PROTECTION OF WAVELET-BASED WATERMARKING SYSTEMS USING FILTER PARAMETRIZATION

Werner Dietl, Peter Meerwald, Andreas Uhl

Department of Scientific Computing, University of Salzburg Jakob-Haringerstrasse 2, 5020 Salzburg, Austria / Europe E-mail: {wdietl, pmeerw, uhl}@cosy.sbg.ac.at

Abstract

We discuss wavelet filter parametrization as a means to add security to wavelet based watermarking schemes. It turns out that our proposed scheme is resistant to unauthorized detection and unauthorized removal attacks and is equally robust as compared to the use of standard wavelet filters in case the underlying watermarking scheme is spread-spectrum based. Quantization-based watermarking techniques are not suited to be enhanced by this approach.

We assess the security, quality and robustness of our approach in the light of four selected watermarking algorithms. Furthermore, we present an attack tailored to exploit a common weakness in image-adaptive watermarking and propose filter parametrization as a suitable countermeasure.

1 Introduction

Fast and easy distribution of content over the Internet is a serious threat to the revenue stream of content owners. Watermarking has gained high popularity as a method to protect intellectual property rights on the Internet. For introductions to this topic see [17,8,15,4,9].

Over the last several years the wavelet analysis was developed as a new method to analyze signals [6,36,21]. Wavelet analysis is also used in image compression, where better energy compaction, the multi-resolution analysis and many other features make it superior to the existing discrete-cosine based systems like JPEG. The new JPEG2000 compression standard [14,31] uses the wavelet transformation and achieves higher compression rates with less perceptible artifacts and other advanced features.

Preprint submitted to Elsevier Science

With the rising interest in wavelets also the watermarking community started to use them. Many watermarking algorithms have been developed that embed the watermark in the wavelet transform domain, for an overview see [22].

Cryptanalysis is the subfield of cryptography that analyses cryptographic algorithms and tries to circumvent them. There has been a fruitful interaction between the development of new algorithms and the defects found through cryptanalysis. For the field of watermarking and steganography this subfield is called steganalysis. In part it builds on the technologies developed for cryptanalysis and adds its own methods to detect and disable hidden information. The resilience of a watermarking system can be separated into robustness and security. In the following we will use the terminology from [4].

Robustness means the resistance against common signal distortions that are known beforehand. For example, a system that includes a transmission over a noisy communication channel needs to be robust against the noise. It is known in advance that the channel is noisy and without robustness against the noise the watermarking system is useless.

Security means the resistance against malicious, intentional modifications of the watermarked signal. Depending on the application scenario we can distinguish four types of attacks. Unauthorized detection allows the attacker to detect the existence of a watermark or extract the watermark information. Unauthorized removal tries to remove the embedded watermark information. Unauthorized embedding attempts to embed a watermark without the correct authorization. Finally a system attack tries to exploit weaknesses in the overall system [4].

One popular tool that combines simple signal distortions, like additive noise or low-pass filtering, with geometric attacks, like small rotations or translations, is Stirmark (http://www.cl.cam.ac.uk/~fapp2/watermarking/stirmark/ index.html) developed by Fabien Petitcolas [26,27]. Many early watermarking systems were broken by the Stirmark attacks, which led to the development of more advanced watermarking systems.

In previous work the following techniques to enhance the security of watermarks have been proposed. Pseudo-random skipping of coefficients has been proposed by Wang [34] or Kundur [19], but skipping significant coefficients reduces the capacity of the systems. Our approach can be combined with coefficient skipping, but can also improve the security without a reduction of the capacity. Fridrich [12] introduced the concept of key-dependent basis functions in order to protect a watermark from hostile attacks. By embedding the watermark information in a secret transform domain, Fridrich's algorithm can better withstand attacks such as those described by Kalker [16] employing a public watermark detector device. However, Fridrich's approach suffers from the computational complexity and the storage requirements for generating numerous orthogonal patterns of the size of the host image. In a later paper Fridrich reduced the computational complexity of the system [11].

In this paper we propose the use of parametrized wavelet filters as a method to protect wavelet-based watermarks against unauthorized detection and to increase the resilience against malicious removal attacks. We apply this approach to four wavelet-based watermarking algorithms, two non-blind algorithms and two blind algorithms: The algorithms by Kim [18] and Wang [34] both use spread-spectrum techniques and need the original image to extract the watermark. In contrast, the methods introduced by Kundur [19] and Xie [38] are quantization-based and only need the watermarked image to extract the watermark.

In section 2 we shortly review wavelet-based watermarking and describe the four watermarking algorithms which are subsequently enhanced by our approach. Section 3 introduces wavelet filter parametrization as security concept for wavelet-based watermarking and evaluates security against unauthorized detection attacks and quality/robustness issues. In section 4 we discuss our scheme's resilience against unauthorized removal attacks. Finally we discuss possible application scenarios for our approach in section 5 and conclude our paper in section 6.

2 Wavelet-based Watermarking

Most watermarking algorithms transform the host image into a domain that facilitates embedding of the watermark information in a robust and imperceptible way. Early approaches often employed the discrete cosine transform (DCT) to mark perceptually significant coefficients in the mid- and low-frequency spectrum [3].

In this paper, we focus on watermarking algorithms operating in the wavelet domain. We can distinguish between algorithms that use additive embedding of a spread-spectrum sequence or, on the other hand, follow a quantize-andreplace strategy. Either the approximation image or the detail subbands of the wavelet domain can be modified. For an extensive overview of wavelet-based watermarking see [22].

Hsu [13] states that the choice of the wavelet filter is a critical issue for the quality of the watermarked image and the robustness to compression attacks. However, the filter criteria for watermarking purposes might be different compared to image compression applications. Filters that pack most energy of the original image in the lowest resolution approximation image give best com-

pression performance because information in the detail subbands can be easily discarded without severe perceptible image distortion. However, watermarking applications using such filters to embed watermark information in the detail subbands will seriously suffer from compression attacks.

The choice of embedding domain versus attack domain (e.g. due to image compression) has been controversially discussed by Wolfgang [37], Kundur [20] and Fei [10]. However, no clear answer to the problem of choosing the embedding domain was found.

