A Secure Semi-fragile Watermarking Algorithm for Image Authentication in the Wavelet Domain of JPEG2000

T. Palfner, M. Schlauweg and E. Müller,
Institute of Communications Engineering, University of Rostock, 18119 Rostock
Correspondence Email: torsten.palfner@uni-rostock.de

ABSTRACT

In this paper, we present a new authentication method in the wavelet domain of JPEG2000, which is robust to re-compression, resolution changes and noise. We show that, because of its bit plane oriented processing, JPEG2000 is more suitable for the integration of an authentication watermarking algorithm than JPEG. The disadvantages of JPEG-based authentication schemes are discussed. It is shown that our authentication scheme is secure in contrast to most of the authentication schemes proposed so far in the wavelet domain. The security of our authentication algorithm depends only upon hash algorithm and encryption. Since these algorithms can be updated easily in our authentication scheme, the proposed authentication algorithm will also be secure in the future.

Keywords: Authentication, Watermarking, DWT, JPEG-2000

1. INTRODUCTION

Nowadays it is rather difficult to tell if a picture has been taken by a camera or if it is a fake. Since it is so easy to tamper pictures with the help of a computer, pictures cannot be considered trustworthy anymore. This credibility crisis is getting bigger and bigger in the future, because the production of analog cameras will sooner or later stop. The prices of digital cameras are coming down. Graphic software becomes more sophisticated. Less and less skill is required by the user to tamper an image without leaving a visible trace.

To avert this credibility crisis, efficient authentication algorithms are needed.

1.1 Requirements

A modern image authentication algorithm should have the following properties:

- **Integrity**: The algorithm should be able to detect malicious modifications of the image data.
- **Embedding**: The embedding of the authentication data into the image allows file onversions.
- **Robustness**: The embedded data should be robust to non-malicious alterations of the image.
- **Visibility**: The embedding induced image modification should not be visible to a human being.
- **Image dependence**: The authentication data should be image dependent to prevent tampering.
- **Blindness**: The integrity verification algorithm should not require the original image file.
- **Verification**: Public verification must be allowed. No third party should be needed.
- **Security & Updatability**: The integrity of the image data should not have an expiration date. Since the security of an encryption scheme cannot be guaranteed forever, it must be possible to update the encryption scheme without degrading the quality of the image in the future.

The secure authentication algorithm has to be implemented into a chip of a *secure camera* introduced by Friedman [1]. Such a secure camera has to fulfill the following additional requirements:

- **Reproduction**: The camera should record the position (longitude, latitude, altitude), direction, date and time of a shot (see Fig. 1). This information has to be added to the image authentication data and encrypted. This would deter the forger from photographing fake sceneries.
- **Access**: The attacker should be denied access to the raw image data or the embedding algorithm.

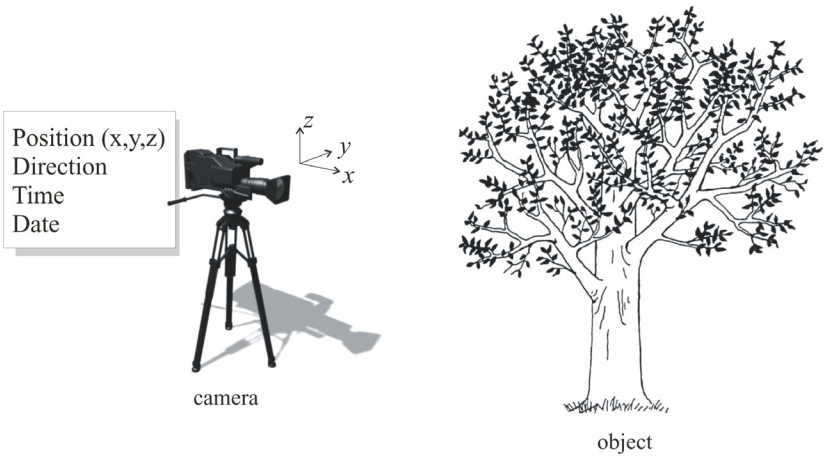

Figure 1: Additional information required for authenticity check of a picture.

1.2 Previous Work

In this section, we review different wavelet-based approaches to verify the authenticity of images. For further information on other image authentication algorithms, we would like to refer the reader to the surveys [2]-[8].

We would like to point out that all these authentication schemes discussed in this section do not fulfill the requirements of Sec. 1.1. Therefore we propose a new approach in Sec. 2.

1.2.1 External authentication data

The location of authentication data could be extern or intern. If the authentication data is stored in a separate file, two files have to be managed.

