Error Resilient Content-based Image Authentication Over Wireless Channel

Shuiming Ye [1, 2]*, Qibin Sun* [1]*, and Ee-Chien Chang* [2]
[1] Institute for Infocomm Research, Singapore 119613
[2] School of Computing, National University of Singapore, Singapore 117543
{shuiming, qibin}@i2r.a-star.edu.sg, changec@comp.nus.edu.sg

Abstract—**The pervasive distribution of digital images triggers an emergent need of authenticating degraded images by lossy compression and transmission. This paper proposes a robust content-based image authentication scheme for image transmissions over wireless channels. A robust multiscale based feature is used to generating digital signature. The feature vector is reduced, encoded, and encrypted as signature of the image. This signature is then embedded into the image using a blind quantization based watermarking algorithm. Typical distortion patterns of malicious image manipulations and image transmissions over lossy channels are used for image authenticity verification. This scheme is also able to detect tampering areas in the attacked image. Experiments demonstrate the effectiveness and validity of our scheme.**

I. Introduction

Image authentication based on digital signature aims to verify the source and the integrity of an image, which is desirable to be robust to normal image processing, compression and transmission errors, while able to detect malicious tampering on the image. Particularly, image transmission is always affected by the errors in wireless channel (environmental noises fading, multipath and Doppler frequency shift), or packet loss due to congestion when using UDP over IP protocol in Internet network.

In order to be robust to some acceptable manipulations and compression, several robust image authentication algorithms have been proposed [1]. However, these techniques failed in surviving image transmissions over lossy channels. The reason is that it is difficult to predict where and when the errors will be introduced. When errors occur during transmissions, there is no constraint to ensure the correctness of every bit of received images. Furthermore, compressed images are sensitive to these errors since compression techniques such as variable length coding and entropy coding result in error propagations.

Authenticating image over lossy channels is a challenging task. There are several solutions proposed for authenticating data stream, such as efficient multi-chained [2] or augmented chain stream signature [3], which are robust against packet losses by sending multiple hashes of other packets within the current packet. When a packet is lost, its hash will be found in other packets unless total packet loss of a segment exceeds a threshold. However, directly applying this solution to image transmission has several drawbacks: (1) with the increase of Bit Error Rate (BER) and the need of time synchronization, the transmission overhead will be unavoidably large; (2) in image transmission, the importance and the size of packets vary in different environments; (3) treating data as bit stream, it does not take advantage of the fact that images are tolerable to some degree of errors.

Content-based authentication is an efficient robust image authentication, which passes images as authentic when the content does not change [4, 5]. In this paper, we propose a robust content-based authentication scheme, which is robust to transmission errors. In order to distinguish the images damaged by transmission errors from tampered ones, the distortion patterns of them are analyzed and employed in our proposed scheme. It is the image content to be authenticated, instead of image data itself. This proposed scheme is able to detect malicious modifications of the image, while robust to transmission errors and other common acceptable image processing. The major differences that distinguish this paper from our previous paper [6] are: (1) the scheme proposed in this paper can be more generally applied, suitable for both wavelet-based images and DCT-based images; (2) more robustness is achieved by exploiting the distortion patterns of acceptable operators and malicious operators.

II. Content based Signature and Watermarking

Image content is generally represented by a feature vector extracted from it. Edges in a natural image have important effects on the subjective visual quality, since edges are always associated with the boundary of an object, or with marks on the object [7]. It is also shown that high edge preserving ratios can be achieved by our error concealment algorithm [7]. Therefore, image edge is selected as the feature in our authentication scheme. We adopt the fuzzy reasoning based edge detector [8], because no thresholds are needed in it, which can improve the feature robustness.

Feature map is reduced before embedding into the image, because the watermarking capacity is limited. In consideration of robustness, no compression or entropy coding would be used, since they will cause error propagation. To reduce feature map, one bit of the resulting feature vector **F** represents one 8×8 block of the feature map by marking the positions of the K_S most prominent blocks.