In the following we shortly describe the algorithms used in this present paper.

Spread-spectrum Watermarking. Based on the work of Cox [3] in the DCT domain, Kim [18] utilizes DWT coefficients of all subbands including the approximation image to equally embed a random Gaussian distributed watermark sequence in the whole image. Perceptually significant coefficients are selected by level-adaptive thresholding to achieve high robustness. However, the location of the watermark information is not protected and open for malicious attacks.

Following the design of his multi-threshold wavelet coding scheme, Wang [34] proposes a watermarking algorithm that refines Kim's thresholding scheme and selects significant coefficients on a per subband basis. Here, random skipping of significant coefficients is discussed as a mean to achieve non-invertibility [5] and improve watermark security although this will also limit the robustness and capacity of the scheme. Additionally, it is proposed to keep the wavelet transform structure and filters secret in order to protect the location of embedded watermark information.

Quantization-based Watermarking. A different approach to watermark embedding is taken by Xie [38] and Kundur [19]. Instead of adding a signal to selected coefficients, sets of coefficients are quantized to represent the binary watermark information. Embedded and extracted watermark sequences are compared by computing the Hamming distance between the two sequences normalized by the length of the sequence.

Xie's approach selects non-overlapping coefficient sets of size $3 \ge 1$ from the wavelet domain's approximation image. The median coefficient of each set is then quantized to embed one bit of watermark information.

The algorithm proposed by Kundur is similar but selects three coefficients from the detail subbands: one from the LH, HL and HH subband at the same spatial location and decomposition level. In Kundur's scheme, the much higher



Figure 1. Overview of watermark embedding procedure using parametrized wavelet filters.

number of embedding locations (three detail subbands versus one approximation subband of the same size) are partially used for a reference watermark. The reference watermark helps to weight redundant watermark bits, and hence improves detection accuracy.

3 Protection of Wavelet-based Watermarking Systems

In this section, we present a solution to construct a key-dependent transform domain using wavelet filter parametrization. Based on two or more parameters we can create a whole family of wavelet filters that we can use as keyspace. The basic idea of this technique has been presented in our earlier papers [23,7].

Classically, the security of many watermarking schemes depends on using a pseudorandom sequence to prevent unauthorized detection and removal in spread-spectrum systems. For quantization-based as well as for some spread-spectrum systems, random sequences play an important role where they are used for selecting the embedding locations. In addition to employing pseudo-random sequences to secure the watermarking systems our approach operates in a key-dependent wavelet domain.

Other approaches that aim to improve security by employing a key-dependent wavelet transform domain include the work of Seo [29], Wang [35] and Suhail [30]. Seo proposes to weight intermediate coefficient in the wavelet lifting process according to a key parameter before a binary pseudo-random noise sequence is added as a watermark. Seo claims to reenforce security against alteration of the embedded watermark but the key-space seems to be limited looking at the data provided. The system is limited to lifting-based wavelet implementations.

Wang [35] proposes randomly generated orthonormal filter banks depending on the image as a major part of the private key. The goal is to prevent reverse engineering of a private key that can be used to 'detect' a watermark and claim ownership given an arbitrary unwatermarked image. The authors present many robustness results, but do not assess the security of their proposed system.

Suhail [30] uses filter parametrization for feature watermarking and proposes a concept based on our previous work [23] while providing less experimental results.

3.1 Wavelet Filter Parametrization

In order to construct compactly supported orthonormal wavelets, solutions for the dilation equation

$$\phi(t) = \sum_{k \in \mathbb{Z}} c_k \phi(2t - k),$$

with $c_k \in \mathbb{R}$, have to be derived, satisfying two conditions on the coefficients c_k [6]. Schneid [28] describes a parametrization for suitable coefficients c_k based on the work of Zou [39] to facilitate construction of such wavelets. Given Nparameter values $-\pi \leq \alpha_i < \pi, 0 \leq i < N$, the recursion

$$c_0^0 = \frac{1}{\sqrt{2}} \text{ and } c_1^0 = \frac{1}{\sqrt{2}}$$
$$c_k^n = \frac{1}{2}((c_{k-2}^{n-1} + c_k^{n-1}) \cdot (1 + \cos \alpha_{n-1}) + (c_{2(n+1)-k-1}^{n-1} - c_{2(n+1)-k-3}^{n-1})(-1)^k \sin \alpha_{n-1})$$

can be used to determine the filter coefficients c_k^N , $0 \le k < 2N + 2$. We set $c_k = 0$ for k < 0 and $k \ge 2N + 2$.

We propose to decompose the host image using wavelet filters constructed with the above parametrization. The parameter values α_i used for construction and the resulting wavelet filter coefficients are kept secret. Hence, the watermark information can be embedded in a secret multi-resolution transform domain, making it difficult to mount a hostile attack that seeks to destroy or remove watermark information at specific locations. Our concept is illustrated in figure 1. Figure 2 compares a standard Daubechies 6 filter with a parametrized filter with N = 2 that was generated using $\alpha_0 = -0.4815$ and $\alpha_1 = 2.6585$, resulting in a 6-tap filter.

Our approach to generating key-dependent wavelet filters is, in principle, applicable to all wavelet-based watermarking systems and can also be integrated



Figure 2. Comparison of standard Daubechies 6 and a parametrized wavelet filter using $\alpha_0 = -0.4815$ and $\alpha_1 = 2.6585$.

with the JPEG2000, Part 2, standard for image compression.

3.2 Baseline System: Two Filter Parameters

For this first system we use two filter parameters and analyze the security against unauthorized detection and the behavior under compression with JPEG and JPEG2000. Using two filter parameters results in 6-tap filters. We present the results for all four watermarking algorithms described in section 2.

3.2.1 Protection Against Unauthorized Detection

The goal of this assessment is to analyze the effectiveness of the filter parameters to protect against unauthorized detection and to determine the smallest difference in parameter values that still allows a clear distinction between correct and incorrect parameter values.

The filter parametrization is the value that we use as key for the embedding process. Other possible keys, like the seed for the spread spectrum sequence or a seed for significant coefficient skipping, are constant for the following experiments.