For example, strict authentication algorithms are based on conventional *cryptographic hash functions* (e.g. MD2, MD5, SHA-1, SHA-256, RIP-MED-160). These hash functions are sensitive to single bit changes. If even one bit of the input signal is modified, the output of a classical hash function alters dramatically and hence no verification is possible. Therefore, they are only suited for strict authentication. The hash value is usually stored externaly in a separate file.

To allow signal processing operations, which preserve the content of an image, non-strict authentication algorithms are required. One possibility is that, before the hash is calculated, features of the image are extracted. These features must represent the image content and be invariant to global content-preserving signal processing operations. Another often used solution is the use of *robust hash functions* (see e.g. [9]).

Since it is quite easy to lose this external authentication data, it is a better solution to store it inside the image file for easy storage and maintenance. This can be done by adding authentication data as metadata to the image file or by embedding the authentication data as a watermark inside the image.

1.2.2 Embedding authentication information as metadata in JPEG2000 data stream

The advantage to store the authentication data intern as metadata is, that the image quality is not degraded. The drawback of this approach used in JPEG2000 Part 8 (JPSEC) and [10]-[15] is, however, that the authentication data is usually lost after the image is converted into different file formats. Since it is quite common to convert the images into different file formats, it is better to insert the authentication data as a watermark directly into the image.

1.2.3 Embedding authentication information in the wavelet domain

As opposed to the other approaches described before, the authentication data becomes an integral part of the image. The authentication data does not get lost during format conversion operations. The image is only slightly modified. Therefore, the image cannot be reconstructed perfectly. The

small degradation of the image quality due to the embedding process should not be visible to the human eye. Hence, we embed the data in the wavelet domain and so use one special effect of the human visual system (HVS) that the human eye is less sensitive to changes of higher image frequencies. As we will see in the following section, the discrete wavelet transform (DWT) works as a kind of frequency decomposition of an image.

2. PROPOSPED METHOD

The problem of all robust hash functions proposed so far is that the security of these functions has never been proven. They can be attacked, because the attacker can take advantage of the robustness of the feature map. The only functions which can be considered as secure at the moment are the cryptographic hash functions.

Therefore, we propose an image authentication scheme which applies a cryptographic hash function to a good approximation of the original image in this paper. Our authentication scheme is non-strict or, in other words, semi-fragile. The authentication embedding process can take place simultaneously when the image is JPEG2000 compressed. No additional data is embedded as metadata in the JPEG2000 data stream.

2.1 Invariant DWT-Properties

After the DWT transform of image **I** the transformed image **W** is approximated by bit planes:

$$\mathbf{I} \overset{\text{DWT}}{\mapsto} \mathbf{W}, \qquad \text{where} \quad \mathbf{W} = \{\mathbf{W}_b, -\infty \le b \le b_{\max} \le \text{sign}\}.$$

Each coefficient of the 2d-DWT-transformed image is represented by a binary number (see Fig. 2). $\mathbf{W}_{\text{sign}}$ is the binary matrix containing the signs of these binary numbers. If the filters of a 5/3-wavelet (lossless 2d-DWT transform) are used, the coefficients of the transformed image can be represented by a finite number of bit planes ($0 \le b \le b_{\max} \le \text{sign}$).

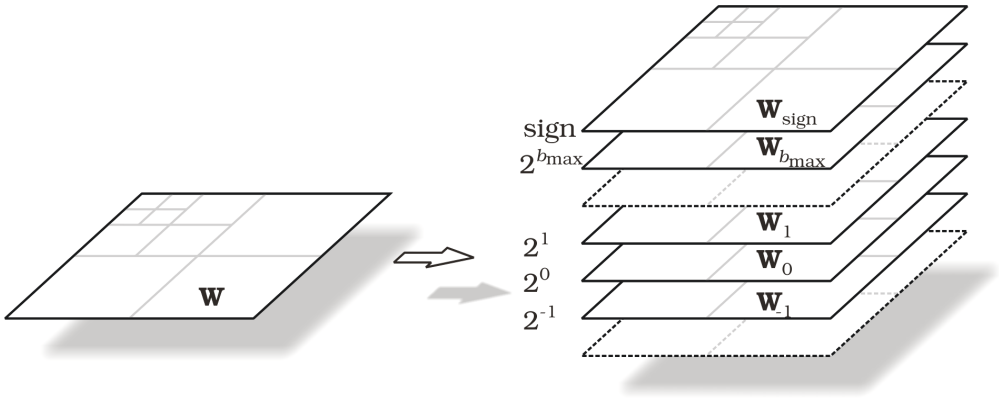

Figure 2: Bit-planes of a 2d-DWT-transformed image.