Error correct coding (ECC) is employed to encode the feature vectors to enhance its embedding robustness. After reducing the feature map, the resulting vector **F** is then encoded with BCH(5,1) and encrypted with the private key K_E of the signing entity to produce the final signature **S**, using the encryption function of public key infrastructure (PKI).

The content-based signature **S** is then embedded into the image as watermarks. We use the same quantization based watermarking algorithm as used in [9] for several reasons. Firstly, the watermarking can embed several bits into one 8×8 block, which make it suitable for embedding the large volume of the signature. Secondly, this scheme has a robustness inherit to JPEG compression, and is compatible with our previous work of error resilient authentication for JPEG images [6]. Finally, it does not modify the DC coefficients, which are identical to the approximate subband coefficients of wavelet transform.

It is important that embedding the signature into the image should not affect the signature verification process, since the watermarking would introduce some distortions. For excluding the adverse impact of watermarking on feature extracting, a compensation function Cps(x) is adopted before feature extraction and watermarking. Let the total selected DCT coefficients form a set **P**. For each coefficient c in **P**, it is replaced with c_w to embed a bit of **S**:

$$c_w = \begin{cases} QC, \text{ if } \mathrm{LSB}(C) = w \\ Q\left(C + \mathrm{sgn}\left(c - QC\right)\right), \text{ else} \end{cases} \tag{1}$$

where C = round (c/Q), Q is the quantization table, and w is the watermark bit to be embedded. Function round(x) returns the nearest integrate of x, sgn(x) returns the sign of x, and function LSB(x) returns the least significant bit of x.

The compensation function Cps(x) is designed according to the watermarking algorithm, as defined below:

$$\mathrm{Cps}(I) = \mathrm{IDCT}\left\{\mathrm{IntQuan}\left(d_i, 4Q, \mathbf{P}\right)\right\} \tag{2}$$

where d_i is the ith DCT coefficient of I, and IDCT is inverse DCT transform. The IntQuan(c, Q, **P**) is defined by:

$$\mathrm{IntQuan}\left(c, Q, \mathbf{P}\right) = \begin{cases} c, \text{ if } c \notin \mathbf{P} \\ Q\,\mathrm{round}(c/Q), \text{ else} \end{cases} \tag{3}$$

From Eq. (1), (2) and (3), we can get Cps(I_w) = Cps(I), thus the feature extracted from the watermarked image I_w is the same as the one from the original image I. This compensation operating ensures that watermarking does not affect the feature extraction.

III. Content Authenticity Verification

In content-based authentication applications, it is often difficult to distinguish distortions caused by acceptable manipulations from those by malicious manipulations. However, it is observed that the acceptable manipulations are usually global distortions while the illegal manipulations tend to be localized [10]. What's more, the attacked areas are always connected due to changing the image content. Based on these observations, two rules are applied during content verification. They are effective to distinguish the malicious attacks from lossy transmissions and other acceptable manipulations.

During image authenticity verification, error detection and concealment will be carried out if the image is received over a lossy channel. The feature vector **F'** of the image is recalculated using the same method as used in image signing procedure. The original feature **F** is got by extracting watermarks, decoding, and decrypting using a public key (K_D) of the sender. The authenticity verification is deduced based on the difference map (**M**) between **F'** and **F**. If the image is deemed as unauthentic, the attacked areas are then detected using the feature aided attack detector.

A. Signature Extraction

A blind watermarking algorithm is used in this paper, that is, the watermark detection does not need the original image or watermarks. The watermark W is extracted using the quantization table Q and key K_W. Then we decrypt the extracted watermarks using the public key K_D. The signature **S** of the image is decoded from the decrypted watermarks. The BCH(5,1) code we use can correct some errors in the signature and detect errors which are beyond the capacity of the correctness.

B. Error Concealment

It is worth noting that it is efficient and advisable for error concealment to be applied before image authentication at the receiver, because the concealed image could be still securely reusable for other applications without re-doing error concealment every time. We adopt an edge directed filter based error concealment algorithm proposed in [7] for wavelet based images, and content-based error concealment algorithm proposed in [11] for DCT based JPEG images.

