# Targeted Attacks on Quantization-based Watermarking Schemes

Peter Meerwald, Christian Koidl, Andreas Uhl

Department of Computer Sciences, University of Salzburg,
Jakob-Haring-Str. 2, A-5020 Salzburg, Austria
E-mail: {pmeerw,ckoidl,uhl}@cosy.sbg.ac.at

## Abstract

*While many watermarking methods show good robustness against common signal processing operations, security of the watermarking schemes under intentional attack exploiting knowledge of the implementation has been widely neglected. In this paper, we demonstrate straightforward, targeted attacks for a number of quantization based watermarking methods and provide implementations. The attacks require only one watermarked image and retain the fidelity of the image. The watermarking methods discussed are therefore not suitable for copyright protection applications.*

## 1 Introduction

Copyright protection is an important watermarking application where information identifying the copyright owner is imperceptibly embedded in multimedia data such that this watermark information is detectable even in degraded copies. Quantization-based watermarking is an attractive choice as it combines high watermark capacity with robustness against manipulation of the cover data. The ability to embed many watermark bits (in the range of $256$ to $1024$ bits) allows to hide a small black-and-white logo image. An extracted logo image can be used to visually judge the existence of a particular watermark. Alternatively, the normalized correlation measure between the embedded and extracted watermark provides for numerical evaluation.

Many watermarking schemes demonstrate good robustness for a wide variety of signal processing attacks such as JPEG compression, median filtering, sharpening and mild rotation. However, in the copyright protection scenario, a watermarking method must not only withstand unintentional processing of the cover data but also intentional, targeted attack by a malicious adversary [4]. For the attack scenario in this paper, we assume that we have access to only a single watermarked image but possess full knowledge of the implementation details of the watermarking scheme. According to the classification suggested by Cayre et al. [1], this constitutes a watermark-only-attack (WOA). Following Kerckhoffs' principle [7], a watermarking system should be 'secure' even if everything except the key is known. Watermark 'security' versus robustness is a controversial topic. Kalker [6] states that 'security refers to the inability by unauthorized users to have access to the raw watermarking channel'.

While general signal processing, geometric and protocol level attacks [3, 11, 15] have received ample attention in the literature, only few works investigate targeted attack directed towards the weakness of a particular watermarking algorithm. The attacks mounted on the proposed scheme during the 'Break Our Watermarking System' (BOWS) contest [13] expose vulnerabilities and indicate design guidelines for robustness and security to be incorporated in new watermarking schemes. It is thus worthwhile to consider attacking a particular watermarking method. Benchmarking may provide a robustness evaluation [12], however in the copyright protection scenario a detailed analysis for potential weaknesses is required.

In Section 2 we describe attacks on six quantization based watermarking schemes in the wavelet domain [2, 8, 9, 14, 16, 17]. We review the security techniques employed and suggest modifications to the watermarking methods in Section 3. In Section 4 we discuss the experimental attack results before we conclude the paper with remarks in Section 5.

## 2 Targeted Attacks

In the following we outline the principles of six quantization-based watermarking methods in order to motivated the attacks and discuss the security weaknesses. Due to lack of space we cannot describe these watermarking systems in detail but instead make our implementations and the corresponding attack code publicly available (see Section 4). Refer to the original papers for details.

**Quantization of Middle Wavelet Detail Coefficients (QMWDC)** is one of the first quantization-based watermarking schemes proposed by Kundur et al. [8] which embeds a binary watermark in wavelet-domain detail subband coefficients. A secret key $K$ selects the embedding positions where for each location the wavelet image components with horizontal, vertical and diagonal orientation are sorted according to their magnitude. The