The quantization of the DCT coefficients in JPEG compression differs from the quantization of the DWT coefficients in JPEG2000. As opposed to JPEG, the approximation interval is always a multiple of two in JPEG2000 (see Fig. 3). This means that a quantized DWT coefficient cannot fall out of its coarse quantization interval after several re-compression cycles. The approximated coefficient always stays in it. In contrast, the DCT coefficients in JPEG can alter arbitrarily due to subsequent JPEG re-compression, as we have shown in [16]. We found out that interfering errors due to rounding processes in the spatial domain have much less effect on single DWT coefficients when the image is re-compressed. This property of the DWT coefficients can be used to construct and embed an image content dependent cryptographically secure hash value.

2.2 The Secure Authentication Scheme

In our authentication scheme, the image **I** is transformed with the filter of the 9/7-wavelet at first. Afterwards the 2d-DWT-transformed image **W** is pre-quantized. Therefore, every wavelet coefficient w of the transformed image **W** is quantized to an invariant value, which is not changed due to the signature embedding process.

The pre-quantization can be considered as a refining bit plane quantization (see Fig. 3). Except for the sign and b_{max} the binary state of the corresponding last refinement bit plane $\mathbf{W}_s$ is set fuzzy to "0,5". Not until after the signature embedding process a binary "0" or "1" is chosen for this bit.

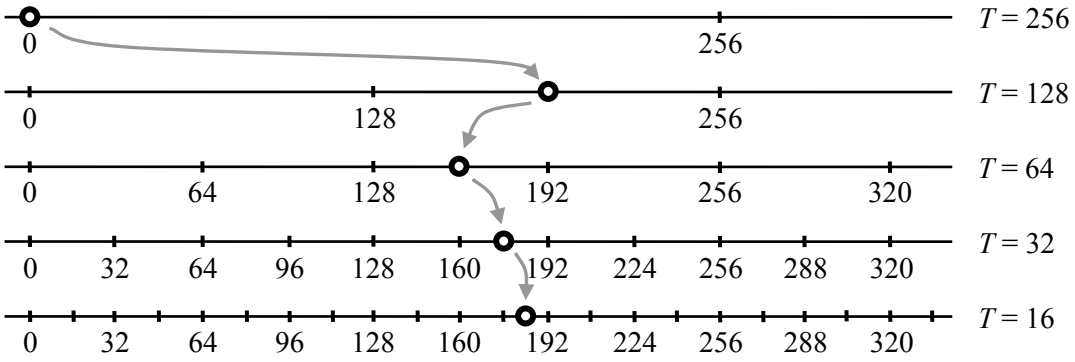

Figure 3: Refinement steps of coefficient $w = 180$ with halved threshold T after each step.

After the pre-quantization the hash value $h = \mathfrak{H}(\tilde{\mathbf{W}})$ is calculated using a cryptographic hash function $\mathfrak{H}(\cdot)$. This hash function maps the pre-quantized 2d-DWT-transformed image $\tilde{\mathbf{W}}$ of arbitrary size to a short fixed length binary number (e.g. SHA-1: 160 bit, SHA-256: 256 bit). It is easy to calculate this hash value, but the computation of image $\tilde{\mathbf{W}}$ from hash value h is computational infeasible. Hence, finding two images **I** and $\mathbf{I}'$, resulting in the same hash value, is computational infeasible too.

Since we hash all coefficients of $\tilde{\mathbf{W}}$, the embedding area is also protected. Therefore, an attack on any coefficients of the signature generation as well as embedding space can be easily detected.

Afterwards, the hash value h is asymmetrically encrypted with a private key K_{private}, which is only known to the authentication watermark embedder. To decrypt the signature, a public key K_{public} is only required. Hence, public verification is allowed to everybody without being able to insert another authentic watermark.

Considering a digital camera embedding the authentication watermark the private key has to be located inside the camera protected against reading. Even the photograph must not know this private key, because he also could be interested in tampering the original image by inserting another authentic watermark into a faked image. Hence, the overall encryption process must be protected against reading.

To deter the forger from photographing fake sceneries with the same camera using the same private key the date D, the time T, the position P, and the direction V is added to the hash. For example, date, time and position could be taken from an integrated GPS receiver. To handle the direction of the camera sensor during the shot we suggest integrating an electronic compass. Only with these additional features a digital camera system can really be considered trustworthy.

Since our authentication scheme relies on the security of the encryption algorithm $\mathfrak{S}(\cdot)$, the pair of keys K_{private}, K_{public} and the hash function $\mathfrak{H}(\cdot)$, this authentication scheme is secure. To improve the security, only the encryption algorithm and/or the hash function has to be replaced.

