SEMI-FRAGILE IMAGE AUTHENTICATION USING GENERIC WAVELET DOMAIN FEATURES AND ECC

Qibin Sun and Shih-Fu Chang

Department of Electrical Engineering, Columbia University, New York City, NY10027, USA

ABSTRACT

In this paper we present a generic content-based solution targeting at authenticating image in a semi-fragile way, which integrates watermarking-based approach with signature-based approach. Robust signatures are cryptographically generated based on invariant features called *significance-linked connected component* (SLCC) extracted from image content and are then signed and embedded back into the image again as watermarks, all in the wavelet domain. De-noising and morphological filtering are applied as pre-processing to tame some small perturbations on extracted features caused by various incidental distortions introduced in acceptable manipulations such as lossy compression, common image processing (blurring, sharpening, etc.) as well as watermarking. Error correcting coding is employed to further bridge between generated signatures and watermarks in a novel way: message bits are formed based on SLCC features, and parity check bits are taken as the seeds of watermarks. The generated signature is hashable and can be incorporated into a PKI framework.

Keywords digital signature, PKI, watermarking, image authentication, error correction coding

1. INTRODUCTION

Semi-fragile image authentication concerns with verifying authenticity of a received image while allowing some acceptable manipulations. Lossy compression is a typical acceptable manipulation. Some common imaging applications often involve multiple cycles of decompressing the image, performing some common image manipulation tasks (e.g., filtering, sharpening, etc.) and then re-compressing the edited image using default or user-defined parameters. Designing such authentication system which can tolerate incidental distortions but reject malicious modifications is the objective of this paper.

A successful semi-fragile authentication system should have a clear definition of acceptable manipulations. It should allow the acceptable manipulations passing the authentication while alert the malicious modifications on the content. In this paper, we define acceptable manipulations as follows, more detailed analysis on their associated distortion distributions are given in [1].

- The first type of acceptable manipulations comes from some common imaging and editing applications including multiple cycles of lossy compressions, image quality enhancement processes such as filtering and sharpening, etc.
- The second type of acceptable manipulations is due to variations in implementing image codecs. For instance, in JPEG2000 codec, precision of wavelet transform (WT) filters may vary. Even in some cases [3], the WT filters used for decoding could be different from those used for encoding.
- The third type of acceptable manipulations is from the procedure of watermarking and other noise sources. The procedure of watermarking is actually a "noise" adding procedure.

In [1][4], we have proposed a generic semi-fragile image authentication framework combining invariant feature based signature, ECC and PKI. The framework is general and does not restrict the use of any specific invariant feature. In this paper, we shall describe new methods in deriving invariant features from the wavelet domain. Unlike earlier solutions developed for specific compression standards (JPEG or JPEG2000), the new methods are based on generic wavelet transform domains, preprocessing, and ECC. The rest of the paper is organized as follows. In section 2, we shall describe what the invariant features are used for signature generation and how the pre-processing procedures contribute to stabilize feature set under defined acceptable manipulations. Some experimental results are shown in Section 3, followed by conclusions and future work in Section 4.

2. PROPOSED SOLUTION

In our proposed solution[1][4], signature generation / verification modules are mainly employed for the role of content signing and authentication. Watermark embedding / extraction modules are used for storing signatures. Refer to Figure 1., in the procedure of content signing, the input original image is firstly decomposed by wavelet transform followed by adaptive de-noising operation.

2.1 Content adaptive de-noising

The purpose of de-noising aims at stabilizing extracted feature set at authentication site under various pre-defined

acceptable manipulations. In [5], S.G. Chang *et al* proposed solution for adaptive wavelet thresholding for image de-noising and compression called BayesShrink. Their algorithm is summarized as follows. Firstly the noise variance σ^2 is estimated empirically or directly based on the image:

$\hat{\sigma} = Median(|Y_{ij}|)/0.6745$, where $Y_{ij} \in$ subband $HH1$ and assuming it is noise-corrupted. The "clean" signal standard deviation σ_X could be computed from:

$\hat{\sigma}_X = \sqrt{\max(\hat{\sigma}_Y^2 - \hat{\sigma}^2, 0)}$, where $\hat{\sigma}_Y^2 = \left(\sum_{i,j=1}^{n} Y_{ij}^2\right)/n^2$, and *nxn* is the total number of coefficients in a subband. Finally de-noising is conducted by soft-thresholding to obtain the "clean" wavelet coefficients:

$X_t = \mathrm{sgn}(Y) \cdot \max(|Y| - \hat{T}, 0)$, $\hat{T}(\sigma_X) = \hat{\sigma}^2/\hat{\sigma}_X$.

We have tested several sources of noises such as multi-iteration lossy compression, WT filter change and direct noise adding / watermarking. Figure 2 compares distortions introduced by such noises to the images with and without de-noising. The top curve with higher PSNR (1-8 dB, thus much less changes) is achieved when images are pre-processed with denoising before manipulations are applied. We can see that de-noising is helpful in reducing (by 1-8 dB) some incidental distortions caused by acceptable manipulations. In other words, such denoising operation is useful in obtaining "stable" representation of image content that's used for deriving authentication features. The manipulations (listed as 1-12 on the x axis in Figure 2) are: adding noise, full bit-rate JPEG2000 compression, JPEG with quality factor (QF) 8, WT change from 9x7 float to 5x3 integer, JPEG2000 transcoded from 2bpp to 0.8bpp in two types of progression orders: resolution based and SNR based, change quantization step size, 10 times JPEG2000 compression with 0.8bpp, 10 times JPEG compression with QF8, Sharpening, JPEG compression with quality factor 3 and SLCC based compression[6] etc.

2.2 Morphological operation and SLCC

After de-noising, the wavelet coefficients are thresholded to obtain a binary "significance map", which is composed of all WT coefficients above the thresholds, by a pre-set authentication strength. If we want the authentication strength towards a fragile direction, then the threshold should be low. Morphological filtering is then performed on the significant map for smoothing and clustering. The reason why we employ morphological operation further on significant map is because the observation shows the coefficient perturbation caused by incidental distortions is most likely around significant coefficients and its capability of reducing noises[6][7]. The morphological operation we adopted is called conditioned dilation as adopted in [6], i.e., $(S \oplus B) \setminus S$, where S is the set where the dilation operation will be applied to, B represents some structuring elements, $\oplus$ means the dilation operation and \ means difference operation. Figure 3 shows the contribution of morphological operation. Figure 3(a) is the significant map of original image decomposed with 5 levels. Figure 3(b) is noise corrupted results of Figure 3(a). Figure 3(c) is the difference between original significant map and noise corrupted significant map (without applying Morphological filtering). Figure 3(d) is the difference between original significant map and noise corrupted significant map. (with morphological filtering applied to original and manipulated images). As we can see from Figure 3(d), morphological filtering helps reduce the distortion caused by acceptable manipulations.

It is well known that the "quality" of the feature selected for generating signature plays a key role in a semi-fragile authentication system. The requirement on feature selection is that any changes resulting from malicious attack on the image content should make generated signature change. If the signature missed detection of malicious attacks, it creates security holes. If it rejects acceptable manipulations, then it's too "fragile". To obtain a good trade-off between system robustness and security, we take the clustered significant map as the feature for signature generation. Actually we can understand the formed significant map is a skeleton of image content in wavelet domain and pyramid structure. Instead of only recording the significant coefficients, we also take insignificant coefficients into account: Refer to Figure 4, starting from LL subband, if the coefficient is significant, we record a "1". Go to its child subband, if a child coefficient is significant too, we record a "0"; Otherwise, record another "1". Continue till the resolution level and subband to be protected and obtain binary feature sets based on blocks. For example, if we want to protect the image with 3 levels, pick one coefficient from the lowest LL subband and one from its siblings (LH/HL/HH, let's say HL as shown in Figure 4), then pick up its 4 children in the next level and 16 grandchildren in the 3rd level. Thus we shall obtain 1+1+4+16=22 bits as one set of features for protecting a 4x4 blocks. Note that our SLCC coding differs from those for compression [6][7] because of different purposes.