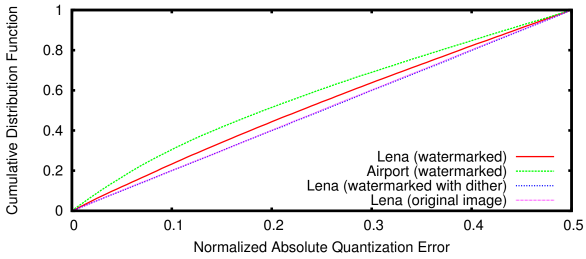


**Figure 1. Normalized absolute quantization error for original and watermarked images**

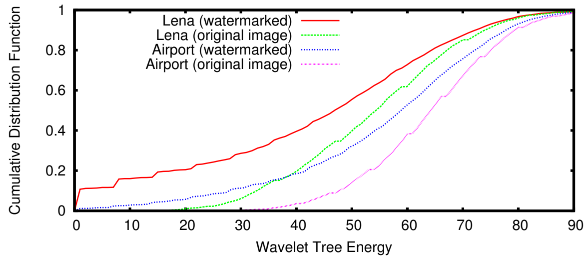


**Figure 2. Wavelet tree energy for original and watermarked images**

middle coefficient $x_d^m[i,j]$ at location $i,j$ and decomposition level $d$ is quantized to fall between the smallest and largest coefficient of the triple, denoted by $x_d^s[i,j]$ and $x_d^l[i,j]$, resp., and encodes one bit of watermark information. The watermark is embedded repeatedly to improve robustness. The absolute quantization error $e_d[i,j] = |\text{round}(x_d^m[i,j]/\Delta_d[i,j]) - x_d^m[i,j]/\Delta_d[i,j]|$ normalized by the corresponding quantization bin width $\Delta_d[i,j] = \frac{x_d^l[i,j]-x_d^s[i,j]}{2Q-1}$ is uniformly distributed for the original image but shows a bias towards smaller quantization errors for the watermarked image, see the cumulative distribution function (CDF) for two original and watermarked host images in Figure 1. Note that real valued wavelet filter coefficients are used and the watermarked image is quantized to integer pixel values in $[0, 255]$. We observe two weaknesses: first, the embedding locations can be guessed due to the bias in quantization error (the scheme leaks information about the key $K$) and second, the quantization bin width $\Delta$ can be derived for each potential embedding location revealing the optimal attack power. In order to minimize the attack power, the attack targets potential embedding locations with small quantization bin width $\Delta$ up to a certain threshold. This attack parameter can be found experimentally with few ($< 10$) detector calls.

The attack first estimates potential embedding locations by selecting all locations where $e_d[i,j] < \Delta_d[i,j]/4$ and then adds or subtracts $\Delta_d[i,j]$ to $x_d^m[i,j]$ in order to flip the encoded bit of information.

**Watermarking Technique based on JPEG2000 Codec (WTJC)** by Chen et al. [2] is a watermarking scheme integrated in the JPEG2000 coding pipeline where a scrambled binary watermark replaces a selected bit plane of the quantized image transform coefficients. A technique called distortion compensation helps to control visible artefacts since the watermark is embedded in the approximation subband. Only the scrambling of the watermark bits is protected by a secret key, hence the attacker has full access to the watermark channel and can choose which bits to flip to remove the watermark while preserving image quality. The fixed and unprotected order of the embedding coefficients makes it possible for the attacker to remove a watermark of known length with minimal modifications to the image.

**Wavelet Tree Quantization (WTQ)** has received particular attention for watermarking purposes. Wang et al. [16] describe the formation of a wavelet tree by concatenating the detail coefficients of all but the highest resolution of a spatial subband location and orientation. For a four-level wavelet decomposition, each tree comprises $1+4+16 = 21$ coefficients. Several trees can be combined into so-called super-trees and two super-trees are used to embed one bit of watermark information: depending of the watermark information, either the first or the second super-tree is quantized (see [16] for details). A secret key is used to permute the order of wavelet trees, therefore the attacker does not know which two wavelet trees make up a super-tree and which two super-trees are use to embed a bit. Nevertheless, it is still possible to estimate the wavelet trees that have likely been quantized and use this information for a simple yet efficient attack. Figure 2 clearly reveals the energy reduction due to embedding.

Das and Maitra [5] attack the WTQ scheme by estimating the location of non-quantized wavelet trees and then perform quantization of this set based on the estimated reference quantization error. In this paper, we propose a slightly different attack which estimates the location of quantized wavelet trees and fills the two least significant bit planes with ones. The attack power can be significantly reduced with this new method.