We embed one specific watermark with one filter parametrization and then try to detect the watermark with other filter parameters. Our goal in introducing the parametrized wavelet filters is to increase the security of the watermarking systems. Therefore the correlation between the embedded watermark and the watermark extracted with the wrong filter parametrization should be very low.

The ideal system would have a normalized correlation of zero everywhere except for the key location where the correlation would be 1.0. Because of

rounding errors and random similarities that perfect system is not possible. In reality you have to decide on a threshold level and say that the watermark is embedded, if the correlation is above the threshold. The selection of the threshold depends on the watermarking system, expected attacks and the desired false-positive and false-negative rates.

If there is high correlation even when the filter parameters are set to wrong values close to the correct key values, then the security improvement is not as high as expected, because even if you are only in the proximity of the correct key you already know that the watermark is embedded.

We look for the smallest difference between parameter values that is necessary to have a clear separation in the correlation values of the correct filter parameters and incorrect ones. If we choose this difference too small, then we might get high correlation even though the wrong filter parameters are used. If we choose it too large, then the number of possible filter parameters is small and therefore limits the key space of the system.

In the following experiments we embed the watermark with the parameters $\alpha_0 = 0.150$ and $\alpha_1 = 0.650$ and select an embedding strength that results in 40dB PSNR. We use the well known "Lena" image as cover data. Then we use different parameter ranges to vary the filter parameters and measure the resulting correlation between the embedded watermark and the watermark that is extracted with this filter.

Kim Algorithm. The results of the first experiment with the Kim algorithm are shown in figure 3.

Here we select $\alpha_0, \alpha_1 \in \{-3.14, -3.13, \ldots, 3.13, 3.14\}$ with $\Delta = 0.01$. This results in 395641 samples, i.e. potentially different filter parameters. There is only one peak with a correlation of 0.90 at the location of the embedding parameters. For the other samples the correlation is always smaller than 0.25, most of the time being below 0.10.

This result is very promising. We have high correlation only with the parameters used for embedding the watermark and the correlation is low for all other filter parameters.

Because the only increase in correlation is around the embedding parameters we will confine our further examinations to this area.

Figure 4(a) shows the correlation when $\alpha_0 \in \{0.000, \ldots, 0.300\}$ and $\alpha_1 \in \{0.500, \ldots, 0.800\}$ with $\Delta = 0.002$ resulting in 22801 measurements. Applied to the whole parameter space using $\Delta = 0.002$ would result in 9865881 possible filter keys. There are two peaks in the correlation — at the embedding position and in very close proximity.



Figure 3. Kim algorithm. $\alpha_0, \alpha_1 \in \{-3.14, -3.13, \dots, 3.13, 3.14\}$ with $\Delta = 0.01$.

To better visualize this result we use the parameter values α_0 and α_1 as Xand Y-axis and map the correlation to a shade of gray. A correlation of zero corresponds to black and a correlation of one to white. In figure 4(b) you can see one white dot at the center of the image.

This picture helps us see that there is a higher correlation around the Y-axis for $\alpha_1 = 0.650$. To further analyze this we look at the sensitivity of one parameter if the other parameter is already set to the right value. Diagram 5(a) shows the correlation when $\alpha_1 = 0.650$ and $\alpha_0 \in \{0.000, \ldots, 0.300\}$; diagram 5(b) sets α_0 to 0.150 and varies $\alpha_1 \in \{0.500, \ldots, 0.800\}$; both diagrams use $\Delta = 0.0005$ which results in 601 measurements each, which would be equivalent to 157778721 different filter keys.

At this fine resolution we see that there are several peaks close to the embedding value. Diagram 5(a) also has a broader region with high correlation than diagram 5(b). This results in the vertical "stripe" in diagram 4(b). With the step size Δ at around 0.0005 we are approaching the finest meaningful resolution, further refinement would lead to a large range of high correlation around the true embedding key.

Wang Algorithm. Figure 6 shows the corresponding results for the Wang algorithm.

Diagrams 6(a) and (b) show the correlation when $\alpha_0 \in \{0.000, \ldots, 0.300\}$ and $\alpha_1 \in \{0.500, \ldots, 0.800\}$ with $\Delta = 0.005$ resulting in 3721 measurements. At this resolution we have 1580049 possible filter keys overall. The correlation has only one clear maximum at the embedding values. The results for the variation of one parameter alone look similar to the results for the Kim algorithm and are not shown again.



Figure 4. Kim algorithm. $\alpha_0 \in \{0.000, \dots, 0.300\}$ and $\alpha_1 \in \{0.500, \dots, 0.800\}$ with $\Delta = 0.002$.



Figure 5. Kim algorithm. Variation of only one parameter. Both diagrams use $\Delta = 0.0005$. (a) $\alpha_1 = 0.650$ and $\alpha_0 \in \{0.000, \dots, 0.300\}$. (b) $\alpha_0 = 0.150$ and $\alpha_1 \in \{0.500, \dots, 0.800\}$.

A good choice for Δ therefore is in the range of 0.005 and 0.01 which results in a large number of possible keys and either one clear peak or a very small area of high correlation.

Kundur Algorithm. Figure 7 shows the results for the first quantizationbased algorithm. We repeat our experiment and embed a watermark with 40dB using the parameters $\alpha_0 = 0.150$ and $\alpha_1 = 0.650$ for wavelet filter construction.

For watermark detection we 'guess' $\alpha_0, \alpha_1 \in \{-3.14, -3.13, \ldots, 3.13, 3.14\}$ with $\Delta = 0.01$ and plot the correlation in diagram 7(a). Unfortunately, high



Figure 6. Wang algorithm. $\alpha_0 \in \{0.000, \dots, 0.300\}$ and $\alpha_1 \in \{0.500, \dots, 0.800\}$ with $\Delta = 0.005$.