The signature $s = \mathfrak{S}(\{h,D,T,P,V\}, K_{\text{private}})$ is embedded in the DWT-coefficients of the upper LL-subband. If the number of signature bits is bigger than the number of coefficients in this subband,

coefficients of other subbands are used. The embedding process is illustrated in Fig. 4. To embed a signature bit the last refinement interval used during pre-quantization is halved again. It depends on the signature bit if the pre-quantized value lies in the center of the first half ($w_s = 160$) or the second half ($w_s = 224$) of the hashed interval. No iterations are necessary to embed the signature.

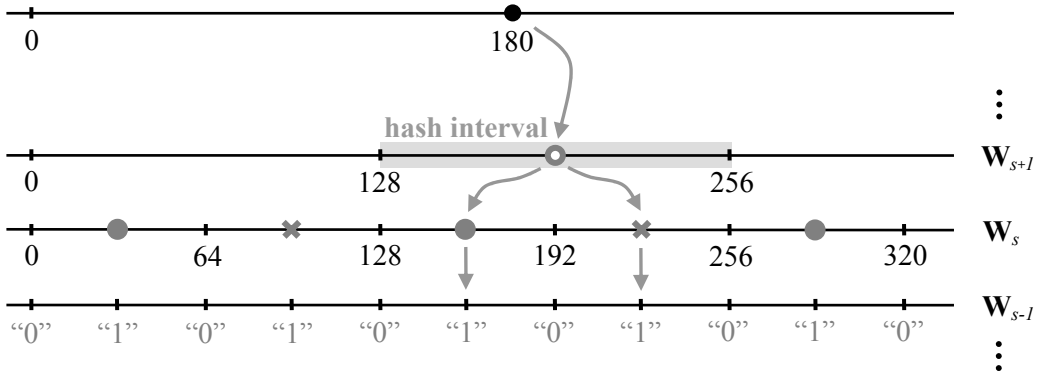

Figure 4: Embedding a signature bit in the pre-quantized and hashed coefficient.

The embedding process described above is a special case of quantization index modulation (QIM) which has been proposed by Chen and Wornell [17, 18]. It is the so-called dither modulation, which can also be interpreted here as a kind of *bit plane embedding* (see Fig. 4 and Fig. 5). The bit plane $\mathbf{W}_s$ contains the signature bits. All other bit planes $\mathbf{W}_b$, $s < b \leq b_{\max}$, contain the image data. Only these bit planes above bit plane $\mathbf{W}_s$ are hashed by the hash function $\mathfrak{H}(\cdot)$.

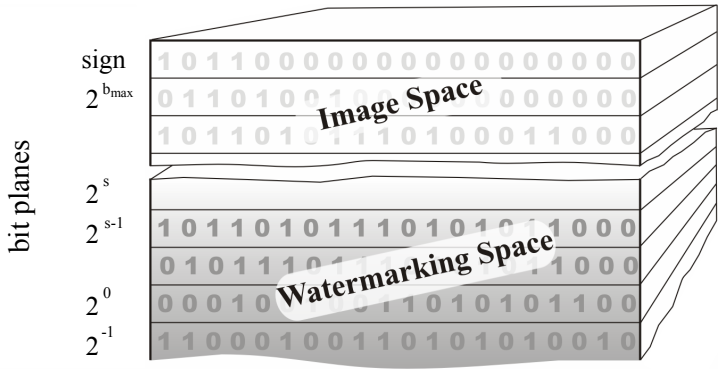

Figure 5: Watermarking space of bit planes.

2.2.1 Reducing the Watermarking Noise

To get a robust watermark, the pre-quantized values must lie in center of the quantization interval. Therefore all bits of bit plane $\mathbf{W}_{s-1}$ are set to "1" and the bits of all the bit planes below bit plane $\mathbf{W}_{s-1}$ are set to "0".

It is also possible to use the bit planes below the bit plane $\mathbf{W}_s$ for watermarking. In that case, the robustness of these additionally embedded information decreases, but more data can be embedded (see Fig. 5).

To reduce the watermarking noise, we propose the following approaches:

- The signature bits should only be embedded into the significant coefficients, because the pre-quantization error of the significant coefficients is much smaller than the pre-quantization error of the insignificant coefficients.

- As in Fig. 4, the signature bit is embedded in the lower or the upper half of the hashed interval. Since the distribution of large coefficients of a 2d-DWT-transformed image can be approximated by a uniform distribution, this position is not optimal.