**Structure-Based Wavelet Tree Quantization (SBWTQ)** proposed by Wu et al. [17] only uses three wavelet decomposition levels and constructs wavelet trees from the two lower resolution detail subbands. Four adjacent wavelet trees are arranged into a super-tree which encodes one bit of watermark information by enforcing a relationship between the two upper and lower wavelet trees. Note that no key is used to obscure the composition of super-trees or the arrangement of wavelet trees within a super-tree. With exact embedding position knowledge it is an easy task to read and modify, e.g. erase, the watermark.

**Double Wavelet Tree Energy Modulation (DWTEM)** is a recent scheme presented by Tsai et al. [14] which takes into account the targeted attack on WTQ [5]. After constructing wavelet trees as in [16], a secret key $K$ is used to randomly shuffle the trees. One or several wavelet trees are combined to form a super-tree and four consecutive super-trees are used to embed one bit of watermark information. The energy of a super-tree, $e(\mathrm{ST})$, is defined as the sum of its absolute wavelet coefficient values. For each watermark bit, four super-trees are grouped into two pairs and the pair with the larger absolute energy difference is called the Check Supertrees (CST), the other pair named Quantized Supertrees (QST). Further, the symbols $d_{\mathrm{CST}}$ and $d_{\mathrm{QST}}$ denote the energy difference between the first and second super-tree within the respective super-tree pair, e.g. $d_{\mathrm{QST}} = e(\mathrm{QST}_1) - e(\mathrm{QST}_2)$. To embed watermark symbol 1, the QST are changed such that $d_{\mathrm{CST}} \cdot d_{\mathrm{QST}} > 0$. For watermark symbol $-1$, the relation $d_{\mathrm{CST}} \cdot d_{\mathrm{QST}} < 0$ is enforced, again by changing the QST. In order to alter the sign of $d_{QST}$, the coefficients of $\mathrm{QST}_1$ and $\mathrm{QST}_2$ are multiplied or divided by a factor $m = \sqrt{e(\mathrm{QST}_2)/e(\mathrm{QST}_1)} + \Delta$ where $\Delta$ controls the embedding strength.

The key-dependent permutation of wavelet trees occludes the embedding locations and the individual embedding power because the QST and CST can not be determined. The distribution of wavelet tree energy is preserved, rendering the attack of Das et al. [5] ineffective. We note that DWTEM multiplies or divides wavelet tree coefficients by a factor $m$. However, the coefficients of the highest resolution detail subband are not part of the wavelet tree. We conjecture that the ratio between the energy of the finest detail wavelet tree coefficients and the energy of the corresponding coefficients in the highest resolution subband reveals the information whether a wavelet tree's energy has been made larger or smaller during watermark embedding. In the case of DWTEM, we have 16 high resolution wavelet tree coefficients $T_{d=2}$ with energy $e(T_{d=2})$ and 64 detail subband coefficients $C_{d=1}$ with energy $e(C_{d=1})$ at the same spatial location and orientation; $d$ denotes the wavelet decomposition level. The energy ratio thus is defined as $f = e(T_{d=2})/e(C_{d=1})$. The CDF of the coefficients' energy ratio is shown in Figure 3. Note the slight deviation between original and watermarked images: small and medium energy ratios are more pronounced in the watermarked image.

The attack selects wavelet trees with little energy and uses the energy ratio $f$ to determine the attack direction: for small values of $f$, the wavelet tree's energy is increased while for large values of $f$, the wavelet tree's energy is decreased. The exact parameters of the attack (energy threshold for wavelet trees, threshold for small and large energy ratio, attack power) depend on the image statistics and have to be found experimentally; a limited number detectors call ($< 10$) is sufficient.

**Significant Difference of Wavelet Coefficient Quantization (SDWCQ)** is another recent proposal by Lin et al. [9]. Adjacent coefficients of one detail subband are grouped into blocks to embed one bit of watermark information. The blocks are shuffled according to a secret, key-dependent permutation. Within each block, the largest and second-largest coefficient, denoted $max$ and $sec$, are selected. Their significant difference $d = max - sec$ encodes one watermark symbol: to encode 1, $T$ is added to coefficient $max$ if $d < \max(\epsilon, T)$; to encode $-1$, the largest coefficient is set to $sec$, i.e. $max' = sec$. $\epsilon$ is the average significant difference over all blocks and $T$ controls the embedding strength.

