

# Watermarking based Image Authentication using Feature Amplification

Shuiming Ye<sup>1,2</sup>, Ee-Chien Chang<sup>1</sup> and Qibin Sun<sup>2</sup>

<sup>1</sup>School of Computing, National University of Singapore, Singapore

<sup>2</sup>Institute for Infocomm Research, A\*STAR, Singapore

## Abstract

In a typical content and watermarking based image authentication approach, a feature is extracted from the given image, and then embedded back into the image using a watermarking method. Since the entropy of the feature might be higher than the capacity of the watermarking scheme, or the feature is represented in a continuous domain, it has to be further quantized before embedding. The loss of information during quantization potentially degrades the overall performance of the authentication scheme. This paper propose a simple but effective approach that avoids the feature quantization by additive feature: the feature is firstly added into the image before watermark embedding, and latterly subtracted from the watermarked image. In our experiments, the proposed approach obtains larger achievable robustness/sensitivity region and has a smaller fuzzy region of authenticity than the typical approach.

## 1. Introduction

The pervasive distribution of digital images triggers an emergent need of authenticating images received by unreliable channels, such as public Internet and wireless networks. Many image authentication techniques have been proposed, which can be divided into two broad categories: labeling based approaches [1, 2, 3], in which authentication data are written in a separate file; and watermarking based approaches [4, 5], in which authentication data are embedded in the images. This paper will focus on the watermarking based methods.

A typical approach of content and watermarking based authentication follows three steps: firstly, a feature is extracted from the image; secondly, the feature is discredited and quantized; finally, the quantized feature is embedded as message into the image. During authenticity verification, the message is detected using the watermarking detector, and the feature is extracted from the watermarked image. A typical authenticity verification decision is based on the comparison between a preset threshold  $T$  and the distance of the extracted feature and the detected watermark.

Assuming the feature representation is a good

choice, the performance of the above scheme depends on the effectiveness of the feature quantization and watermarking code. In practice it is impossible to obtain a perfect code. As a result, there is often a fuzzy region separates the space of authentic images from the unauthentic images [6].

In this paper we propose an alternative approach by avoiding the quantization step. In doing so, potentially we can reduce the size of the fuzzy region. In our approach, the feature is not quantized, but added into the image before watermarking. The feature is then subtracted from the watermarked image to get the output of the embedding procedure. Experimental results show that compared with the typical approaches, our scheme can achieve a smaller fuzzy region of authenticity, and a better achievable robustness/sensitivity region.

## 2. Proposed image authentication using feature amplification

Fig. 1 shows the structure of the proposed approach. It is a general scheme in the sense that different watermarking schemes and feature extractions can be employed. The embedding and verifying procedures require a predefined parameter, the amplifying factor  $k$ , and a predefined message  $m$ .

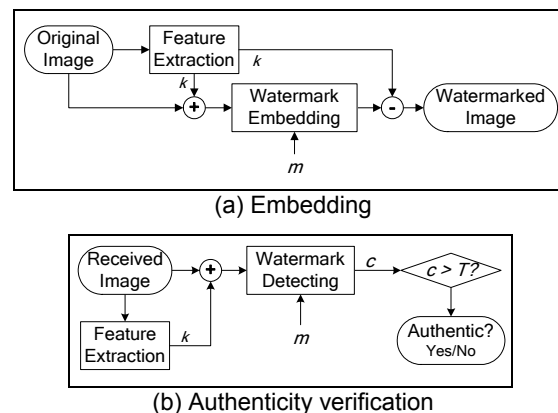


Fig. 1: Proposed feature aided authentication scheme

### 2.1. Embedding

Given an image  $I$ , the embedding procedure outputs

the embedded image  $I_w$ . The relationship between  $I$  and  $I_w$  can be expressed as:

$$F = \text{Ex}(I) \quad (1)$$

$$I_w = \text{Em}(I + kF, m) - kF \quad (2)$$

where  $\text{Ex}(I)$  is the feature extraction function, and  $\text{Em}(x)$  is the watermarking encoder. Thus,  $I_w$  can be obtained from  $I$  by these steps: firstly, the feature  $F$  is extracted using function  $\text{Ex}(I)$ , and is added into the image with an amplification factor  $k$ , and then the predefined message  $m$  is embedded into the image after feature added using function  $\text{Em}(x)$ . After embedding, the feature is subtracted from the watermarked image  $I_w$  as compensation.

