ISSN 1392 – 124X INFORMATION TECHNOLOGY AND CONTROL, 2008, Vol. 37, No. 4

A NEW SCHEME FOR IMAGE AUTHENTICATION FRAMEWORK

Romualdas Baušys, Artūras Kriukovas

Department of Graphical Systems, Vilnius Gediminas Technical University
Saulėtekio al.11, Vilnius, Lithuania

Abstract. The paper proposes a digital signature-based approach for image authentication, pixel-wise tamper localization and iterative restoration. Usage cases for semi-fragile watermarking and proposed digital signature are analyzed, the advantages of the digital signature method are presented. Integration with Public Key Infrastructure (PKI) is outlined and discussed.

1. Introduction

While the rapid development and deployment of new IT technologies has improved the ease of access to digital information, it has also led to fears that identity and authenticity could be eroded by the illegal copying, modification and redistribution of digital media. This presents real threats to, for example, commercial published digital audio and video, whose existence depends on credibility and clarity of their information assets.

Image authentication is a relatively new research area compared to more traditional research areas such as multimedia compression. Scientists in different areas and with different technical backgrounds may use different definitions for "authentication" term. In our paper we define authentication as the process by which it is possible to determine whether meaning of image in question has been altered. Thus we allow minor content preserving modifications and try to identify content changing modifications.

Content-based image authentication methods are historically classified in (a) watermark based and (b) digital signature based.

The main advantage of watermarking is authentication data integration in the image to be authenticated. Theoretically this decreases hassle in authenticity establishment process. It is declared that watermarked image is enough to determine authenticity of the image – in contrast to digital signature methods that require additional signature file. However this holds true only for blind watermarking case. Semi-blind or non-blind watermarks require additional, side information, in order to authenticate the watermarked image in question.

Disadvantages of watermarking are the limited payload – it is image-dependent and cannot be fixed in algorithm design stage – and distortions introduced in original image data [9]. As minor as they may be – in some cases they are not acceptable. Lossless watermarking methods have appeared as a possible solution – with the increased complexity, fragility and processing overhead [2].

Digital signatures, on the other hand, do not possess direct disadvantages of watermarking – because of different authentication data handling. They do require additional file – signature – to be transferred with the image. The content of the file is subject of different research directions – message authentication code (MAC) (with variations – approximate message authentication code – AMAC, approximate image authentication code AIMAC) [5, 18], visual hash [16], robust hash [8, 17] and digital signature itself [13]. All these methods follow the same path, like watermarks – feature extraction and subsequent use of the feature for later authentication – with variations in features chosen, processing and extraction mechanisms.

First methods of digital signature were based on cryptographic digital signature functions. But when the requirement to authenticate content rather than file became evident, the term digital signature has evolved and changed its meaning in multimedia authentication domain. Robust features, extracted from the image, became basis of the new digital signature. Edge characteristics were computed and transformed into feature codes [7], with feature points encryption [4]. Histogram techniques [13], image moments [6], image feature triangulation [14] were also used. Features extraction in various domains followed – a robust hash based on singular vector decomposition, was proposed in [11]. A relationship between DCT coefficients at the same position in different blocks analyzed in [12] as incidental modifications always disturb these coefficients. This feature is robust to multicycle JPEG compression, but it is not robust to image scaling. Wavelet based hashing in [15] uses the idea that the

inter-scale relationship is difficult to be destroyed by content preserving manipulations and hard to be preserved by content changing manipulations. A scheme in [4] is tolerant to lossy image compression but can detect malicious image manipulations.

However all these solutions were not enough to ensure effective digital signature for the content of the image with pixel-wise tamper localization.

Some scientists extend the aforementioned methods by combining them with additional mechanisms, like error correction codes (ECC). In some cases ECC is applied for the initial data, extracted features, digital signature itself or only parities of ECC are used [10, 11] in order to further expand the methods of digital signature.

