Acta Universitatis Tamperensis 2120

YEVHEN ZOLOTAVKIN

# New Methods for Digital Image Watermarking

YEVHEN ZOLOTAVKIN

# New Methods for Digital Image Watermarking

ACADEMIC DISSERTATION
To be presented, with the permission of the
Board of the School of Information Sciences of the
University of Tampere,
for public discussion in the auditorium Pinni B 1096,
Kanslerinrinne 1, Tampere, on 11 December 2015, at 12 o'clock.

UNIVERSITY OF TAMPERE

YEVHEN ZOLOTAVKIN

# New Methods for Digital Image Watermarking

*Acta Universitatis Tamperensis 2120*
*Tampere University Press*
*Tampere 2015*

UNIVERSITY
OF TAMPERE

ACADEMIC DISSERTATION
University of Tampere
School of Information Sciences
Finland

The originality of this thesis has been checked using the Turnitin OriginalityCheck service in accordance with the quality management system of the University of Tampere.



Cover design by
Mikko Reinikka

Distributor:
verkkokauppa@juvenesprint.fi
https://verkkokauppa.juvenes.fi

Acta Universitatis Tamperensis 2120
ISBN 978-951-44-9986-9 (print)
ISSN-L 1455-1616
ISSN 1455-1616

Acta Electronica Universitatis Tamperensis 1617
ISBN 978-951-44-9987-6 (pdf)
ISSN 1456-954X
http://tampub.uta.fi

Suomen Yliopistopaino Oy – Juvenes Print
Tampere 2015

# Abstract

This thesis contributes to the field of protection of Digital Rights of an owner of multimedia content. More specifically, new methods of Digital Image Watermarking (DIW) enabling that kind of protection were developed, tested and represented in the thesis in the form of journal and conference publications. For each of the methods, the goal was to improve the trade-off between robustness and transparency of digital watermark. With that aim, several new watermark embedding techniques were proposed and applied in the domain of Singular Value Decomposition (SVD) of image blocks. For each of the proposed techniques, the trade-off depends on the model of watermark embedding and its parameters as well as the measure of transparency and the kind of possible attacks on watermarked images.

The thesis describes two different types of embedding models in the domain of SVD: the first type assumes embedding of a watermark by modifying left or right orthogonal matrices of the decomposition; the second type assumes watermark embedding by quantizing singular values.

For the first type of embedding models, three different implementations were proposed that are based on Van Elfrinkhof's rotation matrix. The most important characteristics of the mentioned implementations are: a) the result of the modification of original matrix can be adjusted using several parameters; b) the modified matrix remains orthogonal. The outcomes implied by the mentioned characteristics are the opportunity to improve transparency and to provide better robustness in regards to a) and b), respectively.

For the second type of embedding models, new principles of Distortion Compensation (DC) for Quantization Index Modulation (QIM) of different dimensionality were applied. In the case of scalar DC-QIM, the quantization is performed in a way that is defined by several parameters which makes possible: a) asymmetric distribution of quantized samples; b) associate Initial Data Loss (IDL) with some samples. The beneficial outcome of a) is the distinctive feature that helps to recover after Gain Attack (GA), while b) provides better robustness-transparency trade-off under intense Additive White Gaussian Noise (AWGN). In the case of two-dimensional DC-QIM, a new direction of quantization is proposed which takes into account the form of Voronoi cell that the Quincunx lattice quantizer operates on.

In the thesis, the measure of transparency of watermarked images was based on Mean Square Error (MSE) as it is simple, easy to adjust and sensitive to additive watermarking. Robustness of watermarked images was tested under different attacks, such as Additive White Gaussian Noise, JPEG-compression, Gain Attack (GA), Salt&Pepper, Cropping, Rotation, Median Filtering.

**Keywords:** digital rights, watermarking, transparency, robustness, SVD, orthogonal matrix, quantization, AWGN, gain attack

# Contents

# List of Abbreviations

| Abbreviation | Description |
|---|---|
| AWGN | Additive White Gaussian Noise |
| AGN | Additive Gaussian Noise |
| BER | Bit Error Rate |
| BSW | Buyer–Seller Watermarking |
| CCD | Charge-Coupled Devices |
| DC-QIM | Distortion Compensated QIM |
| DCT | Discrete Cosine Transform |
| DFT | Discrete Fourier Transform |
| DIW | Digital Image Watermarking |
| DRM | Digital Rights Management |
| DWR | Document to Watermark Ratio |
| DWT | Discrete Wavelet Transform |
| EZW | Embedded Zero-tree Wavelet |
| FFT | Fast Fourier Transform |
| FWT | Fast Wavelet Transform |
| GA | Gain Attack |
| HH | horizontal Highs/vertical Highs |
| HL | horizontal Highs/vertical Lows |
| HVS | Human Visual System |
| IDL | Initial Data Loss |
| IQM | Image Quality Measure |
| ISVD | Inverse SVD |
| JPEG | Joint Photographic Experts Group |
| LH | horizontal Lows/vertical Highs |
| LL | horizontal Lows/vertical Lows |
| LS | Least Squares |
| MLES | Maximum Likelihood Error Scenario |
| MQ | Modified Quincunx |
| MSE | Mean Square Error |
| NSC-QIM | NS-QIM with Constant parameters |
| NS-QIM | Non-Symmetric QIM |
| PDF | Probability Density Function |
| PSNR | Peak Signal-To-Noise Ratio |
| QIM | Quantization Index Modulation |
| RDM | Rational Dither Nodulation |
| RGB | Red-Green-Blue |
| SMT | Secure Media Technologies |
| SPIHT | Set Partitioning In Hierarchical Trees |
| SS | Spread Spectrum |
| STFT | Short-Time Fourier Transform |
| SVD | Singular Value Decomposition |
| TCQ | Trellis-Coded Quantization |
| VQ | Vector Quantization |
| WNR | Watermark to Noise Ratio |

# List of Original Publications

I. ZOLOTAVKIN, Y. & JUHOLA, M.: SVD-based Digital Image Watermarking on Approximated Orthogonal Matrix. In: Proceedings of *the 10th International Conference on Security and Cryptography* (SECRYPT 2013): *SciTePress*, July 2013, pp. 321—330.

II. ZOLOTAVKIN, Y. & JUHOLA, M.: An SVD-based Transparent Watermarking Method. In: Proceedings of *the International Conference on E-Technologies and Business on the Web (EBW2013)*: *SDIWC*, May 2013, pp. 85—90.

III. ZOLOTAVKIN, Y. & JUHOLA, M.: A new blind adaptive watermarking method based on singular value decomposition. In: Proceedings of *International Conference on Sensor Network Security Technology and Privacy Communication System* (SNS & PCS): *IEEE*, May 2013, pp. 184—192.

IV. ZOLOTAVKIN, Y. & JUHOLA, M.: A New QIM-Based Watermarking Method Robust to Gain Attack. In: *International Journal of Digital Multimedia Broadcasting* Vol 2014 (Sep 2014), Nr. Art 910808, S. pp. 1—14.

V. ZOLOTAVKIN, Y. & JUHOLA, M.: Quantization Based Watermarking Approach with Gain Attack Recovery. In: Proceedings of *IEEE International Conference on Digital Image Computing: Techniques and Applications* (DICTA'14): *IEEE*, Nov 2014, pp. 1—8.

VI. ZOLOTAVKIN, Y. & JUHOLA, M.: A New Two-Dimensional Quantization Method for Digital Image Watermarking. In: Proceedings of *IEEE International Conference on Advanced Communications Technology* (ICACT'15): *IEEE*, July 2015, pp. 155—160.

# Chapter 1

# Introduction

Recent advances in multimedia production, delivery and processing have also created new opportunities for the dissemination and illegal consumption of multimedia content [1, 2, 3, 4, 5]. Digital Rights Management (DRM) is the practice of imposing technological restrictions on actions with regards to digital media [6, 7, 8, 9]. The task of DRM is to provide "remote control" and "persistent protection" for digital content. On the other hand, some information about the protected content should be available to everyone, which limits usage of cryptographic tools [10, 11]. Driven by the requirements of DRM, the last two decades have seen the development of new tools to tackle the problems in media security. Secure Media Technologies (SMT) is the concept encompassing a wide range of diverse technological areas including watermarking, steganography, cryptography, biometrics, fingerprinting, network security and digital forensics [12, 13, 14]. Digital Watermarking is one of the most promising, versatile and fast developing areas of SMT.

Digital Watermarking incorporates means of securing the rights of the owner of the digital data, providing authentication of the source or originality of the digital data [11, 15, 16]. The hidden message (watermark) signifies information that can be detected and retrieved by authorized personnel or systems designed for that purpose. Methods of Digital Watermarking can be applied to many types of content such as text [17, 18, 19], audio [20, 21, 22], images [23, 24, 25], video [26, 27, 28], 3D meshes [29, 30, 31], software programs [32, 33, 34] and network packets [35, 36, 37].

In spite of a great variety of digital watermarking methods, characteristics of robustness and transparency are the most universal and important for any of them. A degree of ***robustness*** of the watermark data defines how immune it is against modifications and/or malicious attacks [38, 39]. Another important characteristic is the perceptual ***transparency*** of the watermark. Artifacts introduced through a watermarking process may reduce the commercial value of the watermarked data [3, 40].

More specific characteristics include: visibility, fragility, blindness, reversibility (invertibility) [4, 16]. Visible watermarks are visual patterns (like logos) which are inserted into or overlaid on images (or video), very similar to visible paper watermarks. The purpose is, for example, to visibly mark preview images available in image databases in order to prevent their commercial use [24, 41, 42]. An invisible watermark is imperceptible, but can be extracted via computational methods which require watermarking key (password) to be known [11, 43, 44]. The fragile watermarking is primarily used for authentication and provides low robustness (high sensitivity) toward modifications so that one can easily detect an attempt of attack [45, 46]. As an opposite to it, robust watermarking should provide a high level of immunity against attacks directed toward removing or damaging of a watermark [47, 48]. A watermark which is sensitive to some attacks and is robust to other is a semi-fragile [49, 50]. Non-blind watermarking systems in addition to the key require at least the original data for watermark extraction [1, 51]. Some types of non-blind systems also require a copy of the embedded watermark in order to confirm

that it has been embedded. Blind (or oblivious) watermarking systems present the most challenging task as only the key is required for watermark extraction [52, 53, 54, 55]. Reversible watermarking is a technique which enables watermarked content to be authenticated and then restored to its original form by removing the digital watermark [56, 57]. This would make the images acceptable for legal purposes. A simple scheme of image watermarking process including stages of embedding, transmitting, attacking, and extracting is presented on Figure 1.1. On the sender side, a watermark is embedded in the cover image in the positions defined by a key. Then, the watermarked image is sent over a channel where an attack might occur. On the receiver side, a possibly corrupted watermark is extracted using the key. As only a key is required for extraction, the process is blind and the watermark is invisible.