## 2.2. Authenticity verification

The procedure of image authenticity verification is the reverse procedure of the embedding:

$$\tilde{F} = \text{Ex}(\tilde{I}) \quad (3)$$

$$c = \text{De}(\tilde{I} + k\tilde{F}, m) \quad (4)$$

where  $\text{De}(x)$  is the watermarking decoder. During verification, the feature  $\tilde{F}$  of  $\tilde{I}$  is extracted using the same technique used in the embedding procedure, and then added into the image. Based on the summarization result, the correlation  $c$  is computed by the watermark decoder function  $\text{De}(x)$ . Based on  $c$ , a decision on the authenticity of the image is made.

Slightly different from common notation, our watermarking detector, given an image and a message, output a correlation value, which is the likelihood that the image is watermarked by the message. Note that the detectors in many watermarking scheme makes its decision based on a correlation value. Here, instead of using a detector that outputs the decision, we want the detector to output this correlation value.

The correlation value  $c$  between the extracted watermark  $w$  and the original message  $m$  is calculated to derive the final decision. If  $c$  is larger than a threshold  $T$ , the image is decided as authentic, otherwise unauthentic.

During embedding and verification, since the feature is to be added to the image, the dimension of the feature  $n_f$  and the number of image coefficients  $n_c$  must be the same. In case  $n_f < n_c$ , we need to expand the feature. This can be achieved by multiplying the feature vector by a predefined random  $n_c$  by  $n_f$  matrix. In our implementation, we simply duplicate the feature vector roughly  $(n_c / n_f)$  times to obtain the expanded n-vector.

## 2.3. Generic scheme for typical approach TIA

This section compares the proposed scheme with its complement authentication scheme that quantized the

feature. Let us call this scheme the typical image authentication (TIA) scheme, which will use the same components on feature extraction and watermarking techniques as our scheme. However, it required an additional quantizer. The embedding and verifying procedures are shown in Fig. 2.

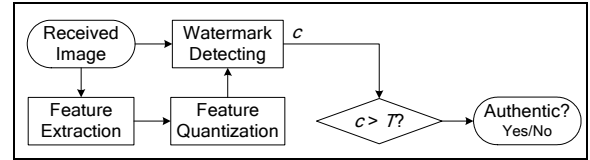
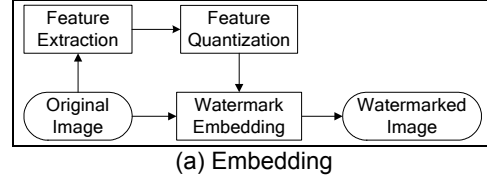


Fig. 2: Typical content based image authentication

## 2.4. The effect of amplification factor $k$

**Effect of  $k$  on robustness:** When the received image is distorted by additive noise  $N$ , i.e.  $\tilde{I} = I_w + N$ , then  $\tilde{F} = \text{Ex}(\tilde{I}) = F + N'$ . If the feature is robust to noises and  $k$  is small, then  $kN' \ll N$ , i.e., the input of watermark detector is approximately as:

$$\tilde{I} + k\tilde{F} = I_w + N + kF + kN' \approx I_w + kF + N \quad (5)$$

Thus, compared with TIA, the signal-to-noise ratio (SNR) of the watermark detector input is  $(I_w + kF)/N$ , compared with  $I_w/N$  in TIA. It shows the robustness is improved by the additive feature.

**Effect of  $k$  on sensitivity:** The sensitivity is defined as the ability to be sensitive to distortions in the image feature. If feature is corrupted by noise  $N$ , then the input of detector is:

$$\tilde{I} + k\tilde{F} = I_w + kF + kN \quad (6)$$

Thus, compared with TIA, the detector input SNR is decreased from  $I_w/N$  to  $(I_w + kF)/kN$ , which shows that our scheme is more sensitive to the modifications of feature than TIA. For a large  $k$ , such noise is amplified and detector SNR decreases. Hence, the variance of  $c$  should increase.

## 3. Performance measure

In this section, we describe the main performance measurements used in our experimental studies.

**Distortion:** Same as watermarking scheme, the distortion between the embedded image and original image should be small. In this paper, we use the *Euclidean*

distance function to measure the distortion.

**Fuzzy region:** The goal of authentication is to determine whether the image has been tampered with and hence induce significant distortion to its feature. From another viewpoint, the aim is to make a hypothesis on the level of feature distortion due to tampering. For the purpose of analysis, we consider the ideal situation where a tampering amount to an AGWN in the feature domain, and the tampering does not affect the pixel. Suppose an AGWN with standard deviation  $N$  and zero mean is begin added to the feature, let us investigate the statistical properties, in particular, the standard deviation and mean of the correlation value  $c$ .

