Applications of Information Hiding and Digital Watermarking

J. J. Eggers and R. Bäuml and R. Tzschoppe and J. Huber

Telecommunications Laboratory, University of Erlangen-Nuremberg, Germany

Abstract

Digital watermarking denotes the hiding of information in general documents, where the embedded information can be exploited to combat copyright infringements or to verify the integrity of a document. We emphasize that blind watermarking, where the original document is not known to the watermark receiver, can be interpreted as communication with side information at the encoder. Significant improvements over state-of-the-art watermarking schemes can be achieved when using this side information. The new watermarking technology broadens the range of possible applications for digital watermarking and information hiding, which is demonstrated for three example applications.

1. Introduction

Digital media has replaced analog media in many applications within the last decade. This is mainly due to improved properties like simple noise-free transmission over general-purpose channels, compact storage, perfect copying and simple editing. Further, the availability of the Internet as an open global network for the transmission of digital data accelerated the use of digital media. These first-view advantages of digital media over analog media turn into disadvantages with respect to the protection of intellectual property rights and the trustworthiness of the content. Simple noise-free transmission of digital data (e. g., images, audio, video) via the Internet enables pirates to reach a huge market, while prosecution is difficult. Even worse, perfect copies are obtained easily and can be stored on CDs or DVDs. Simple editing of digital data can destroy the trust in documents, e. g., by the replacement of a person's face within a digital image.

Digital watermarking has been proposed as an approach to enforce copyrights or to ensure the integrity of generalized documents, containing a variety of digital data types. Here, digital watermarking is considered as the *imperceptible, robust, secure communication* of information by embedding it in and retrieving it from other digital data. Digital watermarking is particularly appropriate for data types consisting of continuous valued elements, e.g., 3-D coordinates, pixel intensities or frequency coefficients. The basic idea is that the embedded information – the watermark message – travels with the data wherever the watermarked document goes. Possible applications are access control, e. g., for DVDs [4], distribution tracing by embedding different watermarks into different copies, broadcast monitoring, or media integrity verification. Digital watermarking is a special case of *information hiding*, where an attacker tries to impair watermark reception as much as possible. Additional applications for information hiding are for instance, covert communication (*steganography*) or the embedding of supplemental information into multimedia data.

This paper reviews recent developments of digital watermarking schemes based on a communications approach and presents example applications for different document types. Sec. 2 describes basic performance differences for general purpose watermarking schemes, including our new SCS watermarking. The application of SCS watermarking to watermarking of chemical structure sets, to the verification of image integrity, and to the robust embedding of many watermark bits into one image are discussed in Sec. 3, Sec. 4 and Sec. 5, respectively. Sec. 6 concludes the paper.

2. Principles of Blind Watermarking

In general, we have to distinguish between the communication of a watermark message and the detection of the existence of an embedded watermark. Here, the corresponding basic concepts are reviewed. We emphasize that blind watermarking should be considered communication with side information, which enables significant improvements over state-of-the-art blind watermarking schemes.

2.1. Communications of a Watermark Message

Fig. 1 depicts a general perspective on watermark communication. A *watermark message* m is embedded into the *original data* (host data) $\mathbf{x}$ to produce the *watermarked data* $\mathbf{s}$. Here, $\mathbf{x}$ denotes a vector of data elements. Depending on the document to be watermarked, an original data element can be for instance a signal sample, the intensity of an image pixel or a transform coefficient. The essential requirements on data being robustly watermarkable are that there is enough uncertainty about the exact realization of the original data and that quality assessments can be made only in a statistical sense.

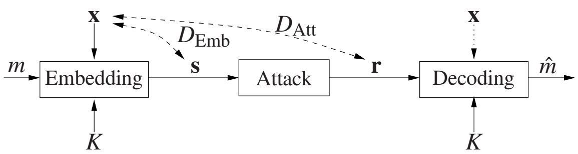

Figure 1: *General model of digital watermarking.*

The embedding process depends on the key K and must be such, that the quality difference between $\mathbf{x}$ and $\mathbf{s}$ (*embedding distortion* D_{Emb}) is not too large. The difference $\mathbf{w} = \mathbf{s} - \mathbf{x}$ is denoted the *watermark signal*. The definition of an appropriate data quality measure is very important for the analysis of digital watermarking. In many cases, a mean squared error (MSE) distortion measure allows a meaningful quality assessment. The data quality after watermark embedding is characterized by the document-to-watermark power ratio $\text{DWR} = 10\log_{10}(\sigma_x^2/\sigma_w^2)$ [dB].

