Localized Lossless Authentication Watermark (LAW)

Mehmet U. Celik[a], Gaurav Sharma[b], A. Murat Tekalp[a] and Eli Saber[b]

[a]University of Rochester, Rochester, NY, USA
[b]Xerox Corporation, Webster, NY, USA

ABSTRACT

A novel framework is proposed for lossless authentication watermarking of images which allows authentication and recovery of original images without any distortions. This overcomes a significant limitation of traditional authentication watermarks that irreversibly alter image data in the process of watermarking and authenticate the watermarked image rather than the original. In particular, authenticity is verified before full reconstruction of the original image, whose integrity is inferred from the reversibility of the watermarking procedure. This reduces computational requirements in situations when either the verification step fails or the zero-distortion reconstruction is not required. A particular instantiation of the framework is implemented using a hierarchical authentication scheme and the lossless generalized-LSB data embedding mechanism. The resulting algorithm, called localized lossless authentication watermark (LAW), can localize tampered regions of the image; has a low embedding distortion, which can be removed entirely if necessary; and supports public/private key authentication and recovery options. The effectiveness of the framework and the instantiation is demonstrated through examples.

Keywords: lossless compression, tamper localization, reversible data embedding, invertible authentication

1. INTRODUCTION

Compared to the analog systems traditionally used for imaging, today's digital imaging systems provide sophisticated processing capabilities, flexibility, and reliability- all at a lower cost and with competitive or better quality. As a result, digital image acquisition, processing, storage, and reproduction systems have been steadily replacing their analog counterparts. Nevertheless, adoption of digital imaging in medical, military and legal fields has been hampered by increasing doubts about the trustworthiness, i.e. authenticity and integrity, of digital images. Due to the limited processing abilities in analog media, malicious manipulation of images has been a tedious task with only low quality results being realized without prohibitively expensive professional equipment. In contrast, digital images can be easily manipulated using a variety of sophisticated signal processing tools that are readily available as commercial packages. Today, photo-realistic manipulations can be created by virtually everyone using low-cost off-the-shelf hardware and software components.

Traditionally, source authentication and integrity verification of digital data have been performed by *digital signatures*.[1,2] Recently, the use of digital watermarks instead of—or in conjunction with—the digital signatures has been proposed for image data.[3–5] Digital watermarks typically afford additional functionality by exploiting the redundancy of the image data and the properties of the human visual system (HVS). One such advantage is the direct embedding of authentication information into the image, wherein the watermark—thus the information it represents—is tightly bound to the image and survives even when the host image goes under a format conversion. In contrast, a digital signature appended in the header of an image file can accidentally be stripped off, when the file is opened and saved in an alternate format, even though the data itself is unaltered. Another important functionality supported by the use of digital watermarks is *tamper localization*. Tamper localization refers to the identification of the image regions that have been tampered (manipulated) after the insertion of the authentication watermark.

Send correspondence to M.Celik: E-mail: celik@ece.rochester.edu, Telephone: 1 585 275-8122, Address: Electrical and Computer Engineering Department, University of Rochester, Rochester, NY, 14627-0126, USA, WWW: http://www.ece.rochester.edu/projects/iplab

Security and Watermarking of Multimedia Contents V, Edward J. Delp III, Ping Wah Wong, Editors, Proceedings of SPIE-IS&T Electronic Imaging, SPIE Vol. 5020 (2003)

It is worth mentioning that, both digital signatures and authentication watermarks are useful only for establishing the source of the image and detecting manipulations occurring after the signature/watermark has been inserted. However, neither technique by itself is capable of certifying that a signal represents an original unaltered scene, unless supported by additional mechanisms.[6]

The additional functionality offered by digital watermarks, however, often comes at the expense of image fidelity. Most watermarking techniques modify, and hence distort, the host signal in order to insert the authentication information, and furthermore, the watermarked image rather than the original is authenticated. The distortion induced on the host image by the watermarking procedure is called the *embedding distortion*. Often, the embedding distortion is small and bounded, yet irreversible, i.e. it cannot be removed to recover the original host image. In many applications, the loss of image fidelity is not prohibitive as long as original and modified images are perceptually equivalent. On the other hand, in medical, military and legal imaging applications, images are often enlarged, enhanced or further processed by image processing algorithms. Potential sensitivity of post processing operations to the embedding distortion and mission critical nature of these applications prohibit the permanent loss of image fidelity during watermarking. The loss of signal fidelity can be remedied by the use of *Lossless Authentication* (also referred as *reversible*, *invertible* or *distortion-free* authentication watermarks) techniques.[7–11] These methods, like their lossy counterparts, insert authentication information by modifying the host signal, thus induce an embedding distortion. Nevertheless, they also enable the removal of such distortions and the exact—lossless—restoration of the original host signal.