One of the distinctive and most beneficial features of watermarking is that protective information is mixed with the original media content. As a result, the watermark is usually unnoticeable and inseparable from the content which means that the watermarked media can be used legally with very little functional limitations in many practical applications while still witnessing an owner. In contrast to that, for instance, media cryptographically encrypted by owner can not be used in any way by a third party if not being decrypted. Additionally, in case of cryptographic protection the fact of encryption is always observable and obvious which might encourage an attacker to break the channel.

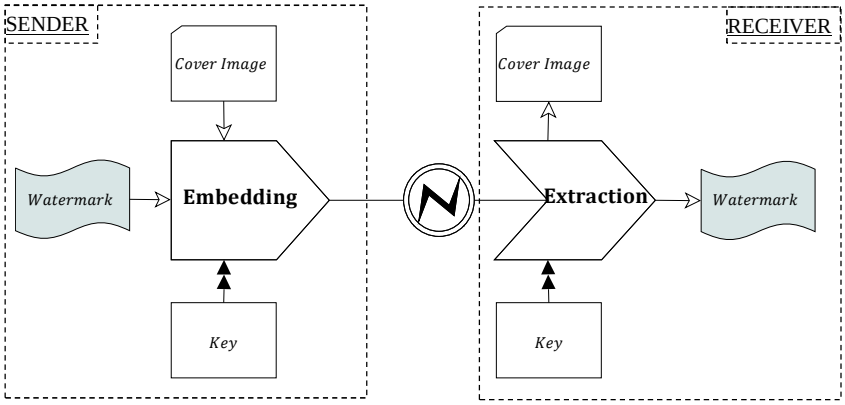



**Figure 1.1.** Digital Image Watermarking Scheme

In watermarking, fingerprinting and labeling are terms that denote special applications. Fingerprinting means watermarking where the embedded information is either a unique code specifying the author of the cover data or a unique code specifying the recipient of the data. Labeling means watermarking where the embedded data may contain any information of interest, such as a unique data identifier [1, 11].

One of the most vivid DRM problems related to Digital Watermarking can be illustrated on digital cinema example [58, 59]. Digital cinema can be defined as the digital electronic distribution and display of theatrical film content. A responsibility of the movie theater owners is to prevent a spectator from filming the projected movie with a handicam at the back of the theater. In this scenario, the most relevant watermarks are forensic tracking watermarks that are called “exhibition fingerprints” and identify the circumstances of the exhibition. The fingerprint should include identification of the content as well as the exhibition and be resistant to the “handicam copy” which means severe image distortions, such as scaling, cropping, affine

transforms, but also nonlinear geometrical transform due to optics. This watermarking scenario is, therefore, characterized by very high quality requirements for the watermarked signal (invisible watermark) and quite complex transform (digital cinema to handicam shot) [60].

Another example of DRM task is control of contribution links which are the liaisons between content providers (like the European Broadcasting Union) and studios [4, 61]. The providers are multicasting content, which is remastered at each studio to be redistributed in secondary links. The main DRM concern for studios is to identify the copyright owner of content when it has gone through several postproduction processes. In this case, a watermark containing the owner's identity should be very robust and protected with a secret key. Only the owner of this secret key can read or detect the watermark and if a dispute is submitted to the court, the key owner will be able to produce the watermark in question.

Many important problems of DRM arise in the context of a patient's data protection in healthcare organizations [62, 63, 64]. "Instagram for doctors" is the application that was designed for doctors to share pictures of their patients, both with each other and with medical students [65, 66]. Only verified healthcare professionals can upload photos or comment on them but are required to follow strict guidelines on what is not permitted: a patient's face, any text or numbers or identifiable marks. Built-in image editing tools can be used to ensure patient privacy. Once uploaded, images are queued before being manually reviewed and can be rejected if have identifiable marks or are not of educational value. The images are also stripped of all metadata that could be used to identify the patient who has to sign a consent form. However, a probable violation of a patient's rights might happen if identifiable marks have not been removed by a doctor and a reviewer has not rejected the image. Such an image can be advised to be removed from the platform later, but during the time it is exposed it can be copied and spread on other platforms. A possible solution in that case is to assign and embed secret watermarks identifying the doctor and the patient (on a hospital site) as well as the reviewer. A suitable watermark should be robust (for instance, to cropping and compression) and does not cause large quality degradation.

The illustrated examples have different requirements, but unified framework solutions for a wider scope of applications are also in demand. During the recent years, many advanced watermarking protocols were proposed in the literature in order to improve DRM [67]. A buyer–seller watermarking (BSW) protocol combines encryption with digital watermarking [68, 69]. It is an asymmetric fingerprinting protocol where the fingerprint is embedded by means of watermarking in the encrypted domain. The basic idea is that each buyer obtains a slightly different copy of the digital content offered by the seller. Such a difference, the watermark (or fingerprint), does not harm the perceptual quality of the digital content and cannot be easily removed by the buyer. Thanks to the latter property, when a malicious buyer redistributes a pirated copy, the seller can associate the pirated copy with its buyer by its embedded watermark. On the other hand, a malicious seller cannot frame an honest buyer because the buyer's watermark and the delivered watermarked content are unknown to the seller [70].

In a multiparty multilevel DRM architecture, each party embeds its watermark signal separately into the digital content. Therefore, the quality of digital content deteriorates with each watermarking session [5, 6, 71]. Hence, effective protection of the owner rights, and satisfaction of quality requirements of consumer needs a smart protocol as well as a balanced compromise between watermarking distortions and robustness.

As it has been mentioned, for the needs of digital watermarking on all levels from a micro level of a specific application to a macro level of a watermarking protocol, aspects of ***transparency*** and ***robustness*** greatly influence effectiveness of protection. Unfortunately, goals of transparency and robustness are contradictory in digital watermarking and it is not feasible to improve both the qualities simultaneously [4]. Instead, it is more reasonable to set the goals in terms of robustness-transparency trade-off that is necessary to improve for a given class of watermarking implementation.

In Digital Image Watermarking (DIW), the robustness-transparency trade-off depends on several aspects such as: a domain chosen to embed a watermark; a technique used to modulate the coefficients in that domain; a type of attack that might occur; a measure of transparency for the watermarked image. Several domains are traditionally used for DIW. Among them there are spatial as well as basis transform domains where each domain has its advantages under different types of attacks and transparency measures [3]. Some of the most popular modulation techniques are taken from the field of digital communication and are, in fact, methods of coding [72, 73]. Their robustness under Additive White Gaussian Noise is usually quite high, but the other types of attacks might be more harmful. Nowadays, a great variety of image processing tools can be applied to editing of watermarked images with the aim of their enhancing or compression. For an embedded watermark, this can be qualified as an attack which might be combined with intentional/unintentional noising (additive or multiplicative in nature) [74, 60]. Definitions of transparency and quality for watermarked images are subjective in principle and may differ depending on the application. However, one out of a number of the known objective measures should be chosen for an unbiased comparison of DIW methods [75, 76].

In order to contribute to the field of DIW, the following research questions and hypothesis should be examined. The thesis will attempt to explain and test them in the following chapters.

- What are the common transform domains that are used in Digital Image Watermarking?
- What are the state of the art encoding techniques that are used to embed a digital watermark?
- Which measures of visual quality can serve purposes for DIW?
- What kinds of attacks on the watermarked images are the most common?
- The hypothesis was that problems associated with the above questions are possible to solve and the new DIW techniques that provide improved ***robustness-transparency*** trade-off can be designed.

The outline of this thesis consists of the following parts: Chapter 2 provides information about transforms commonly used in DIW such as Discrete Cosine Transform, Discrete Wavelet Transform and Singular Value Decomposition. Chapter 3 describes the most popular encoding techniques, namely Spread Spectrum and Quantization Index Modulation with both scalar and multidimensional implementations. Measures of image transparency are discussed in Chapter 4 while information about attacks is provided in Chapter 5. Results of individual publications as well as a generalized insight on the development and progress of the whole research are represented in Chapter 6. Finally, Chapter 7 concludes the thesis.

# Chapter 2

# Transform Domains traditionally used for DIW

## 2.1 Discrete Cosine Transform

Discrete Cosine Transform is one of the most popular in Digital Image Processing [77, 78]. Widely used compression format JPEG utilizes this transform which is also the reason for its adoption by numerous watermarking methods as in that case it is relatively easy to provide robustness to JPEG [79, 80, 81].

### 2.1.1 1D Discrete Cosine Transform

The principle of Discrete Cosine Transform (DCT) is tightly connected with Discrete Fourier Transform (DFT). Coefficients of continuous FT for a continuous signal $f(t)$ are defined as:

$$F(j\omega) = \int_{-\infty}^{\infty} f(t)e^{-j\omega t}dt\ , \tag{2.1}$$

where $\omega$ is the angular frequency.

For a discrete signal $f[n]$, $n \in \{0, 1, \dots, N-1\}$ Fourier coefficients are calculated in the following way:

$$F(j\omega) = \sum_{n=0}^{N-1} f[n]e^{-j\omega nT}\ . \tag{2.2}$$

However, if we treat discrete signal as periodic, it is enough to calculate only $N$ coefficients:

$$F[k] = \sum_{n=0}^{N-1} f[n]e^{-j\frac{2\pi}{N}nk},\ k \in \{0, 1, \dots, N-1\}\ . \tag{2.3}$$

In contrast to DFT, coefficients of DCT are real. There are several types of DCT known [82]. They differ in how the signal (which has clear limits in time domain) is extended in its infinite-time representation. For instance, interpretation of the first and the last samples of the signal can be different. In the most popular type of DCT, coefficients are defined as

$$C[k] = \begin{cases} \sum_{n=0}^{N-1} 2f[n]\cos\left(\frac{\pi}{2N}k(2n+1)\right), 0 \le k < N, \\ 0, \qquad\qquad otherwise. \end{cases} \tag{2.4}$$

Inverse DCT should be performed according to the equation:

$$f[n] = \begin{cases} \frac{1}{N}\sum_{k=0}^{N-1} w[k]C[k]\cos\left(\frac{\pi}{2N}k(2n+1)\right), 0 \le n < N, \\ 0, \qquad\qquad otherwise. \end{cases} \tag{2.5}$$

where

$$w[k] = \begin{cases} 0.5, \quad k = 0, \\ 1, 1 \le k < N. \end{cases} \tag{2.6}$$

There is the relation between DCT and DFT that can be described using the following steps.

1) Create an extended pseudo-sequence $y[n]$ from original discrete $f[n]$:

$$y[n] = \begin{cases} f[n], \qquad 0 \le n < N, \\ f[2N-n-1], N \le n < 2N. \end{cases} \tag{2.7}$$

2) Calculate coefficients of DFT using $y[n]$: $Y[k] = DFT\{y[n]\}$;
3) Obtain DCT coefficients from $Y[k]$:

$$C[k] = \begin{cases} Y[k]e^{-j\frac{\pi}{2N}k}, 0 \le k < N, \\ 0, \qquad otherwise. \end{cases} \tag{2.8}$$

The advantage of using the described relation is that step 2) can be computed using Fast Fourier Transform (FFT). Comparison of the properties of DFT and DCT for original sequence $f[n]$ is equivalent to comparison of the properties of DFT for original sequence $f[n]$ and pseudo-sequence $y[n]$, respectively. A short conclusion is that DCT does not imply signal discontinuities while extending it and this eliminates unnecessary high-frequency components. This also means better energy compaction which explains popularity of DCT in data compression applications.