In this paper, we present a digital signature-based method for image content authentication and tamper localization establishment. Although we do use ECC, the usage is limited to discrete vectors, not on the whole image as in previous works. The method is based on the digital signature strategy combined with wavelet transformation and integration into PKI infrastructure. Generated signature is signed by authorized persons and published in the Internet. Possession of the corresponding public keys allows verifying authenticity of the digital signature. Digital signature itself is used to establish identity of the image in a semi-fragile way.

2. Digital signature workflow analysis with image watermarking

In the latest developments of image authentication and tamper localization methods, we notice a divergence of development direction between watermarking and digital signatures schemes. Image watermarking as well as image signatures are used to reach the following objectives [8]:

1) copyright protection – it is the most analyzed and still discussed problem.
2) data authentication – the main objective is to identify global authentication. Tamper localization is desired, but optional.
3) fingerprinting – used to ensure unique data-owner relationship identification. In this case it is important not to establish owner identification, but to define data path traceability.
4) copy protection – has more specific usage case – only to stop unauthorized copying. This is usually used in hardware solutions – CD/DVD players.
5) broadcast monitoring – is similar to fingerprinting, but with more "neutral" usage – to monitor radio/ television broadcasts, advertising, to ensure correct usage of information channels.

We will focus our attention on image watermarks for data authentication only.

The main advantage of image watermarking over the image signature is the strict image-watermark association. Watermarks are always embedded in image data – spatial or transform domains, fragile or robust way. In any case, watermark is transferred together with the image without additional hassle. Watermark can be extracted and used to process the image as required. In blind watermarking case, this association makes it easier for final customer – there is no need to remember of additional file to transfer – the image has all required data inside. But in case of semi-blind or non-blind watermarks, this advantage against digital signatures disappears.

Furthermore, direct watermark embedding in image data presents additional problems and disadvantages:

1) watermark degrades image quality – there is limited set of images that really have "enough space" to embed the watermark. Majority of images do not have enough space. This means that embedding watermark discards some image data at the same time. As the size of watermark increases, more and more free space in the original image is required. Furthermore, some applications – medical, military – require access to the original image. Reversible (lossless) watermarking methods have appeared as a response [2] at the price of more complicated and vulnerable watermarking process.
2) watermark is declared to be invisible data. But nevertheless it can be detected by cryptanalysis, steganalysis methods – as watermark data embedded differ from ordinary image data. These discrepancies can be analyzed, allowing to detect embedded watermark data and attack them. With the increase of the size of the watermark, more data for steganalysis methods is presented.
3) the amount of data that can be embedded by watermark is limited. There do exist upper limitations – from "free space" available at image, from efficiency of steganalysis methods, from visible distortion of the image. High-performance methods try to deal with this problem, but the more data are embedded into the image, again, the more vulnerable it becomes. This positively affects the efficiency of steganalysis methods and attacks [19].
4) watermark can be damaged by conventional image processing methods. "Red eye" removal process, image histogram modification to improve image brightness – may be enough to damage the watermark, especially lossless or high-performance one. Simple, conventional image processing has undesired effects on embedded watermarks.

In this context, image signature-based methods are more favored, as most watermark-originating problems are solved by definition of image signatures. They embed nothing in image – no image degradation is produced, no limit on the size of signature data is imposed, no impact on image processing operations. Steganalysis methods and approximately half of the watermarking attacks are effectively disabled.

It is obvious that the main disadvantage of the digital signature based methods in compare with blind-watermarking is the usage of digital signature as of separate file. Anyone who needs to authenticate the received image has to request the source to provide the signature. This may be inconvenient in some cases. However it should be noticed, that semi-blind or non-blind watermarking cases also require additional, separate file (watermark) in order to authenticate the image. In these cases we have no advantages of the watermark, only disadvantages.

In this paper we take into account usage processes of watermark and digital signature for average user. We show that at least in these cases the usage of digital signature is more advantageous than that of digital watermarking.