Recall that a decision regarding the level of AGWN in the feature domain is to be made based on  $c$ . Two common hypotheses are: (1) whether the noise energy is less than a certain level  $N$ , and (2) whether the noise energy is more than  $N$ . Each gives a threshold, say  $T_0$  and  $T_1$  respectively, for the hypothesis testing. If we want to achieve high degree of confidence, then there is a large gap ( $T_1 - T_0$ ). This gap gives a fuzzy region: images whose correlation falls into the interval ( $T_0, T_1$ ) are neither authentic nor unauthentic. The size of the gap is closely related to the standard deviation of  $c$ . Since it is desirable to have a small gap, it is desirable to have a scheme that gives small standard deviation of the  $c$ .

**Robustness vs. Sensitivity:** An alternative measurement is to look that the ratio of Sensitivity over Robustness. Permissible operations like lossy compression is usually model as AGWN in the pixel domain. Thus, the choice of feature representation and the authentication should withstand high level of AGWN in the pixel domain.

On the other hand, the scheme should also be sensitive to changes in the feature. In view of the above two requirements, from the correlation value  $c$ , we want to find threshold  $T$ ,  $N_r$  and  $N_s$  satisfying:

- Robustness: If ( $c < T$ ), then we are confident that the level of AGWN in the pixel domain is less than the predefined value  $N_r$ .
- Sensitivity: if ( $c \geq T$ ), then we are confident that the level of AGWN add to the feature is more than a predefined value  $N_s$ .

By changing the value of  $T$ , we will have different ratio  $N_s/N_r$ . Clearly, we seek for a scheme with low  $N_s/N_r$  ratio.

Note that the two measurements, size of the gap and the ratio, although are different, they are closely related. A small gap typically corresponds to a low ratio. The choices of the measurement depend on applications. Nevertheless, we use both measures to evaluate the performance of our scheme.

## 4. Experimental results

The proposed scheme is tested with a variety of images, but here we only give the results of using the gray image *Lena* ( $512 \times 512$ ) for example. We choose the compensation factor  $\alpha=0.55$ , and step size  $\Delta=13$  in the QIM. The message length for the TIA scheme is 1024 bits.

There are two key issues in using this algorithm: the feature extraction algorithm and watermarking algorithm. In case of fairness, we use same watermarking and feature extraction techniques when compared with TIA. In our implementation, we use the simple feature of *Sobel* edge intensity and Quantization index modulation (QIM) watermarking algorithm [7].

Independent additive white *Gaussian* noise attacks, which are good models for at least some types of un-informed and unintentional attacks, are constructed to the watermarked images to test the robustness of the methods. Normally distributed noises with mean 0 and different standard deviations were added to each of the watermarked images. Since the same additive watermarking technique is used in both schemes, the distortions of both schemes are almost the same for each noise level.

The results the effect of noises added into the watermarked image are shown in Fig. 3(mean correlation) and Fig. 4 (standard deviations). The mean correlations of our scheme are much larger than the typical scheme (Fig. 3) and the deviation of our scheme is much smaller than the TIA scheme (Fig. 4), especially for small  $k$ .

The smaller deviations of our scheme also lead to a smaller size of fuzzy region. If the image is distorted by noise at a strength  $N$ , we measured the probability of its acceptance probability during authenticity verification. Here we set the acceptance threshold  $T$  as 0.75. The results of such probability and the corresponding noise strength are shown in Fig. 5 there a smaller fuzzy region between accepted noise level and rejected noise level in our scheme.

To evaluate the achievable robustness and sensitivity region of our scheme, we added noises into the feature domain and image pixel domain independently. Given a required robustness  $N_r$ , we find the smallest  $T$  such that the scheme is robust against AGWN of  $N_r$  in the pixel domain. Next, with the fixed  $T$ , we determine the smallest achievable (and hence the best) sensitivity  $N_s$ . Again, the sensitivity  $N_s$  is determined experimentally by adding various level of noise in the feature domain. The whole experiment is repeated for different robustness  $N_r$ , and the results are shown in (Fig. 6). The regions above the curves are the achievable sensitivity and robustness regions. The curve with larger  $k$  has better choice to achieve larger robustness and sensitivity re-

gion. Combined the results shown in Fig. 3 to Fig. 6, the suitable choice of  $k$  can be determined by a careful tradeoff. In our experimental results we find 4 is a good tradeoff for  $k$ .

## 5. Conclusion

This paper proposes a feature aided image authentication scheme, in which image feature is added into the image before watermarking, and subtracted from the watermarked image after watermarking as compensation. Theory analysis and experimental results show that this scheme has superior performance over the typical content based authentication scheme, offering better choice of robustness/sensitivity region and smaller fuzzy region, at the cost of the same distortion introduced. Future works will construct more experiments to test the proposed scheme, such as robustness against image compressions, and sensitivity against malicious attacks.

