JPEG2000-Based Secure Image Authentication

Mathias Schlauweg, Dima Pröfrock and Erika Müller
Institute of Communications Engineering, Faculty of Computer Science and Electrical Engineering
University of Rostock
Rostock 18119, Germany
+49 (0) 381 498 7304

{mathias.schlauweg, dima.proefrock, erika.mueller}@uni-rostock.de

ABSTRACT

We present an efficient JPEG2000-based image authentication scheme, which is robust to JPEG compression and other allowed signal processing operations. Positive wavelet-based watermarking approaches proposed in recent years are enhanced by image adaptive perceptual modeling and error correction coding. Our new method is secure in contrast to most of the schemes proposed so far. Lots of popular features of the JPEG2000 compression framework are supported, such as quality and resolution scalability, lossless image rotation and flipping. All coefficients of the wavelet decomposition are protected using our new extended scalar quantization and hashing scheme. We show that error correction coding yields impressive robustness improvements of the embedded image content dependent signature without raising a security gap. Further, we introduce a technique to remove the embedded information during the decompression process and thus to improve the image quality during, e.g., visualization. The functionality is not only proved by experimental results but also a real prototype camera implementation and web-based verification.

Categories and Subject Descriptors

H.3.1 [**Information Storage and Retrieval**]: Content Analysis and Indexing – *Indexing Methods;* I.4.9 [**Image Processing and Computer Vision**]: Applications*;* K.6.5 [**Management of Computing and Information Systems**]: Security and Protection – *Authentication, Insurance.*

General Terms

Algorithms, Security, Verification.

Keywords

JPEG2000, wavelet domain, authentication, watermarking, ECC.

1. INTRODUCTION

Today's rapid evolution of multimedia technology and the progress of computer networks along with the development of the Internet bring many advantages in the creation and distribution of image content. But beneath the ability of easy copying, transmitting and editing digital images the need for image content protection increases. Digital images can be modified or forged by a wide variety of available manipulation software, and hence it is rather difficult to tell if a picture is the original one, which has been taken by a camera, or if it has been tampered with.

The integrity of a digital image is pre-eminent in fields such as forensics, medical imaging and military or industrial photography. For instance, courts make decisions affecting an individual's liberty based, in part, on images presented as evidence. The burden of proof of authenticity always lies with the person seeking to admit. He must provide other evidence to support this authenticity. Further, military photographs may determine target locations based on their content and interpretation. Thus, it is important to maintain the integrity of all images from capture to final use.

To prevent illegitimate tampering and fraudulent use of modified images authentication techniques were introduced. As known from the classical cryptography, to verify the exact data integrity, a signature may be generated from the source signal by the use of secure hash functions and encryption. A recipient decrypts the signature and matches it with the hash generated from the received signal. If even one bit of the signal has been modified, it will no longer match the signature, so any tampering can be detected. But this so-called fragile property is sometimes not practical when considering distribution of images. For instance, lossy compression has to be performed to reduce the amount of data or signal processing is applied to correct gamma, to de-noise or to resample an image. These manipulations change the pixel values but not the content and hence not the authenticity.

To tolerate certain kinds of signal processing semi-fragile authentication methods for digital images have been developed. The aim is to allow admissible manipulations such as JPEG compression, but to reject malicious manipulations that change the visual content. Commonly used techniques extract features representing the image content and re-embed these features as watermark information into the host image data [1-7]. Some approaches involve image positions of edges, contours or zero-crossings in the spatial domain whose existence is proved during the verification process. Other methods are based on single coefficients or on relationships between pairs of different coefficients in the transform domain (e.g., DCT, DWT or DFT).

The advantage of directly embedding authentication data as a watermark is, that the signature can not get lost during format conversion operations. No additional data has to be submitted besides the watermarked image except some watermarking parameters and the key used for decryption. But, every single watermark bit that has to be embedded into the host data slightly modifies the image and degrades the visual quality. Hence, the right choice of the feature extraction as well as watermark embedding approach is decisive for a practical authentication system.

MM&Sec'06, September 26–27, 2006, Geneva, Switzerland.

The remainder of this paper is organized as follows. In Section 2, we propose our new semi-fragile image authentication framework in the discrete wavelet domain of JPEG2000. We compare our approach to similar schemes proposed so far. Experimental results and analyses are given in section 3. Section 4 concludes the work.