The watermarked data $\mathbf{s}$ might be further processed or even replaced by some other data. This process, denoted *attack*, produces the *attacked data* $\mathbf{r}$. The goal of the attack is to impair or even remove the embedded watermark information. Watermark reception denotes both, *decoding* of a received watermark message $\hat{m}$ using key K and, watermark *detection*, meaning the hypothesis test whether $\mathbf{r}$ is watermarked or not. Watermark detection is discussed in more detail in the Section 2.2.

In some applications of digital watermarking, the original data $\mathbf{x}$ might be available to the watermark receiver as indicated with the dotted arrow in Fig. 1, however, in many applications it is not available. We call the first case *non-blind watermarking* and the latter case *blind watermarking*. Here, we focus on *blind watermarking*. The unknown host data is considered in the popular blind spread-spectrum (SS) watermarking scheme as unavoidable interference. However, Fig. 1 reveals that the host data is side information to the watermark encoder which can be exploited for improved blind watermarking schemes.

The performance of four different watermarking schemes for an additive white Gaussian noise (AWGN) attack is discussed in order to illustrate the potential gains by exploiting the side information at the encoder. The AWGN attack is a simple attack that can be easily applied to all possible data types. Further, a theoretical analysis of the achievable watermark rate (watermark capacity) is possible and serves often as a basis for the analysis of extended attacks. The strength of an attack by AWGN $\mathbf{v}$ is characterized by the watermark-to-noise power ratio $\text{WNR} = 10\log_{10}(\sigma_w^2/\sigma_v^2)$ [dB], where σ_v^2 denotes the noise power.

The considered watermarking schemes are

- *spread-spectrum* (SS) watermarking as established in the watermarking community for several years [9, 3]
- a theoretically ideal scheme that exploits the side information about the host data, referred to as *Ideal Costa Scheme* (ICS) [1, 2]
- a practical watermarking scheme called *Scalar Costa Scheme* (SCS) [7] that we derived from ICS
- *Spread Transform - SCS* (ST-SCS), which further improves the robustness of SCS especially for strong attacks.

A detailed description of the presented watermarking schemes and the analysis of their capacities is given in our previous publications [7, 8]. Fig. 2 shows that ST-SCS and SCS watermarking do not achieve the capacity of ICS, but are not too far from ICS either. ST-SCS watermarking gives an advantage over SCS watermarking only for a WNR below 0.01 dB. Blind SS watermarking suffers significantly from host data interference. For weak to moderately strong attacks SCS watermarking outperforms SS watermarking by far.

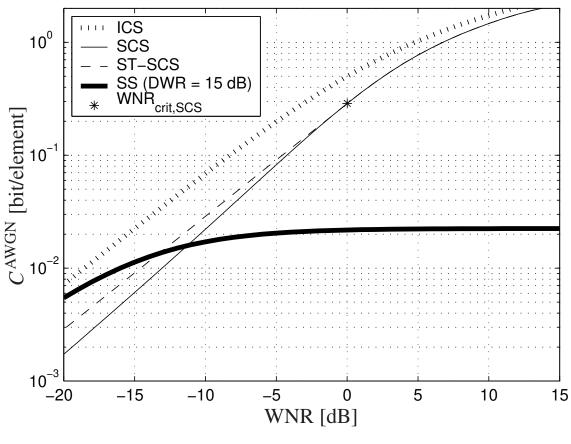

Figure 2: *Capacity of blind watermarking schemes facing an AWGN attack with* $\text{DWR} = 15$ dB.