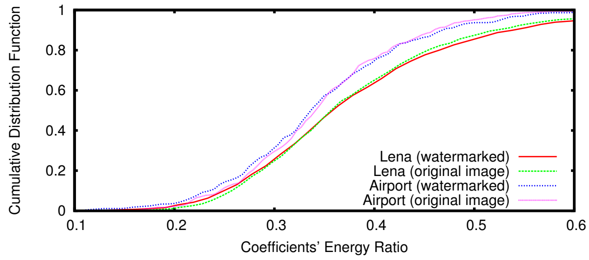


**Figure 3. Coefficients' energy ratio for original and watermarked images**

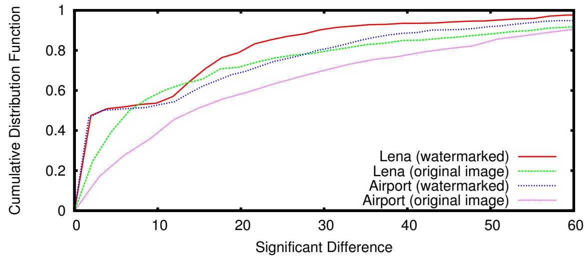


**Figure 4. Significant difference for the original and watermarked image**

The weakness is that the blocks shuffling only encrypts the watermark message but does not protect access to the watermark channel. According the Cox et al. [4, Chapter 2.3.8], the key is a *cipher* key, not a *watermark* key. In Figure 4, we show the CDF of significant differences for all possible blocks of two original and watermarked host images. Due to the shuffling, we do not know which blocks carry watermark information. However, the effect of quantizing the second-largest coefficient of a block and the enforced difference $\max(\epsilon, T)$ become immediately evident and can be used to mount an efficient attack which increases small significant differences and reduces larger differences when below a threshold. The attack is described in detail in [10].

## 3 Security Discussion and Improvements

All weaknesses have in common that they leak information on the watermarking channel used. Thus all discussed schemes violate Kalker's security principle stated in the introduction. This allows the attacker to concentrate the attack on a smaller set of coefficients or permits finely tuned attack vectors resulting in lower overall attack energy.

The QMWDC scheme can be improved by protecting the embedding locations with the use of key-dependent dither modulation, see [4, Chapter 9.2.5]. Even if an attacker does not know the exact embedding positions in the QMWDC watermarking scheme, it is known that only the middle coefficient of the triples is used for embedding. Further, the quantization bin width is revealed. Shuffling the details subband coefficients before constructing the coefficient triples can be used to disguise the coefficients' relationship, however, there might be an impact on the robustness and/or imperceptibility of the scheme and further experiments are needed.

Das el al. [5] describe a modified WTQ (MWTQ) scheme which imposes a energy difference between two super-trees, alleviating the security issue. However, the modification depends on the organization information of super-trees to be transmitted via a secure side-channel, limiting the applicability of the watermarking method and turning MWTQ into a semi-blind watermarking scheme.

To improve the WTJC scheme the fixed selection of embedding locations has to be broken up. Further the embedding and embedding strength bit plane can adjusted in a key-dependent way such that an attacker can not determine the coefficients used for embedding and hence does not know the bit plane to attack.

The weak point of the SBWTQ scheme is that the attacker has full access to the unprotected watermarking channel. An attack would be more difficult if the watermarking channel is hidden for instance by assembling the super-trees not from adjacent trees but from trees at random locations in the subband.

Wavelet trees seem to be a popular choice for watermark embedding (with little justification), although spatial and multi-resolution organization of the watermark is revealed. For the DWTEM scheme, one could include the highest resolution subband coefficients in the wavelet tree and/or pseudo-randomly permute the detail subband coefficients before constructing the wavelet trees, although this demolishes the very idea of this structure.