Figure 7. Kundur algorithm. The watermark was embedded with a PSNR of 40dB at $\alpha_0 = 0.150$ and $\alpha_1 = 0.650$. (a) $\alpha_0, \alpha_1 \in \{-3.14, -3.13, ..., 3.13, 3.14\}$ with $\Delta = 0.01$. (b) $\alpha_0 \in \{0.000, ..., 0.300\}$ and $\alpha_1 \in \{0.500, ..., 0.800\}$ with $\Delta = 0.002$.

correlation is achieved for a wide range of watermark extraction parameters. In diagram 7(b) we zoom in on the region where the watermark was embedded and plot the correlation for $\alpha_0 \in \{0.000, \ldots, 0.300\}$ and $\alpha_1 \in \{0.500, \ldots, 0.800\}$ with $\Delta = 0.002$. As can be seen, correlation is still highest closely to the point where the watermark was embedded, however, there is no sharp peak anymore. This clearly reduces the obtainable keyspace.

Xie Algorithm. Next we look at the results that can be obtained with Xie's watermarking algorithm when repeating the experiment. Diagram 8(a) shows results similar to our other previous quantization-based algorithm. In diagram 8(b) we can see a wild fluctuation between points of high and low correlation in



Figure 8. Xie algorithm. The watermark was embedded with a PSNR of 40dB at $\alpha_0 = 0.150$ and $\alpha_1 = 0.650$. (a) $\alpha_0, \alpha_1 \in \{-3.14, -3.13, ..., 3.13, 3.14\}$ with $\Delta = 0.01$. (b) $\alpha_0 \in \{0.000, ..., 0.300\}$ and $\alpha_1 \in \{0.500, ..., 0.800\}$ with $\Delta = 0.002$



Figure 9. Xie's algorithm. Variation of only one parameter. Both diagrams use $\Delta = 0.0005$. (a) $\alpha_1 = 0.650$ and $\alpha_0 \in \{0.000, \ldots, 0.300\}$. (b) $\alpha_0 = 0.150$ and $\alpha_1 \in \{0.500, \ldots, 0.800\}$.

the region where the watermark has been embedded. When taking a close look by plotting the correlation due to the two parameters α_0 and α_1 separately (holding one parameter constant), compare with diagram 9(a) and (b), we see that varying only parameter α_1 leads to rather low correlation results while varying only the other parameter, α_0 , the correlation stays high.

A very plausible explanation of the results is that the observation is due to the non-linear loss of fine detail in the coefficient values caused by the quantization used by Xie and Kundur. The different coefficient values from different filter parametrizations which are close to each other are lost due to the quantization process and are therefore mapped onto a single coefficient value.



Figure 10. Correlation and PSNR of Kim watermarks under JPEG and JPEG2000 compression.

We have to conclude that filter parametrization for quantization-based embedding does not yield security as high as for spread-spectrum based watermarking schemes.

3.2.2 Quality Assessment and Robustness

Next we explore the difference between the parametrized filters and the standard Daubechies-6 and Biorthogonal 7/9 filters with regard to correlation and PSNR under JPEG and JPEG2000 compression.

We watermark the "Lena" image with each parametrized filter with $\alpha_0, \alpha_1 \in \{-3.00, \ldots, 3.00\}$ and $\Delta = 0.20$. This generates 961 different wavelet filters. The watermark is embedded with an embedding strength that results in a 40dB PSNR. Each picture is JPEG and JPEG2000 compressed with different quality levels. First we measure the PSNR of the compressed image. Then



Figure 11. Correlation and PSNR of Wang watermarks under JPEG and JPEG2000 compression.

we try to extract the embedded watermark from the compressed image and measure the normalized correlation between the embedded watermark and the extracted watermark.

Kim Algorithm. Figure 10 shows the results for the Kim algorithm. Diagram (a) shows the correlation after JPEG2000 compression of the watermarked image, diagram (b) the PSNR after JPEG2000 compression. Diagram (c) shows the correlation, diagram (d) the PSNR after JPEG compression.

The diagrams contain the average of all measured parametrized filters and the minimum and maximum for each compression rate. The average parametrized filters PSNR values are slightly above the standard filters and behave very similarly to them. The influence of parametrization on image quality is therefore minimal. The correlation values vary in a wider range. The average parametrized correlation is close to that of the standard filters.



Figure 12. Correlation and PSNR of Kundur's watermarking algorithm under JPEG and JPEG2000 compression.

Wang Algorithm. Diagram 11 contains the results for the Wang algorithm. Again we see that the average correlation for the parametrized system is very close to the two standard filters. The Wang system is more robust to compression than the Kim system, which can be seen in the higher correlation values after stronger compression. The distance from the average to the minimum parametrized system is larger for the Wang system. But the average behavior is closer to the maximum, indicating that on average the system produces good results.

The PSNR behavior is again a little bit above the two standard systems on average.

Kundur Algorithm. Figure 12 provides the robustness and quality assessment results for Kundur's algorithm. Robustness against JPEG2000 and JPEG compression in terms of watermark correlation is shown in diagram (a)



Figure 13. Correlation and PSNR of Xie's watermarking algorithm under JPEG and JPEG2000 compression.

and (c). The averaged correlation for our parametrized filters is below that of standard wavelet filters (Daubechies 6 and Biorthogonal 7/9) commonly used for image compression.

From diagram (b) and (d) we can observe that image quality remains about the same in terms of PSNR after JPEG2000 and JPEG, regardless if we are using parametrized or standard filters.

Xie Algorithm. Robustness results for Xie's algorithm using parametrized filters are clearly inferior compared to using standard wavelet filters, see figure 13, diagram (a) and (c). The correlation varies widely from our best to worst parametrized filter, making it impossible to predict the achievable robustness. Average image quality in term of PSNR is en par with standard filters, but still there is a variation of up to 4dB as illustrated in diagrams (b) and (d). For quantization-based schemes, the embedding strength can not be easily

adjusted to obtain the desired embedding distortion of 40dB PSNR. Therefore, the PSNR varies in a wider range.

From the security and quality assessment we conclude that parametrized wavelet filters are not suitable for quantization-based watermarking algorithms.

3.3 Multi-Level System

In order to increase the keyspace to a cryptographically reasonable size we have considered two possibilities. Increasing the number of parameters used for filter parametrization and the use of different filters on different decomposition levels.

In previous work [7] we have analyzed the behavior of non-stationary multiresolution analysis [2,32] using different parametrized filters on different decomposition levels of the wavelet transform. We have seen that for the Kim algorithm the security does not improve as much as expected.

