IMPROVED IMAGE AUTHENTICATION USING CLOSED-FORM COMPENSATION AND SPREAD-SPECTRUM WATERMARKING

Sufyan Ababneh, Rashid Ansari, Fellow, IEEE, Ashfaq Khokhar, Senior Member, IEEE

Dept. of Electrical and Computer Engineering, University of Illinois at Chicago, USA

ABSTRACT

This paper presents an image authentication scheme based on compensated watermarking employing a Lagrangian-based closed-form solution to compensate for signature perturbation due to the embedding operation. The proposed scheme uses a spread-spectrum based watermarking technique and a blind detector, thus making it attractive for applications that may not have the original image available at the time of authentication. Existing compensated signature embedding frameworks use an iterative mechanism to reach a desired compensation. The iterative approach is time consuming and less effective than the closed-form approach proposed in this paper, which performs an accurate compensation in one step while meeting the minimum distortion criteria of *image least mean square distortion* to guarantee image fidelity. Simulation results are presented to show the proposed scheme's efficiency and accuracy.

Index Terms— Authentication, Spread spectrum, Watermarking, Compensated signature embedding

1. INTRODUCTION

Due to the fast growing usage of digital multimedia and the availability of many tools to easily manipulate media content, integrity verification and authentication has become an essential requirement for many applications. One of the promising approaches for multimedia content authentication is by using data-hiding or watermarking [1][2][3]. To this end, robust watermarking techniques can be used to embed content-based fragile signatures in the media signal. When the content is tampered or manipulated, the authentication system can extract the embedded fragile signature and reveal tampering details by comparing the extracted original signature with a new signature generated from the media signal being authenticated.

Digital multimedia authentication using compensated signature embedding (CSE) has been introduced in our earlier works [4][5]. In CSE, a compensation operation is performed in order to make up for the signal alteration resulting from signature embedding. The compensation operation can be conducted using an *iterative* mechanism or employing a *closed form* solution as long as the fidelity of the media signal is maintained and the embedded signature is not altered or lost. So far, only an iterative solution has been provided [4]. Since signature embedding compensation is a fundamental part of the CSE framework, it is essential to provide a fast, deterministic and reliable way to conduct this operation. In this paper, we provide a fast closed-form one-step compensation solution that does not require any iterations. We consider this as an important improvement that demonstrate the validity and effectiveness of the closed-form compensation approach which is superior to the iterative approach in terms of speed and in providing a deterministic criteria to control compensation without impacting image quality or fidelity.

Another important aspect in the CSE framework is using a watermarking technique that is robust and flexible. In [4], a quantization based watermarking (i.e. QIM-DM) technique was used. In this paper we demonstrate the closed-form CSE authentication using a robust spread-spectrum wavelet-based watermarking technique [6][7]. Unlike many other spread-spectrum techniques that require the original image at the detector (e.g. [8]), the technique we use does not require an informed-detector. Instead, a blind-detector is used which does not need the original image for watermark detection. Blind detection is often a requirement for digital multimedia authentication and many other applications in which the original image is not available at the detector.

2. COMPENSATED SIGNATURE EMBEDDING

In this section, we briefly review the compensated signature embedding (CSE) [4]. The CSE system is depicted in Figure 1, it consists of an encoder and a decoder units. The encoder generates a fragile signature, embeds a robust watermark and compensates for the perturbation due to embedding. The decoder extracts the embedded signature, generates a new signature and evaluates the results. All these functionalities operate on a signal which may be the raw image itself or its transformed version. The following notation provides a formal description.

The set of integers is denoted by $\mathcal{Z}$. For any positive integer K, let $I_K = \{k \in \mathcal{Z} : 0 \leq k \leq K-1\}$. The set D denotes the domain of the signal. Let M and N be positive integers that are multiples of 2^L, where L is also a positive integer. Define $M_l = M/2^l$ and $N_l = N/2^l$, for $l \in \mathcal{Z}$, $1 \leq l \leq L$. For a raw $M \times N$ image, the set D can be chosen to be $D^o = I_M \times I_N$. We are interested in the subband/wavelet representation of the image using L-level decomposition, where $L > 0$. Although $D = D^o$ defined for the raw image will work, it is defined differently in order to capture the structure of the wavelet decomposition at different levels. In this case,