Similarly, the weakness in SDWCQ can be mitigated by performing a secret, key-dependent permutation on the subband coefficients before constructing the blocks, thus blocking access to the watermark channel. Also this modified SDWCQ schemes is vulnerable [10], but the attack mainly exploits the limited robustness and concentration of the watermark power in one wavelet detail subband.

## 4 Experimental results

The implementation of the discussed watermarking schemes and the related attacks are available as Python code at `http://www.wavelab.at/sources`. For our experiments, we use ten $512 \times 512$ gray-scale image freely available from the USC SIPI image database[1], see Figure 5.

The effectiveness of the attack is measured by the normalized correlation (NC) between the embedded and extracted watermark, $\mathrm{NC}(\mathbf{w}, \mathbf{w}^\star) = \frac{1}{N}\sum_{i=1}^{N} w_i w_i^\star$, and the PSNR (dB) between watermarked and attacked image, denoted (w,a), and the PSNR (dB) between the original and attacked image, denoted (o,a). In addition, we give the PSNR (dB) between the original and watermarked image to illustrate the watermark embedding power. To judge the existence of the watermark, NC is compared against a threshold $T_{\mathrm{NC}}$: if $\mathrm{NC} > T_{\mathrm{NC}}$ the watermark is declared present, otherwise absent. For a watermark of length $N = 512$, equiprobable watermark symbols $w_i \in \{-1, 1\}$ and a desired false-alarm probability of approximately $10^{-7}$, $T_{\mathrm{NC}}$ is set to $0.23$. We try to evaluate the schemes on a common ground. Therefore, we use the popular Daubechies 9/7 wavelet filter for image decomposition and always embed a pseudo-random $512$ bit watermark sequence. The attack experiment is repeated ten times for each image with different watermarks.

We now briefly discuss the attack results. The watermark is completely removed with a NC value close to zero for QMWDC (Table 1), WTJC (Table 3), SBWTQ (Table 4), WTQ (Table 5) and SDWCQ (Table 7). Most interestingly, the attack power (w,a) is significantly smaller than the embedding power (o,w) in terms of PSNR (dB), therefore the attack is unlikely to perceptually degrade the image.

In Table 2 we provide attack results for the QMWDC scheme when using a key-dependent dither vector as an additional security measurement which prevents estimation of potential embedding locations. Compared to the previous results in Table 1, the attack power has to be increased by more than 3 dB, the attacked image looses approximately $0.5$ dB PSNR.

In Table 6 we present our attack on DWTEM which has been designed with the results of an earlier security analysis in mind, see [5]. For all images, we have reduced the NC measure just below the detection threshold $T_{\mathrm{NC}}$ set to $0.23$. The effectiveness of the attack depends on the image characteristics in order to permit estimation of the attack direction based on the energy ratio criterion. For some images, the attack power is well below the embedding power, for others the attack is less effective. As a results, the quality of attacked images is on average 2 dB lower in terms of PSNR than for the watermarked images. This may result in a slightly lower perceptual fidelity for some attacked images compared to the watermarked images. Nevertheless, the attack demonstrates that the security margin for DWTEM is practically zero.

## 5 Conclusion

This paper presents a number of attacks on several published watermarking scheme for copyright protection by exploiting knowledge of the schemes' implementation. The lack of protection of the embedding locations allows to completely remove the watermark while maintaining a high PSNR. We highlight the need for a detailed security analysis, assuming the attacker is familiar with the watermarking scheme's implementation. We successfully analyzed a scheme designed to withstand targeted attacks. We expect

[1] `http://sipi.usc.edu/database/`

| Image | ∅ NC | ∅ PSNR (dB) | | |
|---|---|---|---|---|
| | | (w,a) | (o,a) | (o,w) |
| Lena | 0.021 | 54.29 | 45.79 | 46.13 |
| Goldhill | 0.014 | 52.36 | 44.99 | 45.42 |
| Peppers | 0.056 | 54.64 | 45.31 | 45.61 |
| Man | 0.039 | 51.57 | 43.01 | 43.29 |
| Airport | 0.064 | 51.02 | 42.22 | 42.48 |
| Tank | −0.009 | 53.01 | 47.46 | 48.18 |
| Truck | −0.032 | 52.97 | 47.00 | 47.62 |
| Elaine | 0.073 | 53.55 | 47.17 | 47.79 |
| Boat | −0.036 | 52.28 | 43.39 | 43.69 |
| Barbara | −0.063 | 50.80 | 42.54 | 42.83 |
| Average | 0.013 | 52.65 | 44.89 | 45.30 |