Now we present a system that uses the Wang algorithm. We use filters that are generated by five parameters and use four different filters for the first four decomposition levels. This results in a combined system with 20 parameters as embedding key.

3.3.1 Protection Against Unauthorized Detection

For 20 parameters examining every parameter variation is not possible. Therefore we decided to look at the system and try to "attack" it. The attacker knows the basic design of the system, but does not know the key that was used for embedding. He in turn looks at the different decomposition levels and tries to independently guess the value of the 5 parameters used for that level. The parameters for the other levels are set to zero.

Parameter	1	2	3	4	5
Level 1	-0.5	2.5	-1.0	1.5	0.5
Level 2	-2.5	2.0	-2.0	0.5	1.0
Level 3	2.0	-1.5	0.5	2.5	-2.0
Level $4 + higher$	-1.0	-2.0	1.0	-0.5	2.5

The watermark was embedded using the parameter values:

In the following tests we vary all five parameters for each level at the same time. We take the starting value 0.6 below the correct value for each of the five parameters. Then we increment the parameters by 0.2 until all parameters are 0.6 above the correct value. We therefore have $7^5 = 16807$ measurements for every level.

In diagram 14(a) we try to attack the first level. We vary the five parameters for the first decomposition level and keep the other parameters set to zero. From the correlation response to the different parametrizations there is no way for the attacker to guess the correct values.

If for some reason the attacker knows the correct parameter values for the first decomposition level, then he only needs to search for the remaining 15 parameters. Again we focus on the next set of five parameters used for the second decomposition level. The first decomposition is performed with the correct filter and the third and higher filter parametrizations are unknown to the attacker. In diagram 14(b) we see that the attacker did not get any additional knowledge from knowing the first decomposition filter. The correlation is low over the complete range of guessed parametrizations, although the correct parametrization for the second decomposition level was tested.

The same is true for the third level. If we already have the first and second decomposition level parameters and only need to find the third and fourth level, then we will again first try to find the five parameters for the third decomposition. Diagram 14(c) shows the correlation when the first two levels are set to the correct keys, the fourth level is set to zero and the five parameters for the third level are varied over a set of parametrizations that contain the correct parameter values. Again there is no sign which of the tested parametrizations is the correct one and the attacker has no way of knowing that he has tested the correct parameters for the third level.

Only in case the attacker already knows the decomposition parameters for the first three levels is he able to determine the correct parameters for the fourth level. In diagram 14(d) the attacker already knows the first 15 keys and only varies the last five parameters for the fourth decomposition level. There is a clear peak at the location of the correct embedding parametrization and low correlation everywhere else.

So only after having all 20 parameters right does the attacker get a high correlation.

Next we look at the sensitivity of the different levels to parameter changes. We set all levels to the correct embedding parameters and only vary the five parameters for one level and measure the correlation. In diagram 15(a) we vary the parameters for the first level and have the correct values for the other three levels. There is one peak at the embedding position and low variation



Figure 14. Attacks on all four levels.

everywhere else.

For levels two and three we see the respective results in diagrams 15(b) and 15(c). For the fourth level this attack is the same as the previous attack on the fourth level. The results are shown again in diagram 15(d).

We see that the first level has higher correlation for wrong parameter values. Overall the behavior of the system is very good and there is always one clear peak.

To give a clear view of the security advantage of using the Wang embedding scheme we show the attacks on the first and second levels for the Kim method in diagrams 16(a) and 16(b).

For diagram 16(a) we varied the five parameters for the first levels and set all remaining 15 parameters to zero. There is one clear peak with a correlation of around 0.30. So although we do not get a 1.00 correlation by only guessing the first level we still see a significantly higher correlation for the correct



Figure 15. Variations of all levels.



Figure 16. Kim — Attacks on first and second level.



Figure 17. Correlation and PSNR of multi-level system under JPEG2000 compression.

parameters for this level.

Now by using the correct filters for the first level, setting the third and higher levels to zero and only varying the second level parameters we get diagram 16(b). The first thing to notice is that the correlation is above 0.20 for all tested parametrizations. But again there is a significantly higher correlation for the correct filter parametrization. We see the same behavior for the third and fourth levels, only that the overall correlation gets higher and higher.

We see from this last experiment that only by using the Wang embedding algorithm we get the real security of all 20 parameters.

Using 20 filter parameters we get a vast key space. If we use a resolution of 0.20 we have around $(2 * \pi/0.20)^{20} \approx 2^{99}$ possible filter parametrizations. For a finer resolution of 0.01 we get $(2 * \pi/0.01)^{20} \approx 2^{185}$ filters.

If we choose a parameter resolution between 0.20 and 0.01 we have a very large key space and have very good separation between the correct key and incorrect embedding parametrizations.

3.3.2 Quality Assessment and Robustness

In this investigation we look at the correlation and PSNR behavior of the combined system under JPEG and JPEG2000 compression. We create 768 different filter parametrizations by randomly choosing values for the 20 parameters between -3.14 and 3.14. For each parametrization we embed a given watermark with an embedding strength that results in 40dB PSNR into the Lena image. Then we compress the watermarked image with JPEG and JPEG2000 at different compression rates. We try to detect the watermark in the com-



Figure 18. Correlation and PSNR of multi-level system under JPEG compression.

pressed images and measure the correlation between the extracted watermark and the embedded one. Also the PSNR is measured to determine how strongly distorted the compressed image is.

We calculate the average, minimum and maximum values from all 768 different parametrizations and compare the results for the combined system to the results for the two standard systems using the Daubechies 6 and the Biorthogonal 7/9 filters.

Diagram 17(a) shows the correlation behavior under JPEG2000 compression, diagram (b) shows the PSNR. The corresponding values for JPEG compression are shown in diagram 18(a) and (b). The average behavior of the combined system is very close to the two standard systems.

4 Protection against Unauthorized Removal Attacks

In this section we discuss the resistance of our approach to an unauthorized removal attack. For this purpose a malicious attack is pursued against an embedded watermark.

We embed the watermark with the standard Wang method using standard filters and with the proposed parametrized wavelet filters. The attacker only knows that we use the Wang method to select the coefficients, but does not know which filters we use.