### 2.1.2 2D Discrete Cosine Transform

An image can be seen as two-dimensional signal (Figure 2.1). Therefore, utilization of 2D DCT might be more beneficial in terms of representing information in a more compact way.

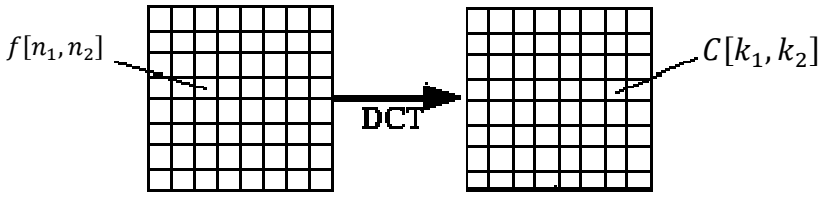



**Figure 2.1.** Transforming spatial domain of image to the domain of DCT

For $k_1 \in \{0, 1, \dots, N_1 - 1\}$ and $k_2 \in \{0, 1, \dots, N_2 - 1\}$ 2D DCT coefficients are calculated as:

$$C[k_1, k_2] = \sum_{n_1=0}^{N_1-1} \sum_{n_2=0}^{N_2-1} 4f[n_1, n_2] \cos\left(\frac{\pi}{2N_1} k_1(2n_1 + 1)\right) \cos\left(\frac{\pi}{2N_2} k_2(2n_2 + 1)\right). \quad (2.9)$$

For other values of $k_1$ and $k_2$ DCT coefficient $C[k_1, k_2] = 0$.

Inverse 2D DCT for $n_1 \in \{0, 1, \dots, N_1 - 1\}$ and $n_2 \in \{0, 1, \dots, N_2 - 1\}$ is calculated according to the following expression:

$$f[n_1, n_2] = \frac{1}{N_1 N_2} \sum_{k_1=0}^{N_1-1} \sum_{k_2=0}^{N_2-1} w_1[k_1] w_2[k_2] C[k_1, k_2] \cos\left(\frac{\pi}{2N_1} k_1(2n_1 + 1)\right) \cos\left(\frac{\pi}{2N_2} k_2(2n_2 + 1)\right), \quad (2.10)$$

where

$$w_1[k_1] = \begin{cases} 0.5, k_1 = 0 \\ 1, 1 \le k_1 < N_1 \end{cases}, \; w_2[k_2] = \begin{cases} 0.5, k_2 = 0 \\ 1, 1 \le k_2 < N_2 \end{cases}. \quad (2.11)$$

For other values of $n_1$ and $n_2$ signal samples $f[k_1, k_2] = 0$.

Alternatively, 2D DCT can be described as two-stage one-dimensional transform (Figure 2.2):

1) Create intermediate 2D sequence $\theta[k_1, n_2]$ by computing 1D DCT of rows;
2) Compute 1D DCT of columns of $\theta[k_1, n_2]$.

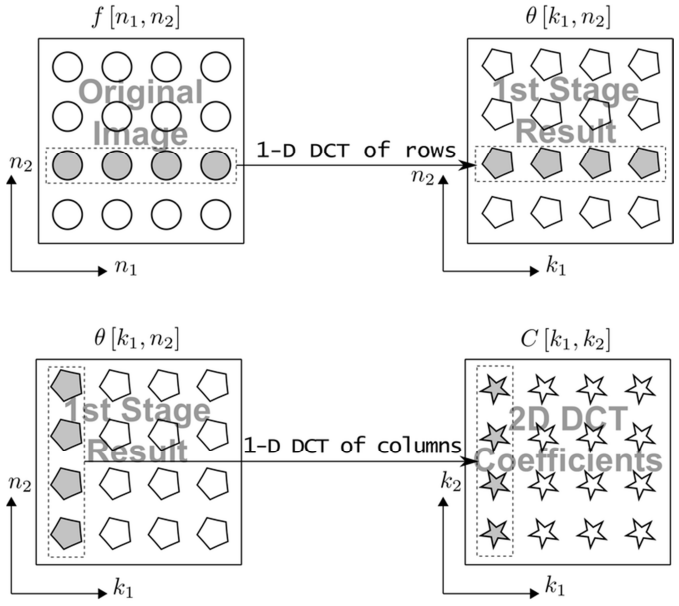



**Figure 2.2.** Two-stage 2D DCT

One of the most popular image formats JPEG is based on 2D DCT. Considerable compression can be achieved thanks to quantization of the coefficients. Different coefficients have different importance and should be quantized with different precision. The quantization

precision is expressed by quantization matrix. It is used in order to provide desirable trade-off between visual quality and compression rate [78].

## 2.2 Discrete Wavelet Transform

Alternatively to DFT or DCT in the field of Signal Processing a choice is often made in favor of DWT. A certain shortcoming of DFT is that it is not possible to tell at what instant a particular frequency rises because DFT is taken over the whole time axis. One of the existing solutions for this problem is to use a sliding window to find spectrogram, which gives the information of both time and frequency [83]. Nevertheless, another limitation for such Short-time Fourier transform (STFT) is the fact that the resolution in frequency domain is limited by the size of the window [84].

As a contrast to the earlier transforms, DWT is based on small wavelets with limited duration. This approach makes possible to analyze signal at certain periods of time using translated-version wavelets. Also, the signal can be analyzed in different scale using scaled-version wavelets. Multiresolution ability of wavelets is beneficial for watermarking. This quality improves flexibility for achievable robustness/transparency trade-off.

### 2.2.1 Principles of multiresolution analysis

In wavelet analysis, scaling function $\phi(t)$ refers to an approximated component of signal $f(t)$ while wavelet function $\psi(t)$ refers to a detailed component of $f(t)$. Orthogonal basis needs to be constructed using $\phi(t)$ and $\psi(t)$. Signal can be analyzed by translation and scaling of the constructed basis along $f(t)$. Translation and scaling are defined by $k, j \in \mathbb{Z}$, respectively:

$$\phi_{j,k}(t) = 2^{j/2}\phi\left(2^j t - k\right), \tag{2.12}$$

$$\psi_{j,k}(t) = 2^{j/2}\psi\left(2^j t - k\right). \tag{2.13}$$

Vector subspaces $V_j$ and $W_j$ will be used further in order to represent important properties of multi-resolution analysis. The mentioned subspaces are defined over corresponding approximated and detailed components, respectively. All the possible translations $k$ for the components with a fixed scaling parameter $j$ create sets $\{\phi_{j,k}(t)\}$ and $\{\psi_{j,k}(t)\}$. For each of these sets, intersection of all of its subspaces is defined as a span:

$$V_j = \mathrm{Span}\{\phi_{j,k}(t)\}, \tag{2.14}$$

$$W_j = \mathrm{Span}\{\psi_{j,k}(t)\}. \tag{2.15}$$

The major requirements for multiresolution analysis are:

1) The scaling function is orthogonal to its integer translates;

2) The subspaces spanned by the scaling function at low scales are nested within those spanned at higher scales: $V_j \subset V_{j+1}$;

3) Any function can be represented with arbitrary precision.

There are many known wavelet families, for instance, Haar, Daubechies, Coiflet, Symmlet, etc. [85]. The simplest is Haar:

$$\phi_{j,k}(t) = 2^{j/2}\begin{cases}1, & if \quad k(0.5)^j \le t < (k+1)(0.5)^j, \\ 0, & otherwise;\end{cases} \tag{2.16}$$

$$\psi_{j,k}(t) = 2^{j/2}\begin{cases}1, & if\ k(0.5)^j \le t < (2k+1)(0.5)^{j+1}, \\ -1, & if\ (2k+1)(0.5)^{j+1} \le t < (k+1)(0.5)^j, \\ 0, & otherwise.\end{cases} \tag{2.17}$$

The second requirement for multiresolution analysis can be shown to be satisfied, for instance, for Haar scaling function. It can easily be seen that the following functional relation keeps between functions of different scale $j$:

$$\phi_{0,0}(t) = \tfrac{1}{\sqrt{2}}\phi_{1,0}(t) + \tfrac{1}{\sqrt{2}}\phi_{1,1}(t)\ , \tag{2.18}$$

The generalized multiresolution analysis equation is given as:

$$\phi(t) = \textstyle\sum_{n=0}^{1} h_\phi[n]\sqrt{2}\phi(2t-n)\ . \tag{2.19}$$

Here, for Haar scaling function, $h_\phi[n] = 1/\sqrt{2}$. A simple interpretation of the equation is that function $\phi(t)$ with lower frequency components can be represented by function $\phi(2t)$ with higher frequency components and discrete lower-pass filter $h_\phi[n]$ is used for that purpose. There is a similar relation for the wavelet functions:

$$\psi(t) = \textstyle\sum_{n=0}^{1} h_\psi[n]\sqrt{2}\phi(2t-n). \tag{2.20}$$

For Haar wavelets $h_\psi[n] = (-1)^n/\sqrt{2}$ and the relation between the filters is $h_\psi[n] = (-1)^n h_\phi[1-n]$.

New subspaces can be formed using direct sum operator $\oplus$ over subspaces $V_j$ and $W_j$ obtained from corresponding scaling and wavelet functions. A union of infinite wavelet sets is equal to the $L^2(\mathbb{R})$:

$$L^2(\mathbb{R}) = V_0 \oplus W_0 \oplus W_1 \oplus W_2 \oplus \ldots \tag{2.21}$$

Requirements 2) and 3) are illustrated on Figure 2.3.

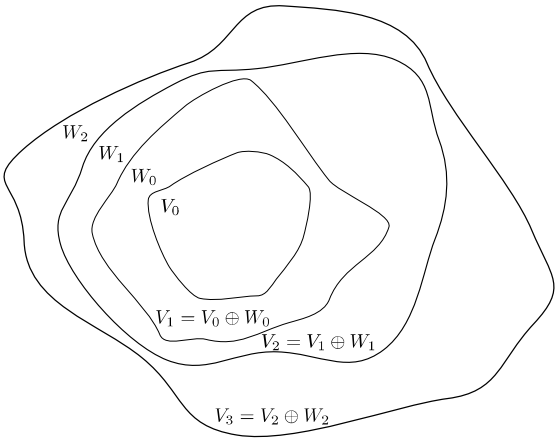



**Figure 2.3.** Scheme of relations between wavelet sets

Hence, any function in $L^2(\mathbb{R})$ can be decomposed using the scaling function and wavelet functions.

