image I and its transpose I^T can be considered as two independent sources for BSS. Comparing with other watermarking techniques, our proposed method has the following advantages: (1) the 'key image', which can be publicly available, is used in the extraction instead of the original image; (2) no knowledge of watermark is needed in the extraction. (3) the watermark is not limited to some specific format; and (4) robust against many salient attacks.

2. BSS AND WATERMARKING

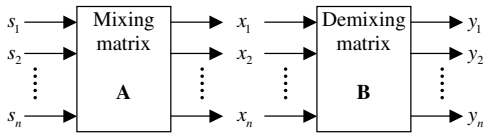

Fig. 1. The BSS mixing and demixing model.

A BSS model (Fig.1) includes two sub-models: mixing model and demixing model. The mixing model, in which the observed signals $\mathbf{x} = [x_1, x_2, ..., x_n]^T$ are assumed to be linear mixtures of n unknown statistically independent source signals $\mathbf{s} = [s_1, s_2, ..., s_n]^T$, can be expressed in matrix form with a mixing matrix $\mathbf{A}_{n\times n}$ as

$$\mathbf{x} = \mathbf{A}\mathbf{s} \tag{1}$$

In demixing model, the unknown sources $\mathbf{s}$ are estimated from the observations $\mathbf{x}$ by maximizing the statistical independent criteria among the outputs $\mathbf{y} = [y_1, ..., y_n]^T$. When converged, the outputs $\mathbf{y}$ will be a permutation of the unknown sources $\mathbf{s}$. The demixing model is formulated as

$$\mathbf{y} = \mathbf{B}\mathbf{x} \tag{2}$$

0-7803-8554-3/04/$20.00 ©2004 IEEE.

where $\mathbf{B}$ is called demixing matrix. Many BSS algorithms have been developed to estimate the demixing matrix $\mathbf{B}$ and the outputs $\mathbf{y}$. More detail on BSS can be found in [2].

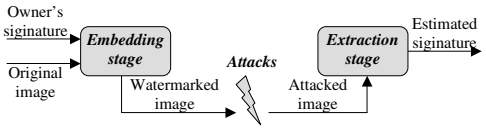

Fig. 2. Scheme of the watermarking problem.

Similarity between watermarking and BSS can be seen from Fig.1 and Fig.2. The original image and the watermark are viewed as unknown independent sources and the watermarked images as their mixtures. The embedding stage and extraction stage are then the BSS mixing and demixing models, respectively. Thus, watermarking, from this particular point of view, can be considered as a BSS problem.

3. THE WATERMARK EMBEDDING SCHEME

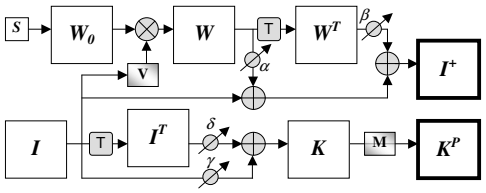

Fig. 3. The proposed embedding scheme.

In embedding scheme, as shown in Fig.3, the watermarked image I^+ is directly computed from the watermark W and the original image I though the following steps:

1. Generate the initial watermark W_0 by duplicating the author signature S up to the size of the image I.

2. Modify the initial watermark W_0 with the help of a visual mask V to get the mark $W = (1 - V)W_0$.

3. Calculate the watermark transpose W^T. Embed W and W^T into I to generate I^+

$$I^+ = I + \alpha W + \beta W^T. \tag{3}$$

4. Calculate 'key image' K from I and I^T. Generate a random mask M from secret key k^s, use M to make a 'public image' K^P

$$K = \gamma I + \delta I^T \tag{4}$$
$$K^P = M \circ K. \tag{5}$$

where $\circ$ is the element-by-element product.

Parameters α and β are called embedding strengths and γ and δ are the 'key image' coefficients. These parameters can be any non-zero values in the range of $[-1, 1]$. The 'public image' K^P can be made public while the secret key k^s is kept for the authorized people only.

Noise Visibility Function (NVF) [6] is applied as the visual mask V to identify the significant areas where the watermark can be more strongly embedded. With visual mask, the watermark strength is increased considerably while the image quality and watermark invisibility are maintained .