  To reduce this embedding noise which is higher than simple quantization noise, the verification algorithm at the receiver side should move the quantized coefficients to the center of the hashed interval after the signature bits are extracted. In that way the mean squared error distortion can be reduced to about 57 percent at the receiver side.

Concerning the later statement, it is obvious that the signature embedding process is reversible. Therefore the number of embedded bits does not affect the quality of the watermarked image, i.e., the signature can be replaced by a longer signature, which is more secure, in the future. So our algorithm fulfills the updateability requirement of Sec. 1.1.

To reduce the noise caused by the pre-quantization process, it is also possible to code the original coefficients without pre-quantization (see Fig. 6). This does not change the robustness of these coefficients with respect to JPEG2000 re-compression. But, however, the embedded signature information is not robust to any other attack (e.g. image noise, JPEG re-compression).

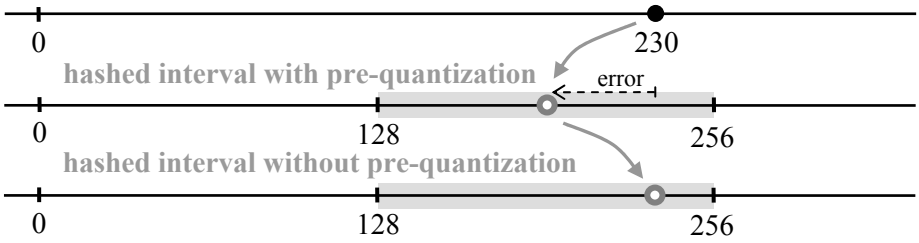

Figure 6: Hashing the interval of a coefficient with/without pre-quantization.

2.2.2 Wavelets: 5/3 vs. 9/7

If the filters of the 5/3-wavelet are used instead of the filters of the 9/7-wavelet, rounding errors can be avoided. There is no rounding noise, which could increase after each DWT/IDWT cycle, as this transform is lossless (reversible).

This is desirable in practice. However, the drawback of the reversible transform implemented in JPEG2000 is the non-linear perturbation introduced by the sequential transform which does not expand the numeric precision. This non-linear transform has a negative effect on the spreading of errors in the DWT-domain, because integer-valued input samples are always mapped to integer-valued subband samples. The error energy is not evenly spread.

Therefore it is hard to make the authentication algorithm robust to other attacks (e.g. noise in the image domain, JPEG re-compression). Thus, it is better to use the 9/7-wavelet filters to transform the images before watermarking.

2.2.3 Robustness vs. Watermarking Noise

The robustness of the embedded data depends on the position of the bit plane $\mathbf{W}_s$. When the index s increases, the robustness of the embedded signature increases because the quantization interval becomes bigger. At the same time, the embedding induced distortion increases in the watermarked image.

If the bit planes of this watermarked image are coded by a bit plane coder (e.g. SPIHT), the result is the rate-PSNR-curve shown in Fig. 7. It is obvious that the quality of the reconstructed image does not improve after the bit rate R_I is reached, because afterwards only the signature bits of bit plane $\mathbf{W}_s$ are coded which do not reduce the error of the approximated coefficients. The PSNR does not get better than P_s. The pre-quantization noise is not reduced if more bit planes are (de)coded.

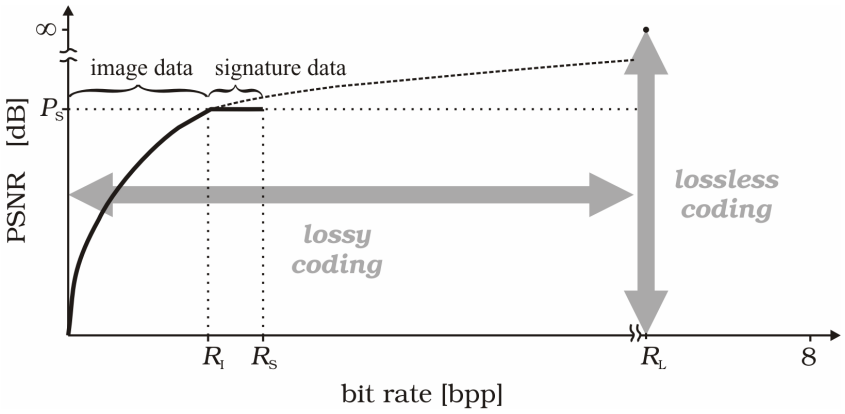

Figure 7: Rate-PSNR-curve of the watermarked image

As long as the bit rate R of the re-compressed image is bigger than R_S, the image is authentic. If the bit rate R is smaller than R_S, the decoded image is a fake.