### 2.2.2 1D Discrete Wavelet Transform

A discrete signal in $L^2(\mathbb{Z})$ can be approximated using wavelets as

$$f[n] = \frac{1}{\sqrt{M}} \sum_k W_\phi[j_0, k] \phi_{j_0,k}[n] + \frac{1}{\sqrt{M}} \sum_{j=j_0}^{\infty} \sum_k W_\psi[j, k] \psi_{j,k}[n] \,, \tag{2.22}$$

where $f[n]$, $\phi_{j_0,k}[n]$, $\psi_{j,k}[n]$ are discrete functions that are defined on $M$ points $\{0, 1, \dots, M-1\}$. Wavelet coefficients are calculated as following:

$$W_\phi[j_0, k] = \frac{1}{\sqrt{M}} \sum_n f[n] \phi_{j_0,k}[n] \,, \tag{2.23}$$

$$W_\psi[j, k] = \frac{1}{\sqrt{M}} \sum_n f[n] \psi_{j,k}[n] \,, j \geq j_0 \,. \tag{2.24}$$

Coefficients $W_\phi[j_0, k]$ are called approximation coefficients while $W_\psi[j, k]$ are called detailed coefficients. There is a fast and convenient way to calculate coefficients using convolution ($*$) with coefficients of a higher order:

$$W_\phi[j, k] = h_\phi[-n] * W_\phi[j+1, n]\big|_{n=2k, k\geq 0} \,, \tag{2.25}$$

$$W_\psi[j, k] = h_\psi[-n] * W_\phi[j+1, n]\big|_{n=2k, k\geq 0} \,. \tag{2.26}$$

This approach lets to find coefficients level by level rather than calculate them directly and is called Fast Wavelet Transform (FWT) [83].

### 2.2.3 2D Discrete Wavelet Transform

In the two-dimensional case, the scaling and wavelet functions are functions of two variables that are denoted as $\phi(x,y)$ and $\psi(x,y)$, respectively, and defined as

$$\phi_{j,m,n}(x,y) = 2^{j/2}\phi\big(2^j x - m, 2^j y - n\big), \tag{2.27}$$

$$\psi^i_{j,m,n}(x,y) = 2^{j/2}\psi^i\big(2^j x - m, 2^j y - n\big),\ i = \{H,V,D\}\ . \tag{2.28}$$

In contrast to the one-dimensional case, there are three different wavelet functions, $\psi^H(x,y)$, $\psi^V(x,y)$ and $\psi^D(x,y)$. Further it is assumed that the wavelet function is separable and

$$\phi(x,y) = \phi(x)\phi(y), \tag{2.29}$$

$$\psi^H(x,y) = \psi(x)\phi(y), \tag{2.30}$$

$$\psi^V(x,y) = \phi(x)\psi(y), \tag{2.31}$$

$$\psi^D(x,y) = \psi(x)\psi(y). \tag{2.32}$$

Wavelet coefficients are calculated as

$$W_\phi(j_0,m,n) = \frac{1}{\sqrt{MN}}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1} f(x,y)\phi_{j_0,m,n}(x,y), \tag{2.33}$$

$$W^i_\psi(j,m,n) = \frac{1}{\sqrt{MN}}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1} f(x,y)\psi^i_{j,m,n}(x,y),\ i = \{H,V,D\}. \tag{2.34}$$

The original signal is synthesized in the following way:

$$f(x,y) = \frac{1}{\sqrt{MN}}\sum_m\sum_n W_\phi(j_0,m,n)\phi_{j_0,m,n}(x,y) +$$

$$+\frac{1}{\sqrt{MN}}\sum_{i=H,V,D}\sum_{j=j_0}^{\infty}\sum_m\sum_n W^i_\psi(j,m,n)\psi^i_{j,m,n}(x,y)\ . \tag{2.35}$$

If the scaling and wavelet functions are separable, the summation can be decomposed into two stages. The first step is along the $y$-axis and then to be calculated along the $x$-axis. For each

axis, FWT can be applied to accelerate the speed. The schematic diagram of this process is shown in Figure 2.4.

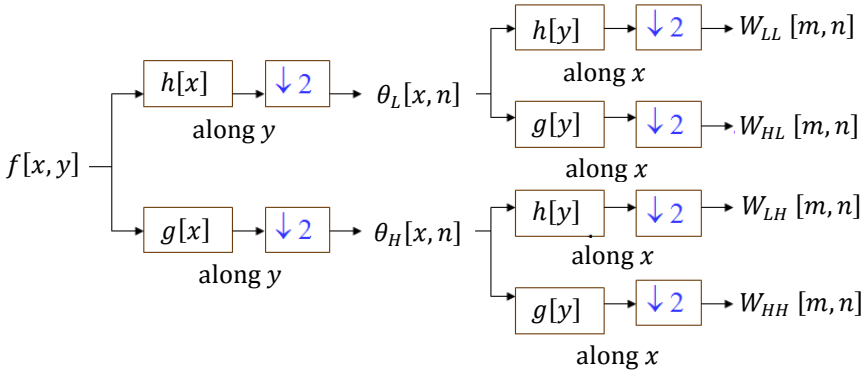



**Figure 2.4.** Diagram of two-step image decomposition, ↓ 2 denotes downsampling

Here, the purpose of considering two-dimensional signal is that it has a direct analogy to images. During the decomposition an image is divided into four bands: LL (left-top), HL (right-top), LH (left-bottom), and HH (right-bottom). The HL band indicates the variation along the $x$-axis while the LH band shows the $y$-axis variation. The power is more compact in the LL band which is of higher importance. In opposite to LL band, HH band contains information of a lowest importance for an observer. An example of wavelet decomposition of an image is shown on Figure 2.5.

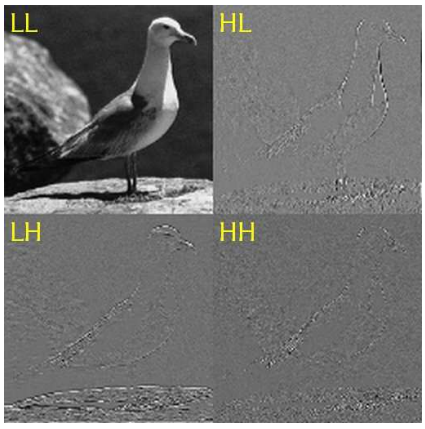



**Figure 2.5.** "Seagull" image decomposed on sub-bands

For the purposes of image processing, a famous compression algorithm, named Embedded Zero-tree Wavelet (EZW) was proposed by Shapiro [77]. Later, some modified versions of EZW have been proposed [78].

In regards to DIW, the wavelet domain is one of the most beneficial [28, 86, 87]. Under many types of popular attacks, robustness of watermarking schemes (utilizing LL, HL, LH sub-bands) is usually better than that in DCT domain. However, in some cases DCT-based watermarking methods are favorable due to the fact that in practice the usage of JPEG format is more frequent than JPEG2000.

## 2.3 Singular value decomposition

In the field of image processing, SVD plays an important role. Pixel intensities of any rectangular fragment can be seen as values of elements of a rectangular matrix. Then, a factorization of that matrix can help to represent such rectangular fragment in a very compact form. The most important components of such representation can be successfully used in DIW because they are robust to various types of noise and processing techniques that are popular in image processing [88, 89, 90].

Matrix $\mathbf{M}$ of size $m \times n$ is factorized according to SVD [91] in the following way:

$$\mathbf{M} = \mathbf{U}\mathbf{S}\mathbf{V}^{\mathrm{T}}, \tag{2.36}$$

where matrix $\mathbf{S}$ is a diagonal $m \times n$ matrix arranged in descending order (non-negative) for singular values, $\mathbf{U}$ and $\mathbf{V}$ are some orthonormal matrices with dimensions $m \times m$ and $n \times n$ (respectively) so that

$$\mathbf{U}^{\mathrm{T}}\mathbf{U} = \mathbf{I},\ \mathbf{V}^{\mathrm{T}}\mathbf{V} = \mathbf{I}\ . \tag{2.37}$$

The decomposition is not possible in case there are no $\mathbf{U}$ and $\mathbf{V}$ satisfying the previous condition. The diagonal matrix $\mathbf{S}$ can be expressed using $\mathbf{M}$, $\mathbf{U}$, and $\mathbf{V}$:

$$\mathbf{S} = \mathbf{U}^{\mathrm{T}}\mathbf{M}\mathbf{V}. \tag{2.38}$$

Calculating the SVD consists of finding the eigenvalues and eigenvectors of $\mathbf{M}\mathbf{M}^{\mathrm{T}}$ and $\mathbf{M}^{\mathrm{T}}\mathbf{M}$. The eigenvectors of $\mathbf{M}^{\mathrm{T}}\mathbf{M}$ make up the columns of $\mathbf{V}$, the eigenvectors of $\mathbf{M}\mathbf{M}^{\mathrm{T}}$ make up the columns of $\mathbf{U}$. Also, the singular values in $\mathbf{S}$ are square roots of eigenvalues from $\mathbf{M}\mathbf{M}^{\mathrm{T}}$ or $\mathbf{M}^{\mathrm{T}}\mathbf{M}$.

One of the most popular ways of performing SVD for relatively small matrices requires two equations:

$$\mathbf{M}^{\mathrm{T}}\mathbf{M} = \mathbf{V}\mathbf{S}^{\mathrm{T}}\mathbf{S}\mathbf{V}^{\mathrm{T}},\ \mathbf{U} = \mathbf{M}\mathbf{V}\mathbf{S}^{-1}. \tag{2.39}$$

The first equation is equivalent to the problem of eigen-decomposition of $\mathbf{M}^{\mathrm{T}}\mathbf{M}$. The problem of finding a column $\mathbf{v}_i$ in $\mathbf{V}$ that is an eigenvector of square $n \times n$ matrix $\mathbf{M}^{\mathrm{T}}\mathbf{M}$ is defined as

$$\mathbf{M}^{\mathrm{T}}\mathbf{M}\mathbf{v}_i = \lambda_i\mathbf{v}_i\ , \tag{2.40}$$

where $\lambda_i$ is eigenvalue. Taking into account that more than one eigenvector are needed to form $\mathbf{V}$, the following requirement needs to be satisfied:

$$p(\lambda) = \det(\mathbf{M}^{\mathrm{T}}\mathbf{M} - \lambda\mathbf{I}) = 0\ , \tag{2.41}$$

where $p(\lambda)$ is characteristic polynomial. When all the roots of $p(\lambda)$ are found, $\mathbf{S}$ is known. Columns $\mathbf{v}_i$ can be defined by solving linear systems $\mathbf{M}^{\mathrm{T}}\mathbf{M} - \lambda_i\mathbf{I}$ and normalizing the solution vectors. Calculation of $\mathbf{U}$ is straightforward.

In spite of the fact that the described procedure of SVD has just few steps, some of them might be computationally hard for large $m$ and $n$. For instance, finding roots of $p(\lambda)$ and solving $n$ linear systems are potentially heavy steps. Therefore a QR-algorithm is used in practice for eigen-decomposition. There are several modifications of QR-algorithm that are widely used [91].