2.2. Watermark detection

Watermark detection refers to the decision whether the received data $\mathbf{r}$ is not watermarked with key K (H_0) or is watermarked with key $K(H_1)$, which can be formulated as an hypothesis test. We assume that the *Probability Density Functions* (PDF) $p_r(\mathbf{r}|H_0)$ and $p_r(\mathbf{r}|H_1)$ for receiving $\mathbf{r}$ depending on hypothesis H_0 or H_1, respectively, are known. Then, Bayes' solution to the hypothesis-testing problem can be applied, which is

$$\frac{p_r(\mathbf{r}|H_1)}{p_r(\mathbf{r}|H_0)} \begin{cases} > T & \Rightarrow \quad \text{accept } H_1 \\ \leq T & \Rightarrow \quad \text{accept } H_0, \end{cases} \tag{1}$$

with T being the decision threshold. Assuming equal a-priori probabilities, (1) can be reformulated so that H_1 is accepted if

$$P_r = \frac{p_r(\mathbf{r}|H_1)}{p_r(\mathbf{r}|H_1) + p_r(\mathbf{r}|H_0)} > 0.5. \tag{2}$$

Here, P_r, with $P_r \in [0,1]$, denotes the reliability that the received data elements $\mathbf{r}$ are watermarked.

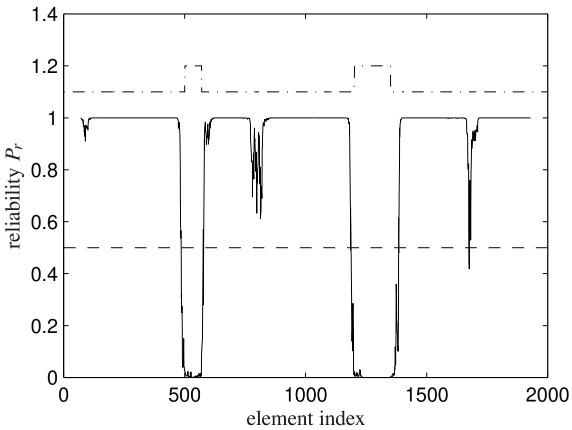

Figure 3: *Sliding window watermark detection. Two data blocks of length 70 and 150 were replaced (dash-dotted line). The solid line depicts P_r at the position of the window center. $P_r > 0.5$ (dashed line) indicates watermarked data.*

Reliable SCS watermark detection requires only a relatively small number of received data elements compared to SS watermarking due to the lack of host data interference. This can be exploited to detect local replacements of data blocks by non-watermarked data. Fig. 3 shows an example for a data vector of length 2000 with embedded SCS watermark after an AWGN attack (WNR = 3 dB). Two watermarked data blocks were replaced by random data after the AWGN attack. A sliding window of length 140 is moved over the data elements and for each window position SCS watermark detection is applied. We observe that the non-watermarked regions could be located quite accurately. In this example, no false-positive and only one false-negative error occurred when the window covered completely non-watermarked or completely watermarked data, respectively.

3. Watermarking of Chemical Structure Sets

The information about 3-D atomic coordinates of chemical structures is valuable knowledge in many respects. Therefore, the producer of such a data set is interested in enforcing his intellectual property rights. The computer chemistry center of the University Erlangen-Nuremberg has already reported cases where the copyright of some of their data has been infringed. Thus, the embedding of copyright information via digital watermarking of 3-D atomic coordinate data is considered. Such a system based on ST-SCS watermarking has been developed[6]. A short overview of the system and its performance is given in the following.

The intellectual property of the considered data set resides only in the 3-D atomic coordinates which have been derived from sophisticated rule-driven model builders. Taking into account the limited precision of the model builder, a variation of the coordinates is acceptable and can be used for watermarking purposes. Since we are mainly interested in identifying the origin of large data sets, e.g., including 100,000-200,000 structures, the watermark message is distributed over several molecule structures.

In an attempt to embed a watermark in 3-D atomic coordinates, the representation of the molecule structure needs to be normalized and identified. For this, a unique 3-D orientation and canonic order of the atoms is generated. Often the number of watermarkable data elements per molecule is smaller than the number of coded watermark bits. Thus, only a specific part of the binary encoded watermark message is embedded into a single structure. The embedded code bits are determined by a hash key generated from the structure. Once the encoded message part has been identified, ST-SCS watermark embedding into the ordered atomic coordinates is applied.