In this paper, after reviewing earlier methods, we propose a novel lossless authentication framework, which provides greater flexibility and improved computational performance (Sec. 2). The new framework, in contrast with the earlier methods, verifies the authenticity and integrity *before* recovering the original, unwatermarked image. This reduces the computational requirements in cases where *i)* the image is not authentic, i.e., it either does not bear a watermark or has been tampered after the watermark insertion, or *ii)* the watermarked image is of sufficient quality and reconstruction of the original is not necessary. In Sec. 3, we present a particular instantiation of this framework, which utilizes the Hierarchical Authentication Watermark[4] and the Lossless Generalized-LSB data embedding mechanism.[10,11] The new algorithm is called Localized Lossless Authentication Watermark (LAW). Localized LAW improves earlier methods with its tamper localization capability, low and optionally reversible embedding distortion, and flexibility. We demonstrate the effectiveness of Localized LAW through examples in Sec. 4, before drawing conclusions in Sec. 5.

2. NOVEL LOSSLESS AUTHENTICATION FRAMEWORK

The concept of using a lossless data embedding method for authentication watermarking has been proposed earlier in the literature[7–9,12–14] and is commonly referred as *invertible* or *reversible* authentication *. A general block diagram which is representative of the prior techniques † is seen in Fig. 1. All these methods are based on calculating the authentication information, and inserting this information using a lossless (reversible) data embedding method. The authentication information may be a hash, message authentication code, or digital signature computed over the unwatermarked image. The methods are differentiated by the particular reversible data embedding scheme used. In particular, Fridrich et al. propose substituting least significant bit (LSB) plane(s) of the image by a bit-string containing the authentication information and the compressed form of the original LSBs.[7] In this method, additional capacity is created through the lossless compression of the LSBs, which also allows for reconstruction of the original LSBs, thus the original image. This method is later replaced with the RS-Embedding scheme, which improves the capacity—or equivalently reduces the embedding distortion—in comparison with the earlier method.[5] Honsinger et al.[8] proposed using an additive spread spectrum watermark for data embedding. Reversibility of the scheme is guaranteed through the use of modulo arithmetic. Recently, Tian[9] explored the integer wavelet transform. In his method, called difference expansion, detail coefficients of the transform are modified in an invertible manner. In these methods, the integrity and authenticity of the image is verified by *i)* extracting the embedded authentication information, *ii)* reversing

*In this paper, we limit our scope to *fragile* authentication watermarks, which provide exact, i.e. bit per bit, integrity verification. Earlier lossless authentication watermarks have also been of this type.

†With the possible exception of.[13]

the watermarking procedure, thus reconstructing the original image, *iii)* comparing the reconstructed image with the extracted signature. If the extracted signature matches to the signature that is calculated from the reconstructed image, the image is deemed authentic. Note that the image reconstruction, which often is the most computation intensive process, is required for verification, even when the image is not authentic.

Dittmann et al.[13] proposed an alternative protocol for the LSB compression technique.[7] They replace the signature of the whole image with two signatures that are computed from the most significant bits (MSB) of the image and the compressed version of the LSBs, respectively. Their approach allows for validation before image reconstruction. In addition, encryption of the compressed LSBs facilitates reconstruction of the original by authorized parties—who hold the secret key—only, without affecting the public key validation process. Nevertheless, the protocol is dependent on the particular reversible data embedding mechanism and is not compatible, for instance, with Honsinger's method. Moreover, a second signature consumes additional capacity and thus increases embedding distortion. Note also that none of the lossless authentication methods in the literature offers tamper localization capability, which is one of the major advantages of authentication watermarks over conventional digital signatures.