2. THE AUTHENTICATION SCHEME

Efficiency should be a major design criterion for an image authentication algorithm, since the target device for implementation will be, e.g., a digital camera, mobile phone or Pocket PC with integrated camera. But the security of the overall framework has to be explicitly considered as well. Many authentication frameworks lay to much emphasis on robustness, which brings into question security issues for authentication applications. Often, the image content is pretended to be secured by protecting only the correct existence of mean values of extensive pixel areas. An attack, intended to change the image content, can maliciously operate on these local pixel areas as long as the mean values are not changed. For example, a forger is able to insert edges without raising an alarm when the authenticity is verified as long as he maintains the mean values of these extensive pixel areas.

Another crucial security gap exists if a simple logo is embedded for authentication purpose instead of an image content dependent signature. Often the embedded logo or bit sequence is only proved by correlation and large tolerance margins, such as in [6-7]. In this case, especially if the attacker has knowledge about the used watermark detector, the framework is not practical.

A secure alternative to strive after should be the use of cryptographically secure hash functions mapping all important content dependent features to a small amount of bit information, which can be encrypted and embedded. To do so, invariant properties for signature generation as well as bit embedding are needed, because the output of classical hash functions alters dramatically if even one input bit is changed and hence no verification is possible.

In [8], we analyzed authentication systems in the DCT domain of JPEG compression and found out that large tolerance margins are required resulting in dramatic security gaps. The coefficients of the wavelet domain turn out to be much better suited for authentication watermarking purpose. Because DWT is a global transform, watermark embedding in the low frequency coefficients does not result in block artifacts in watermarked images, as shown in [9]. Hence, we generate and embed the image content dependent features in the wavelet domain of JPEG2000.

Our proposed system works very efficiently, since it is directly integrated in the JPEG2000 compression process. The framework is structured modularly so that single components such as the used hash function or the encryption scheme can be replaced without influencing the overall functionality.

2.1 Feature Extraction in the DWT-domain

As opposed to the JPEG compression framework, the quantization in JPEG2000 is bit-plane oriented. The quantization interval is always an exact power of two and hence simple JPEG2000 re-compression does not require the coefficients to be re-quantized. This is also the basis for the well known "encode once; decode many" strategy of the JPEG2000 specification [10]. But even if small allowed operations are applied to the watermarked image in the spatial domain, the upper bit-planes of the quantized DWT-coefficients turn out to be very stable. Thus, for our authentication scheme, we define an extended version of the scalar dead-zone quantization technique used in the JPEG2000 coding framework. This extension is completely new and outperforms existing quantization-based watermarking approaches in the wavelet-domain of JPEG2000, such as the one in [7].

The image $\mathbf{I}$ is DWT-transformed with the bi-orthogonal 9/7-wavelet filter, at first, and approximated by a finite number of bit-planes ($0 \le b \le b_{\max} \le \text{sign}$).

$$\mathbf{I} \overset{\text{DWT}}{\mapsto} \mathbf{W} = \left\{ \mathbf{W}_b,\ 0 \le b \le b_{\max} \le \text{sign} \right\} \tag{1}$$

Afterwards, the transformed image $\mathbf{W}$ has to be pre-quantized using the step size δ. Therefore, every wavelet coefficient $\boldsymbol{w}$ of the transformed image is quantized to an invariant value, which is not changed during the signature embedding process [5]. Except for the dead-zone, the magnitude of every coefficient is set to the centre of the according δ-sized hash interval, in Figure 1 marked with the symbols **II**. Coefficients in the ranges $[-\delta \le \boldsymbol{w} \le -\delta/2]$ or $[\delta/2 \le \boldsymbol{w} \le \delta]$ are set to $-\delta/2$ or $\delta/2$, respectively. The other coefficients of the dead-zone are left unchanged resulting in higher image quality. We justify this approach by the fact that DWT coefficients are most often very small due to the very good energy concentration capability of the wavelet transform. Finally, all coefficients could be hashed to one overall hash value using a cryptographic hash function, such as MD5, SHA-1 or SHA-256.