$$D = \{\mathbf{n} = (l, n_l, i, j) : l - 1 \in I_L, n_l \in I_{3+\lambda}, i \in I_{M_l}, j \in I_{N_l}\}, \quad (1)$$

where $\lambda = \delta_{lL}$, where δ_{kl} is the Kronecker's delta. The signal w (i.e. the image or its transformed version) is defined as a mapping $w : D \longrightarrow \mathbb{R}$, with some additional structure such as bounded real valued signal. Let S denote the set of these signals. The signature can be generated using some or all of the signal samples. It is desired not to exclude from the signature generation process those samples that are used for embedding since the latter could be a substantial fraction of all samples. The signature is then embedded by modifying a subset of the samples. The embedding usually perturbs the signature. Therefore, a different subset of the sample values is adjusted to compensate for the signature perturbation due to embedding. Some more notation is introduced to explain this pro-

cess. Let $D_1 \subseteq D$ denote the subset of the signal domain used for generating the signature. As mentioned before, in our study we set $D_1 = D$. The domains of embedding and compensation are denoted by $D_2 \subset D$ and $D_3 \subset D$, respectively. The cardinality of the set D_j is denoted by N_j, $j = 1, 2, 3$. Preferably, D_2 and D_3 are disjoint, i.e. $D_2 \cap D_3 = \phi$. The signals w_1, w_2, and w_3, defined as the restrictions of w to D_1, D_2, and D_3, respectively, are used to conduct the system's operations for signature generation and embedding. Here $w_j \in S_j$.

The operations of signature generation, embedding, and compensation, are represented by σ_1, σ_2, and σ_3, respectively. The operation σ_j, $j = 1, 2, 3$, uses w_j with an optional key $\{k_j\}$ from a set K_j to support a secure operation, and an optional parameter p_j from a set P_j that could support user preferences, such as the level of robustness or other performance measures. An example of using a key for added security is to define options in using w_j for the operation σ_j based on a private key from a set K_j. The signature generation operation σ_1 creates a vector $\mathbf{F}$ of n_1 bits. These bits represent a fragile signature that is obtained from suitable signal features such as signal energy or coefficient histogram. The operation can be defined by $\sigma_1 : S_1 \times K_1 \times P_1 \longrightarrow \{0,1\}^{n_1}$, and $\mathbf{F} = \sigma_1(w_1, k_1, p_1)$. The signature embedding operation σ_2 consists of modifying the signal samples over the embedding domain with a user-defined process such that the corresponding extraction process performed on the signal containing the embedded signature vector reproduces the vector $\mathbf{F}$. Again, optional keys and performance parameters may be included in the operation. The operation σ_2 represents the embedding, and it is defined by $\sigma_2 : S_2 \times K_2 \times P_2 \times \{0,1\}^{n_1} \longrightarrow S_2$, where the signal $\hat{w}_2 = \sigma_2(w_2, k_2, p_2, \mathbf{F})$ with the embedded signature yields $\mathbf{F}$ upon signature extraction. The extracted signature $\bar{\mathbf{F}}$ is obtained through the signature extraction procedure ψ_2, $\bar{\mathbf{F}} = \psi_2(\hat{w}_2, k_2, p_2)$, where $\hat{w}_2$ is obtained as a result of introducing embedding noise to w_2. Since the embedding technique is robust to common tampering (i.e. the embedded signature can be extracted without errors even after applying these operations), this means that $\bar{\mathbf{F}} = \mathbf{F}$. After performing the compensation process, the new signature $\hat{\mathbf{F}}$ is generated again by operating on $\hat{w}_1$, where $\hat{w}_1$ is the modified w_1 as a result of adjusting w_3 and $\hat{\mathbf{F}} = \sigma_1(\hat{w}_1, k_1, p_1)$. The embedding compensation operation σ_3 consists of modifying the signal samples over the embedding compensation domain D_3. $\sigma_3 : S_3 \times K_3 \times P_3 \times \{0,1\}^{n_1} \times \{0,1\}^{n_1} \longrightarrow S_3$, where the signal $\hat{w}_3 = \sigma_3(w_3, k_3, p_3, \mathbf{F}, \hat{\mathbf{F}})$ represents the sample values after compensation. Since $D_3 \subset D$ and $D_1 = D$, the signal changes that led to $\hat{w}_3$ will also lead to changing w_1 to $\hat{w}_1$. The goal of the compensation operation σ_3 is to obtain $\hat{w}_3$ such that the newly generated signature $\hat{\mathbf{F}}$ is identical to the embedded signature $\mathbf{F}$ (i.e. $\hat{\mathbf{F}} = \mathbf{F}$). The fragility of the signature means that a minor distortion introduced in the signal will lead to a different signature (i.e. $\hat{\mathbf{F}} \neq \mathbf{F}$)