For the attack we only use the watermarked image. By applying the wavelet coefficient selection on the already watermarked image we are likely to select different coefficients from the ones that were used for embedding. The attack



Figure 19. Fixed quantization of coefficients after Biorthogonal 7/9 decomposition.

therefore will have to modify more coefficients and hope that the correct coefficients are attacked. On the other hand the quality of the image must not be degraded severely, otherwise making the image useless.

We embed a watermark of length 1000 and use a 7 level decomposition. The "Lena" image is used and all watermarks are embedded with a strength that results in 40dB PSNR. As reference we use the standard Wang algorithm with the Daubechies 6 and the Biorthogonal 7/9 filters.

For the parametrized filters we use 2, 3 and 5 parameters to generate the wavelet filters. We also test the multi-level system with a total of 20 parameters discussed in section 3.3. For each system we select 40 parameter combinations at random and embed the watermark with each combination. After the attack we calculate the average, minimum and maximum remaining correlation of the different parametrizations.

For all attacks we use a 7 level pyramidal decomposition with either the Daubechies 6 or the Biorthogonal 7/9 filters. Between 100 and 20000 coefficients are attacked. We developed several different attacks on the coefficient values, but only present results of one attack here. The other attacks did either impact the image quality too severely or did not result in a clear difference between the systems.

This attack applies a fixed quantization with a step size of 100 to the selected coefficients. We tested step sizes of 1, 10, 20, 50 and 100. A step size of 100 is effective at removing the watermark information and still preserving the image quality.

Figure 19 compares the correlation and PSNR of the different systems when we use the Biorthogonal 7/9 filter for the attack. In (a) we see that the standard system that also uses the Biorthogonal 7/9 filter for embedding has a



Figure 20. Fixed quantization of coefficients after Daubechies 6 decomposition.



Figure 21. Fixed quantization of coefficients: details of Multi-Level 20 system.

correlation of around 0.40 after 3000 coefficients have been attacked. After attacking 3000 coefficients no further correlation decrease can be seen because almost all significant coefficients carrying watermark information have been quantized. The parametrized systems all perform more than 0.30 better because the embedding domain is different from the attack domain.

The average multi-level system performs best with a correlation of nearly 90 percent even after 20000 coefficients have been attacked. The advantage of this attack is that the PSNR is not degraded badly. After attacking 20000 coefficients we still have a PSNR above 33dB. But because we only need to attack 2000 or 3000 coefficients to significantly reduce the correlation for the standard system we can have an attacked image quality of 36 or 35dB respectively.

Figure 20 shows the results when the Daubechies 6 filter is used for the at-

tack decomposition. The results for the parametrized systems are nearly unchanged. The average, minimum and maximum of the Multi-Level 20 system are shown in figure 21 for both the Biorthogonal 7/9 and the Daubechies 6 filters.

Considering the impact of this attack we see a clear advantage of our proposed system. With the standard Biorthogonal 7/9 and Daubechies 6 embedding methods the correlation drops below 40 percent while the quality of the attacked image is still reasonable. This means that an un-watermarked image can be obtained without a severe quality reduction.

Using secret parametrized filters we maintain significantly higher correlation under attack and could still proof ownership of the attacked images. In order to remove the watermark we need to reduce the image quality below an acceptable level. This clearly demonstrates that embedding the watermark in a key-dependent wavelet transform domain provides superior resilience against unauthorized removal attacks.

5 Application Scenarios

When investigating application scenarios, we want to determine which aims we might achieve by using parametrized wavelet filterbanks or secret transform domains in general. On the one hand, we clearly target onto enhancing security, on the other hand we want to analyze whether other additional functional features may be identified.

When discussing security issues in the sense of attack resistance, we distinguish between two types of hostile attacks: first, "custom" attacks, which take advantage from the exact knowledge about the embedding procedure and are therefore tailored against a specific watermarking system, and second, "general purpose attacks", which try to be efficient against as many systems as possible (e.g., Stirmark and Checkmark [24]). If the embedding domain is not known, mounting a custom attack might be more difficult. In particular, it is not possible to determine exactly which coefficients have been selected to embed the watermark even if the embedding procedure is publically known. Therefore, a direct manipulation of those coefficients to destroy the mark is not possible in an efficient manner (see section 4). On the other hand, the efficiency of general purpose attacks is not reduced by the parametrization (see the results on robustness against JPEG and JPEG2000 compression in section 3). However, there is an advantage even in this case. It is not possible for the attacker to verify the success of his attack and therefore the attack parameters can not be adjusted and optimized iteratively based on the success of previous attacks (compare also the results in [12,16]).

In the context of identifying supplementary features we need to discuss in which cases we apply identical parametrizations and in which we apply different ones. In particular, it is possible to employ different parametrizations for each individual media object, for each individual user, or for both. The first case covers the classical copyright embedding application where for example each image file in a database is marked using a different filter parametrization. Note that all experimental evaluations and results in this paper cover this case only. Apart from being more secure against removal attacks, are there additional benefits? In general, the watermark can only be extracted if the parametrization is known. Therefore, a user can not verify if a retrieved image contains copyright information at all if the copyright holder does not want him to. As a consequence, a user employing material without having properly acquired it can never be sure about the possibility of secretly embedded information. As a second benefit, a company can protect its customer database. In case all media objects of a media distributor are marked identically, a competitor only needs to perform a search across the web to reveal the identity of a good share of the first company's customers. If secret parametrizations have been used, this is not possible. For proofing rightful ownership in a court of law it might be necessary to reveal the spread spectrum sequence that was used to watermark an image. If all media objects are signed with the same sequence this would allow attackers to find other images that belong to this person. By using different filter parametrizations for different images we can reveal the spread spectrum sequence and filter parametrization that is used to watermark one image without allowing an attacker to find other images that are signed with the same sequence, but another filter parametrization.

As a third advantage, only the copyright owner is able to trace his copyrighted material — others can not detect illegitimate use of copyrighted material and try to blackmail companies or individuals doing so. As a final advantage, illegitimate use of a specific media object may be traced by a corresponding search through the web using the one and only correct parametrization for that object.