2.2.4 Variable Resolution

To support various resolutions, we propose to divide the 2d-DWT-transformed image $\mathbf{W}$ into Z regions $\mathbf{W}_1^H, \mathbf{W}_2^H, \ldots, \mathbf{W}_Z^H$, as shown in Fig. 8. From every single region $\mathbf{W}_i^H$ a hash value $\mathfrak{H}(\mathbf{W}_i^H)$ is calculated.

Figure 8: Different hash regions to support three different resolutions ($Z = 3$).

This multi-hash approach makes the watermarked image very robust to resolution changes, because the hash $\{\mathfrak{H}(\mathbf{W}_2^H), \mathfrak{H}(\mathbf{W}_3^H)\}$ does not change, if all coefficients of $\mathbf{W}_1^H$ are set to zero.

$$h = \{\mathfrak{H}(\mathbf{W}_1^H), \mathfrak{H}(\mathbf{W}_2^H), \ldots, \mathfrak{H}(\mathbf{W}_Z^H)\}$$

The drawback of this approach is that the length of the overall hash h increases. The length of h in conjunction with the number of bits used for additional information such as date, position and direction must be less than the maximum signature key length of the encryption scheme.

It is obvious that the resolution change, which can be applied without affecting the authenticity, must be the full image resolution divided by a multiple of 2, otherwise this approach is fragile.

3. RESULTS

In this section, we provide some results. In Fig. 9, the rate-PSNR-curves of the images "Lena" and "Barbara" are shown. We used a 1024 bit signature, which was embedded in the upper subband at 4 levels wavelet decomposition. To implement our JPEG2000-based authentication watermarking algorithm, we chose the well-known "Kakadu" reference software written by Taubman [19].

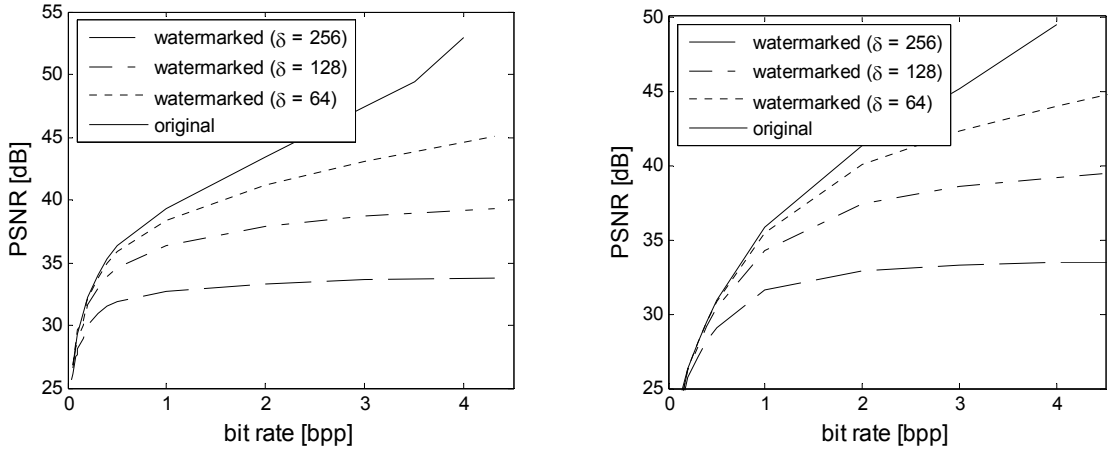

Figure 9: Rate-PSNR curves of image "Lena" (left) and image "Barbara" (right).

The PSNR of the watermarked image is shown in Fig. 10 (left) for the 512 x 512 pixel images "Lena", "Barbara" and "Mandrill". Since at low bit rates JPEG2000 seems to bear resemblance to a low-pass system, the images "Barbara" and "Mandrill" having more high-frequency textured regions are stronger affected. Obviously, there is a coefficient pre-quantization interval step size δ, which is commonly good for most natural images. Evidently, the authentication algorithm is more robust to re-compression if the pre-quantization interval is bigger. We suggest $\delta = 64$ for the upper subband as a tradeoff between robustness and embedding induced distortion. The δ-values for the other subbands have to be weighted differently because of the gain of the wavelet filters.

In Fig. 10 (right) and Fig. 11, we demonstrate the robustness against JPEG as well as JPEG2000 compression and noise in the spatial domain, respectively. As can be seen, if an image is watermarked with $\delta = 64$, JPEG re-compression can take place without losing authenticity as long as JPEG quality factor QF > 55 (approximately).