3. IMPROVED IMAGE AUTHENTICATION

The improved image authentication system presented here is a demonstration of how a closed-form compensation can be applied in the context of the CSE platform. Any CSE closed-form solution is dependant on the features used for signature generation. In this paper, as also in [4], the signal energy is used to generate the signature. We first describe the signature generation operation and then we describe the spread-spectrum embedding operation. After that, we describe the closed-form compensation. We will adopt the notation in terms of 1-D vectors. If $\mathbf{v}^* \in \mathbb{R}^{MN}$ denote a given array of image wavelet coefficients, we define $\mathbf{v} = vec(\mathbf{v}^*) \in \mathbb{R}^{MN}$.

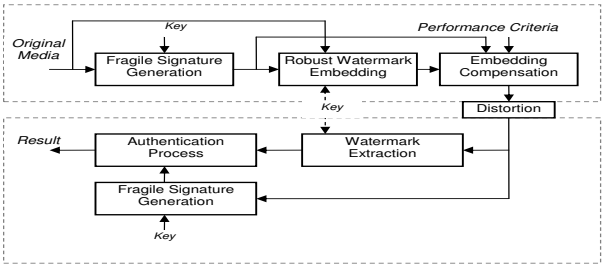


Fig. 1. CSE block diagram

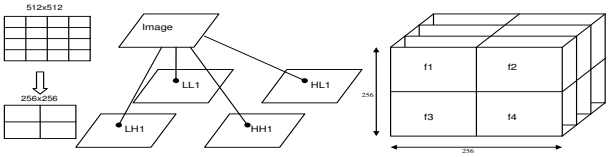


Fig. 2. DWT decomposition and signature generation

we use $\mathbf{v}$ to denote the vector of original wavelet coefficients (i.e. before watermarking), $\mathbf{v}'$ to denote the vector of wavelet coefficients immediately after watermarking and $\hat{\mathbf{v}}$ to denote the vector of desired wavelet coefficients after compensation.

3.1. Signature Generation

The signature generation is illustrated in Figure 2 for a 1-level discrete wavelet transform (DWT) which generates four subbands (LH_1, HL_1, HH_1 and LL_1). Each subband is a two-dimensional array of wavelet coefficients of a certain spatial frequency range. The subbands spatially represent a reduced version of the original image. For example, in the case of 512×512 image, the subband planes take the dimension of 256×256. The signature generation process is performed by operating on samples ($v_i \in \mathbb{R}$) obtained from $\mathbf{v}$. The samples could be selected based on a security key, but for illustration purposes, we choose all the samples. To generate a content-based fragile signature with localization support, we stack the four wavelet planes and partition them to 4 blocks $j = 1, \ldots, 4$. Choosing the number of blocks is influenced by the embedding capacity, with an increase in the latter generating an increase in the former. Each sub-signature value is obtained by calculating the average energy of the samples in each block j of size N_j samples. We will use V_j to denote those samples in block j,

$$f_j = \frac{1}{N_j} \sum_{i=1}^{N_j} {}^j v_i^2, \tag{2}$$

where ${}^j v_i$ denotes sample i that belongs to the group of samples V_j, with $i \in \{1, \ldots, N_j\}$. Each sub-signature value f_j is then rounded to a 10-bit integer value, which forms a subvector $\mathbf{f}_j \in \{0,1\}^{10}$. The final signature is comprised of four concatenated subvectors $\mathbf{F} = [\,\mathbf{f_1} \mid \ldots \mid \mathbf{f_4}\,]$, where $\mathbf{F} \in \{0,1\}^{n_1}$, $n_1 = 40$.