4. THE WATERMARK EXTRACTION SCHEME

The following scheme (Fig.4) describes the process to create four mixtures from the test image I^* (attacked image), 'public image' K^P and secret key k^s, then applies BSS algorithm to extract the hidden watermark.

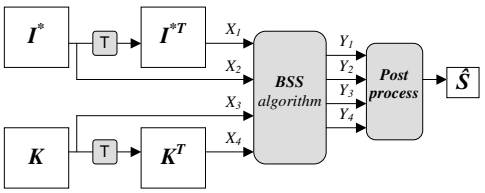

Fig. 4. The BSS-based extraction scheme.

1. Use k^s to regenerate the mask M. Retrieve the key image K by using the formula $K = K^P \circ / M$ (where $\circ /$ represents an element-by-element division). From I^* and K, two transposes I^{*T} and K^T are created. $[I^*, I^{*T}, K, K^T]$ is then served as the input of the BSS algorithm (the $[X_1, X_{,2}, X_3, X_4]$ in Fig.4).

2. Apply BSS algorithm to extract four images $Y_1, ..., Y_4$.

3. Run the post process algorithm on $Y_1, ..., Y_4$ to produce the estimated signature $\hat{S}$.

The watermark is derived from the BSS outputs $Y_i (i = 1..4)$ by using the correlation coefficient between Y_i and I^*. The correlation coefficient r between two image $X_{M\times N}$ and $Y_{M\times N}$ is defined as

$$r = \frac{\sum_i \sum_j (x_{(i,j)} - \bar{x})(y_{(i,j)} - \bar{y})}{\sqrt{(\sum_i \sum_j (x_{(i,j)} - \bar{x})^2)(\sum_i \sum_j (y_{(i,j)} - \bar{y})^2)}} \tag{6}$$

where $\bar{x} = \frac{1}{MN} \sum_i \sum_j x_{(i,j)}$, $\bar{y} = \frac{1}{MN} \sum_i \sum_j y_{(i,j)}$, $i = 1, 2, ..., 4$ and $j = 1, 2, ..., 4$.

Let's assume that $Y_1, ..., Y_4$ are the corresponding estimates of I, I^T, W and W^T. And let r_i and r'_i $(i = 1, 2, 3, 4)$ be the correlation coefficients between Y_i and I^*, and between Y_i and I^{*T}, respectively. Since I^* is highly correlated to I, the correlation coefficient r_1 will have a high

value ($|r_1| \approx 1$). Similarly, we have $|r'_2| \approx 1$. On the contrary, both r_k and r'_k ($k = 3, 4$) are nearly zero, since W and W^T are not related to either I^* or I^{*T}. Thus, by selecting the two outputs those have lowest value of $|r_i| + |r'_i|$, we can generate the estimated watermark $\hat{W}$ by the formula $\hat{W} = (Z_1 + Z_2^T)/2$ (see Fig.5). Z_1 and Z_2 represent for two selected outputs (Y_3 and Y_4 as in this case).

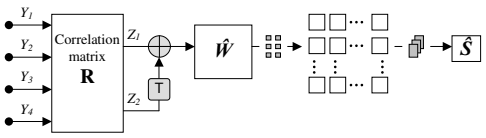

Fig. 5. The post process scheme.

Assuming the size $M_S \times N_S$ of the signature S is known, we can retrieve a better estimate of S from $\hat{W}$ by applying a refining process. The image $\hat{W}$ is partitioned into n small images of size $M_S \times N_S$, $\hat{w}_1, ...\hat{w}_n$, and thereafter, these small images are summed up to get the final result

$$\hat{S} = (\hat{w}_1 + \hat{w}_1 + \ldots + \hat{w}_n)/n. \quad (7)$$

Detailed explanation of the scheme may be found in [7].

5. PERFORMANCE ANALYSIS

Three experiment were simulated to study the robustness of the proposed method, and its effect on the image and watermark. The first experiment was on a medium-textured Lena (512x512) image with a black and white (b/w) university's logo (128x128) as the watermark. The second experiment was on the same image but with a simple 'NTU' (64x64) watermark. The third one used the 'NTU' watermark but on a highly-textured Baboon (512x512) image.