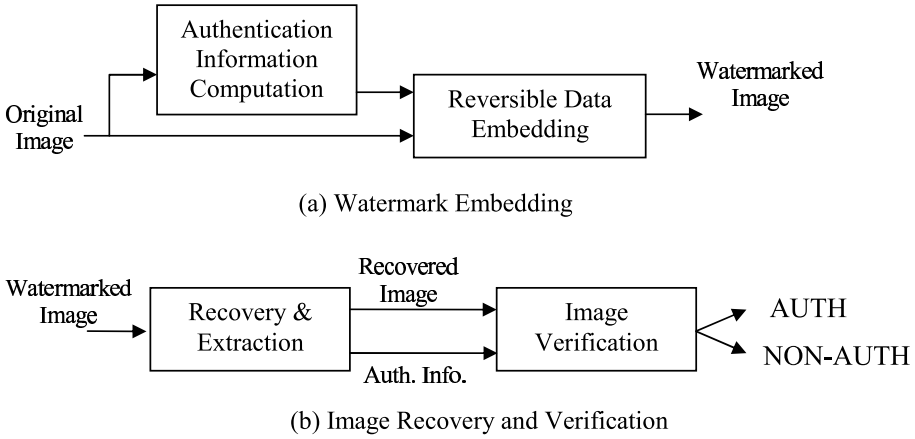


Figure 1. Prior lossless authentication watermarking methods: (a) Signature calculation and embedding. (b) Original image reconstruction, signature extraction and image verification.

The lossless authentication framework proposed herein provides an alternative construction for authentication watermarking with lossless (reversible) data embedding. As seen in Fig. 2, our approach differs from the prior techniques in the order of authentication and lossless data embedding phases. The prior schemes calculate the authentication information first, and then embed this information with a reversible data embedding method. As a better alternative, we propose a reversible pre-embedding step to prepare the image for the authentication watermark, following which the authentication information is calculated and inserted (Fig. 2). In this framework, the pre-embedding step creates the necessary capacity for the authentication information, which is computed over the pre-embedded image. At the receiver, the authentication watermark validates the integrity of the pre-embedded (watermarked) image, which is slightly different than the original due to the reversible embedding distortion. If the verification step is successful and the user wishes to recover the original image, pre-embedding procedure is reversed and the original image is reconstructed. It is worth emphasizing that the original image is not authenticated directly; instead, the integrity is inferred from the authentication of the pre-embedded image and the uniqueness of the pre-embedding mechanism. This idea is analogous to authenticating the compressed version of a file, rather than the file itself and does not introduce any weakness in the authentication scheme.

Our framework provides a number of advantages over the existing methods and protocols. Its modular design enables the use of a wide variety of lossless data embedding and authentication watermarking algorithms. For instance, a spread spectrum based modulo arithmetic method or a wavelet transform based data embedding method may be utilized. Nevertheless, this framework imposes some restrictions on the design of the modules.

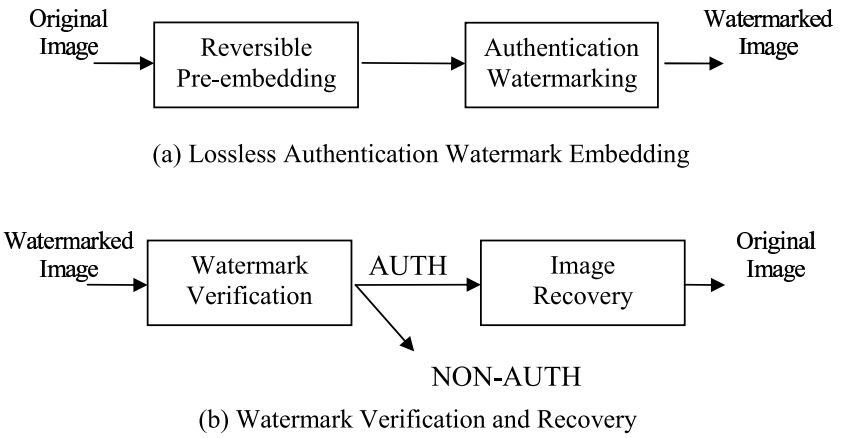


Figure 2: Lossless authentication watermarking: (a) Embedding (b) Detection & Recovery.

In particular, the authentication watermark— and the insertion of authentication information thereof—should not conflict with the reversibility of the pre-embedding scheme. In general, a potential conflict may be resolved by requiring each method to operate on different pixel subsets. An example construction is provided in our implementation in Sec. 3.