The mentioned QR-algorithm is an iterative approach based on QR-decomposition. The task of QR-decomposition is to represent a real square matrix $\mathbf{A}$ (substitution $\mathbf{A} = \mathbf{M}^{\mathrm{T}}\mathbf{M}$ is utilized for further simplicity) in a way that

$$\mathbf{A} = \mathbf{QR}\,, \tag{2.42}$$

where $\mathbf{Q}$ is an orthogonal matrix and $\mathbf{R}$ is an upper triangular matrix. This can be done in various ways, for example, using Gramm-Schmidt process. Each $(k+1)$-th iteration of QR-algorithm consists of two steps: a) obtaining a new matrix $\mathbf{A}_{k+1}$

$$\mathbf{A}_{k+1} = \mathbf{R}_k\mathbf{Q}_k, \tag{2.43}$$

and, b) QR-decomposition of $\mathbf{A}_{k+1}$. According to a) it can be seen that

$$\mathbf{A}_{k+1} = \mathbf{R}_k\mathbf{Q}_k = \mathbf{Q}_k^{-1}\mathbf{Q}_k\mathbf{R}_k\mathbf{Q}_k = \mathbf{Q}_k^{\mathrm{T}}\mathbf{A}_k\mathbf{Q}_k\,, \tag{2.44}$$

which means that eigenvalues of $\mathbf{A}_k$ are the same for all iterations $k$. This is an important property because it has been shown that $\mathbf{A}_k$ converges (under some circumstances) to a triangular matrix which eigenvalues are on the diagonal.

Despite that eigen-decomposition for an arbitrary matrix can not be guaranteed, the SVD works well in practice. In most cases of DIW applications, SVD is applied to relatively small fragments of natural images. The brightness of pixels inside small regions is usually quite consistent which is a good condition for eigen-decomposition. Also, for the majority of DIW applications it is not absolutely necessary that SVD followed by inverse SVD (ISVD which is a composition of a matrix using two orthonormal and one diagonal matrices) produce the fragment identical to the original fragment of an image. Small deviations can be tolerated by most of coding approaches used in DIW. This kind of deviation might be needed, for instance, to correct a defective matrix, which can usually be done by modern SVD procedures.

As during SVD a fragment of an image is decomposed on several orthogonal matrices it can be considered a basis transform. However, in contrast to DWT and DCT, the basis is not standardized and is different for different fragments. It turns out that this feature provides good transparency for DIW applications [43, 92, 93]. Therefore, due to successful trade-off between robustness and transparency SVD is chosen as a main domain for watermark embedding in this thesis.

# Chapter 3

# Encoding approaches popular in DIW

## 3.1 Spread Spectrum encoding

Spread Spectrum (SS) watermarking is one of the oldest known in DIW encoding approaches which idea is analogous to jamming counteraction in radio (or TV) communication systems [20, 25, 94]. In a standard radio or TV communication system, the transmitter sends a signal in a relatively narrow frequency band. This technique would be inappropriate in a communication problem with a jammer, because the jammer would allocate all his power to that particular band of frequencies.

Therefore, a watermark in SS watermarking is sent by modulating secretly allocated sequences (with a broad frequency spectrum). The receiver demodulates the data using a filter matched to the secret sequences. In order to corrupt the watermark the jammer must spread his power over a broad frequency range. However, only a small fraction of that power will have an effect on communication performance.

Starting from the first SS schemes, the most significant spectrum components **s** were recommended for watermarking. For instance, in [20] the authors use 1000 largest coefficients of the DCT (excluding the DC term) as the host sequence **s**. The coefficients are then modulated by a secret sequence **p** which is generated randomly in accordance with normal distribution with zero mean. The inverse DCT is applied to the coefficients in order to obtain watermarked image in spatial domain.

In the later works, authors assume that the secret pattern $\mathbf{p}^{(m,k)}$ is defined by message $m$ and a secret key $k$. The process of modification of the host sequence **s** is

$$x_n = s_n + \gamma p_n^{(m,k)}\,, \tag{3.1}$$

where $\gamma$ is the parameter of watermark strength and defines the embedding distortion. The diagram of SS watermarking procedure is depicted on Figure 3.1. The mean-square embedding distortion is $\gamma^2\left\|\mathbf{p}^{(m,k)}\right\|^2$ and is usually the same for all $m, k$. The attacker can possibly add some noise **v** to the watermarked sequence **x** which will result in a corrupted sequence **y**:

$$\mathbf{y} = \mathbf{x} + \mathbf{v}\,. \tag{3.2}$$

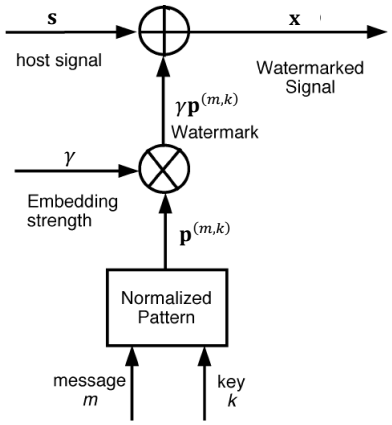



**Figure 3.1.** Spread Spectrum embedding scheme

The secret key $k$ is known to the receiver and pattern $\mathbf{p}^{(m,k)}$ can be generated for any message $m \in \mathcal{M}$ because the generator is synchronized. Due to noise, the extracted message $\hat{m}$ might be different from the original $m$. If the original (host) sequence $\mathbf{s}$ is not known, the extraction is blind and performed in the following way:

$$\hat{m} = \arg \max_{m \in \mathcal{M}} t_m(\mathbf{y}, k) , \tag{3.3}$$

where $t_m(\mathbf{y}, k)$ is the index dependent on the correlation between $\mathbf{y}$ and $\mathbf{p}^{(m,k)}$:

$$t_m(\mathbf{y}, k) = \sum_{n=1}^{N} y_n p_n^{(m,k)} , m \in \mathcal{M} . \tag{3.4}$$

In case of non-blind extraction, the following index is used:

$$t_m(\mathbf{y}, \mathbf{s}, k) = \sum_{n=1}^{N} (y_n - s_n) p_n^{(m,k)} , m \in \mathcal{M}. \tag{3.5}$$

Even for the earliest SS methods it has been shown that non-blind watermarking is quite robust to various types of processing such as additive noise, low pass filtering, cropping, lossy JPEG compression, etc. [20, 95].

One of the main disadvantages of blind detection is that component $(\mathbf{s} + \mathbf{v})$ is unknown and its energy is much higher than that of $\gamma \mathbf{p}^{(m,k)}$. This fact reduces reliability of detection procedure. Several improvements were proposed in the literature in order to diminish such a negative effect.

In some applications, the parameter of embedding strength $\gamma$ is adjusted depending on index $n$ and the embedding equation is:

$$x_n = s_n + \gamma_n(\mathbf{s}) p_n^{(m,k)}, \tag{3.6}$$

where $\gamma_n$ might depend on frequency or temporal characteristics of the host. Another solution that can be beneficial for blind SS watermarking is to pre-process original sequence **s** in a way that increases $t_m(\mathbf{x}, k)$ for right $m$ [94]. However, it is necessary to emphasize that the efficiency of blind watermarking is greatly dependent on the right expectations about parameters of the noise added by the jammer. The performance might be drastically decreased in case jammer changes his strategy and applies, for example, colored Additive Gaussian Noise (AGN) instead of white.

## 3.2 Encoding Based on Quantization Index Modulation

As a contrast to SS watermarking, QIM does not require any secret pattern **p** for information transmission. Instead, the watermarking according to QIM can solely be seen as a special kind of processing of **s** in accordance to some codebook. Sometimes this scheme is called "informed embedding" because in order to choose the right (optimal) **x** we need to know host **s** as well as message $m$. On the other hand, SS watermarking is called "blind embedding" because the choice of **p** does not depend on **s**.

### 3.2.1 Scalar Quantization Index Modulation

Further we will use the following notation for the watermark **w**:

$$\mathbf{w} = \mathbf{x} - \mathbf{s}\,. \tag{3.7}$$

It has been shown by Costa [73] that the achievable capacity of watermarking channel under AWGN **v** is

$$C = 0.5 \log_2 \left(1 + \frac{\sigma_{\mathrm{w}}^2}{\sigma_{\mathrm{v}}^2}\right). \tag{3.8}$$

Here $\sigma_{\mathrm{v}}^2$ and $\sigma_{\mathrm{w}}^2$ denote variance for noise and watermark, respectively. Further in the text, even in case the distribution in the codebook is not Gaussian, $\sigma_{\mathrm{w}}^2$ should be interpreted as variance. On the other hand, distribution of the host sequence **s** for most of the real-life signals is quite close to normal with variance $\sigma_{\mathrm{s}}^2$. The result for $C$ shows that the capacity is independent of the variance of **s**, which is a contrast to SS watermarking. For such performance it is required that the codebook is designed using parameter $\alpha$:

$$\mathcal{U}^{L_{\mathrm{s}}} = \left\{\mathbf{u}_l = \mathbf{w}_l + \alpha \mathbf{s}_l | l \in \{1, 2, \ldots, L_{\mathcal{U}}\}\right\}, \tag{3.9}$$

where $\mathbf{w}\sim\mathcal{N}\left(0,\sigma_{\mathrm{w}}^2 I_{L_s}\right)$ and $\mathbf{s}\sim\mathcal{N}\left(0,\sigma_{\mathrm{s}}^2 I_{L_s}\right)$. Here, $I$ is the identity matrix, $L_s$ is the length of $\mathbf{s}$ sequence, $L_{\mathcal{U}}$ is the total number of entries in $\mathcal{U}^{L_s}$. There exists at least one codebook such that for $L_s \to \infty$, the capacity $C$ in (3.8) can be achieved and the value of parameter $\alpha$ (providing that $C$) is defined as:

$$0 \le \alpha^* = \frac{\sigma_{\mathrm{w}}^2}{\sigma_{\mathrm{w}}^2+\sigma_{\mathrm{v}}^2} = \frac{1}{1+10^{-0.1\mathrm{WNR}}} \le 1\,. \tag{3.10}$$

In the last expression WNR is a contraction of Watermark to Noise Ratio and is commonly used to express the relation between watermark and noise energies:

$$\mathrm{WNR} = 10\log_{10}\left(\frac{\sigma_{\mathrm{w}}^2}{\sigma_{\mathrm{v}}^2}\right). \tag{3.11}$$

Watermark message $m$ can be seen as a number (or index). Also, several indices of the codebook entries can be associated with the same $m$. Such indices form group $\mathcal{L}_m = \{l_m^1, l_m^2, \dots, l_m^N\}$ so that any $l_m^i$ $(1 \le i \le N)$ has meaning of $m$. Given a particular host sequence $\mathbf{s}_0$ sending of message $m$ assumes two steps: 1) find $l_m \in \mathcal{L}_m$ such that the pair $\left(\mathbf{u}_{l_m}, \mathbf{s}_0\right)$ is jointly typical which is equivalent to the following

$$l_m = \arg\min_{l\in\mathcal{L}_m} |(\mathbf{u}_l - \alpha^*\mathbf{s}_0)^{\mathrm{T}}\mathbf{s}_0|\,; \tag{3.12}$$

2) if $\mathbf{w}_0 = \mathbf{u}_{l_m} - \alpha^*\mathbf{s}_0$ satisfies power constraint $\sigma_{\mathrm{w}}^2$ transmit $\mathbf{x}_0 = \mathbf{w}_0 + \mathbf{s}_0$. An affected by AWGN sequence $\mathbf{y}_0$ is received on the decoder side. Using the codebook $\mathcal{U}^{L_s}$, the decoder finds an index $l_{\widehat{m}}$ (associated with message $\widehat{m}$) such that the pair $\left(\mathbf{u}_{l_{\widehat{m}}}, \mathbf{y}_0\right)$ is jointly typical. Unfortunately, the described random codebook $\mathcal{U}^{L_s}$ has an infinite size and is not practical [96]. In order to solve this problem, several suboptimal, but practical implementations have been proposed. The core idea of those approaches is to construct a structured codebook $\mathcal{U}^{L_s}$.