The embedding strengths α and β were controlled so that the watermarked images have high quality in term of the Peak Signal-to-Noise Ratio ($PSNR$). $PSNR$ between an image $X_{M\times N}$ and its modification $\hat{X}_{M\times N}$ is

$$PSNR = 20log_{10}\frac{255}{\sqrt{\frac{1}{MN}\sum_i\sum_j(x_{(i,j)} - \hat{x}_{(i,j)})^2}} \quad (8)$$

where $i = 1, ..., M$, $j = 1, ..., N$. The parameters of the test configuration are provided in Table 1.

Table 1. Configuration of the three experiments.

	$I + W$	α	β	$PSNR(dB)$
Exp.1	Lena+Logo	0.073	-0.01	41.45
Exp.2	Lena+NTU	0.06	0.015	43.99
Exp.3	Baboon+NTU	0.04	-0.01	42.91

With properly selected embedding strengths, the watermarked images are almost identical to the original images ($PSNR > 40dB$), the embedded marks are invisible to normal view. The values of γ and δ were -0.73 and 0.4, respectively, for all the experiments. Fig.6 shows the signature S, the watermark W, the public image K^P and the watermarked image I^+ used in Exp.1 and Exp.3.

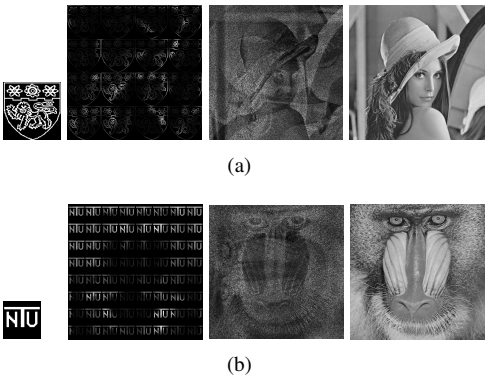

Fig. 6. From left to right: The signatures S, the watermark W, the public key K^P and the watermarked image I^+. (a) Experiment 1. (b) Experiment 3.

The Second Order Blind Identification (SOBI) algorithm [8] was used for source separation. The correlation coefficient r (Eq.6) between the estimated signature $\hat{S}$ and the original signature S was used as the performance index for comparing the experimental results.

5.1. JPEG compression test

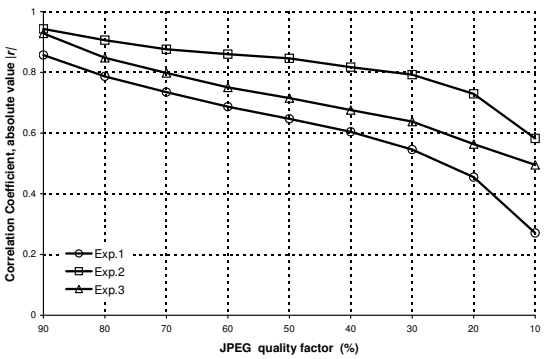

Fig. 7. Comparison of results against JPEG compression.

The proposed scheme provided very good performance on all simulated experiments. The quality of the estimated signatures remains high even when the JPEG compression quality factor is reduced awfully (Fig.7). Only in the first experiment where the watermark is complex, and only when the compression quality factor is lowered to 10% that the estimated signature is considered unrecognizable ($|r| < 0.4$). In this test, the medium-textured Lena image embedded by

a simple watermark (Exp.2) seems to be the most resistant against the compression. The poorer performance of the other two experiments is probably because of the quantization in the compression. It destroyed the details in the texture region where the watermark is mostly embedded.

5.2. Gray level reduction test

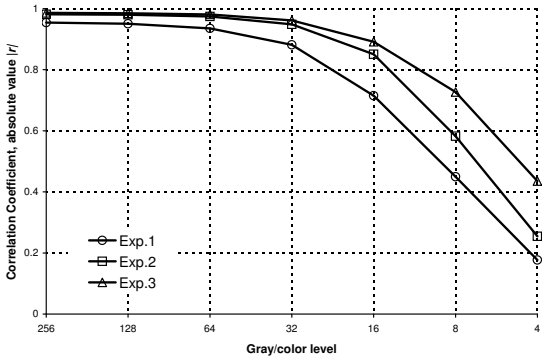

Fig. 8. Comparison of results against gray level reduction.