2.3 Signature generation and watermarking

We apply ECC encoding on the binary feature sets derived by the above procedure. Such ECC process generates corresponding parity check bits (PCBs) of each block-based feature set. (We refer it as local or block signature) An example of ECC scheme is BCH[8]. Taking PCBs as the seed of watermark (without the original feature set) to form the block-based watermark. The requirement on watermarking is that the embedded watermark should be robust enough for extraction from received images under

acceptable manipulations. Since distortion may be caused by acceptable manipulations, the embedded watermarks may undergo some minor changes. To solve this problem, we apply another ECC: PCBs are encoded by another ECC scheme and then the ECC encoded output is embedded as watermark data. In our system we adopt a quantization based watermarking proposed in [9]. The watermarks could be embedded either into other subbands at the same resolution level or a lower decomposition level. As another layer of protection, all codewords (features together with their corresponding PCBs) are concatenated and hashed by a cryptographic hashing function such as MD5[10]. Finally, content owner's private key is used to sign the hashed value (We refer it as the global signature). The encrypted hashed value can be stored in a place external to the image content such as its file header or embedded into the image again as another watermark.

Refer to Figure 1 (Lower part). The authentication procedure is the inverse procedure of signing except using content owner's public key for signature verification. Given the image to be authenticated, we repeat feature extraction described in content signing procedure. From the embedded watermarks, we also extract the corresponding PCBs. Note the features are computed from the received image, while the PCBs are recovered from the watermarks that are generated and embedded at the signing site. After we combine the feature set and the corresponding PCBs to form codewords, the whole authentication decision could be made orderly. First, we calculate the syndrome block by block to see whether there exist any blocks whose codewords are uncorrectable. If yes, then we could claim that the image is unauthentic and indicate alteration locations where the extracted features extracted from received image do not match the extracted PCBs from the watermarks. If all codewords are correctable, we replace any erroneous codewords with their PCB corrected ones. Then we repeat the same process at the signing site: concatenate all corrected codewords into a global sequence and cryptographically hash the result sequence. By using the owner's public key, the authenticator can decrypt the hashed value that's generated at the signing site. The final authentication result is then concluded by bit-by-bit comparison between these two hashed sets: if there is any single bit different, the authenticator will report that the image unacceptable ("unauthentic").

Now we have a clear understanding on the roles of block signature and global signature: The global signature can be hashed to enhance system security. Local signatures cannot be hashed but serve to detect local changes and locate alteration areas.

3. EXPERIMENTAL RESULTS

Figure 5 shows an experimental example of image authentication. Figure 5(a) is the original image. Figure 5(b) is noise corrupted image done by adding Gaussion noise with the strength 10 with Photshop5™. Figure 5(c) is the result of original image compressed by JPEG2000[11] with bit-rate 0.8bpp. Figure 5(d) is the result of original image compressed by JPEG with quality factor 3 done by Photshop5™. (The file size is comparable to (c)). They all can pass authentication. An image after content altering attacks is shown in Figure 5(e) – including attacks of crop/paste, insertion, deletion, intensity changes etc (e/g/. see the time change on the clock). Successful detection and location of attacks are shown in Figure 5(f). In our testing scheme we did not embed global signature into image. ECC schemes for signature generation is BCH (31,21,2) and for watermarking is BCH (15,11,1). On-site demonstrations will be further given at the conferences.

It is noted that we did not address watermarking issues in more details such as capacity, robustness evaluation etc in this paper due to paper size limits. More detailed implementation issues and analysis will be presented in our future publications.

4. CONCLUSIONS AND FUTURE WORK

We have introduced a generic wavelet based solution for robustly and securely authenticating images in a semi-fragile way. Significant-linked connected components are taken as invariant features for signature generation. Wavelet based adaptive de-noising and morphological operation are adopted for stabilizing extracted features under incidental distortions. Furthermore, the proposed solution can be incorporated into PKI by employing ECC in a novel way: PCBs are taken as the seeds of watermarks not the message bits.