All received molecule structure sets are considered for watermark reception. First, the information from all received structures is combined to obtain reliability information for the coded watermark bits. Next, error correction decoding gives the received watermark message bits. Some of the watermark bits are known to the receiver so that the validity of the decoded watermark message can be verified.

The implemented watermarking scheme has been investigated for synthetical data and molecule data, where in both cases highly reliable watermark reception could be demonstrated. Here, we present some experimental results for synthetical data that show the achievable watermark bit error rate (BER) depending on the number of received data elements. Reliable reception from as few data elements as possible is desired. The discussion is restricted to an AWGN attack with WNR = -3 dB. A rate 1/3 convolutional code was used to encode 96 watermark bits (the copyright information) into 315 coded watermark bits. Communication of 20000 random watermark messages was performed so that BER $\approx 10^{-5}$ could be measured reliably. Fig. 4 shows the measured BER for 800, 1200, 1600, and 2000 received data elements. We observe that BER $< 10^{-5}$ is achieved when receiving 2000 data elements. The BER increases slowly when less data elements are received. Note that 2000 watermarked data elements can be easily obtained from 30-50 single molecules, where the exact number depends on the size of the molecules. Thus, our system performs good enough to receive the watermark from small subsets of large structure data sets.

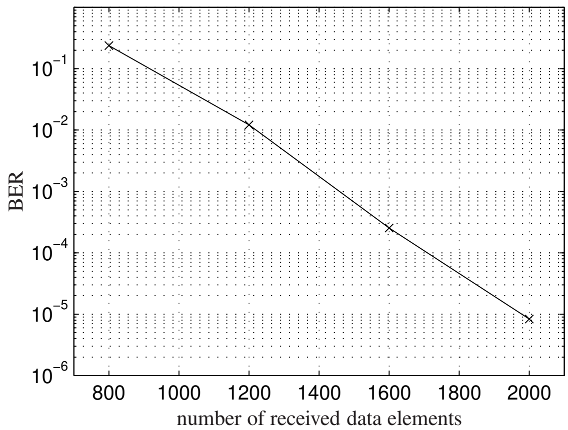

Figure 4: BER *for receiving 96 watermark message bits after AWGN attack (WNR = -3.0 dB) and different number of received data elements.*

4. Image Integrity Verification Based on SCS Watermarks

This section focuses on image integrity verification, i. e. modifications of the image content, like exchanging or erasing image objects, should be detectable. It is assumed that modification of the content requires the replacement of the pixels in an entire region of the image. The image is considered original (or authentic) if no such modifications are detectable.

The most important requirements for image integrity verification are

- good localization of image regions with modified content,
- robustness against common processing distortion, e. g., lossy compression or D/A and A/D conversion
- and security, which means that only authorized parties can create data which is classified to be original.

When using digital watermarking for image integrity verification, watermark information generated with a *key* is spread all over the original image. The image marked in such a way is subject to operations like D/A and A/D conversion, lossy compression or additive noise, leading to a distorted image. The watermark is designed such that it is reliably detectable as long as the distorted image has a sufficiently high quality and the correct key is known. However, if some image region is replaced by somebody not knowing the key, the watermark information will not be detectable from the modified image region. Therefore, reliability of watermark detection can be used as a measure of integrity.

The usage of SCS watermarking is proposed[5] due to its good detection reliability for detection from a small number of received data elements as shown in Sec. 2.2. For image data, local dependencies between pixels have to be considered. Thus, SCS watermarking of image data in the 8×8-block DCT domain is applied. High frequency DCT coefficients are not watermarked due to the vulnerability of watermarks in these coefficients against JPEG compression. We present here simulation results obtained for watermarking of the grayscale test image "Girl" with $\mathrm{PSNR_E} > 40$ dB, where $\mathrm{PSNR_E} = 10\log_{10}(255^2/D_{\mathrm{Emb}})$ denotes the peak-signal-to-noise ratio after watermark embedding.