We took an independent look at the problem and analyzed the situation from the process workflow point of view.

Semi-blind watermarking can be defined as a process, which takes original image and watermark as input parameters and returns watermarked image as an output. Authenticity verification process analogically can be defined as taking suspected image and watermark and returning boolean answer (or probability) of authenticity of the image in question. Standard semi-blind watermarking process model, where the user has to provide to a computer the image in question and the watermark itself and the boolean answer for the user is generated, is presented in Figure 1.

If we analyze digital signature case, we observe that no serious differences exist. The user still has to provide to a computer the image in question and additional file (signature file) and the same boolean answer is generated.

Representation of the same process for digital signature case is provided in Figure 2.

But although there is no difference from the usage perspective, there are important differences on the quality and effectiveness of image authentication and tamper localization establishment. Therefore we propose an advanced digital signature method.

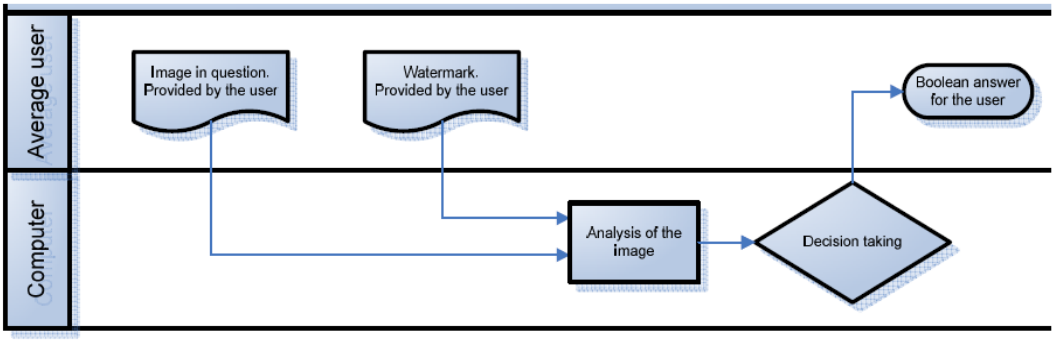

Figure 1. Authenticity verification process, watermarking

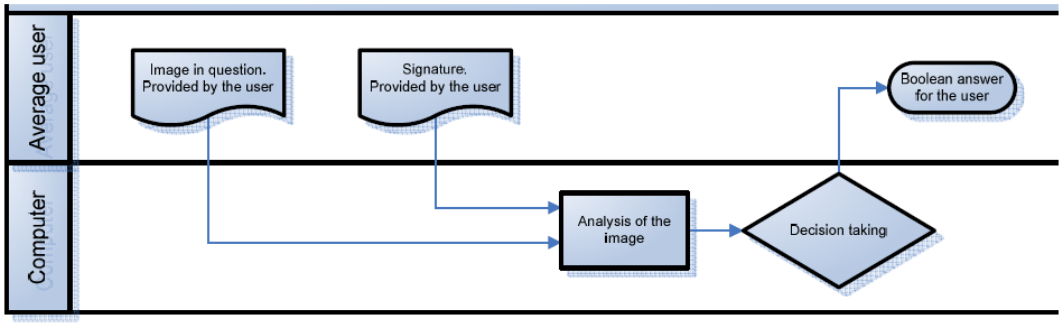

Figure 2. Authenticity verification process, digital signature

3. Digital signature-based method integration with public key infrastructure

Hard authentication ensured by PKI scheme is not quite suitable for semi-fragile image authentication. Therefore our proposal is to integrate our digital signature method into PKI.

Legacy PKI (Public Key Infrastructure) plays an important role in asserting the ownership of a user's public key by CA (Certificate Authority) as a Trusted Third Party. In general, PKI infrastructure involves the following participants:

1) certificate authority CA, which can issue and revoke a PKI certificate;
2) registration authority RA, which handles identity verification;
3) directories to store sensitive information;
4) customers that define business needs.