Another advantage of the proposed framework is the reduction in the average computational requirements. By validating the authenticity of the image first, we avoid the computationally expensive image reconstruction step when *i)* the watermarked image is of sufficient quality and the original image reconstruction is not necessary, *ii)* the image under inspection is not authentic, i.e. it has been tampered after watermarking, or *iii)* the image has not been watermarked at all. If the original image is reconstructed, its authenticity is inferred from that of the pre-embedded image without any additional computation or sacrifice of security. Earlier methods either require the reconstruction of the original[7, 9] or require verification of multiple watermarks (digital signatures).[13] In the latter case, the use of a single authentication watermark also saves valuable reversible data embedding capacity. At the embedder, performing the reversible data embedding step before authentication watermarking offers similar computational advantages. For instance, an application may require inserting current time-stamps as a part of the authentication watermark. In this case, pre-embedding is performed only once and different authentication information is inserted at each time a client requests the image. Avoiding multiple pre-embedding processes reduces the load on the server.

The framework also facilitates public key verification of the watermarked image while restricting the access to the original image. Since the authentication process is independent of image reconstruction, the later step can be dependent on a private key without disturbing the public authentication process. All parties use the first step to validate the authenticity of the image, but only authorized parties, who hold the private key, can access to the original image. In this respect, the framework is similar to the proposal by Dittmann et al..[13]

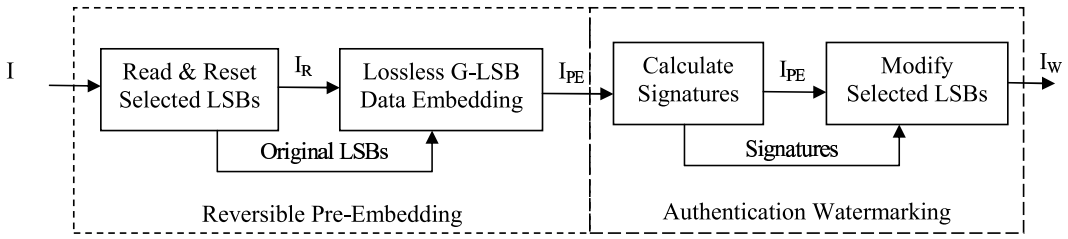
3. LOCALIZED LOSSLESS AUTHENTICATION WATERMARK (LAW)

In the preceding section, we described a novel lossless authentication framework. We now present an example construction, which we call *Localized Lossless Authentication Watermark* or *Localized LAW*. In particular, we utilize the hierarchical authentication watermark[4] in conjunction with the lossless generalized-LSB data embedding algorithm.[10, 11] Brief descriptions of these algorithms are given below and readers are referred to earlier publications[4, 10, 11] for particular details.

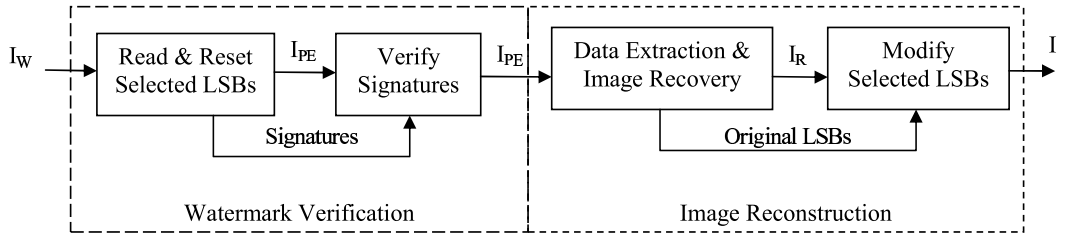
The hierarchical authentication watermark is a secure extension of the Wong's scheme[3] and provides excellent tamper localization accuracy with the ability to employ public key authentication. This method, like Wong's scheme, inserts authentication information to the LSBs of selected pixels. Hierarchical authentication watermark is inserted by *i)* setting LSBs of selected pixels to zero; *ii)* dividing the image into blocks in a multi-level hierarchy; *iii)* computing the digital signature—or the message authentication code— of each block; *iv)* rearranging these

signatures according to their position in the hierarchy; and *v)* replacing the LSBs of selected pixels by the signatures. Tamper localization is provided by the block-based nature of the algorithm as in the original Wong's scheme.