Fig. 5 shows the measured total detection error rates (mean of false-positive and true-positive error rates) when detecting SS and SCS watermarks from image blocks with 32×32 pixels. The watermarked image had been JPEG compressed with different quality factors, before half of the 32×32 blocks were replaced. Detection error rates $p_e < 10^{-3}$ could be achieved for SCS watermark detection and an image quality of $\mathrm{PSNR_A} > 36$ dB, where $\mathrm{PSNR_A} = 10\log_{10}(255^2/D_{\mathrm{Att}})$ denotes the peak-signal-to-noise ratio after the attack. The reliablity of SS watermark detection is limited to $p_e \approx 0.05$ due to large interference from the original image. This demonstrates again the superiority of SCS watermarking.

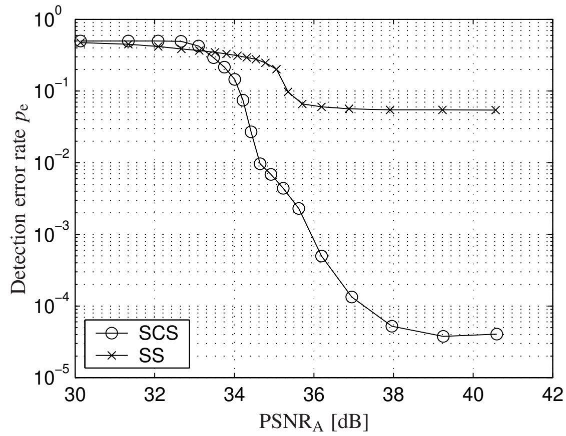

Figure 5: *Detection error rates p_e after JPEG compression for SCS and SS watermark detection.*

Fig. 6 shows the detection results for example manipulations. The watermarked image was manipulated in four different regions and JPEG compressed with quality factor 70. The watermarked and the manipulated images are shown in the upper row of Fig. 6. Next, sliding window detection with a window of size 32×32 has been applied. The detection results are shown in the lower row of Fig. 6 with and without the manipulated image. Dark spots indicate detected modified regions. All manipulations have been detected, even the small extension of the band in the hair. One region was falsely classified to be modified. In this region, the original image was almost flat so that robust watermarking is not possible.

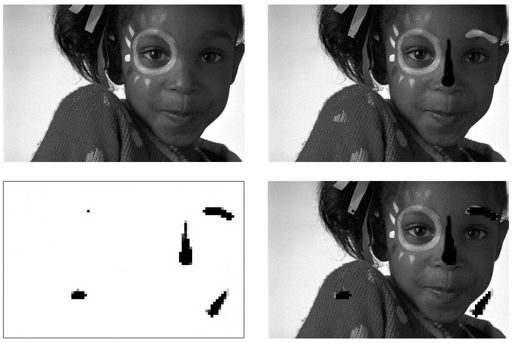

Figure 6: *Upper left: watermarked; upper right: manipulated and JPEG compressed; lower left: detected regions with modified content; lower right: detected regions with modified content on top of the manipulated image.*

5. Communication of Watermark Messages via Image Data

In this section, the applicability of ST-SCS watermarking to image data for the embedding of copyright information is demonstrated. We present an experimental watermarking system which operates under the assumptions that the decoder is synchronized and has perfect knowledge about the attack parameters. Our analysis relates theoretically computed watermark capacities to practically implementable watermark payloads which achieve a sufficiently low bit error rate (BER), e.g., BER $< 10^{-5}$, for certain attack scenarios. Thus, the influence of practical constraints like the constraint on the codeword length by the image size is investigated. Optimization of the watermarking scheme is performed for a *linear filtering and additive colored Gaussian noise (FACGN) attack*, though a JPEG compression attack is considered as well to demonstrate the practical usefulness of this watermarking scheme.

We discuss the robust embedding of a sequence of watermark message bits into a grayscale image of size 512×512. As in Sec. 4, embedding in the 8×8 block-DCT domain is considered. Each 8×8 block is transformed into 64 DCT coefficients. Next, the coefficients with identical frequency index from all 8×8 blocks compose a subchannel. Thus, there are 64 subchannels, all having the same length, which is identical to the number of 8×8 blocks in the given image. The DC component cannot be utilized for embedding due to it's unfavourable statistics, so the watermark is embedded only into the AC components. The transform coefficients of each subchannel compose a signal which is ST-SCS watermarked, where turbo coding with rate 1/3 is applied. Throughout the section, the watermark embedding is such that the image quality after embedding, measured in $\mathrm{PSNR_E}$, is about 40.5 dB.