In case we apply different parametrizations for different users this may be interpreted as a case of fingerprinting. Note that fingerprinting can also be based on the use of different seed numbers. Also, our approach may be combined with this technique thereby increasing its security significantly. Based on the knowledge about the correct parametrization, the purchaser of a specific copy of a media object may be identified. As an example, we consider a web-shop selling stock photography. In the case of direct fingerprinting, the specific wavelet filter parametrization used for embedding is computed by applying e.g. a MAC onto the customer ID. The advantage is a personalization of the media, i.e. the identity of the user is inherently embedded. The disadvantage is that watermark embedding has to be performed on-line causing a high load on the server system of the web-shop. In the case of indirect fingerprinting, the filter parametrization is associated with the user ID via an entry in a database. The obvious disadvantage is that the owner can only be identified with help of the database, on the other hand pre-embedding is possible in this case. However, we suspect that the collision resistance is not sufficient if filter parametrization is employed as the only means of fingerprinting (without having empirical evidence for that).

Using different parametrizations for both, different media objects and different users, may have additional advantages from a functional point of view. First, if a user is not trustworthy, he may try to attack the secret filter parametrizations by using the fact that all media objects sold to him have been marked using the same embedding parametrization. Using additionally different parametrizations for different media objects can avoid this attack. Second, if the media distributor suspects that a specific customer has used his copy of the media object in question in a non-proper way, again a search through the web for that specific media object sold to that specific customer may give evidence for that.

Entirely different applications are the embedding of several different watermarks into one media object using different parametrizations (where additional security elements like different coefficient selection etc. need to be performed) and the embedding of reference watermarks.

Reference or template watermarks were introduced to estimate the channel distortion due to attacks or processing. They do not add information hiding capacity but help to estimate global geometric distortion [25] or local manipulation [19].

In previous schemes such as [25], the template watermark equally spaced peaks in the DFT spectrum of the image. The pattern of peaks can be used to estimate and revert the global geometric transformations. Even if the actual watermark is protected by a key and cannot efficiently be tampered with, the template watermark is publically accessible since the transform domain, in the case of our example the DFT, is know. Therefore, a desynchronization attack can be mounted as shown in [33] to render the template watermark and thus also the capacity watermark undetectable. Using a key-dependent transform domain, the template watermark can be protected.

Finally, we would like to point out two entirely different possible applications of parametric wavelet filters in the area of multimedia security. First, this technique can be potentially used for data hiding by embedding the data to be secretly transmitted in a secret transform domain. Second, a very fast partial encryption scheme might be developed by encoding visual data using secret parametric wavelet filters. In this scenario, only the parameters (corresponding to header data) need to be encrypted whereas the entire visual data may be stored or transmitted without being further protected.

6 Conclusion

In this work we have proposed to use wavelet filter parametrization as a secret transform domain in order to enhance wavelet based watermarking schemes from a security viewpoint. Experiments reveal that spread-spectrum based watermarking schemes can be made resistant to unauthorized detection and unauthorized removal attacks by our approach. A combination of filter parametrization and non-stationary wavelet decomposition achieves a keyspace of cryptographically reasonable size. Also, robustness against JPEG and JPEG2000 compression is on an equal level as compared to the use of standard wavelet filters. On the other hand, quantization-based watermarking techniques are not suited to be combined with filter parametrization since the quantization process destroys differences among different transform domains during watermark embedding. Whereas the possible impact of our approach on the purely copyright embedding oriented application has been shown in this work, we will focus on possible applications like fingerprinting, embedding of multiple watermarks, and embedding of reference watermarks in future work.

References

- C.K. Chui, L. Montefusco, and L. Puccio. Wavelets: Theory, Algorithms and Applications. Academic Press, San Diego, 1994.
- [2] A. Cohen. Non-stationary multiscale analysis. In [1], pages 3–12. Academic Press, 1994.
- [3] Ingemar J. Cox, Joe Kilian, Tom Leighton, and Talal G. Shamoon. Secure spread spectrum watermarking for multimedia. In *Proceedings of the IEEE International Conference on Image Processing, ICIP '97*, volume 6, pages 1673– 1687, Santa Barbara, California, USA, October 1997.
- [4] Ingemar J. Cox, Matthew L. Miller, and Jeffrey A. Bloom. *Digital Watermarking*. Morgan Kaufmann, 2002.
- [5] Scott A. Craver, Nasir Memon, Boon-Lock Yeo, and Minerva M. Yeung. On the invertibility of invisible watermarking techniques. In *Proceedings of the IEEE International Conference on Image Processing, ICIP '97*, volume 1, page 540, Santa Barbara, California, USA, October 1997.
- [6] Ingrid Daubechies. Ten Lectures on Wavelets. Number 61 in CBMS-NSF Series in Applied Mathematics. SIAM Press, Philadelphia, PA, USA, 1992.

- [7] Werner Dietl, Peter Meerwald, and Andreas Uhl. Key-dependent pyramidal wavelet domains for secure watermark embedding. In Edward J. Delp and Ping Wah Wong, editors, *Proceedings of SPIE Volume 5020*, Santa Clara, CA, USA, January 2003.
- [8] Jana Dittmann, editor. Digitale Wasserzeichen: Grundlagen, Verfahren, Anwendungsgebiete. Springer Verlag, 2000.
- [9] J. J. Eggers and B. Girod. *Informed Watermarking*. Kluwer Academic Publishers, 2002.
- [10] C. Fei, Deepa Kundur, and R. Kwong. The choice of watermark domain in the presence of compression. In Proceedings of the IEEE International Conference on Information Technology: Coding and Computing, Special Session on Multimedia Security and Watermarking Applications, Las Vegas, NV, USA, April 2001.
- [11] Jiri Fridrich. Key-dependent random image transforms and their applications in image watermarking. In Proceedings of the 1999 International Conference on Imaging Science, Systems, and Technology, CISST '99, pages 237–243, Las Vegas, NV, USA, June 1999.
- [12] Jiri Fridrich, Arnold C. Baldoza, and Richard J. Simard. Robust digital watermarking based on key-dependent basis functions. In David Aucsmith, editor, *Information hiding: second international workshop*, volume 1525 of *Lecture notes in computer science*, pages 143–157, Portland, OR, USA, April 1998. Springer Verlag, Berlin, Germany.
- [13] Chiou-Ting Hsu and Ja-Ling Wu. Multiresolution watermarking for digital images. *IEEE Transactions on Circuits and Systems II*, 45(8):1097–1101, August 1998.
- [14] ISO/IEC JPEG committee. JPEG 2000 image coding system ISO/IEC 15444-1:2000, December 2000.
- [15] Neil F. Johnson, Zoran Duric, and Sushil Jajodia. Information Hiding: Steganography and Watermarking - Attacks and Countermeasures. Kluwer Academic Publishers, 2000.
- [16] Ton Kalker. A security risk for publicly available watermark detectors. In Proceedings of the 18th Symposium on Information Theory in the Benelux, Veldhoven, The Netherlands, 1998.
- [17] Stefan Katzenbeisser and Fabien A. P. Petitcolas. Information Hiding Techniques for Steganography and Digital Watermarking. Artech House, December 1999.
- [18] Jong Ryul Kim and Young Shik Moon. A robust wavelet-based digital watermark using level-adaptive thresholding. In *Proceedings of the 6th IEEE International Conference on Image Processing, ICIP '99*, page 202, Kobe, Japan, October 1999.