C. Content Authenticity Verification

In order to distinguish distortions by transmission errors from those of attacks, typical distortion patterns of acceptable distortions and malicious attacks are used for verification. No thresholding process is introduced in this algorithm. The basic idea is to use distortion patterns of these manipulations, convert them into rules, compute the degree of membership for error map, calculate the degree of the authenticity and unauthenticity, and then get the final authentication result.

The distortion of the attacked image is often concentrated on some content of interest, whereas the distortion from transmission errors is much more randomly distributed over the whole image. Furthermore, the attack areas are more likely to be connected (e.g., the objects modified). Based on these observations, two rules are applied:

- Rule 1: Acceptable manipulations cause global distortion, whereas tampering operations cause local distortion.

- Rule 2: The maximum size of the connected modified areas of acceptable manipulations is small, whereas the one of tampering operations is large.

The difference map **M** of the extracted feature vector **F** from watermarks and the recalculated feature **F**' of the image is used in the verification. We use sigmoid and Gaussian membership functions to calculate the degree of membership for each rule. The degree of rule 1 for the image is defined as:

$$\mu_1 = \frac{1}{1+\exp\left(\dfrac{\alpha N}{XY}-\beta\right)} \tag{4}$$

where X and Y are the number of differences in the histogram of horizontal and vertical projection of **M**, and N is the total number of differences in **M**. Experimentally we let α be 100 and β be 10.

The degree of rule 2 for the image is defined as:

$$\begin{cases} \mu_2^L = \begin{cases} 1, & \text{if } m \geq L \\ \exp(-(m-L)^2/2\sigma^2), & \text{else} \end{cases} \\ \mu_2^S = \begin{cases} 1, & \text{if } m \leq S \\ \exp(-(m-S)^2/2\sigma^2), & \text{else} \end{cases} \end{cases} \tag{5}$$

where $\sigma^2 = (L-S)^2/8\ln2$, and m is size of the maximum connected areas in **M**, L and S denote the large and small size respectively.

These two rules are combined to calculate the degree of authenticity and unauthenticity:

$$\begin{cases} D_Y = \min(\mu_1, \mu_2^S) \\ D_N = \min(1-\mu_1, \mu_2^L) \end{cases} \tag{6}$$

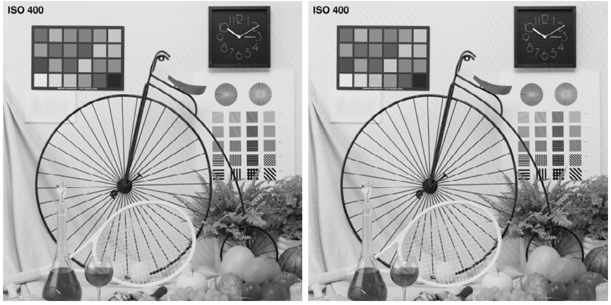
where the D_Y is the degree of authenticity, and D_N for degree of unauthenticity. If $D_Y > D_N$, then the image is classified as authentic; otherwise, tampering areas are detected.

D. Feature Aided Attack Location

If the image is deemed as unauthentic, it is helpful and convincing that the tampered areas can be detected. We call our attack location as feature aided because the attack areas are detected using information combining watermarks and image feature. Firstly, we calculate the difference map between the watermark extracted and the feature extracted. Morphological operations are used to compute connected areas and remove the isolated blocks and little connected areas. Then the difference map is masking by the union of the watermark and feature. The masking operation can refine the detected areas by concentrate these areas around the objects in the attacked image or in the original image. Those areas in the difference map which do not belong to an object are removed, which may be false alarm of some noises or acceptable image manipulations.