The Lossless Generalized-LSB data embedding algorithm[10, 11] is a reversible data embedding algorithm which is similar to LSB-plane embedding.[7] In this method, the lowest levels (bit-planes) of the pixel values are replaced with their compressed description and additional watermark payload. The particular compression mechanism exploits the correlation between different bit-planes and neighboring pixels by utilizing the higher pixel levels as side information. Very good compression efficiency is achieved through prediction, context modeling and adaptive arithmetic coding.



(a) Watermark Embedding



(b) Watermark Verification and Recovery

Figure 3: Localized Lossless Authentication Watermark: (a) Embedding (b) Detection & Recovery.

Overview of the embedding and verification procedures for the Localized LAW algorithm are seen in Fig. 3. Given an image I, the reversible pre-embedding step first reads the original values of LSBs of pixels at selected positions. These LSBs are later set to zero. Modified image is passed to the Lossless G-LSB algorithm, which embeds original LSB values in a reversible manner, thus creates additional capacity. Note that the lossless data embedding avoids modifying pixels at selected locations. These locations are determined a priori and they are shared with the authentication watermark. Let us denote the image after pre-embedding by I_{PE}. In the second phase, I_{PE} is divided into blocks in a multi-level hierarchy and block signatures are computed. The block signatures are inserted into the image by replacing the LSBs of pixels that have been selected and reset earlier. The watermarked image is denoted by I_W. Note that I_W differs from I_{PE} at only those LSB positions.

At the receiver end, first the watermark verification step tries to authenticate the image. The LSBs at pre-determined positions are read and set to zero. If the image under inspection is a watermarked image, the LSBs represent the block signatures and the modified image is exactly equal to the pre-embedded image I_{PE}. In this case, the signatures validate the authenticity and integrity of I_W and I_{PE} is passed to the image reconstruction phase. If the image under inspection is tampered, signature verification step fails and a tamper localization map is generated by the hierarchical authentication procedure. If the image is authentic and image reconstruction is desired, lossless data embedding procedure is reversed and original values of the selected LSBs are extracted. After restoring those LSBs, the original image is reconstructed exactly, i.e. without any distortion.

In our implementation, a 320 bit DSA signature, which is a public key digital signature scheme, is used at the top of the hierarchy. This allows for public validation of the watermarked images. At the lower levels of the hierarchy, 64 bit message authentication codes, MD5 HMAC in particular, are used. The private key

nature of the MACs reserves the tamper localization capability for authorized parties. A primary block size of 64×64 pixels is selected, as the parameter for tamper localization accuracy. A key based extension of the Lossless G-LSB algorithm is used to restrict public access to the original image. In particular Advanced Encryption Standard (AES) (128 bit key) is used to encrypt the compressed description during Lossless G-LSB data embedding.

As a result, our implementation of Localized LAW supports the following functionality:

- A secure, well-known public key authentication of the watermarked image.
- Exact (lossless) recovery and authentication of the original unwatermarked image.
- Low embedding distortion, through the use of Lossless G-LSB method.
- Reduced computational requirements, when the image is not authentic/watermarked.
- Private key tamper localization ability.
- Public validation (authentication) with private recovery using a single digital signature.

The functionality of our implementation reflects a number of design decisions. The algorithm, however, supports greater flexibility and may be adjusted for a different set of design criteria. If desired, public key signatures may replace the MACs at the lower levels of the hierarchy and allow for public-key tamper localization. The encryption step may be skipped in Lossless G-LSB algorithm and public reconstruction of the original may be allowed. It is possible to reduce the primary block size, hence increase the tamper localization accuracy of the method. Nevertheless, dividing the image into smaller blocks increases the number of blocks, thus the number of MACs. In turn, more reversible data embedding capacity is required to convey this information. Although Lossless G-LSB embedding can accommodate the increased payload, its embedding distortion increases accordingly. In short, there is a trade-off between the tamper localization ability and the embedding distortion of Localized LAW and the operating point has to be selected based on the particular application.

4. EXPERIMENTAL RESULTS

We test the effectiveness of Localized LAW algorithm on the *Aerial* image (gray-scale, 1024×1024 pixels). *Aerial* image is watermarked using Localized LAW with default parameters described in the preceding section. The watermarked image (see Fig. 4) has a peak signal to noise ratio (PSNR) of 50.85 dB and carries 2224 Bytes of authentication information . If the image is not altered after watermarking, the watermark detector verifies its authenticity and recovers the original image exactly without any loss.