3.2. Spread-spectrum watermarking in the wavelet domain

There are two types of spread-spectrum watermarking techniques. One is non-blind, sometimes called oblivious, which requires the availability of the original signal at the detector [8]. The other type

is blind, meaning that the original image is not required at the detector [6][7]. We need to use a blind technique because the original image signal is not available at the detector for authentication purposes. In order to demonstrate the usage of spread-spectrum watermarking in the context of CSE, we use a watermarking technique similar to those described in [6][7]. In addition, we show how the technique can be used to embed a signature of n bits. The watermarking is conducted based on symbols. Each symbol ξ_i is mapped to a unique sequence (x_i) of m real numbers, where each number is chosen independently according to $\mathcal{N}(0,1)$ (where $\mathcal{N}(\mu,\sigma^2)$ denotes a normal distribution with mean μ and variance σ^2). The value of symbol ξ_i can be used as a seed to generate a unique watermark normal sequence (x_i) with zero mean and unit variance. In our case, each symbol value represents 10 bits of the signature. The number of symbols that can be used depend on the partitioning strategy used to divide the image space for embedding. The embedding operation is conducted by partitioning the image space in the wavelet domain into k blocks (e.g. k=4). The coefficients that are used for embedding are chosen based on a threshold criteria, for a block j with N_j coefficients, $(v_i : v_i \in V_j, v_i > T_1, i = 1 \dots N_j)$. The watermark sequence is spread according to

$$v_i{}' = v_i + \beta\,|v_i|\,x_i,\;\; i = 1 \dots N_j, \tag{3}$$

where β is a scaling factor set to 0.2 (other empirical values of β are possible). The watermark detection operation is conducted by measuring the detection response r which represents the *correlation* between the test coefficients $(\hat{v}_i : \hat{v}_i \in V_j, \hat{v}_i > T_2, i = 1 \dots N_j)$ and a possibly different watermark x'_i.

$$r = \frac{1}{N_j}\sum_{i=1}^{N_j} \hat{v}_i x'_i, \tag{4}$$

where $T_2 > T_1$ to guarantee that the correlation is performed on coefficients casted with a watermark. This excludes coefficients that are originally below T_1 and, due to manipulations, that might become greater than T_1. It should be noted that there exist a trade-off between the number of symbols that can be embedded and the watermarking robustness, this is the case in all the watermarking schemes. Watermark visibility is usually very low with spread-spectrum techniques because of the normal distribution nature of the watermark. Also, robustness to attacks such as cropping or filtering is very high due to the spreading of the watermark over a wide range of frequencies and over a wide spatial space.

3.3. Closed-Form Compensation

The closed-form compensation performs an accurate compensation in one step while meeting the minimum distortion criteria of *image least mean square distortion* to guarantee image fidelity. As mentioned earlier, the signature of the original image is defined as $\mathbf{F} = [\,\mathbf{f_1} \mid \dots \mid \mathbf{f_4}\,]$. Let the signature generated immediately after embedding be $\mathbf{F'} = [\,\mathbf{f'_1} \mid \dots \mid \mathbf{f'_4}\,]$. We perform the compensation for each block independently (i.e. each sub-signature value is independent of the other values). For each block j, the goal is to compensate for the difference between f'_j and f_j. This entails modifying selected wavelet coefficients within V_j to yield $\hat{f}_j$ such that $\hat{f}_j = f_j$. Let $\delta = f'_j - f_j$. To completely compensate for δ, each coefficient $(v'_i : v'_i \in V_j, i = 1 \dots N_j)$ is modified with value $(\triangle v'_i)$ such that

$$\frac{1}{N_j}\sum_{i=1}^{N_j} [(v'_i + \triangle v'_i)^2 - v_i'^2] = \delta, \tag{5}$$

where N_j is the total number of coefficients in block j. Equation (5) can have many solutions. We consider the solution that guarantee *image least mean square distortion* as an optimal solution to (5).