- [19] Deepa Kundur. Improved digital watermarking through diversity and attack characterization. In *Proceedings of the ACM Workshop on Multimedia Security* '99, pages 53–58, Orlando, FL, USA, October 1999.
- [20] Deepa Kundur and Dimitrios Hatzinakos. Mismatching perceptual models for effective watermarking in the presence of compression. In *Proceedings of the SPIE Conference on Multimedia Systems and Applications II*, volume 3845, Boston, MA, USA, September 1999.
- [21] S. Mallat. A wavelet tour of signal processing. Academic Press, 1997.
- [22] P. Meerwald and A. Uhl. A survey of wavelet-domain watermarking algorithms. In Ping Wah Wong and Edward J. Delp, editors, *Proceedings of SPIE, Electronic Imaging, Security and Watermarking of Multimedia Contents III*, volume 4314, San Jose, CA, USA, January 2001. SPIE.
- [23] P. Meerwald and A. Uhl. Watermark security via wavelet filter parametrization. In Proceedings of the IEEE International Conference on Image Processing (ICIP'01), volume 3, pages 1027–1030, Thessaloniki, Greece, October 2001. IEEE Signal Processing Society.
- [24] Peter Meerwald and Shelby Pereira. Attacks, applications and evaluation of known watermarking algorithms with Checkmark. In Ping Wah Wong and Edward J. Delp, editors, *Proceedings of SPIE, Electronic Imaging, Security and Watermarking of Multimedia Contents IV*, volume 4675, San Jose, CA, USA, January 2002. SPIE.
- [25] Shelby Pereira and Thierry Pun. An iterative template matching algorithm using the chirp-z transform for digital image watermarking. *Pattern Recognition*, 33(1):173–175, January 2000.
- [26] Fabien A. P. Petitcolas and Ross J. Anderson. Weaknesses of copyright marking systems. In Multimedia and Security Workshop at the 6th ACM International Multimedia Conference, pages 55–61, Bristol, England, 1998.
- [27] Fabien A. P. Petitcolas and Ross J. Anderson. Evaluation of copyright marking systems. In *Proceedings of IEEE International Conference on Multimedia Computing and Systems '99*, volume 1, pages 574–579, Florence, Italy, June 1999.
- [28] J. Schneid and S. Pittner. On the parametrization of the coefficients of dilation equations for compactly supported wavelets. *Computing*, 51:165–173, May 1993.
- [29] Yong-Seok Seo, Min-Su Kim, Ha-Joong Park, Ho-Youl Jung, Hyun-Yeol Chung, Young Huh, and Jae-Duck Lee. A secure watermarking for JPEG-2000. In Proceedings of the IEEE International Conference on Image Processing, ICIP '01, Thessaloniki, Greece, October 2001.
- [30] M. A. Suhail and M. M. Dawoud. Watermarking security enhancement using filter parametrization in feature domain. In *Proceedings of the 9th* ACM Multimedia 2001 Conference, pages 53–58, Ottawa, Ontario, Canada, September 2001.

- [31] D. Taubman and M.W. Marcellin. JPEG2000 Image Compression Fundamentals, Standards and Practice. Kluwer Academic Publishers, 2002.
- [32] A. Uhl. Image compression using non-stationary and inhomogeneous multiresolution analyses. *Image and Vision Computing*, 14(5):365–371, 1996.
- [33] Sviatoslav Voloshynovskiy, Alexander Herrigel, and Y. B. Rytsar. Watermark template attack. In *Proceedings of SPIE, Security and Watermarking of Multimedia Contents III*, volume 4314, San Jose, CA, USA, January 2001.
- [34] Houng-Jyh Wang and C.-C. Jay Kuo. Watermark design for embedded wavelet image codec. In *Proceedings of the SPIE's 43rd Annual Meeting, Applications* of Digital Image Processing, volume 3460, pages 388–398, San Diego, CA, USA, July 1998.
- [35] Y. Wang, J. F. Doherty, and R. E. Van Dyck. A wavelet-based watermarking algorithm for copyright protection of digital images. *IEEE Transactions on Image Processing*, 11(2):77–88, February 2002.
- [36] M.V. Wickerhauser. Adapted wavelet analysis from theory to software. A.K. Peters, Wellesley, Mass., 1994.
- [37] Raymond B. Wolfgang, Christine I. Podilchuk, and Edward J. Delp. The effect of matching watermark and compression transforms in compressed color images. In Proceedings of the IEEE International Conference on Image Processing, ICIP '98, Chicago, IL, USA, October 1998.
- [38] Liehua Xie and Gonzalo R. Arce. Joint wavelet compression and authentication watermarking. In Proceedings of the IEEE International Conference on Image Processing, ICIP '98, Chicago, IL, USA, October 1998.
- [39] H. Zou and Ahmed H. Tewfik. Parametrization of compactly supported orthonormal wavelets. *IEEE Transactions on Signal Processing*, 41(3):1423– 1431, March 1993.