IV. Experimental Results and Discussions

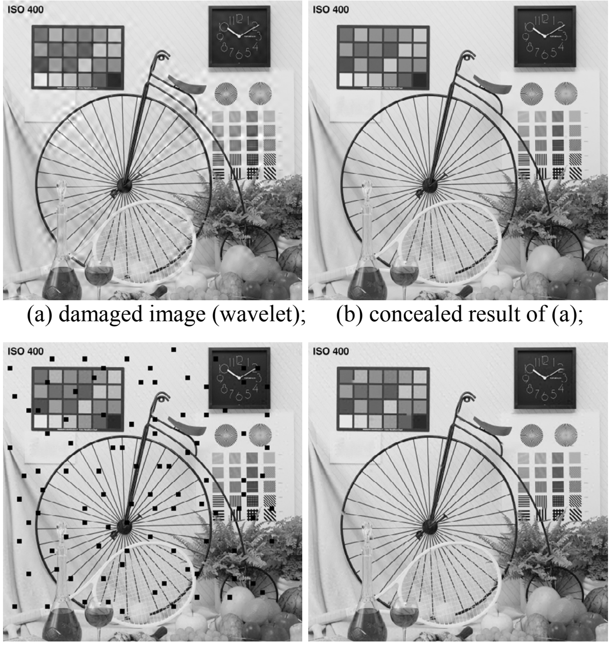
We implemented our proposed authentication scheme and tested its watermarking fidelity, robustness against transmission errors, robustness against some acceptable manipulations, and ability to detect tampered areas.



(a) original image *Bike*; (b) watermarked image (44.34 dB)

Fig. 1. Subjective fidelity

This proposed scheme got good robustness and fidelity by exploiting the distortion patterns of image manipulations during authenticity verification. We can embed a large volume of watermarks (64×64×5 bits in 512×512 grayscale image), and only introduce small distortions into the image. Our watermark embedding system did not introduce visual artifacts in the images to be protected (Fig. 1).With the set of images (Table I), we have obtained an average peak signal-to-noise ratio (PSNR) of 44.42 dB (Table I), which is above the usually tolerated degradation level of 40 dB [12], and also better than the 42.47 dB obtained by the paper [12].



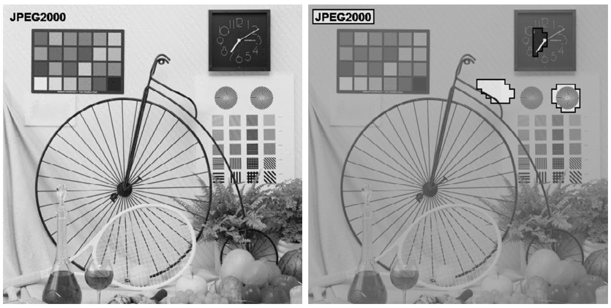
(a) damaged image (wavelet); (b) concealed result of (a);

(c) damaged image (DCT); (d) concealed result of (c)

Fig. 2. Robustness against transmission errors

Our scheme can also obtain a good robustness against transmission errors (Fig. 2), where the damaged images are firstly error-concealed using corresponding error concealment algorithm. We simulated the transmission errors based on the Rayleigh model with BER at 10^{-4}, and used *Daubechies* 9/7 wavelet filter for the wavelet transform. Fig. 2(a) is a damaged wavelet-based image with transmission errors, and Fig. 2(b) is its error concealed result. Fig. 2(c) is a damaged DCT-based image with transmission errors, and Fig. 2(d) is its error concealed result. Both error concealed images are authentic.

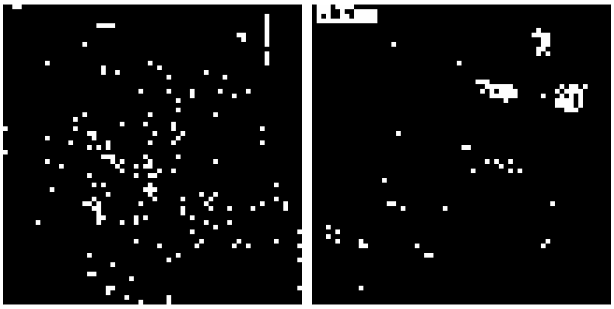
We also tested our scheme with some acceptable manipulations: (a) histogram normalization; (b) brightness adjustment (-40); (c) contrast adjustment; (d) JPEG compression (10:1); (e) JPEG2000 compression (8:1); (f) noise addition (*Gaussian* 20). All these images are authentic.