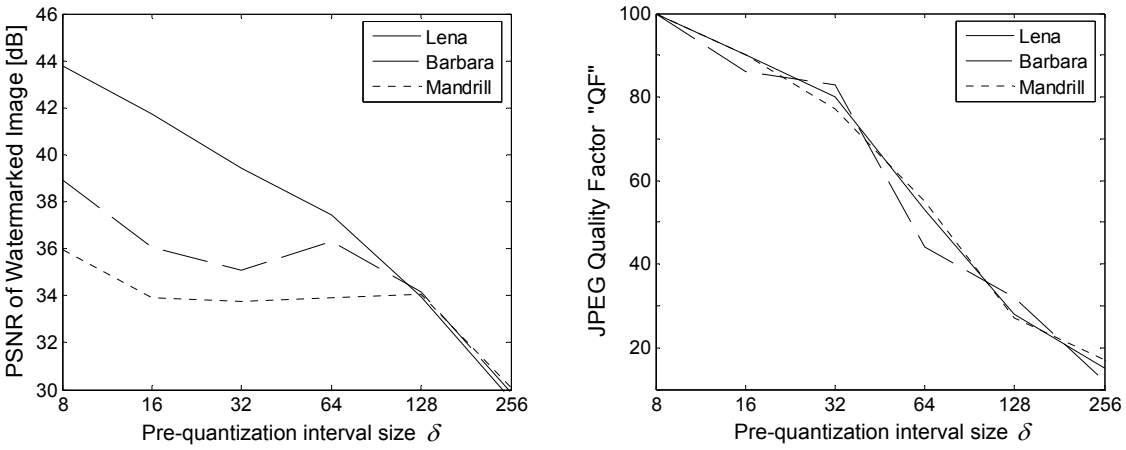

Figure 10: PSNR curves of watermarked image "Lena", "Barbara" and "Mandrill" (left) and their robustness against JPEG re-compression (right).

Since malicious or non-malicious attacks on watermarked images are often modeled as noise, we tested the robustness against Gaussian noise in the spatial domain. In Fig. 11 (right), the strength of the added pixel noise is given by the watermark-to-noise-ratio WNR = $20*\log_{10}(255/NL)$, where NL is the maximum pixel noise signal amplitude.

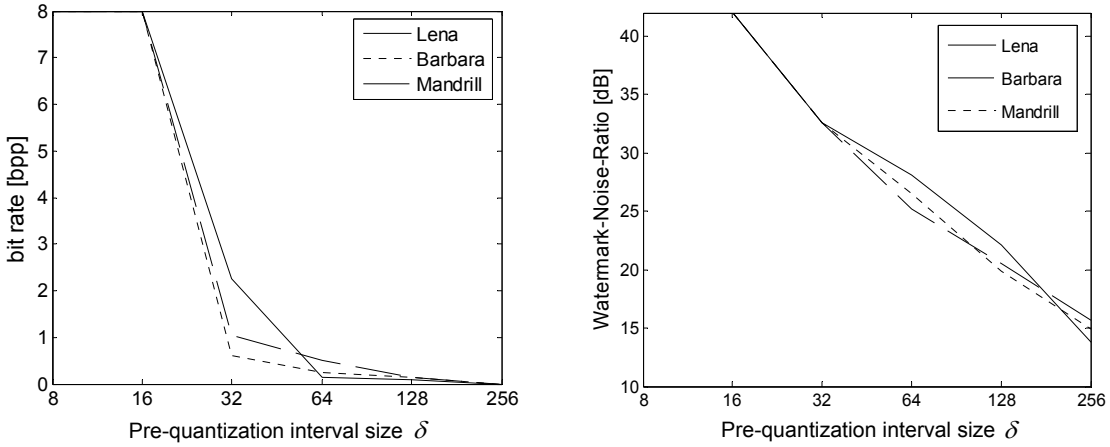

Figure 11: Robustness of the watermarked test images against JPEG2000 re-compression (left) and Gaussian noise in the spatial domain (right).

Fig. 12 gives an example of the embedding induced distortion ($\delta = 64$). The well-known original image "Lena" is not shown. The difference of the watermarked image to the original one is demonstrated at the right side. Its pixel values range from -53 to 35 and are typical for wavelet/quantization-based compression.

Figure 12: Watermarked image "Lena" (left), $\delta = 64$, PSNR = 37.42 dB, and difference image to the original one (right).