The embedding scheme has been designed for a certain $\mathrm{PSNR_A}$ that should be survived. Here, the design for two differently severe attacks, characterized by $\mathrm{PSNR_{A,0}} \approx 35$ dB and $\mathrm{PSNR_{A,1}} \approx 30$ dB, is considered. We consider three different embedding strategies. (E1) denotes the strategy being optimized for attacks with $\mathrm{PSNR_{A,0}}$. (E2) denotes the strategy that is optimized for attacks with $\mathrm{PSNR_{A,1}}$. In (E1) and (E2) watermarking of all 63 AC components is considered. The third strategy, denoted by (E3), is a heuristic that works well for FACGN attacks *and* for JPEG compresssion attacks. (E3) considers for watermark embedding only the first 20 AC components in zigzag-scan.

Tab. 1 shows the theoretic capacity and the achieved watermark payload for three different embedding strategies. The differences between the capacity and the actual payload are due to the unavoidable losses introduced by practical coding techniques. The losses are mainly due to a finite interleaver length of the turbo code and due to spread transforms with integer spreading factor. For the embedding strategy (E1), optimized for the attack strength $\mathrm{PSNR_{A,0}}$, the achieved payload is about 80 % of the watermark capacity. For the embedding strategies (E1) and (E2), designed for the attack strength $\mathrm{PSNR_{A,1}}$, the achieved payload is about 70 % of the watermark capacity. Note that a practical scheme can operate closer to the watermark capacity for weak attacks, since an increased capacity allows longer codewords in the ST-domain.

Embedding strategy	Capacity	Payload
(E1) - $\mathrm{PSNR_{A,0}}$	13839	11085
(E2) - $\mathrm{PSNR_{A,1}}$	1320	912
(E3) - $\mathrm{PSNR_{A,1}}$	925	645

Table 1: *Theoretically derived watermark capacity and practically implemented watermark payload for ST-SCS watermarking of the grayscale image "Lenna" of size* 512×512. *Optimized embedding distortion allocation (E1/E2) for FACGN attacks with* $\mathrm{PSNR_{A,0}}$ *and* $\mathrm{PSNR_{A,1}}$, *and the heuristics (E3) for an FACGN attack with* $\mathrm{PSNR_{A,1}}$ *is considered.*

Fig. 7 depicts the measured BER for the differently designed ST-SCS watermarks for the Lenna image. The FACGN attack is considered when using the embedding strategies (E1) and (E2), where BER $< 10^{-5}$ is achieved for $\mathrm{PSNR_A} = \mathrm{PSNR_{A,0}}$ and $\mathrm{PSNR_A} = \mathrm{PSNR_{A,1}}$, respectively. Thus, the embedded watermarks achieve accurately the design goals. The simulation results for the embedding

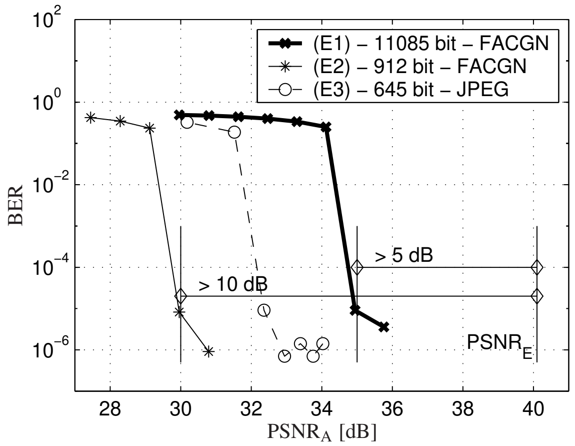

Figure 7: *BER after turbo decoding of the watermark bits for ST-SCS watermarking after FACGN and JPEG attack. The watermarks have been designed for* $\mathrm{PSNR}_{\mathrm{A},0}$ *(thick lines) and* $\mathrm{PSNR}_{\mathrm{A},1}$*(thin and dashed lines).*