## References

[1] M.P. Queluz, "Authentication of Digital Images and Video: Generic Models and a New Contribution", *Signal*

*Processing: Image Communication*, Vol.16, Jan. 2001, pp.461-475.

[2] C. Y. Lin and S.F. Chang, "A Robust Image Authentication Method Distinguishing JPEG Compression from Malicious Manipulation", *IEEE Transaction on Circuits and Systems of Video Technology*, Vol.11, 2001, pp.153-168.

[3] E.C. Chang, M.S. Kankanhalli, X. Guan, Z.Y. Huang, and Y.H. Wu, "Robust Image Authentication Using Content-based Compression", *ACM Multimedia Systems Journal*, Vol. 9, No. 2, 2003, pp.121-130.

[4] Q. Sun and S.F. Chang, "Semi-fragile Image Authentication using Generic Wavelet Domain Features and ECC", *IEEE International Conference on Image Processing (ICIP)*, Rochester, USA, Sep. 2002.

[5] C.W. Tang and H.M. Hang, "A Feature based Robust Digital Image Watermarking Scheme", *IEEE Transaction on Signal Processing*, Vol. 51, No. 4, Apr. 2003, pp. 950-959.

[6] C. W. Wu, "On the Design of Content-Based Multimedia Authentication Systems", *IEEE Transactions on Multimedia*, Vol. 4, No. 3, Sep. 2002, pp.385-393.

[7] B. Chen and G. W. Wornell, "Quantization Index Modulation: A Class of Provably Good Methods for Digital Watermarking and Information Embedding", *IEEE Transaction on Information Theory*, Vol. 47, No. 4, May 2001, pp.1423-1443.

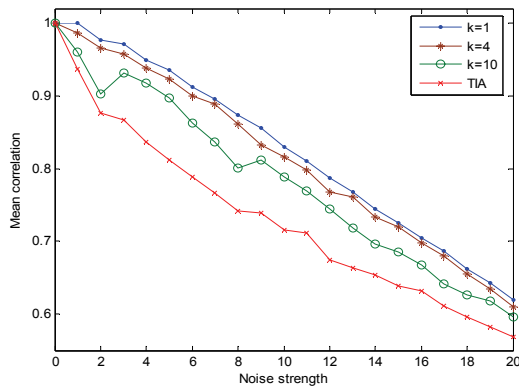


Fig. 3: Correlation average versus Gaussian noises

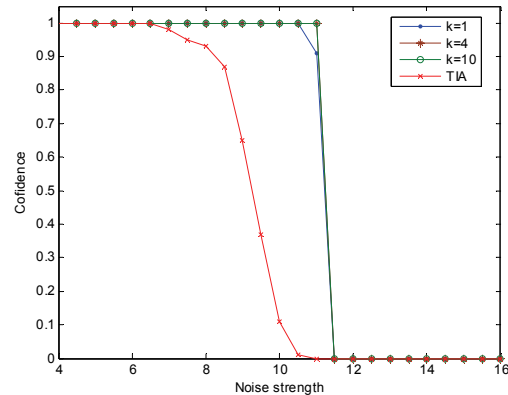


Fig. 5: Acceptance confidence versus Gaussian noises

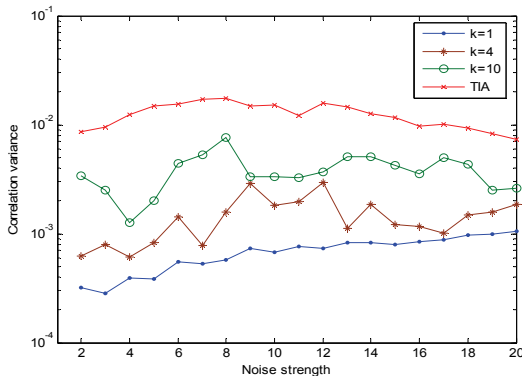


Fig. 4: Correlation standard deviation versus Gaussian noises

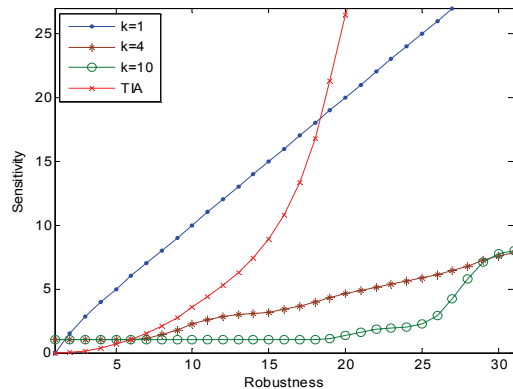


Fig. 6: Robustness and sensitivity curve