The proposed algorithm continues to offer excellent results in the next attack, the gray level reduction test. The values of the performance index in all simulations are close to each other and the estimated signatures are highly correlated to the original ones (Fig.8). For all the three experiments, the quality of the extracted signature is quite good ($|r| > 0.4$) when the image gray level reduced from 256 to 8. The first experiment's results are inferior to the others because of the complicated structure of the signature.

5.3. Image resizing test

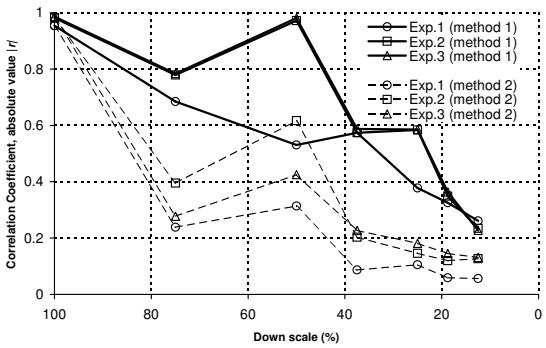

Fig. 9. Comparison of results against image resizing.

Since this test modifies the image size, we need to synchronize the size of the 'key image' K and the test image I^* before watermark extraction. For this, we propose two synchronization methods. In the first method, we scale the key image K to the size of I^*, then use the newly-scaled image K^* and I^* as inputs in the extraction scheme. The estimated watermark $\hat{W}$ is then re-scaled to the size of K before applying the refining process. Whereas, in the second method, the test image I^* is resized to the key image's size and the extraction process is executed normally. As it is shown in Fig.9, the performance of the first method is much better than the second one. The quality of the estimated signature is quite good ($|r| > 0.4$) upto a downscale of 25%, for all the three experiments.

6. CONCLUSION

A new simple watermarking method using only a single key image for extracting have been introduced. Besides its unique characteristics, the method is robust against many striking attacks and provide excellent results.

Since the 'public key' K^P plays an important role in the security of the method, the users may apply some other more sophisticate encoding algorithms (such as the public key encoding method) to replace the random M in (5).

We are working on applying the method to embed multiple watermarks. The use of more than one mark will help to improve security of the watermarked image and make the marks harder to be identified and removed by an unauthorized person. However, it may degrade the quality of the watermarked image.

Finally, another modified version is being developed to reduce the amount of side information, i.e. the 'key image'. The new method only requires a small-sized 'key image' as the support information.

7. REFERENCES

[1] I. J. Cox, M. L. Miller, and J. A. Bloom, *Digital Watermarking*, Morgan Kaufmann, 1st edition, Oct. 2001.

[2] A. Cichoki and S. Amari, *Adaptive blind signal and image processing*, John Wiley & Sons Ltd, 2002.

[3] S. Zhang and P. K. Rajan, "Independent component analysis of digital image watermarking," in *Proc. of IEEE International Symposium on Circuits and Systems (ISCAS'02)*, May 2002, vol. 3, pp. 217–220.

[4] M. Shen, X. Zhang, L. Sun, P. J. Beadle, and F. H. Y. Chan, "A method for digital image watermarking using ICA," in *Proc. 4th International Symposium on Independent Component Analysis and Blind Signal Separation (ICA2003)*, Nara, Japan, Apr. 2003, pp. 209–214.

[5] D. Yu and F. Sattar, "A new blind image watermarking technique based on independent component analysis," *Springer-Verlag Lecture Notes in Computer Science*, vol. 2613, pp. 51–63, Jan. 2003.

[6] S. Voloshynovskiy, A. Herrigel, N. Baumgaertner, and T. Pun, "A stochastic approach to content adaptive digital imagewatermarking," in *Proc. of International Workshop on Information Hiding*, Dresden. Germany, Oct. 1999, pp. 212–236.

[7] T. V. Nguyen and J. C. Patra, "Blind source separation-based digital image watermarking," Submitted to EURASIP Journal on Applied Signal Processing, 2004.

[8] A. Belouchrani, K. Abed-Meraim, J. F. Cardoso, and E. Moulines, "A blind source separation technique using second-order statistics," *IEEE Trans. Signal Processing*, vol. 45, no. 02, pp. 434–444, Feb. 1997.