**Table 1. Attack result on the QMWDC scheme with** $Q = 4$

| Image | ∅ NC | ∅ PSNR (dB) | | |
|---|---|---|---|---|
| | | (w,a) | (o,a) | (o,w) |
| Lena | 0.028 | 50.05 | 45.06 | 46.11 |
| Goldhill | −0.054 | 48.54 | 44.15 | 45.32 |
| Peppers | −0.018 | 51.02 | 44.73 | 45.49 |
| Man | −0.005 | 47.21 | 42.27 | 43.24 |
| Airport | 0.009 | 47.84 | 41.73 | 42.48 |
| Tank | −0.037 | 50.34 | 46.71 | 48.17 |
| Truck | −0.023 | 49.62 | 46.06 | 47.52 |
| Elaine | −0.043 | 51.04 | 46.58 | 47.76 |
| Boat | −0.012 | 48.66 | 42.87 | 43.70 |
| Barbara | 0.018 | 48.27 | 41.99 | 42.71 |
| Average | −0.013 | 49.26 | 44.22 | 45.25 |

**Table 2. Attack result on the QMWDC scheme employing dither quantization;** $Q = 4$

| Image | ∅ NC | ∅ PSNR (dB) | | |
|---|---|---|---|---|
| | | (w,a) | (o,a) | (o,w) |
| Lena | −0.007 | 47.18 | 39.74 | 40.30 |
| Goldhill | −0.024 | 47.95 | 41.02 | 41.68 |
| Peppers | 0.023 | 48.10 | 40.38 | 40.88 |
| Man | 0.118 | 50.57 | 41.56 | 41.92 |
| Airport | 0.048 | 49.55 | 42.43 | 43.02 |
| Tank | −0.152 | 42.81 | 39.11 | 41.27 |
| Truck | 0.071 | 48.83 | 39.70 | 40.02 |
| Elaine | −0.029 | 46.25 | 39.19 | 39.82 |
| Boat | −0.073 | 45.73 | 39.63 | 40.55 |
| Barbara | −0.021 | 47.46 | 40.36 | 40.98 |
| Average | −0.005 | 47.44 | 40.31 | 41.04 |

**Table 3. Attack results on WTJC;** $\alpha = 0.6$ **and distortion reduction**

| Image | ∅ NC | ∅ PSNR (dB) | | |
|---|---|---|---|---|
| | | (w,a) | (o,a) | (o,w) |
| Lena | 0.000 | 54.76 | 44.57 | 44.73 |
| Goldhill | 0.000 | 51.15 | 42.12 | 41.31 |
| Peppers | 0.000 | 53.49 | 41.40 | 41.38 |
| Man | 0.000 | 51.95 | 42.02 | 41.68 |
| Airport | 0.000 | 51.14 | 41.37 | 40.92 |
| Tank | 0.000 | 51.24 | 44.63 | 44.08 |
| Truck | 0.000 | 50.66 | 42.27 | 41.53 |
| Elaine | 0.000 | 53.08 | 44.90 | 44.87 |
| Boat | 0.000 | 54.17 | 42.43 | 42.46 |
| Barbara | 0.000 | 53.03 | 42.77 | 42.56 |
| Average | 0.000 | 52.47 | 42.85 | 42.55 |