Future work include exploring new image processing tools to tame more incidental distortions, finding new features, analyzing models for incidental and intentional distortions, fine-tuning individual modules among signature generation, ECC and watermarking to further reduce security risks, etc.

5. REFERENCES

[1] Qibin Sun and Shih-Fu Chang, Semi-Fragile Authentication of JPEG-2000 Images with a Bit Rate Control, *Columbia University ADVENT Technical Report*, 2002-101.

[2] Information Technology—JPEG2000 Image Coding System, *ISO/IEC International Standard 15444-1*, ITU Recommendation T.800, 2000.

[3] Margaret Lepley, Effects of decoding JPEG2000 9x7 with an irreversible 5x3, *ISO/IEC JTC 1/SC 29/WG1 N2281*, 2001.

[4] Qibin Sun, Shih-Fu Chang, Maeno Kurato and Masayuki Suto, "A new semi-fragile image authentication framework combining ECC and PKI infrastructure", *ISCAS02*, Phoenix, USA, May, 2002.

[5] S. Grace Chang, Bin Yu and Martin Vetterli, Adaptive wavelet thresholding for image denoising and compression, *IEEE Trans on Image Processing*,Vol.9, No.9, pp1532-1546, 2000.

[6] Sergio D. Servetto, Kannan Ramchandran and Michael T. Orchard, Image coding based on morphological representation of

wavelet data, *IEEE Trans on Image Processing*, Vol.8, No.9, pp.1161-1174, 1999

[7] B.-B. Chai, Jozsef Vass and Xinhua Zhuang, Significant linked connected component analysis for wavelet coding, *IEEE Trans on Image Processing*, Vol.8,No.6,pp.774-784, 1999

[8] S. Lin and D. J. Costello, JR., Error control coding: Fundamentals and applications, Prentice-Hall, 1983

[9] L. Xie and G. R. Arce, Joint wavelet compression and authentication watermarking, Proceedings of ICIP'98. 1998.

[10] B. Schneier, *Applied Cryptography*, New York:Wiley, 1996.

[11] JJ2000: An implementation of JPEG2000 in JAVA™, available at http://jj2000.epfl.ch.

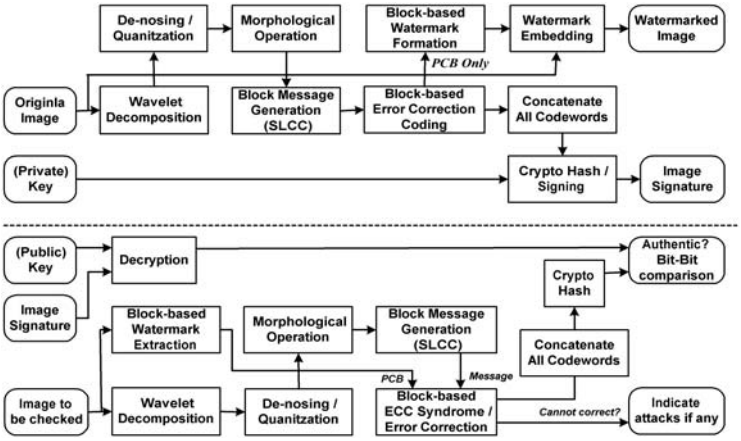


Figure 1. Proposed system diagram for image authentication
Upper: content signing Lower: content authentication