Before we describe the derivation that leads to the optimal solution, we present the result first and then proceed with the detailed steps in the derivation. The optimal solution will be obtained by adding a constant proportional rate value α where $(\triangle v'_i = \alpha v'_i)$. The result can, therefore, be reached by substituting for $\triangle v'_i$ in (5), which leads to

$$\sum_{i=1}^{N_j} \left\{[v'_i(1+\alpha)]^2 - v_i'^2\right\} = \delta N_j. \tag{6}$$

We now outline the steps that lead to the result stated in (6). The *image least mean square distortion* can be defined as

$$J(\triangle \mathbf{v}') = \sum_{i=1}^{N_j} \triangle v_i'^2. \tag{7}$$

We can now obtain the optimal solution using the *Lagrange multiplier* approach, by minimizing (7) subject to the following *difference* constraint given by (5). The *difference* constraint D is

$$D(\triangle \mathbf{v}') = \sum_{i=1}^{N_j} [(v'_i + \triangle v'_i)^2 - v_i'^2] - \delta N_j = 0. \tag{8}$$

Having the *difference* constraint is intuitive, since the goal is to make the compensation effect cancel out the difference between the two signature values generated before and after embedding. The Lagrangian Λ can be defined as

$$\Lambda(\triangle \mathbf{v}') = J(\triangle \mathbf{v}') + \lambda D(\triangle \mathbf{v}'), \tag{9}$$

where λ is the Lagrangian multiplier. By substituting for J and D, we get

$$\Lambda(\triangle \mathbf{v}') = \sum_{i=1}^{N_j} [(1+\lambda)\triangle v_i'^2 + 2\lambda v'_i \triangle v'_i] - \lambda\delta N_j. \tag{10}$$

The minimization of the Lagrangian Λ is achieved by taking a partial derivative of Λ with respect to $\triangle v'_i$ and setting the result to zero,

$$\frac{\partial \Lambda(\triangle \mathbf{v}')}{\partial(\triangle v'_i)} = 0, \tag{11}$$

this leads to

$$\triangle v'_i = \frac{-\lambda v'_i}{1+\lambda} = \alpha v'_i,\;\; i = 1 \dots N_j, \tag{12}$$

where $\alpha = \frac{-\lambda}{1+\lambda}$. By substituting the value of $\triangle v'_i$ in (8), we get

$$\sum_{i=1}^{N_j} \left\{[v'_i(1+\alpha)]^2 - v_i'^2\right\} = \delta N_j. \tag{13}$$

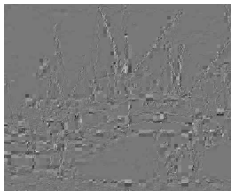
By solving for α, we get

$$\alpha = \sqrt{\left(\frac{\delta N_j}{\sum_{i=1}^{N_j} v_i'^2} + 1\right)} - 1. \tag{14}$$

The above approach, with some variation, has been also used to address a different application in [9].

(a) original image

(b) compensated image



(c) difference (b)-(a)

Compensation type	PSNR *Boat* Image	PSNR *Lena* Image
Closed-form compensation	66.23	66.59
Iterative compensation	56.79	55.05

(d) PSNR Comparison

Fig. 3. Demonstration of image fidelity

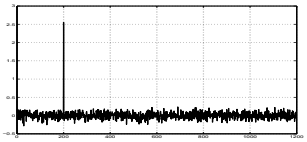
Compared Images	PSNR	CORR
Watermarked vs. Original	34.85	0.9962
Compensated vs. Watermarked	66.23	~1.0
(Compensated and Watermarked) vs. Original	34.84	0.9961

Table 1. PSNR and Correlation

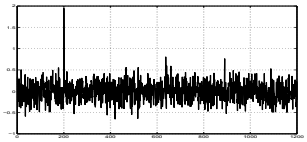
4. EXPERIMENTAL RESULTS

Simulation experiments were focused on verifying that the closed-form compensation approach delivers an accurate tamper detection, while maintaining high image fidelity. In addition, the robustness of the spread-spectrum watermarking is tested. As an objective measure, Table 1 lists PSNR and normalized correlation values to show that image fidelity is maintained after applying the closed-form compensation and watermarking. PSNR value of 66.23 in the table indicates that the compensation operation did not add any significant distortion and kept an optimal correlation between the watermarked and the compensated images. The PSNR value of 34.84 is due to the spread-spectrum watermarking which can be enhanced in the future by means such as visual human system adaptation. Figure 3(a) through 3(c) reflect those objective measures. Figure 3(d) shows the improvement on image fidelity as a result of using closed-form compensation in comparison with the iterative one using both *Boat* and *Lena* example images.