**Table 4. Attack results on SBWTQ;** $\Delta = 10$

| Image | ∅ NC | ∅ PSNR (dB) | | |
|---|---|---|---|---|
| | | (w,a) | (o,a) | (o,w) |
| Lena | −0.049 | 49.55 | 40.90 | 41.49 |
| Goldhill | 0.063 | 51.13 | 44.92 | 45.82 |
| Peppers | −0.121 | 49.83 | 43.51 | 44.54 |
| Man | 0.122 | 51.52 | 45.49 | 46.30 |
| Airport | 0.116 | 51.89 | 45.93 | 46.81 |
| Tank | −0.036 | 51.54 | 46.22 | 47.24 |
| Truck | 0.002 | 51.20 | 45.80 | 46.85 |
| Elaine | −0.177 | 50.31 | 45.29 | 46.68 |
| Boat | 0.023 | 50.63 | 43.39 | 44.12 |
| Barbara | 0.073 | 50.45 | 42.51 | 43.11 |
| Average | 0.001 | 50.81 | 44.40 | 45.30 |

**Table 5. Attack result on the WTQ scheme with** $E = 100$, $q_{max} = 336$ **and** $\epsilon = 0.1$

| Image | ∅ NC | ∅ PSNR (dB) | | |
|---|---|---|---|---|
| | | (w,a) | (o,a) | (o,w) |
| Lena | 0.228 | 44.93 | 39.77 | 41.08 |
| Goldhill | 0.222 | 42.44 | 39.60 | 41.90 |
| Peppers | 0.217 | 43.94 | 40.07 | 41.92 |
| Man | 0.229 | 39.07 | 36.75 | 39.38 |
| Airport | 0.229 | 38.24 | 36.63 | 39.92 |
| Tank | 0.222 | 44.99 | 43.39 | 47.16 |
| Truck | 0.225 | 43.23 | 41.40 | 44.80 |
| Elaine | 0.225 | 45.18 | 41.89 | 44.31 |
| Boat | 0.224 | 38.06 | 36.04 | 39.54 |
| Barbara | 0.229 | 36.90 | 35.23 | 39.35 |
| Average | 0.225 | 41.70 | 39.08 | 41.93 |

**Table 6. Attack results on DWTEM;** $\Delta = 0.15$

(a) *Lena* (b) *Goldhill* (c) *Peppers* (d) *Man* (e) *Airport*

(f) *Tank* (g) *Truck* (h) *Elaine* (i) *Boat* (j) *Barbara*

**Figure 5. Ten $512 \times 512$ gray-scale test images**

| Image | ∅ NC | ∅ PSNR (dB) | | |
|---|---|---|---|---|
| | | (w,a) | (o,a) | (o,w) |
| Lena | $0.020$ | $54.42$ | $46.42$ | $46.63$ |
| Goldhill | $-0.109$ | $53.36$ | $45.79$ | $45.91$ |
| Peppers | $-0.023$ | $54.08$ | $45.02$ | $45.05$ |
| Man | $0.025$ | $51.94$ | $42.70$ | $42.85$ |
| Airport | $-0.108$ | $53.00$ | $45.00$ | $45.10$ |
| Tank | $-0.112$ | $54.22$ | $48.81$ | $48.97$ |
| Truck | $-0.121$ | $52.43$ | $44.79$ | $44.96$ |
| Elaine | $-0.066$ | $54.39$ | $47.01$ | $47.37$ |
| Boat | $-0.040$ | $53.79$ | $45.69$ | $45.82$ |
| Barbara | $-0.014$ | $53.96$ | $46.04$ | $46.19$ |
| Average | $-0.055$ | $53.56$ | $45.73$ | $45.88$ |

**Table 7. Attack results on the SDWCQ scheme; $\gamma$ unrestrained, block-size $7$, $T = 12$ and $\alpha = 0.9$**

several more quantization based watermarking schemes to be vulnerable to similar attacks. Evaluation of the robustness against common signal processing operations is insufficient for watermarking schemes in the copyright protection scenario.

## Acknowledgements

Supported by Austrian Science Fund project FWF-P19159-N13.

## References

[1] F. Cayre, C. Fontaine, and T. Furon, "Watermarking security: theory and practice", *IEEE Transactions on Signal Processing*, 53(2), Oct. 2005, pp. 3976–3987.