In order to test the tamper localization capabilities, the watermarked *Aerial* image was manipulated on a personal computer using commercial image editing software. The manipulated image is shown in Fig. 5. In particular, the cars around the buildings at the center of the image have been removed as a result of manipulation. When the manipulated image is presented to the watermark detector, the manipulation (tampering) is detected and tampered regions are marked at the detector output. This is illustrated in Fig. 6, where the shading of an area represents the level at which a block is authenticated for the hierarchical authentication watermark. Darkest areas correspond to blocks that have not been authenticated at any level and cover all of the tampered regions. The region of interest bearing the manipulations has been enlarged in Fig. 7 to clearly demonstrate the results.

In addition to *Aerial* image, we use six standard test images (gray-scale, 512×512) to test the variation in the embedding distortion of the algorithm. Although this distortion is reversible, it is preferable to create a high quality watermarked image. As expected, performance of the lossless data embedding technique—thus the overall algorithm—depends on the image content. The PSNR of the watermarked images range from 46 dB for the highly textured Mandrill image to 56 dB for the F-16 image with large, smooth areas (see Table. 1).

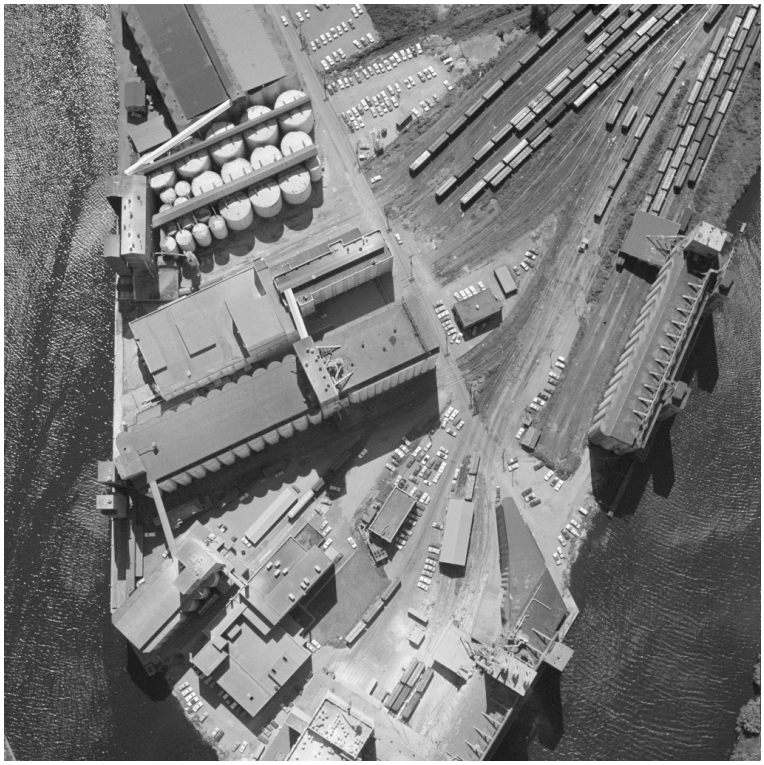


Figure 4: Watermarked "aerial" image. 2224 Byte authentication code is embedded losslessly. (PSNR=50.85dB)

Image	PSNR (dB)	ΔFilesize (Bytes)
F-16	56.54	2876
Mandrill	46.51	1440
Boat	52.91	2658
Barbara	52.91	2019
GoldHill	50.83	1819
Lena	52.83	1510
Average	52.09	2054

Table 1. Embedding distortion and increase in compressed file size for 512x512 gray-scale images. JPEG-LS[15] with default parameters is used for lossless compression.

Medical and military images are often compressed using a lossless codec , such as JPEG-LS,[15] in order to reduce the storage requirements. The effect of Localized LAW on the compression efficiency of JPEG-LS codec[‡] has been measured by comparing the compressed file size before and after watermarking. As seen in Table. 1, on the average, the file size increases by 2054 Bytes, which is significantly larger than the 584 Byte authentication code inserted during watermarking. The increase in file size can be attributed to the properties of the lossless data embedding scheme. Lossless G-LSB embeds the additional information by compressing and replacing parts of the image content. Substitution of image content with an uncorrelated signal disturbs the image statistics. As the secondary compression, i.e. JPEG-LS, is not optimal and inherently assumes a particular model for the image statistics, its performance is degraded beyond the inserted information. (A hypothetical encoder that achieves the entropy limit would not suffer an additional degradation.) Including the digital signature of the image in the header of the compressed file has the advantage of limiting the increase in the file size. Nevertheless,

[‡]In our simulations, we used lossless mode of the JPEG-LS algorithm with default parameters.