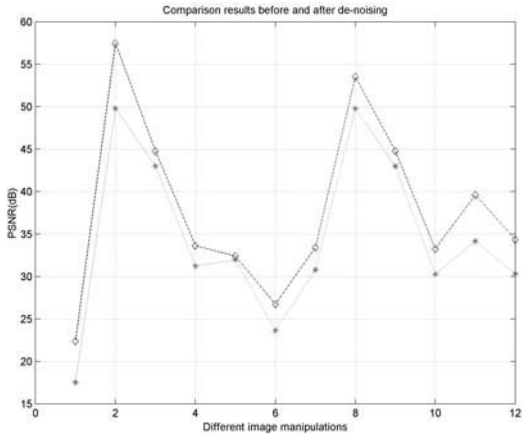
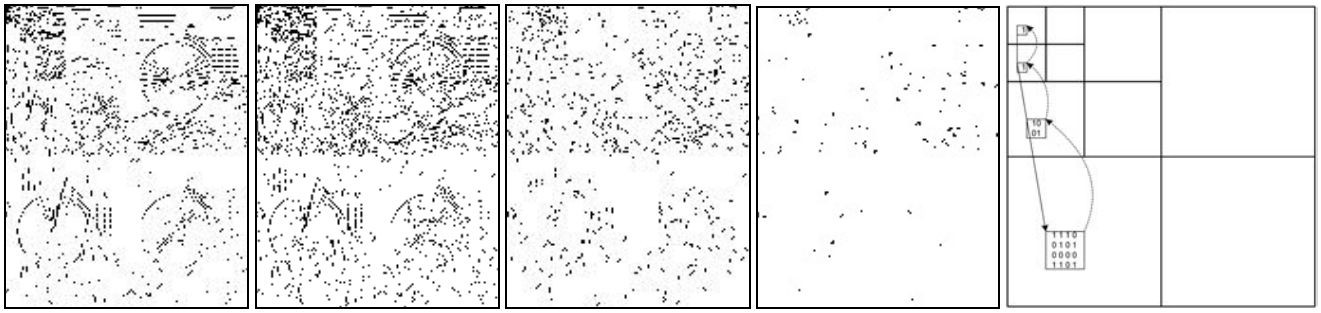


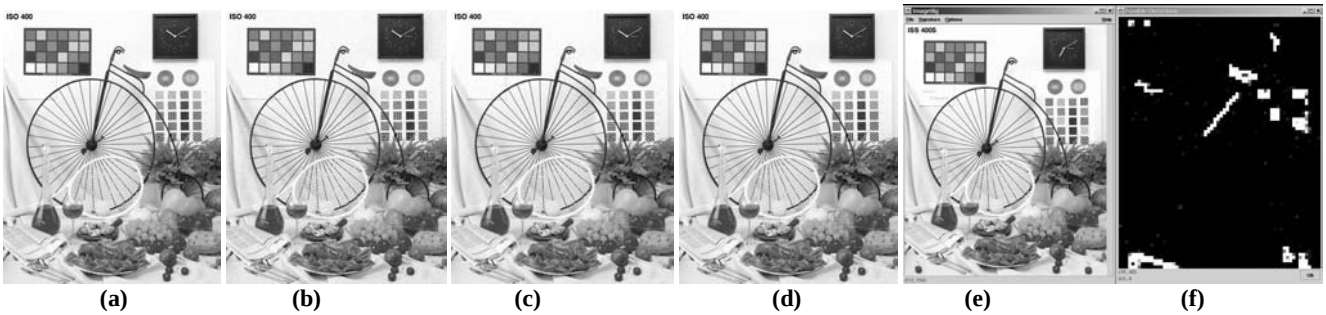
Figure 2.Comparison results without (lower line) and with de-noising (upper line)



(a) **(b)** **(c)** **(d)**

Figure 3. Illustration on the contribution of morphological operation
(a) Original significant map (b) Noise corrupted significant map
(c) Difference between (a) and (b) without morphological operation
(d) Difference between (a) and (b) with morphological operation

Figure 4. Illustration on coding SLCC for block signature



(a) **(b)** **(c)** **(d)** **(e)** **(f)**

Figure 5. Examples of authentication results (image size: 512x640)
(a) Original image (b) Noise corrupted by Gaussian 10 (c) JPEG2000 compressed image with bit-rate 0.8bpp
(d) JPEG compressed image with quality factor 30 (e) attacked image (f) Authentication result of (e).
Note that (b) (c) (d) can all pass the authentication