PKI digital signature scheme should not be mixed with asymptotic public-key cryptography, as PKI integrates not only cryptographic functions, but additional members in PKI scheme – users, certificate

and registration authorities and certificate directories (Figure 3).

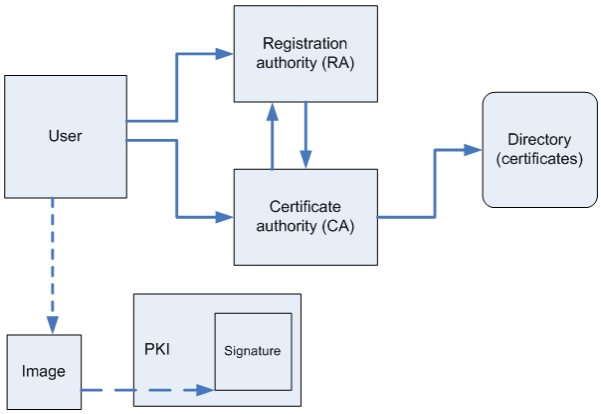

Figure 3. PKI scheme for image authentication

Our proposed method for image authentication with pixel-wise tamper localization meets the following PKI objectives:

a) integrity. Data integrity is implemented in two stages. The first stage is based on security provided by public-private key scheme itself. Both malicious and accidental alterings are detected by public-private key scheme. Subsequent image analysis, malicious manipulation detection and tamper localization is performed in second stage.
b) authentication. Data authentication establishment allows identifying author (owner) of the data, based on standard procedures of PKI/CA.
c) non-repudiation is based on procedures of PKI and ensures protection for the author (owner). This protection is not the objective for image signature we propose.

Proposed integration of digital signature bear some similarity to the standard PKI signing procedure. Difference from PKI comes from efficient image analysis method – instead of cryptographic hash, digital signature of the image is hashed. This provides additional advantages required in image processing – robustness to common image processing operations and tamper localization – that results in fast and robust image authenticity establishment.

It should be noticed that possible digital signature leakage from CA would present no real commercial threat for the author. As it is impossible to regenerate original high value image from the signature, for unauthorized parties possession of low-value signature gives no advantages.

4. Proposed method

The method we propose is integrated with PKI infrastructure. It allows identifying image authentication and locating tampered regions with optional restoration ability. There are two main stages of the method – stage A (authentication) and stage TL (tamper localization). As tamper localization is separated from authentication, any changes in tamper localization part do not affect data authentication – therefore oracle attack is completely disabled [1].

We use DWT decomposition to generate low value version of the image for authentication purposes:

$$Wf(u,s) = (f, \phi_{u,s}) = \int_{-\infty}^{\infty} f(t) \frac{1}{\sqrt{s}} \phi^*(\frac{t-u}{s}) dt \qquad (1)$$

$$\phi_{u,s}(t) = \frac{1}{\sqrt{s}} \phi(\frac{t-u}{s}) \qquad (2)$$

where f(t) is a given signal in time domain t, u is a translation factor and s is the scale factor.

The decomposition was chosen because it is designed to retain the most important features of the image. In the algorithm, DWT functions as a semi-fragile one-way function, i.e. it is mathematically impossible to restore high value version of the image from the signature. The size of the data in the 3rd decomposition level is 1% of the initial size of the image, time to calculate DWT is O(n). DWT is semi-fragile in this case, because the 3rd (or higher) decomposition level is not influenced by minor modifications arising, for example, from file format changes.

In order to proceed to digital signature stage TL, original image is down sampled at first. The down sampled image data are interleaved and forwarded to ECC process as initial data field F_q. F_q is partitioned into sets $V_H(:)$ and $V_V(:)$. For a vector in each set, the corresponding Parity Check Bits (PCB) are calculated.