One of the most popular watermarking methods has been proposed by Chen and Wornell [97]. The watermark encoding is done in accordance with a simple scalar quantization rule which can be shown to be easily derived from the encoding procedure described by Costa [73].

The following interpretation is proposed by the authors in [96]. They demonstrate that an appropriate $L_s$-dimensional codebook can be constructed as a concatenation of $L_s$ scalar codebooks $\mathcal{U}^1$. Moreover, they show that encoding/decoding can also be done as sample-wise (scalar) operations. Random sequence $\mathbf{k}$ from the key $K$ ($k_n \in [0,1)$) is an important component of this codebook. It is used to protect the watermark from malicious attacks (where an attacker analyses and deliberately modifies the watermarked signal) the codebook choice must be dependent on key $K$. In that case, without $\mathbf{k}$ it is almost impossible to reconstruct the codebook. Scalar codebook $\mathcal{U}^1$ is defined according to the following expression:

$$\mathcal{U}^1(k_n) = \left\{u_n = (l+k_n)\alpha\Delta + \frac{d_n}{D}\alpha\Delta \,\middle|\, d_n \in \mathcal{D}, l \in \mathbb{Z}\right\}. \tag{3.13}$$

Here, $\mathcal{D}$ is the alphabet of the watermark message and $D$ is the total number of different symbols in it. Further we assume binary alphabet where $\mathcal{D} = \{0,1\}$ and $D = 2$.

Following the encoding described by Costa, in order to embed a sequence of watermark symbols $\mathbf{d}$ a corresponding $\mathbf{u}_0$ has to be found that turns the pair $(\mathbf{u}_0, \mathbf{s})$ to be jointly typical. This is equivalent to finding a sequence $\mathbf{q} = \mathbf{w}/\alpha = \mathbf{u}_0/\alpha - \mathbf{s}$ that is nearly orthogonal to $\mathbf{s}$. This search can be considered to be a quantization of $\mathbf{s}$ with an $L_s$-dimensional quantizer. Each quantizer representative is derived from the codebook entries $\mathbf{u} \in \mathcal{U}^{L_s}$ where $\mathcal{U}^{L_s} = \mathcal{U}^1 \circ \mathcal{U}^1 \circ \ldots \circ \mathcal{U}^1$. Therefore this process can be seen as sample-wise operation

$$q_n = \mathcal{Q}_\Delta\left\{s_n - \Delta\left(\frac{d_n}{D} + k_n\right)\right\} - \left(s_n - \Delta\left(\frac{d_n}{D} + k_n\right)\right), \tag{3.14}$$

where $\mathcal{Q}_\Delta\{\cdot\}$ denotes scalar uniform quantizer with step size $\Delta$. The transmitted watermark sequence is given by

$$\mathbf{w} = \mathbf{u}_0 - \alpha\mathbf{s} = \alpha\mathbf{q}\ , \tag{3.15}$$

and the watermarked data is

$$\mathbf{x} = \mathbf{s} + \mathbf{w} = \mathbf{s} + \alpha\mathbf{q}\ . \tag{3.16}$$

Quantization error $\mathbf{q}$ (and the watermark $\mathbf{w}$) are almost orthogonal to the quantizer input $\mathbf{s}$ under an assumption that the original samples are distributed uniformly inside quantization bin. It can also be shown [96, 97] that for uniformly distributed key sequence $\mathbf{k}$ watermark $\mathbf{w}$ is statistically independent from $\mathbf{s}$. This condition is similar to the condition of the scheme proposed by Costa [73].

The decoder receives data $\mathbf{y} = \mathbf{s} + \mathbf{w} + \mathbf{v}$ and looks for an index of $\mathbf{u}(\mathbf{k})$ in codebook $\mathcal{U}^{L_s}(K)$ such that the sequences $(\mathbf{y}, \mathbf{u}/\alpha)$ are jointly typical. The index interprets the watermark message $\hat{m}$. In case of hard-decision decoding, this procedure is also equivalent to sample-wise operations:

$$z_n = \mathcal{Q}_\Delta\{y_n - k_n\Delta\} - (y_n - k_n\Delta)\ . \tag{3.17}$$

Here $|z_n| \leq 0.5\Delta$ and the value of the bit $d_n$ is defined in the following way

$$\hat{d}_n = \begin{cases} 0, if\ |z_n| \leq 0.25\Delta \\ 1, \qquad otherwise \end{cases}. \tag{3.18}$$

Soft-decision decoders can also be used to extract the watermark from $\mathbf{z}$. This may provide better results, but the decoding procedure is more complex in that case [96, 72].

### 3.2.2 Multidimensional Quantization Index Modulation

The presented QIM approach describes a scalar quantization case when a bit $d_n$ of a watermark message $m$ is embedded into a sample $s_n$. One can achieve sufficiently high capacity with scalar Distortion Compensated QIM (further referred as DC-QIM), but the limit estimated by Costa is far from being reached [73, 96]. Chen and Wornell [97] were the first who presented an extension of the scalar QIM scheme to the vector case. The idea is to replace the scalar quantizer $\mathcal{Q}_\Delta\{\cdot\}$ with $L$-dimensional VQ quantizer. We omit key $K$ for simplicity. In that case

$$\mathbf{q}_n^L = \mathcal{Q}_\Delta^L\{\mathbf{s}_n^L - \boldsymbol{\lambda}_{d_n}^L\} - (\mathbf{s}_n^L - \boldsymbol{\lambda}_{d_n}^L), \tag{3.19}$$

where, for instance, for $L = 2$ and $D = 2$ one can use $\boldsymbol{\lambda}_0^2 = (0,0)$, $\boldsymbol{\lambda}_1^2 = (0.5\Delta, 0.5\Delta)$. Quantizer $\mathcal{Q}_\Delta^L\{\cdot\}$ quantizes vector $\mathbf{s}_n^L - \boldsymbol{\lambda}_{d_n}^L$ in a sample-wise manner. The transmitted pair of samples $\mathbf{x}_n^2$ is

$$\mathbf{x}_n^2 = \mathbf{s}_n^2 + \alpha\mathbf{q}_n^2 . \tag{3.20}$$

Finally, the whole watermarked sequence is constructed by concatenation of all the pairs $\mathbf{x}_n^2$ together:

$$\mathbf{x} = \mathbf{x}_1^2 \circ \mathbf{x}_2^2 \circ \ldots \circ \mathbf{x}_N^2. \tag{3.21}$$

The decoder receives sequence $\mathbf{y}$ that might be corrupted by noise. For each pair of samples $\mathbf{y}_n^2$ the decoder's output is

$$\hat{d}_n = \arg \min_{d_n \in \{0,1\}} \mathrm{dist}(\mathbf{y}_n^2, \Lambda_{d_n}^2), \tag{3.22}$$

where $\Lambda_{d_n}^L = \boldsymbol{\lambda}_{d_n}^L + \Delta\mathbb{Z}^L$ and $\mathrm{dist}(\mathbf{y}_n^L, \Lambda_{d_n}^L) = \min_{\mathbf{p}\in\Lambda_{d_n}^L} \|\mathbf{y}_n^L - \mathbf{p}\|$. The efficiency of watermarking (e.g. robustness-transparency trade-off) greatly depends on the properties of the lattice $\Lambda$.

A general construction of $\Lambda$ is based on nested lattices and has been discussed in numerous papers [94, 98]. A lattice $\Lambda$ in $N$-dimensional Euclidean space is defined as a set of points in $\mathbb{R}^N$ such that $\mathbf{x} \in \Lambda$ and $\mathbf{y} \in \Lambda$ implies $(\mathbf{x} + \mathbf{y}) \in \Lambda$ and $(\mathbf{x} - \mathbf{y}) \in \Lambda$. Hence, $\Lambda$ has the structure of an additive subgroup of $\mathbb{R}^N$. Using $N \times N$ generator matrix $G$, lattice $\Lambda$ can be defined as a set of all possible integral combinations of basis vectors: $\Lambda = \{\mathbf{u} = \mathbf{a}G, \mathbf{a} \in \mathbb{Z}^N\}$. The choice of $G$ is not unique for a given $\Lambda$. A nested lattice consists of an $N$-dimensional lattice partition $\Lambda_f/\Lambda_c$ where $\Lambda_f$ and $\Lambda_c \subset \Lambda_f$ are fine and coarse lattices, respectively. The relation between corresponding generator matrices $G_c$ and $G_f$ for a pair of nested lattices $(\Lambda_c, \Lambda_f)$ can be expressed as

$$G_c = JG_f , \tag{3.23}$$

where $J$ is an $N \times N$ integer matrix used for subsampling. There is the following interpretation of the role of $(\Lambda_c, \Lambda_f)$ for the watermarking process: $\Lambda_c$ is used for embedding of $d_n$ while $\Lambda_f$ is used for $\hat{d}_n$ extraction.

Better performance requires higher dimensionality for $\Lambda_f$ and $\Lambda_c$. Unfortunately, this would considerably complicate quantization operations. One of the possible solutions is to utilize lattices with a special structure that can be reproduced from some low-dimensional lattices. A different idea is to use recursive quantization in a way it is done by trellis-coded quantization.

### 3.2.3 Trellis-Coded Quantization

Trellis-coded quantization (TCQ) is a special case of trellis coding [99, 100]. A set of trellises is partitioned on several subsets in order to embed a watermarking message.

According to TCQ principle, a codebook is produced using a trellis. A 4-state trellis can be defined using a finite state machine (Figure 3.2).

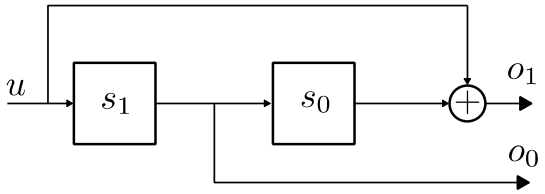



**Figure 3.2.** Finite state machine defining 4-state trellis

Here, $u$ is the input bit of the finite state machine, $s_0$ and $s_1$ are the bits that define the current state while $o_0$ and $o_1$ are the output bits. Hence, four different states are possible. There are two possible transitions from every state depending on the value of $u$. The distance between the two possible outcomes is 2. Each transition between the states is labeled by a corresponding label $\mathrm{D}_i$, $i \in \{0,1,2,3\}$ where current $i$ is defined by $o_1 o_0$ (Figure 3.3). The transitions that correspond to $u =$ "1" are depicted with dotted lines.