[2] T.-S. Chen, J. Chen, and J.-G. Chen, "A simple and efficient watermark technique based on JPEG2000 codec", *ACM Multimedia Systems Journal*, 10(1), June 2004, pp. 16–26.

[3] P. Comesana, L. Perez-Freire, and F. Perez-Gonzalez, "Blind newton sensitivity attack", *IEE Proceedings on Information Security*, 153(3), Sept. 2006, pp. 115–125.

[4] I. J. Cox, M. L. Miller, J. A. Bloom, J. Fridrich, and T. Kalker, *Digital Watermarking and Steganography*, Morgan Kaufmann, 2007.

[5] T. K. Das and S. Maitra, "Analysis of the 'Wavelet Tree Quantization' watermarking strategy and a modified robust scheme", *Multimedia Systems*, 12(2), Aug. 2006, pp. 151–163.

[6] T. Kalker, "Considerations on watermarking security", In *Proceedings of the IEEE Workshop on Multimedia Signal Processing, MMSP '01*, pp. 201–206, Cannes, France, Oct. 2001.

[7] A. Kerckhoffs, "La cryptographie militaire", *Journal des sciences militaires*, 9, Jan. 1883, pp. 5–83.

[8] D. Kundur and D. Hatzinakos, "Digital watermarking using multiresolution wavelet decomposition", In *Proceedings of the 1998 International Conference on Acoustics, Speech and Signal Processing (ICASSP'98)*, volume 5, pp. 2969–2972, Seattle, WA, USA, May 1998.

[9] W.-H. Lin, S.-J. Horng, T.-W. Kao, P. Fan, C.-L. Lee, and Y. Pan, "An efficient watermarking method based on significant difference of wavelet coefficient quantization", *IEEE Transactions on Multimedia*, 10(5), Aug. 2008, pp. 746–757.

[10] P. Meerwald, C. Koidl, and A. Uhl, "Attack on 'Watermarking Method Based on Significant Difference of Wavelet Coefficient Quantization'", *IEEE Transactions on Multimedia*, 2009, accepted.

[11] F. A. P. Petitcolas, R. J. Anderson, and M. G. Kuhn, "Attacks on copyright marking systems", In D. Aucsmith, editor, *Information Hiding: Second International Workshop*, volume 1525 of *Lecture Notes in Computer Science*, pp. 218–238, Portland, OR, USA, Apr. 1998. Springer Verlag, Berlin, Germany.

[12] F. A. P. Petitcolas, C. Fontaine, J. Dittmann, M. Steinebach, and N. Fatès, "Public automated web-based evaluation service for watermarking schemes: Stirmark benchmark", In *Proceedings of SPIE, Security and Watermarking of Multimedia Contents III*, volume 4314, pp. 575–584, San Jose, CA, USA, Jan. 2001.

[13] A. Piva and M. Barni, "The first BOWS contest: break our watermarking system", In E. J. Delp and P. W. Wong, editors, *Security, Steganography, and Watermarking of Multimedia Contents IX*, volume 6505 of *Proceedings of SPIE*, San Jose, CA, USA, Jan. 2007. SPIE.

[14] M.-J. Tsai, C.-T. Lin, and J. Liu, "A wavelet-based watermarking scheme using double wavelet tree energy modulation", In *Proceedings of the 2008 IEEE International Conference on Image Processing, ICIP '08*, pp. 417–420, San Diego, CA, USA, Oct. 2008. IEEE.

[15] S. Voloshynovskiy, S. Pereira, T. Pun, J. K. Su, and J. J. Eggers, "Attack on digital watermarks: Classification, estimation-based attacks and benchmarks", *IEEE Communications Magazin*, 39(8), Aug. 2001, pp. 118–126.

[16] S.-H. Wang and Y.-P. Lin, "Wavelet tree quantization for copyright protection watermarking", *IEEE Transactions on Image Processing*, 13(2), Feb. 2004, pp. 154–165.

[17] G.-D. Wu and P.-H. Huang, "Image watermarking using structure based wavelet tree quantization", In *Proceedings of the 6th IEEE/ACIS International Conference on Computer and Information Science, 2007. ICIS 2007*, pp. 315–319. IEEE, July 2007.