In our scheme we chose an approach based on convolutional codes where each m-bit information symbol to be encoded is transformed into an n-bit symbol. m/n is the code rate ($n \geq m$) and the transformation is a function of the last k information symbols, where k is the constraint length of the code. Block codes of length n and rank k are defined in a linear subspace C with dimension k of the vector space $F_q(n)$ where F_q is the finite field with q elements. In this case, we define the block codes as a fixed length channel code – a block code takes a k-digit information word, and transforms it into an n-digit codeword.

We use $n = 17$ to generate a digital signature of the image for tamper localization. However, we propose to exploit the advantage of ECC capability to identify tampered blocks, not to restore them. ECC ability to restore tampered regions gave rise to another interesting effect – we named it iterative restore process. The process is based on the fact that after the first ECC pass, the values of restored pixels can be used for the second pass, thus increasing the total amount of identified and corrected pixels. We exploit this advantage in our algorithm.

Digital signature stage A is designed to achieve two objectives.

First, it addresses the problem with digital signature based image authentication – it provides a computationally efficient way to establish correct

image-signature pair having a lot of different signatures from different images. Second, it allows determining image authentication in semi-fragile way. It is insensitive to image content preserving modifications and sensitive to operations that modify the image in a major way. Furthermore, design of the authentication mechanism integrates a backup option – human interaction.

Tamper localization procedure is based on the following logic: let x be an element in F_q. Tampered $x`$ may be recognized by identifying discrepancies in PCBs and exploiting the correlation of sets $V_H(:)$ and $V_V(:)$, without the expensive search of nearest word within the Hamming distance from $x`$.

Any $x`$ from the image in question can acquire one of three possible states: (1) trusted; (2) uncertain; (3) damaged. These states are correlated from the definition in generation process. Combining these states, the third, final state can be generated. In this case identification of the final state helps to finish processing of uncertain vectors in state #1 or state #2. Details of the algorithm are presented in [3].

5. Algorithm

Proposed digital signature generation process involves the following steps:

1. Original image I is provided by the user. If secret behavior is required, secret key K has to be provided by the user as well. If public behavior is expected, K is initialized to a known constant value.
2. Low value image I_L is generated from I using DWT as a semi-fragile one-way function.
3. Down sampled image I_d is generated from I. The image I_d is interleaved according to a pseudo-random number generator initialized by K.
4. I_d is transfomed into set VH.
5. I_d is transformed into another set VV. Transformation order for VV has to be different from transformation order of VH.
6. For each vector in VH/VV, ECC parities are calculated.
7. I_L and PCBH/PCBV are combined into a digital signature.

Based on numerical experiments with the defined n, k parameters of Reed-Solomon code, the size of the PCBs is approximately 30% of the original image.

Image authentication process involves the following steps:

1. Suspected image $I`$ is provided by the user. If secret key K was used, it should be provided too.
2. Digital signature S may be provided by the user. Alternatively, digital signature S may be found in the database of digital signatures (in case of authentication center).
3. Low value image $I`_L$ is generated from $I`$. Trust level of image-signature pair is established.
4. If $I`$ was tampered, tamper localization procedure is executed and damage map is generated.

In order to check image authentication, a digital signature should be provided. If authentication center participates in the process, corresponding digital signature has to be found. In both cases low value image $I`_L$ is generated at first. Then it is used to find the corresponding digital signature and to establish trust level of image-signature pair.

When sufficient trust level between image in question and digital signature has been established, it is possible to run complete tamper localization process. We would like to notice that correct image authentication is not a requirement for tamper localization part, i.e. image authentication helps to locate the corresponding digital signature efficiently, to prevent oracle attack, to determine authentication of the image but, if required, tamper localization can be run without determining image authentication.

The efficiency of authentication establishment depends on the efficiency of wavelet decomposition process – $O(n)$.