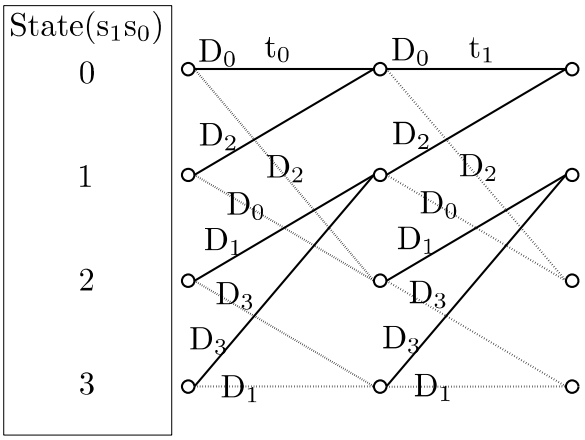



**Figure 3.3.** Transitions between states

The labels have real-valued interpretation and can be periodically reproduced on a real number line using an ensemble of quantizers (Figure 3.4).

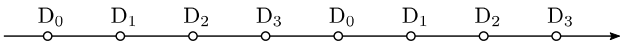



**Figure 3.4.** Labels on real number line

One of the earliest TCQ-based watermarking methods was proposed in [101]. The initial state is always assumed to be $\{0,0\}$. For each transition, $u$ is the watermark bit. A path consisting of three transitions encoding watermarking message $w = \{0,1,1\}$ is shown on Figure 3.5. The decoding should be done according to Viterbi algorithm and using the whole watermarked sequence.

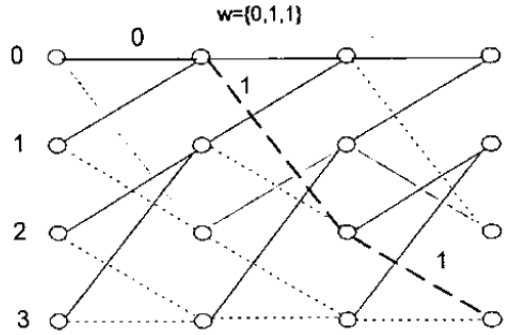



**Figure 3.5.** An example of trellis path for TCQ method in [101]

The main idea of another TCQ approach described in [99] is to associate the initial state of the machine with a watermarking message, e.g. $w = \{s_1, s_0\}$. Then, the message $w$ is encoded (by Viterbi algorithm) using labels $\{D_{i_1}, D_{i_2}, \dots, D_{i_m}\}$ of $m$ transitions. In order to label each transition, each corresponding sample of a host is quantized according to the quantizer on Figure 3.4. On Figure 3.6 it is shown how a watermark message $w = \{0,1\}$ might be encoded using one

of the possible sequences of labels $\{D_2, D_2, D_1\}$. Decoding is done using Viterbi algorithm [72]. Therefore, the decoded sequence of labels $\{\widehat{D}_{i_1}, \widehat{D}_{i_2}, \dots, \widehat{D}_{i_m}\}$ defines the extracted watermarking message $\widehat{w} = \{\hat{s}_1, \hat{s}_0\}$.

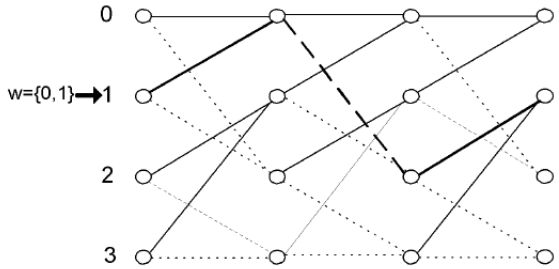



**Figure 3.6.** An example of trellis path for TCQ method in [99]

The robustness of such embedding increases with the number of transitions. In terms of robustness-transparency trade-off, the advantage of TCQ is that a given initial state can be encoded by several different sequences of labels. This enables selection of the sequence (code-word) that is closer to the host and introduces less distortion.

A different insight on watermarking and distributed source coding problems has been investigated in [100] with a particular focus on TCQ. A new design of trellis-based encoder embedding the information in the middle of input of TCQ is considered there. In comparison with previous versions of TCQ, this modification increases robustness of the embedded watermark under AWGN.

# Chapter 4

# Image quality measures

Quality of the watermarked image is an important parameter that in combination with robustness defines the efficiency of a particular DIW approach [4]. However, in contrast to robustness the question about image quality is non-trivial. Humans are the end-consumers of multimedia content and digital images as a part of it. The feedback might also be different depending on the target field. For instance, requirements for entertaining or aesthetical purposes can differ from the purposes of use in healthcare or educational organizations [62]. Moreover, it is difficult to develop a single measure that is equally successful under different kinds of image processing techniques, noise types and malicious attacks on the watermarked image. Nevertheless, the scores of different Image Quality Measures (IQMs) can be averaged and compared for different perceptual aspects. The most popular and adequate IQMs will be reviewed further in short.

IQMs can be separated into two classes: full-reference and no-reference [76]. Full-reference measures compare a watermarked image to known original directly, while no-reference methods assign a score to the watermarked image without knowing the original (in a blind manner). For instance, the simplest full-reference IQMs directly measure the pixel-by-pixel differences between two images and quantify them as the average of the squares or using a logarithm function. Such metrics have the advantage of being computationally efficient, effective for optimization purposes, and having a clear physical meaning. Despite their relative simplicity, though, these algorithms have the drawback of being unable to assess similarity across distortion types.

## 4.1 Full-reference IQMs

There are three attributes that are traditionally chosen for estimation of efficiency of IQM: accuracy, monotonicity and consistency [75]. They can be briefly explained in the following manner. First, for the same parameters controlling a particular type of noise, processing or attack the variance of the IQM estimates among the variety of images should be small. Second, for each type of distortion, IQM estimates should demonstrate a monotonic behavior in respect to parameters controlling the severity of the distortion. Finally, a suitable IQM should provide consistently accurate estimations for all types of images and not to fail badly for a subset of images. Further, some of the most adequate IQMs will be described according to the principle of comparison they utilize: pixel differences, spectral distance, Human Visual System (HVS). Notations $C_k(i,j)$ and $\hat{C}_k(i,j)$ will appear across various sub-sections further. They represent intensities of $k$-th subband pixel in the position $(i,j)$ of original and modified images, respectively.

### 4.1.1 IQM based on pixel differences

The most common IQMs in digital image processing calculate the distortion between two images on the basis of their pixelwise differences or certain moments of the difference image (alternatively called error image). Minkowski average $\varepsilon^{\gamma}$ ($\gamma \geq 1$) is one of the canonical measures in this category:

$$\varepsilon^{\gamma} = \frac{1}{K}\sum_{k=1}^{K}\left\{\frac{1}{N^2}\sum_{i,j=0}^{N-1}\left|C_k(i,j) - \hat{C}_k(i,j)\right|^{\gamma}\right\}^{\frac{1}{\gamma}}. \tag{4.1}$$

It is assumed (for simplicity) that both original and watermarked images are of square form, $N \times N$ pixels. Depending on the application, the interpretation of image sub-bands might differ, but for RGB images one can assume R, G, B sub-bands.

Mean Square Error (MSE) is another well-known and widely used IQM:

$$MSE = \frac{1}{KN^2}\sum_{k=1}^{K}\sum_{i,j=0}^{N-1}\left\|C_k(i,j) - \hat{C}_k(i,j)\right\|^2. \tag{4.2}$$

The main advantages of $MSE$ are the simplicity as well as the fact that it remains the best IQM for additive noise.

### 4.1.2 IQM based on spectral distance

This measure uses matrices $\mathbf{\Gamma}_k(u,v)$ and $\hat{\mathbf{\Gamma}}_k(u,v)$ of 2D Discrete Fourier Transform (DFT) coefficients calculated for the $k$-th sub-band of original and watermarked images, respectively. The spectra are defined as

$$\Gamma_k(u,v) = \sum_{m,n=0}^{N-1} C_k(m,n)\exp\left[-2\pi i m\frac{u}{N}\right]\exp\left[-2\pi i n\frac{v}{N}\right]. \tag{4.3}$$

Usage of the weighted phase and magnitude spectra is proven to be especially efficient. Phase and magnitude spectral coefficients $\varphi(u,v)$ and $M(u,v)$ are calculated as following

$$\varphi(u,v) = \arctan[\Gamma(u,v)], \tag{4.4}$$

$$M(u,v) = |\Gamma(u,v)|. \tag{4.5}$$

The weighted spectral distortion $S$ is

$$S = \frac{1}{N^2}\left(\lambda\sum_{u,v=0}^{N-1}\left(\varphi(u,v) - \hat{\varphi}(u,v)\right)^2 + (1-\lambda)\sum_{u,v=0}^{N-1}\left(M(u,v) - \widehat{M}(u,v)\right)^2\right). \tag{4.6}$$

For the IQM, $S$ needs to be calculated using the whole image and the authors of [75] have chosen parameter $\lambda$ equal to $2.5 \times 10^{-5}$.

However, a case with local image distortion is quite natural. Therefore, instead of the whole image, the spectral distortion can be calculated for image blocks. The obtained block measures can then be averaged using Minkowsky average. For this purpose, an image is partitioned into $L$ non-overlapping blocks of size $b \times b$ and $\Gamma_k^l(u,v)$ is calculated for $l$-th block according to

$$\Gamma_k^l(u,v) = \sum_{m,n=0}^{b-1} C_k^l(m,n) \exp\left[-2\pi i m \frac{u}{N}\right] \exp\left[-2\pi i n \frac{v}{N}\right]. \tag{4.7}$$

For the $l$-th block, measures $J_M^l$ and $J_\varphi^l$ can be defined:

$$J_M^l = \frac{1}{K} \sum_{k=1}^{K} \left( \sum_{u,v=0}^{b-1} \left[ \left| \Gamma_k^l(u,v) \right| - \left| \hat{\Gamma}_k^l(u,v) \right| \right]^\gamma \right)^{\frac{1}{\gamma}}, \tag{4.8}$$

$$J_\varphi^l = \frac{1}{K} \sum_{k=1}^{K} \left( \sum_{u,v=0}^{b-1} \left[ \varphi_k^l(u,v) - \hat{\varphi}_k^l(u,v) \right]^\gamma \right)^{\frac{1}{\gamma}}. \tag{4.9}$$

Here, $\varphi_k^l(u,v)$ and $\hat{\varphi}_k^l(u,v)$ are calculated based on spectra of $l$-th block from $k$-th sub-band of original and watermarked images, respectively.
Finally, for $l$-th block $J^l$ is computed as

$$J^l = \lambda J_M^l + (1-\lambda) J_\varphi^l. \tag{4.10}$$

Then, all the calculated $J^l$ need to be arranged in ascending order $J^{(1)}, \dots, J^{(L)}$ so that $J^{(L)} = \max_l\{J^l\}$. The median block distortion represents the distortion for the whole image and is defined as

$$S_{block} = 0.5\left(J^{(\lfloor 0.5L \rfloor)} + J^{(\lfloor 0.5L \rfloor + 1)}\right). \tag{4.11}$$

For $S_{block}$, it has been experimentally proven that $\gamma = 2$ and block size of either 32 or 64 result in a quite adequate IQM [75].