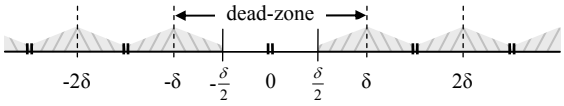

Figure 1. Extended scalar quantization with dead-zone

2.1.1 Variable resolution

To support various resolutions, we propose to divide the DWT-transformed image $\mathbf{W}$ into Z regions $\mathbf{W}^1$, $\mathbf{W}^2$, ..., $\mathbf{W}^Z$, as shown in Figure 2. Now, from every region $\mathbf{W}^i$ or, in other words, from every DWT-level a hash value $\mathfrak{H}(\mathbf{W}^i)$ is calculated.

Figure 2. Different hash regions to support resolution

$$h = \left\{ \mathfrak{H}(\mathbf{W}^1), \mathfrak{H}(\mathbf{W}^2), ..., \mathfrak{H}(\mathbf{W}^Z) \right\} \tag{2}$$

If the coefficients of $\mathbf{W}^1$ are set to zero, the hash values of all other DWT-levels $\left\{ \mathfrak{H}(\mathbf{W}^2), ..., \mathfrak{H}(\mathbf{W}^Z) \right\}$ do not change.

This multi-hash approach makes the watermarked image robust to resolution changes that are powers of two. The "encode once; decode many" strategy of the JPEG2000 specification directly benefits from this approach.

2.1.2 Signature generation

For a secure authentication scheme it is vital that the feature hash value is signed directly inside the camera using a private key that cannot be read out. Only if this key, hard-wired during camera production, is used to encrypt the authentication plaintext during image capture the public key verification can succeed. In other words, a forger, not knowing the appropriate key, is not able to insert another feature hash into the watermark signature.

The bit length of the above described concatenated hash *h* must be less than the maximum signature key length of the encryption procedure. For example, in our tests, we used $Z = 4$ levels for the DWT decomposition, SHA-1-hashing (160 bits per level) and 728 bit RSA-encryption. Besides the bits used for the hash value *h*, we also include position, date and time of the image capture process in the data to be encrypted. This should deter a forger from photographing fake sceneries at a different position or different time using the same camera.

2.1.3 Security aspects of the multi-hash approach

In [4], Fridrich et al. analyzed the security of fragile authentication frameworks. They pointed out that schemes that generate multiple block-wise hashes to localize tampered image areas are insecure. A forger could substitute hash-dependent blocks by exploiting databases of images all protected with the same key. The larger the database, e.g., due to block-wise hashing, the more easy it is for the attacker to find substitutions. Since in our scheme only few, e.g., $Z = 4$ different hash values are concatenated and encrypted per image, the security is unaffected. But due to this security request our scheme is unable to spatially localize malicious modifications. It works at a global level, and it can be verified only to which resolution or quality level the image is authentic.

2.2 Watermark Embedding

To embed the encrypted signature as a watermark we use the well-known quantization index modulation technique called dither modulation. Roughly speaking, the used host signal samples are mapped bit-wise to the elements of a set of two different quantizers, as can be seen in Figure 3. Except for the dead-zone, in every hash interval there are two quantization points, marked with ✖'s and ○'s. If a binary information bit "0" has to be embedded, the coefficient magnitude is set to the point marked with ✖, otherwise, the point marked with ○ is chosen. For more detailed descriptions, we refer to [5].

A special case occurs if the used coefficient is in the dead-zone and a binary bit "1" has to be embedded. In this case we suggest applying the following extended equations to achieve lower embedding induced distortions:

$$\begin{aligned} &\text{if} \quad -\delta \le w \le -\delta/4 \quad \text{then} \quad w = -\delta/4 \\ &\text{if} \quad \delta/4 \le w \le \delta \quad \text{then} \quad w = \delta/4 \\ &\qquad\qquad \text{else don't change } w \end{aligned} \tag{3}$$

Since the quantization in JPEG2000 is bit-plane oriented, the embedding process only affects one single bit-plane, which we identify by $\mathbf{W}_s$. The bit-planes below $\mathbf{W}_s$ remain unchanged, since they were set to zero due to the pre-quantization process.