(a) tampered image *Bike*; (b) attack areas detected

Fig. 3. Attack location

An extra feature of our scheme is its ability to localize the tampered content of the attacked image. The attack location results are shown in Fig. 3. Fig. 3(a) is a result of modifying the image Fig. 1(b), by changing logo at the left top, modifying the time at right top, deleting the saddle, and replacing the right circle with the left one (below the clock). We found that the ability of our system to detect tampering is excellent, even in the presence of multiple tampered areas.



(a) for error concealed image; (b) for attacked image

Fig. 4. Typical distortion patterns

Fig. 4(a) shows the error map for the error concealed image shown in Fig. 2(b), and Fig. 4(b) for the tampered image shown in Fig. 3(a). The errors in Fig. 4(a) is much more evenly distributed than those in Fig. 2(b), and the maximum connected area size of Fig. 4(a) is much smaller than that in Fig. 4(b). These results validate our rules used in the content authenticity verification.

V. Conclusions

This paper proposes a content-based image authentication scheme for image transmission over lossy channels. Content authenticity of the image is verified by exploiting the typical distortion patterns of acceptable image manipulation and malicious content modifications. Instead of hard decision as in data authentication, we provide assistance for image content authentication: the authentication scheme can verify as much as possible the integrity of the received images without assuming the availability of all the original data. No side information is needed, and no original images or watermarks are needed during verification. The analysis and the experimental results validate that our proposed scheme can get a good robustness against transmission errors and some acceptable manipulations, at the cost of only small distortions introduced into the images by embedding the signature into the image.

References

[1] A. M. Eskicioglu and E. J. Delp, "An Overview of Multimedia Content Protection in Consumer Electronics Devices", *Signal Processing: Image Communication*, Vol.16, No.7, pp. 681-699, 2001.

[2] P. Golle and N. Modadugu, "Authenticating Streamed Data in the Presence of Random Packet Loss", In *Proceedings of the Symposium on Network and Distributed Systems Security*, pp. 13-22, 2001.

[3] A. Perrig, R. Canetti, D. Song, and J. D. Tygar, "Efficient and Secure Source Authentication for Multicast", in Proceedings of *Network and Distributed System Security Symposium*, pp. 35-46, 2001

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

[5] C.S. Lu and H.Y. Liao, "Structural Digital Signature for Image Authentication: An Incidental Distortion Resistant Scheme", *IEEE Transaction on Multimedia*, Vol. 5, No. 2, pp. 161-173, 2003.

[6] Q. Sun, S. Ye, L.Q. Lin, and S.F. Chang, "A Crypto Signature Scheme for Image Authentication over Wireless Channel", *International Journal of Image and Graphics*, Vol. 5, No. 1, pp. 1-14, 2005.

[7] S. Ye, Q. Sun, and E.C. Chang, "Edge Directed Filter based Error Concealment for Wavelet-based Images", *IEEE International Conference on Image Processing*, Singapore, 2004.

[8] W. Chou, "Classifying Image Pixels into Shaped, Smooth and Textured Points", *Pattern Recognition*, Vol. 32, No. 10, pp. 1697-1706, 1999.

[9] D. Kundur and D. Hatzinakos, "Digital Watermarking using Multiresolution Wavelet Decomposition", *IEEE International. Conference On Acoustics, Speech and Signal Processing*, Seattle, Washington, 1998.

[10] 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, pp. 121-130, 2003.

[11] S. Ye, X. Lin and Q. Sun, "Content Based Error Detection and Concealment for Image Transmission over Wireless Channel", *IEEE International Symposium on Circuits and Systems*, Thailand, 2003.

[12] A.H. Paquet, R.K. Ward, and I. Pitas, "Wavelet Packets-based Digital Watermarking for Image Verification and Authentication", *Signal Processing, Special Issue on Security of Data Hiding Technologies*, Vol. 83, No. 10, pp. 2117-2132, 2003.

TABLE I. PSNR (dB) of Watermarked Images

Image	*Actor*	*Bike*	*Chart*	*Flight*	*Fruits*	*Hotel*	*Lake*	*Lena*	*Pepper*	*Woman*	Average
Watermarked	44.34	44.40	44.50	44.54	44.17	44.14	44.25	44.60	44.46	44.79	44.42