CONCLUSION

In this paper, we have presented a new JPEG2000-based authentication method, which is robust to re-compression, resolution changes, and noise. It was shown that our authentication scheme is secure in contrast to most of the authentication schemes proposed so far in the wavelet domain. Since hash as well as encryption algorithm, responsible for the security, can be updated easily, the proposed authentication scheme will also be secure in the future. Furthermore, we have enhanced the model of the trustworthy digital camera in such a way as to deter a forger from photographing fake sceneries. Our authentication algorithm also integrates date, time, position and direction of a camera shot. Only with these additional features a camera system can really be considered trustworthy.

REFERENCES

[1] G. L. Friedman, „*The trustworthy digital camera: Restoring credibility to the photographic image*", IEEE Trans. on Consumer Electronics, vol. 39, pp. 905-910, 1993.

[2] P. Meerwald and A. Uhl, „*A survey of wavelet-domain watermarking algorithms*", Proc. of SPIE, Electronic Imaging, Security and Watermarking of Multimedia Contents, vol. 4314, pp. 505-516, San Jose, CA, 2001.

[3] M. Albanesi, M. Ferretti and F. Guerrini, „*A taxonomy for image authentication techniques and its application to the current state of the art*", Proc. of 11th Int. Conf. Image Analysis and Processing (ICIAP), Palermo, pp. 535-540, 2001.

[4] T. Liu and Z. ding Qiu, „*The survey of digital watermarking-based image authentication techniques*", Proc. of IEEE 6th Int. Conference Signal Processing, vol. 2, pp. 1556-1559, 2002.

[5] C. Rey and J. L. Dugelay, „*A survey of watermarking algorithms for image authentication*", EURASIP Journal on Applied Signal Processing (JASP), pp. 613-621, 2002.

[6] Y. Song and T. Tan, „*A brief review on fragile watermarking based image authentication*", Journal of Image and Graphics, vol. 8A, pp. 1-7, 2003.

[7] B. B. Zhu, M. D. Swanson and A. H. Tewfik, „*When seeing isn't believing - current multimedia authentication technologies and their applications*", IEEE Signal Processing Magazine, vol. 21, pp. 40-49, 2004.

[8] C. Y. Lin, „*Issues on Multimedia Authentication*" Multimedia Security: Steganography and Digital Watermarking Techniques for Protection of Intellectual Property (IGP), pp. 173-206, 2004.

[9] C. J. Skrepth and A. Uhl, „*Robust hash functions for visual data - An experimental comparison*" Proc. of IbPRIA, pp. 986-993, 2003.

[10] R. Grosbois, P. Gerbelot and T. Ebrahimi, „*Authentication and access control in the JPEG 2000 compressed domain*", Proc. of SPIE, Applications of Digital Image Processing, vol. 4475, San Diego, CA, pp. 95-104, 2001.

[11] C. Peng, R. H. Deng, Y. Wu and Shao, W, „*A flexible and scalable authentication scheme for JPEG2000 image codestreams*", Proc. of 11th ACM Int. Conf. on Multimedia, pp. 433-441, 2003.

[12] Z. Zhang, et al., „*A unified authentication framework for JPEG2000*", Proc of IEEE Int. Conf. on Multimedia and Expo (ICME), Taipei, pp. 915-918, 2004.

[13] Y. Wu, D. Ma and R. H. Deng, „*Progressive authentication of JPEG2000 codestream*", ISO/IEC JTC1/SC29/WG1, Proposal N3086, 2003.

[14] Y. Wu, D. Ma and R. H. Deng, „*Imtrust: Design and implementation*", ISO/IEC JTC1/SC29/WG1, Proposal N3075, 2003.

[15] Y. Wu, D. Ma and R. H. Deng, „*Progressive protection of JPEG2000 codestreams*", Proc. of IEEE Int. Conf. on Image Processing (ICIP), Singapore, pp. 3447-3450, 2004.

[16] M. Schlauweg, T. Palfner, D. Pröfrock and E. Müller, „*The Achilles' Heel of JPEG-based image authentication*", Proc. of IASTED Int. Conf. on Communication, Network and Information Security (CNIS), Phoenix, AZ, 2005.

[17] B. Chen and G. W. Wornell, „*Digital watermarking and information embedding using dither Modulation*", Proc. of IEEE 2nd Workshop on Multimedia Signal Processing, pp. 273-278, 1998.

[18] B. Chen and G. W. Wornell, „*Quantization index modulation: A class of provably good methods for digital watermarking and information embedding*", IEEE Trans. Information Theory, vol. 47, pp. 1423-1443, 2001.

[19] D. S. Taubman and M.W. Marcellin, „*JPEG2000: Image Compression Fundamentals, Standards and Practice*", Kluwer Academic Publishers, 2002.