For tamper localization additional steps are performed:

1. Down sampled image I_d is generated from I. The image I_d is interleaved according to a pseudo-random number generator initialized by K.
2. I_d is partitioned into sets VH, VV.
3. Each vector in VH, VV is checked for tampering, additional logic is applied.
4. If $I`$ was tampered, damage map and restored image I_R are generated.

6. Numerical experiments

Numerical experiments were performed with standard images. We present results for the Mandrill and Cameraman images. The images were affected by local attacks – "LNK" and a picture of the crow were added as copyright signs (Figure 4, Figure 7). As we see, the method we propose performs successfully. The results of the analysis, despite similar attacks, reveal the advantages of iterative restore procedures. As we see in Figure 5, Cameraman tamper localization performed perfectly – clearly marking contours of "LNK" and crow. Figure 6 presents restored image. Some differences from the original one can be observed and attacked regions are easily identifiable, but the original image, before the attack, can be seen and understood easily as well. PSNR of the image is 49 – very high, if we take into account the type of the attack.

Image of Mandrill was the object of similar attack. The image was chosen because of hard to manipulate structure – and we managed to ensure protection of the image. A third iteration was required to restore the

image and the structure of the image does not allow identifying tampered regions in restored picture with naked eye. The black pixels in Figure 8 are still lost in the current iteration. They may be restored in the following iterations.

Analysis of the results is provided in Table 1.

Table 1. Results from iterative process

	After iteration #1	After iteration #2
V_H, detected as trusted	3326	5104
V_H, detected as recovered	1755	16
V_H, detected as damaged	39	0
Amount of damaged pixels	2174	5
Amount of lost pixels	6	0
Max amount of damaged px in one V_X	5	5
Avg amount of damaged px in one V_X	0.01700	0.00097

Figure 4. Attacked Cameraman

Figure 5. Damage map

Figure 6. Restored image. PSNR=49

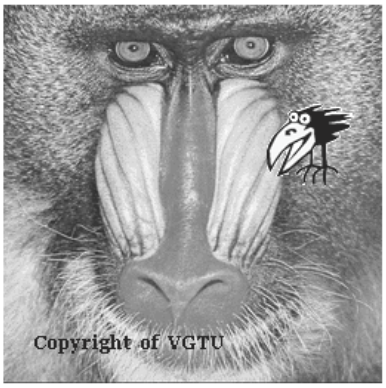

Figure 7. Attacked Mandrill

Figure 8. Damaged pixels after #2 iteration

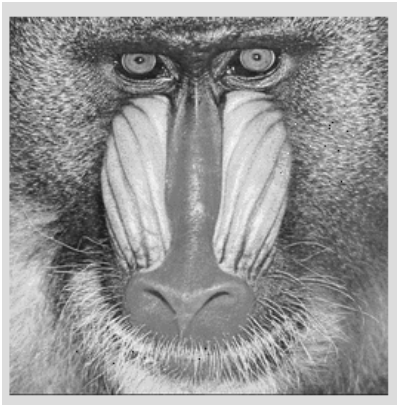

Figure 9. Restored image after #2 iteration

7. Conclusions

In this paper, an advanced semi-fragile digital signature method is presented. The proposed approach enabled us to authenticate image in question – simple but effective design is capable to withstand algorithmic attacks like oracle attack. Innovative 2D analysis, that mimics 2D image structure, allows identifying tampered regions with resolution up to one pixel, using block-based scheme. ECC principles gave rise to iterative restore procedures, thus enabling restoration of damaged image from the digital signature.

References

[1] **R. Baušys, A. Kriukovas.** A Secure Watermarking Scheme for Image Authentication in the Frequency Domain. *Information sciences,* ISSN 1392-0561. *Vol*. 34, *Vilnius University*, 2005, 252-256.

[2] **R. Baušys, A. Kriukovas.** Reversible watermarking scheme for image authentication in frequency domain. 48th *International Symposium ELMAR*-2006 *focused on Multimedia Signal Processing and Communications* 07-09 *June* 2006, *Zadar, Croatia*, ISSN 1334-2630, 53-56.