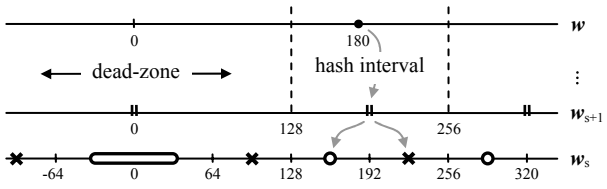

Figure 3. Watermark embedding process using an extended dither modulation (QIM)

We suggest applying error correction coding (ECC) to the signature before embedding takes place. In our tests, we used BCH(511,367,16) coding and embedded the resulting 1022 bits in the upper LL-subband of the wavelet decomposition. ECC with soft decision decoding like Turbo codes or LDPC codes, which have been shown to be very efficient for digital watermarking applications, could be used as well. Finally, the watermarked image may either be transferred as JPEG2000-compressed file or be transformed back to the spatial domain.

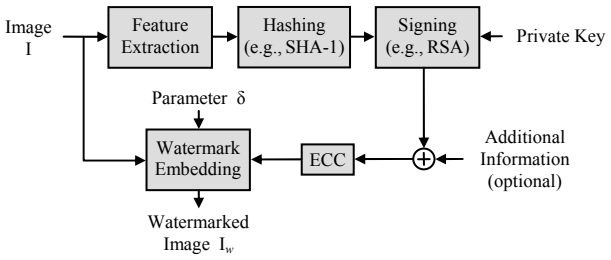

Figure 4. Watermark generation and embedding

At the verification site, the watermark bits are extracted by nearest neighbor quantization to one of the two quantizer subsets. Afterwards, error correction coding is used to get the submitted signature bits. If the watermarked image has been distorted by, e.g., lossy file format conversions, error correction coding can help to reconstruct the distorted signal samples. Without ECC, the allowed distortion to the watermarked coefficients would be only δ/4 but using ECC errors with an absolute value δ/2 can be reconstructed, as in Figure 6. The quantization cell may be thought of as a shifted overlapped version of the original cell. Hence, the range of accepted channel distortion is raised without security loss, as we demonstrated in [5].

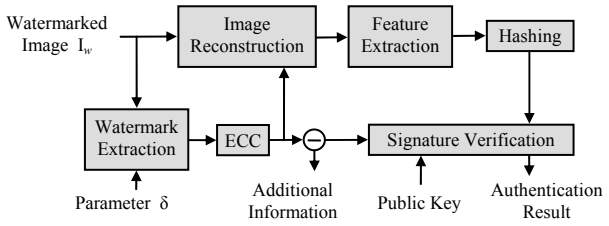

Figure 5. Watermark extraction and verification

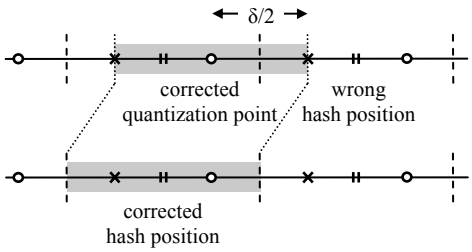

Figure 6. Hash interval reconstruction using ECC

2.3 Image Adaptive Perceptual Modeling

In [9], a watermarking method was proposed that embeds information in the upper DWT-subband similar to our approach but not for authentication purpose. As opposed to our scheme, watermark embedding is done by adding information to the wavelet coefficients, instead of quantization. The authors presented a way to reduce the embedding induced visual distortions by applying a block-based texture classification. Wavelet blocks with strong texture are separated from blocks with weak texture by simply searching for blocks with larger coefficients. In this way, information bits can be embedded with different strength due to the fact that the human visual system (HVS) is less sensitive to changes in regions with edges or transitions.

To adapt this texture masking approach to our authentication scheme, feature extraction as well as watermark embedding has to be modified as follows.

2.3.1 Block organization

After the image has been DWT-transformed, each pixel block in the spatial domain corresponds to several blocks in the DWT domain. For example, if we use $Z = 4$ for the decomposition, one coefficient in the LL^4-subband spatially occupies 16 x 16 pixel.