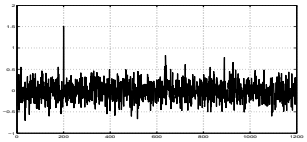
To test the robustness and capacity of the spread-spectrum watermarking technique, we used 1200 seed values with a test watermark seed of 200, and $T_1 = 10, T_2 = 15$ (higher threshold values are possible). Figure 4 shows the correlation response as defined in equation (4). Figure 4(a) shows the response of the detector to a watermark seed with no manipulations and using one block partitioning (i.e. the whole image is one block), and Figure 4(b) shows the same seed but with 4-block partitioning, as expected the response is lower with higher capacity. 4(c) shows the response with 4-block partitioning after applying a 0.6bit/pixel JPEG2000 compression. Figure 4(d) shows the response after applying localized white gaussian noise, (20x20) pixel area, with 4-block partitioning. As expected, we observe that the correlation response is stronger when using fewer blocks which means lower capacity, and stronger response means higher robustness. The manipulations that were applied included white gaussian noise (WGN), JPEG2000 compression, localized WGN in (20x20) pixel areas. In all the test cases, tamper detection was achieved with 100% success rate.



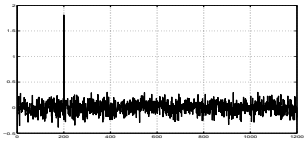
(a) with no manipulations, one block



(b) with no manipulations, 4 blocks



(c) with J2K compression, 4 blocks



(d) with localized editing, 4 blocks

Fig. 4. Watermark detection response for 1200 symbols

5. CONCLUSION

An improved content-based image authentication is proposed by using a closed-form approach to perform the compensation operation in the context of compensated signature embedding (CSE) system [4]. In addition, a robust spread-spectrum watermarking technique is used for signature embedding and detection. Discrete wavelet transform domain is used for watermarking and compensation. Test results are presented to show system's effectiveness.

6. REFERENCES

[1] P. Lin, P. Huang, A. Peng, "A Fragile Watermarking Scheme for Image Authentication with Localization and Recovery", IEEE Int. Symp. on Multimedia Soft. Eng., pp. 146-153, Dec. 2004.

[2] H. Liu, J. Lin, J. Huang, "Image Authentication Using Content Based Watermark", IEEE Int. Symp. on Circuits and Systems, pp. 4014-4017, May 2005.

[3] C. Rey, J. Dugelay, *"A Survey of Watermarking Algorithms for Image Authentication"*, EURASIP Jour. on Applied Signal Proc., Issue 6, pp. 613-621, 2002.

[4] S. Ababneh, A. Khokhar, R. Ansari, "Compensated Signature Embedding Based Multimedia Content Authentication System", IEEE Int. Conf. on Image Proc., pp. I-393-I-396, Sept. 2007.

[5] S. Ababneh, R. Ansari, A. Khokhar, "A Set-Theoretic Approach for Compensated Signature Embedding Using Projections onto Convex Sets", SPIE Visual Comm. and Image Proc., Jan. 2008.

[6] R. Dugad, K. Ratakonda, N. Ahuja, "A New Wavelet-Based Scheme for Watermarking Images", IEEE Int. Conf. on Image Proc., pp. 419-423, Oct. 1998.

[7] A. Piva, M. Barni, F. Bartolini, V. Cappellini, "DCT-based Watermark Recovering without Resorting to the Uncorrupted Original Image", IEEE Int. Conf. on Image Proc., pp. 520-523, Oct. 1997.

[8] I.J. Cox, J. Kilian, F.T. Leighton, T. Shamoon, *"Secure Spread Spectrum Watermarking for Multimedia"*, IEEE Trans. on Image Proc., Volume 6, Issue 12, pp. 1673-1687, Dec. 1997.

[9] H. Yuan, X. Zhang, *"Multiscale Fragile Watermarking Based on the Gaussian Mixture Model"*, IEEE Trans. on Image Proc., Volume 15, Issue 10, pp. 3189-3200, Oct. 2006.