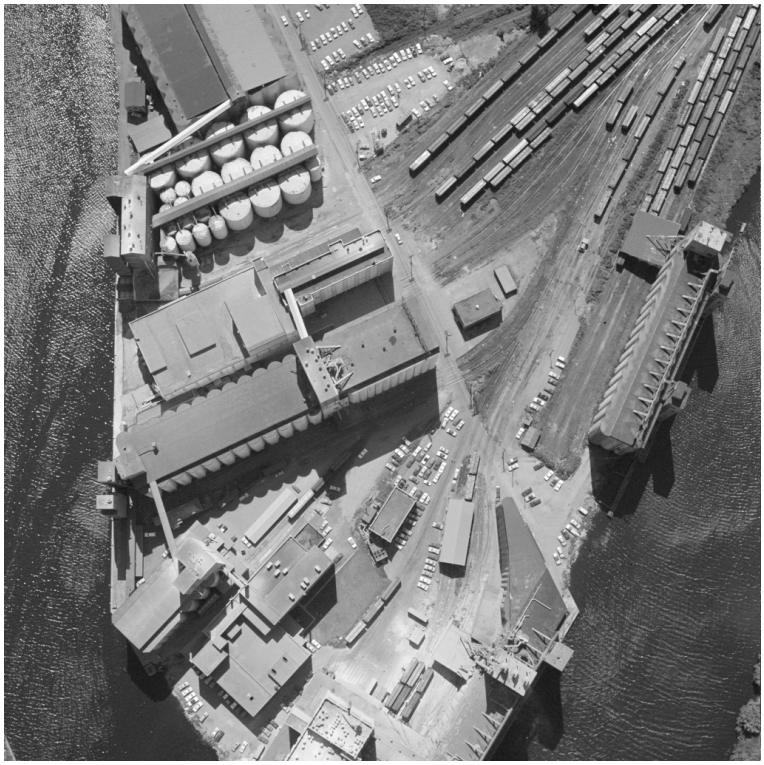


Figure 5: Manipulated image. Cars around the buildings at the center are removed.

this method requires modification of the codec's file syntax and limits the use of other formats and independent conversion utilities. Since the increase in the file size due to authentication watermarking accounts for less than 1% of the total file size, the additional storage cost is often justified by the convenience and the additional functionality provided by the authentication watermark.

5. CONCLUSIONS

A novel lossless (reversible) authentication watermarking framework is proposed. The framework facilitates the use of various lossless data embedding methods and authentication watermarks in a flexible and computationally efficient manner. A particular instantiation of the proposed framework, called Localized lossless authentication watermark (LAW), is implemented to demonstrate the flexibility of the scheme. Localized LAW has the additional capability for tamper localization, which is not found in prior lossless authentication watermarks.

One of the open research problems in this area is the use of content authentication methods, i.e. semi-fragile watermarks, in the context of lossless (reversible) authentication. As the exact recovery of the original image is practical only when the image has not been manipulated, semi-fragile lossless authentication watermarks can be constructed using fragile reversible data embedding techniques in conjunction with the semi-fragile authentication watermarks.

REFERENCES

1. A. Menezes, P. van Oorchot, and S. Vanstone, *Handbook of Applied Cryptography*, CRC Press, Florida, USA, 1997.
2. D. Stinson, *Cryptography: Theory and Practice*, CRC Press, Florida, USA, 1995.
3. P. Wong, "A public key watermark for image verification and authentication," in *Proc. of IEEE ICIP*, pp. 425–429, (Chicago, USA), October 4-7, 1998.