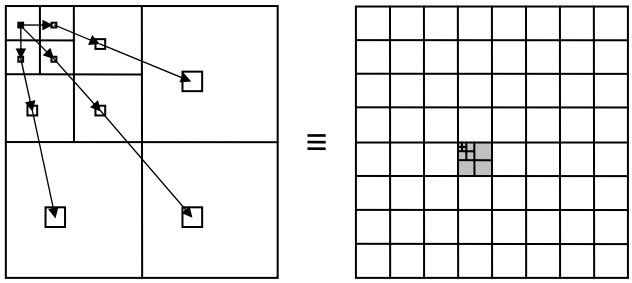

Figure 7. Wavelet block tree organization

2.3.2 Texture classification and pre-distortion

Except for the LL^4-subband, wavelet coefficients have large amplitude where the image has transitions and strong texture. This means blocks composing of coefficients with large amplitude in the corresponding subbands LH^i, HL^i, and HH^i are better suited for embedding than blocks with small coefficients. To separate strong blocks from weak blocks from several tests on natural images we found out that the following approach is suited best:

1. Look for coefficients composing the same block with absolute values larger than a given threshold T.
   - $T_1 = 9$ for the coefficients of LH^4, HL^4
   - $T_2 = 18$ for the coefficients of HH^4
2. Use a *Closing*-Operation for all three resulting 32x32 binary masks to eliminate small gaps
3. If at least two of three threshold decisions at the same spatial position are positive, the corresponding block is strong
4. Use an *Erosion*-Operation for the final bit mask and a pre-distortion if the absolute coefficient values are close to the threshold to lower recognition errors during verification.

2.3.3 Adaptive watermark embedding

The texture block classification yields a LL^4-subband-sized mask which can be used during the signature embedding process. For example, in strong textured blocks bits can be embedded with larger step size δ_1. In other blocks δ_2 is used, where $\delta_1 > \delta_2$.

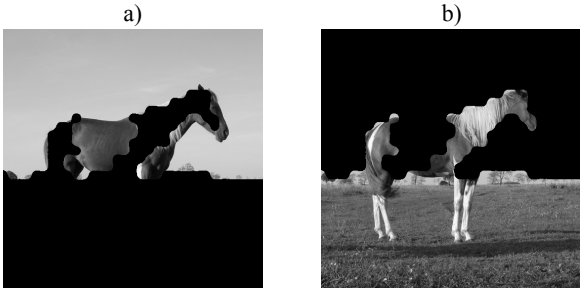

Figure 8. Weak textured region a), strong textured region b)

As can be seen in Figure 9 b), the visual embedding induced distortions are lower in homogenous areas using this new content adaptive approach. We tested the occurrence of recognition errors in the bit mask due to image distortions in the spatial domain. For example, Gaussian lowpass filtering, manipulations of contrast or brightness and even strong JPEG compression yield acceptable error rates, lower than 0.6 percent.

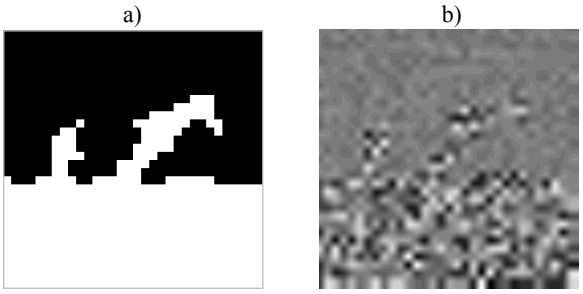

Figure 9. Binary LL^4-mask a), difference of adaptively watermarked image to original image b)

2.3.4 Security aspects of the perceptual adaptation

The DWT-coefficients in our implementation range from -2048 to 2048. We recommend the use of $\delta_1 = 3 \cdot \delta_2$ not larger than 40 for perceptually acceptable image quality using natural images.

Since from every DWT-subband after the quantization process a cryptographic hash value is calculated, it can be recognized if only one single coefficient has left its δ-sized quantization interval due to admissible or malicious modifications. An attacker has to consider this while searching for two images or fabricating two images that have the same features.

For example, as in Figure 10 b), adding a stubbly beard to the face of the test image "Lena" raises the alarm for all DWT-levels. But if a slight scar (deviation of 20 gray levels) is painted on her cheek (Figure 10 c)), the alarm is raised only for the detail levels 1 and 2. DWT-levels 3 and 4 remain authentic during verification.