[3] **R. Baušys, A. Kriukovas.** Digital signature approach for image authentication. *Electronics and electrical engineering*, 2008, *No*.6 (86), 13.

[4] **S. Bhattacharjee, M. Kutter.** Compression Tolerant Image Authentication. *Proceedings of the International Conference on Image Processing, Chicago, USA*, 1998, 435-439.

[5] **Ch. Boncelet.** Image authentication and tamperproofing for noisy channels. *International Conference on Image Processing IEEE, Italy, September* 2005, 677-680.

[6] **A. Datta, N.D.V. Lobo, J.J. Leeson.** Novel feature vector for image authentication. *Proc. IEEE International Conference Multimedia and Expo*, 2003, 221-224.

[7] **J.A. Dittmann, R. Steinmetz.** Content-Based Digital Signature for Motion Pictures Authentication and Content-fragile Watermarking. *Proceedings of the International Conference on Multimedia Computing and Systems, Vol*.2, 1999, 209-213.

[8] **J. Fridrich, M. Goljan.** Robust hash functions for digital watermarking. *International Technology: Coding and Computing* 2000, *Las Vegas, Nevada, March* 2000, 173–178.

[9] **K. Gerūta, J. Eugenijus.** Public key watermarking in wavelet domain. *Information technology and control, Kaunas, Technologija*, 2003, *Vol*.28, 33-41.

[10] **M. Johnson, K. Ramachandran.** Dither-based secure image hashing using distributed coding. *Proc. IEEE International Conference Image Processing, Barcelona, Spain, Sep*. 2003, *Vol*.2, 751–754.

[11] **S.S. Kozat, R. Venkatesan, M.K. Mihcak.** Robust hashing via matrix invariances. *Proceedings of IEEE International Conference on Image Processing* (ICIP), *Singapore*, 2004, 677–680.

[12] **C.-Y. Lin, S.-F. Chang.** A robust image authentication method surviving JPEG lossy compression. *Storage and Retrieval of Image and Video Databases* VI, *I. K. Sethi and R. C. Jain, Eds*., *Vol*.312 *of Proc. SPIE, San Jose, CA, USA, December* 1998, 296–307.

[13] **M. Schneider, S. Chang.** A robust content based digital signature for image authentication. *IEEE International Conference on Image Processing* (ICIP'96), 1996, *Vol*.3, 227-230.

[14] **C.S. Lu, C.Y. Hsu, S.W. Sun, P.C. Chang.** Robust mesh-based hashing for copy detection and tracing of images. *Proc. IEEE International Conference Multimedia and Expo: special session on Media Identification*, 2004, *Vol*.1, 731-734.

[15] **Ch. Lu, H.M. Liao.** Structural Digital Signature for Image Authentication: An Incidental Distortion Resistant Scheme. *Proceedings ACM Multimedia and Security Workshop at the* 8th *ACM International Conference on Multimedia, Los Angeles, California, USA, Nov* 4, 2000, 115-118.

[16] **R. Norcen, A. Uhl.** Robust Visual Hashing Using JPEG 2000. IFIP TC6/TC11 *Conference on Communications and Multimedia Security* (CMS'04), *Lake Windermere, GB, September*, 2004, 223-235.

[17] **Q. Sun, S.Roy.** Robust hash for detecting and localizing image tampering. *ICIP* 2007, *Vol*.6, 117-120.

[18] **L. Xie, G.R. Arce, R.F. Graveman.** Approximate Image Message Authentication Codes. *IEEE Transactions on Multimedia*, 2001, 242–252.

[19] **X. Yu, T. Tan, Y. Wang.** Extended optimization method of LSB steganalysis. *IEEE International Conference on Image Processing*, 2005, *Vol*.2, 1102-1105.

Received June 2008.