strategy (E3) and JPEG compression attacks with quality factors $10, 15, 20, \ldots, 35, 40$ show that low BERs can be achieved for a quality factor 20 or higher which corresponds to $\mathrm{PSNR}_{\mathrm{A}} > 32$ dB. However, this robustness could be obtained only due to the reduced watermark payload of (E3) as shown in Tab. 1. Further, JPEG compression is still more severe than the FACGN attack considered during the watermark design. A detailed analysis of this result reveals that the performance difference in case of FACGN attacks and JPEG compression attacks is mainly due to an attack distortion allocation which has been not expected during the watermark design, where FACGN attacks are assumed. Future research has to show whether more flexible coding techniques can cope with such modified attack strategies.

6. Conclusion

A communication perspective on digital watermarking has been presented. It has been revealed that the host data can be considered side information to the encoder, which should be exploited during watermark embedding, in particular in blind watermarking scenarios. State-of-the-art blind watermarking schemes are mainly based on the principles of spread spectrum (SS) watermarking where the side information about the host data is not exploited. We developed a new blind watermarking technique, called scalar Costa scheme (SCS), which exploits the side information and thus outperforms blind SS watermarking.

The new blind watermarking technology can be applied to many different data types and in different application scenarios. This is illustrated for watermarking of chemical structure sets, for the verification of image content based on an embedded watermark, and for the embedding of copyright information into image data. The high watermark rates achieved with SCS watermarking make this scheme also attractive for steganography or the embedding of supplemental information about the document. However, the implementation of these applications is left for future work. Further, there are still some important practical problems to be solved. Especially the problem of watermark synchronization requires further research activities. Another important topic for future research is the optimization of watermarking for variable attack strategies.

Acknowledgements

The authors thank Jonathan Su, Wolf-Dietrich Ihlenfeldt, Marco Breiling, Robert Fischer and Simon Hüttinger for their active help and fruitful discussions.

References

1. B. Chen and G. W. Wornell. Achievable performance of digital watermarking systems. In *Proceedings of the IEEE Intl. Conference on Multimedia Computing and Systems (ICMCS '99)*, volume 1, pp. 13–18, pages 13–18, Florence, Italy, June 1999. 2

2. M. H. M. Costa. Writing on dirty paper. *IEEE Transactions on Information Theory*, 29(3):439–441, May 1983. 2

3. I. Cox, J. Kilian, T. Leighton, and T. Shamoon. Secure spread spectrum watermarking for multimedia. *IEEE Transactions on Image Processing*, 6(12):1673–1687, 1997. 2

4. I. J. Cox and J.-P. Linnartz. Some general methods for tampering with watermarks. *IEEE Journal on Selected Areas in Communications*, 16(4):587–593, May 1998. 1

5. J. J. Eggers and B. Girod. Blind image watermarking applied to image authentication. In *Proceedings of the IEEE Intl. Conference on Speech and Signal Processing 2001 (ICASSP 2001)*, Salt Lake City, Utah, USA, May 2001. 4

6. J. J. Eggers, W.-D. Ihlenfeldt, and B. Girod. Digital watermarking of chemical structure sets. In I. S. Moskowitz, editor, *Proceedings of 4th Information Hiding Workshop 2001*, volume 2137, pages 198–212, Pittsburgh, PA, USA, April 2001. Lecture Notes in Computer Science, Springer. 3

7. J. J. Eggers, J. K. Su, and B. Girod. A blind watermarking scheme based on structured codebooks. In *Secure Images and Image Authentication, Proc. IEE Colloquium*, pages 4/1–4/6, London, UK, April 2000. 2, 2

8. J. J. Eggers, J. K. Su, and B. Girod. Performance of a practical blind watermarking scheme. In *Proc. of SPIE Vol. 4314: Security and Watermarking of Multimedia Contents III*, San Jose, Ca, USA, January 2001. 2

9. F. Hartung and B. Girod. Digital watermarking of raw and compressed video. In *Proceedings EUROPTO/SPIE European Conference on Advanced Imaging and Network Technologies*, Berlin, Germany, October 1996. 2