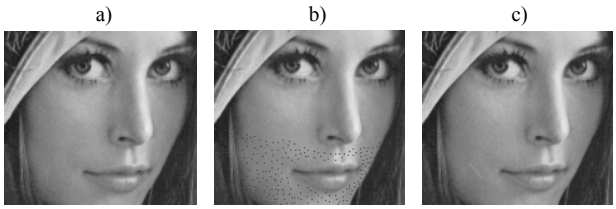

Figure 10. Part of watermarked image "Lena" a), stubbly beard added (not authentic) b), and a slight, almost invisible scar painted on the cheek (partially authentic) c)

2.4 Watermark Removal

As in Figure 3, a signature bit is embedded in the lower or the upper half of the hash interval. Since the distribution of the coefficients of a 2d-DWT-transformed image in each hash interval can be approximated by a uniform distribution, this position is not optimal. To reduce this noise, which is higher than simple quantization noise, the verification algorithm could move the quantized coefficients back to the centre of the hash intervals after the signature bits are extracted.

A common JPEG2000 decoder does not need to extract the embedded watermark information to visualize the compressed image. But, if our technique of watermark removal is applied, the embedding induced distortions can be lowered to approximately 90 percent at the receiver site (see Figure 12).

3. EXPERIMENTAL RESULTS

In this section, we provide some results. For the sake of visibility, in our figures, we only show the curves for the well-known test images "Clown" and "Goldhill". Further, we tested our algorithms with similar results for a large set of test images from the uncompressed colour image database UCID [11].

To implement our JPEG2000-based authentication watermarking algorithm, we chose the well-known "Kakadu" reference software [10]. This implementation enabled us, to also consider the entropy coding stage with bit truncation EBCOT. In addition to the demonstrated curves this implementation was ported to a full Camera-Pocket PC application and web-based verification.

3.1 Perceptual Image Fidelity

Most of early watermarking techniques have focused on embedding the watermark information applying a global power constraint such as the Peak-Signal-to-Noise-Ratio (PSRN) to satisfy fidelity constraints. But, the PSNR value is unparticular reflecting human's visual system, because local image properties such as edges or textures are not considered. Since we use image content adaptive quantization, the PSNR is not optimal here. But to give an impression of the distortions induced by our technique, in Figure 11, the rate-PSNR curves of the watermarked 512 x 512 pixel image "Clown" are shown.

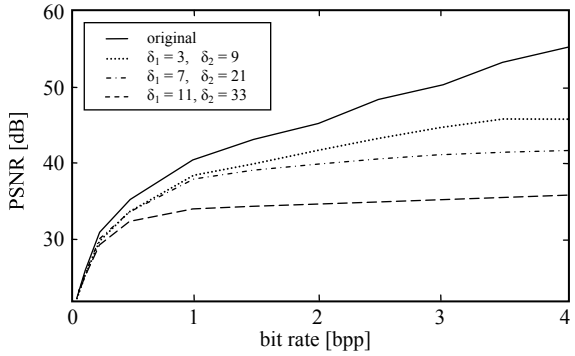

Figure 11. Rate-PSNR curves of watermarked image "Clown"

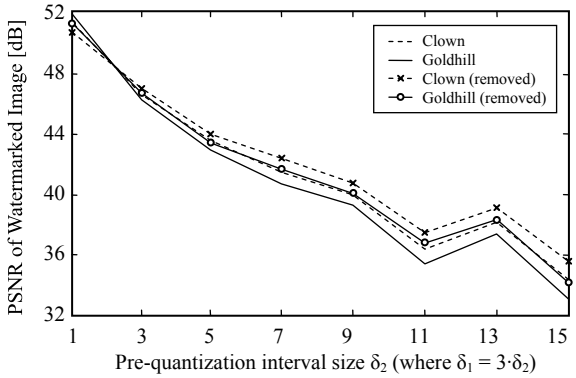

Figure 12. PSNR curves of watermarked images with and without watermark removal at the verification site

3.2 Robustness to Allowed Modifications

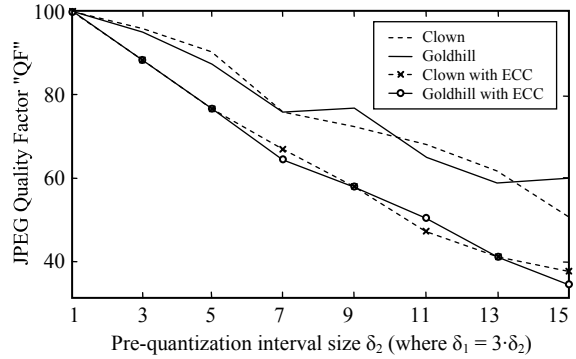

Figure 13. Robustness against JPEG re-compression with and without error correction coding (ECC)