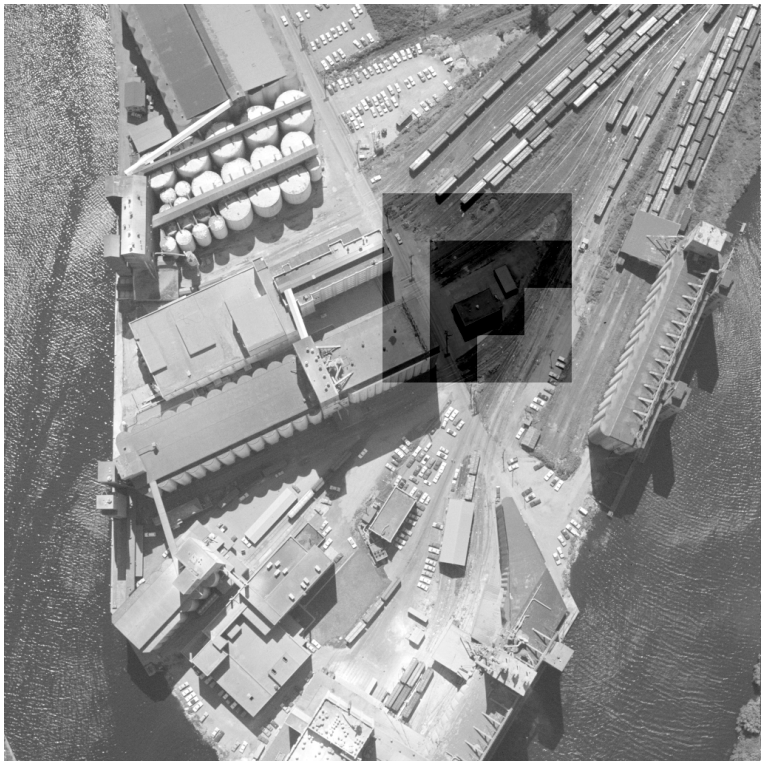


Figure 6. Watermark detection output. Shading indicate the lowest level signature verified. Tampered regions of the image are correctly identified.

Figure 7. Tampered are of the image: Watermarked original (left), tampered forgery (center), detection output (right).

4. M. U. Celik, G. Sharma, E. Saber, and A. M. Tekalp, "Hierarchical watermarking for secure image authentication with localization," *IEEE Trans. on Image Proc.* **11**, June 2002.
5. J. Fridrich, M. Goljan, and A. Baldoza, "New fragile authentication watermark for images," in *Proc. of IEEE ICIP*, (Vancouver, Canada), September 10-13, 2000.
6. G. Friedman, "The trustworthy digital camera: Restoring credibility to the photographic image," *IEEE Trans. on Consumer Electronics* **39**, pp. 905–910, November 1993.
7. J. Fridrich, M. Goljan, and R. Du, "Invertible authentication," *Proc. of SPIE Sec. and Watermarking of Multimedia Cont. III* , pp. 197–208, Jan 2001.
8. C. Honsinger, P. Jones, M. Rabbani, and J. Stoffel, "Lossless recovery of an original image containing embedded data," *US Pat. #6,278,791* , Aug 2001.

9. J. Tian, "Wavelet-based reversible watermarking for authentication," *Proc. of SPIE Sec. and Watermarking of Multimedia Cont. IV* **4675**, pp. 679–690, Jan 2002.
10. M. Celik, G. Sharma, A. M. Tekalp, and E. Saber, "Lossless Generalized-LSB data embedding." submitted to *IEEE Trans. on Image Proc.*, July 2002.
11. M. Celik, G. Sharma, A. M. Tekalp, and E. Saber, "Reversible data hiding," in *Proc. of IEEE ICIP*, pp. II–157–160, (Rochester, NY, USA), Sept. 2002.
12. C. D. Vleeschouwer, J. Delaigle, and B. Macq, "Circular interpretation of histogram for reversible watermarking," in *IEEE Fourth Workshop on Multimedia Signal Proc.*, pp. 345–350, Oct 2001.
13. J. Dittmann, M. Steinebach, and L. Ferri, "Watermarking protocols for authentication and ownership protection based on timestamps and holograms," *Proc. of SPIE Sec. and Watermarking of Multimedia Cont. IV* , pp. 240–251, Jan 2002.
14. J. M. Barton, "Method and apparatus for embedding authentication information within digital data, US Pat. #5,646,997," 1997.
15. Weinberger, Seroussi, and Sapiro, "The LOCO-I lossless image compression algorithm: Principles and standardization into JPEG-LS," *IEEE Trans. on Image Processing* **9**, pp. 1309–1324, Aug. 2000.