In Figure 13, we demonstrate the robustness against JPEG compression, whereas in one case error correction is used at the veri-

fication site and in the second case it is not used. For the latter, of course, less information bits have to be embedded yielding slightly lower visual distortions. For detailed considerations on the efficiency of error correction coding for watermark embedding, we refer to our work in [5]. Here, in the case of not using ECC, we filled the signature with random bits to be able to compare both methods. What can be seen clearly is that ECC makes the authentication work more constant and reliable at the same amount of embedding induced distortions. Small disturbances in the spatial domain, such as noise or compression, have less effect on the verification.

3.3 Properties

In this subsection, we summarize the properties of our proposed semi-fragile authentication framework:

- The authentication signature embedding is imperceptible
- If an image is unmarked or if the embedded signature is lost, the verification always fails
- If one applies another key for signature generation than the one that matches with the public key used for verification, the verification always fails
- A cropped watermarked image cannot be verified correctly
- To fulfill today's security requests the authentication scheme is unable to localize manipulations. It works at a global level
- Image resolution changes that are powers of two are allowed
- Lossy image re-compression, such as JPEG, JPEG2000, is allowed (see Figure 13)

4. CONCLUSION

In this paper, we presented a very efficient JPEG2000-based image authentication watermarking scheme, which is robust to JPEG as well as JPEG2000 re-compression and other allowed signal processing operations. Our new method supports lots of popular features of the JPEG2000 compression framework such as quality and resolution scalability, lossless image rotation and flipping. The scheme is secure in contrast to most of the authentication schemes proposed so far in the wavelet domain. All coefficients of the wavelet decomposition are quantized and hashed using secure cryptographic hash functions. We showed that error correction coding yields impressive robustness improvements of the embedded image content dependent signature without raising a security gap. Further, we introduced a new extended version of the scalar quantization in contrast to the one commonly used in JPEG2000, which lowers the overall visual distortions during the embedding process. Finally, we presented a technique to remove the embedded watermark information during the decompression process and thus to improve the image quality during, e.g., visualization.

5. REFERENCES

[1] Ekici, Ö., Sankur, B., Coşkum B., Nazi U., and Akcay, M. Comparative evaluation of semifragile watermarking algorithms. *Journal of Electronic Imaging, 13*, Jan. 2004, 209-216.

[2] Zhu, B. B., Swanson, M. D., and Tewfik, A. H. When seeing isn't believing. *IEEE Signal Processing Magazine, 21*, 2004, 40-49.

[3] Rey, C. and Dugelay, J.-L. A Survey of Watermarking Algorithms for Image Authentication. *EURASIP Journal of Applied Signal Processing, 6*, March 2002, 613-621.

[4] Fridrich, J. Security of Fragile Authentication Watermarks with Localization. In *Proc. of SPIE, 4675*, Jan. 2002, 691-700.

[5] Schlauweg, M., Pröfrock, D., Palfner, T., and Müller, E. Quantization-based semi-fragile public-key watermarking for secure image authentication. In *Proc. of SPIE, 5915*, July 2005, 41-51.

[6] Lin, C. Y. and Chang, S.-F. Semi-fragile Watermarking for Authenticating JPEG Visual Content. In *Proc. of SPIE, 3971*, Jan. 2000, 140-151.

[7] Meerwald, P. Quantization Watermarking in the JPEG2000 Coding Pipeline. In *Proc. of Int. Conference on Communication and Multimedia Security*, May. 2001, 69-79.

[8] Schlauweg, M., Palfner, T., Pröfrock, D., and Müller, E. The Achilles' Heel of JPEG-based Image Authentication. In *Proc. of IASTED Int. Conference on Communication, Network and Information Security, 499*, Nov. 2005, 1-6.

[9] Huang, D., Liu, J., Huang, J., and Liu, H. A DWT-based Image Watermarking Algorithm. In *Proc. of IEEE Int. Conference on Multimedia and Expo*, Aug. 2001, 313-316.

[10] Taubman, D. S. and Marcellin, M. W. JPEG2000: Image Compression Fundamentals, Standards and Practice, Kluver Academic Publishers, 2002.

[11] Uncompressed Colour Image Dataset [Online] - Available: http://vision.doc.ntu.ac.uk