### 4.1.3 IQM based on HVS

Despite the fact that the Human Visual System (HVS) is a complex concept, there were many relatively successful attempts to estimate image quality with simplified models of HVS [75, 102, 103]. In order to obtain a closer relation with the assessment by the human visual system, both the original and coded images can be preprocessed via filters that simulate the

HVS. One of the models for the human visual system is given as a band-pass filter for 2D DCT domain with the following transfer function:

$$H(u,v) = \begin{cases} 0.05e^{\rho^{0.554}}, & if \quad \rho < 7 \\ e^{-9[|\log_{10}\rho - \log_{10} 9|]^{2.3}}, & otherwise \end{cases}, \tag{4.12}$$

where $\rho = (u^2 + v^2)^{0.5}$.

Then, calculated in that way spectral mask **H** should be applied to an image in DCT domain. The preprocessed image needs to be converted back to the spatial domain using inverse DCT. These steps can be expressed via operator $U\{\cdot\}$:

$$U\{C(i,j)\} = \text{DCT}^{-1}\{H(u,v)\Omega(u,v)\}, \tag{4.13}$$

where $\Omega(u,v)$ denotes a DCT coefficient. Several IQM can be applied to the preprocessed images. One possible choice is normalized absolute error:

$$HVS_1 = \frac{1}{K}\sum_{k=1}^{K} \frac{\sum_{i,j=0}^{N-1} |U\{C_k(i,j)\} - U\{\hat{C}_k(i,j)\}|}{\sum_{i,j=0}^{N-1} |U\{C_k(i,j)\}|}. \tag{4.14}$$

Another natural choice is $L2$ norm:

$$HVS_2 = \frac{1}{K}\sum_{k=1}^{K} \left[\frac{1}{N^2}\sum_{i,j=0}^{N-1} \left|U\{C_k(i,j)\} - U\left\{\hat{C}_k(i,j)\right\}\right|^2\right]^{0.5}. \tag{4.15}$$

Both presented HVS-based IQMs demonstrate good performance for a wide range of distortions [75]. [75]Under assumption that the impact of watermarking algorithms on original (host) image can be modelled by the means of those distortions, one can conclude that the previously detailed HVS-based IQMs can also be used for benchmarking of watermarking applications.

## 4.2 No-reference IQMs

Assessment of the quality of the watermarked image without referring to an original is a difficult task. However, most images are considered to contain some redundant information which can be recovered even after sufficient degradation. The obvious limitation for this procedure is that a complex model has to be used. Nevertheless, some simple and relatively efficient solutions are known in the literature. A suitable no-reference IQM was proposed in [102] and utilizes Fechner's Law. This IQM averages across an image the scores obtained from non-overlapping square blocks which create $k_1 \times k_2$ splitting pattern. The authors use the name "measure of enhancement" and propose several modifications of it:

$$ME^{\{l\}} = \frac{1}{k_1 k_2} \sum_{i=1}^{k_1} \sum_{j=1}^{k_2} 20 \log_{10} \left( F_{i,j}^{\{l\}} \right), \tag{4.16}$$

$$MEE^{\{l\}} = \frac{1}{k_1 k_2} \sum_{i=1}^{k_1} \sum_{j=1}^{k_2} F_{i,j}^{\{l\}} \log_{10} \left( F_{i,j}^{\{l\}} \right), \tag{4.17}$$

where $l$ can take value of either 1 or 2. Index $F_{i,j}^1$ is based on Weber's Law and can be expressed as

$$F_{i,j}^1 = \frac{\max_w I_{i,j}^w}{\min_w I_{i,j}^w}. \tag{4.18}$$

Index $F_{i,j}^2$ is motivated by Michelson contrast and can be expressed as

$$F_{i,j}^2 = \frac{\max_w I_{i,j}^w - \min_w I_{i,j}^w}{\max_w I_{i,j}^w + \min_w I_{i,j}^w}. \tag{4.19}$$

Here, $I_{i,j}^w$ is the intensity of $w$-th pixel in the block which position in the image is $(i, j)$. In total, four measures $ME^1, MEE^1, ME^2, MEE^2$ can be obtained by substituting $F_{i,j}^1$ or $F_{i,j}^2$ into (4.16) and (4.17).

# Chapter 5

# Common attacks on watermarked images

Ability to withstand attacks is the key requirement to DIW methods used in DRM systems [40, 5]. Watermark attacks are aimed at removing or destroying any watermark signals in the host data. In order to design a better and more robust watermarking technique it is important to take into account the models of attacks. A classification of the different types of watermark attacks will be provided in the next section.

## 5.1 Classification of attacks

Watermark attacks commonly classify into four distinct categories namely removal attacks, geometric attacks, cryptographic attacks and protocol attacks [4, 74, 104].

The goal of removal attacks is to distort or cancel the watermark signal in the watermarked image without breaking security of the watermarking algorithm. However, some removal attacks might be unintentional. Their harmful effect can be caused by a casual image processing procedure like, for instance, JPEG compression, but not the desire to destroy a watermark [78]. This situation can occur when a person who processes an image does not assume that it might be watermarked. Removal attack results in a damaged watermarked image as well as watermarked signal. Unfortunately in most cases no simple post processing can recover the watermark signal from the attacked data. Traditionally, noising, histogram equalization, blur and sharpen attacks as well as JPEG, SPIHT and JPEG2000 compression algorithms are in this category.

The most common geometry attacks are rotation, scaling, translation, skewing and cropping. The effect of geometry attack is in principal different from that of removal attack. If only distorted image is available, removal attack can be seen as a random impact. Hence, the amount of information that is needed to describe this impact is large. As an opposite to this, such geometry attack as, for instance, rotation can be described compactly and can be compensated at the decoder. However, the exact parameters for compensation can be difficult to estimate. Another aspect that distances removal attacks from geometry attacks is the quality of perceived image after an attack. While the effect of a removal attack can be quite noticeable (and even disturbing) to human eyes, a geometry attack can be invisible because human vision tries to compensate it (for instance, small angle rotation). This is one of the reasons why simple measures like $MSE$ are not effective for geometry attacks.

The aim of cryptographic attacks is to break the security of watermarking schemes which usually requires finding a key that was used for watermark embedding. Then, most of the watermarks can be easily removed with a very little (or no at all) harm for image quality. Alternatively, a misleading watermark can be inserted in the positions defined by the key. One of techniques in this category is a method based on the brute-force search. This technique tries to find the key by using a large number of known possible measures for meaningful secret

information. A different technique is called the Oracle attack and is used to create a non-watermarked signal when a watermark detector device is available [104].

In opposite to previously mentioned types of attack, protocol attacks do not aim at distorting the watermark signal. Instead, their goal is to add the attacker's own watermark signal in order to be able to claim the ownership. Another protocol attack is the copy attack: instead of destroying the watermark, the copy attack estimates a watermark from watermarked data and copies it to some other data, called the target data. The estimated watermark is adapted to the local features of the target data to satisfy the transparency requirement. The classification scheme for watermark attacks is shown on Figure 5.1.

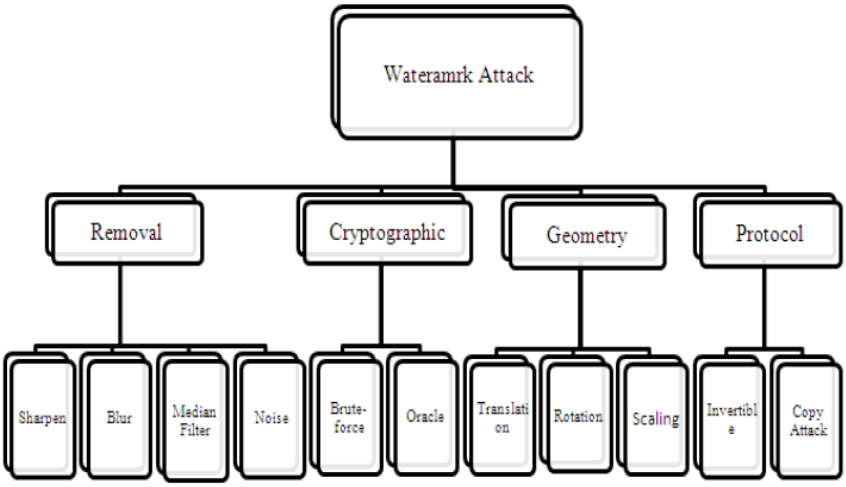



**Figure 5.1.** Classification scheme of attacks in digital watermarking

Analysis of cryptographic and protocol attacks requires assumption about the watermarking algorithm. The effect of a geometry attack also depends on the compensation ability of the decoder. On the other hand, some of the removal attacks are easy to analyze as their model is relatively simple and the effect can be measured by $MSE$. A brief analysis of several removal attacks is given in the following section.

### 5.1.1 Removal attacks

Additive White Gaussian (AWGN) noising is a process that adds a noise signal to an image in order to deliberately corrupt the image, hence reducing its visual quality [77, 78]. The name of this particular type of noise incorporates the property that it is being added to a signal as well as the feature that it has uniform power across different frequencies. Also, the property that AWGN has a normal distribution in the time domain with mean value of zero is reflected in term "Gaussian". The statistical property of this noise follows a Gaussian probability density function (PDF):

$$f(v;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(v-\mu)^2}{2\sigma^2}} . \quad (5.1)$$

Here $v$ is the realization of random variable $V$, mean $\mu$ is equal to 0 and variance $\sigma^2$ controls the severity of AWG noising. Traditionally, AWGN has been used as a channel model in which the only impairment to communication is a linear addition of wideband or white noise with a constant spectral density and a Gaussian distribution of amplitude. It produces simple and tractable mathematical models which are useful for derivations of channel capacity in case of some canonical watermarking approaches like, for instance, Spread Spectrum or Quantization Index Modulation.

In contrast to AWGN, Salt and Pepper noise is not additive. Also, it has a very different probability distribution function compared to AWGN. Its PDF takes the form of two impulse functions at two discrete locations. The effect in, for instance, grayscale images is two kinds of spots: white ("salt" noise) and black ("pepper" noise). For the case with equal fractions of "salt" noise and "pepper" noise the model can be described as:

$$p^*_{i,j} = 255\xi_{i,j}\big((i+j) \bmod 2\big) + \big(1-\xi_{i,j}\big)p_{i,j} , \quad (5.2)$$

where $p_{i,j}$ is the intensity of the original pixel in the position $(i,j)$, $p^*_{i,j}$ is the intensity after "salt and pepper", $\xi_{i,j} \in \{0,1\}$. This kind of noise occurs naturally in the sensors based on Charge-Coupled Devices (CCD) which is a major piece of technology in digital imaging. CCD image sensors are widely used in professional, medical, and scientific applications where high-quality image data is required.

Median Filtering is another kind of image processing that can be used as an attack. Commonly, it aims at reducing the presence of noise in an image, hence enhancing the image quality. However, it might cause a significant degradation of the watermark signal. The Median Filter is well-known order-statistic filter and is non-linear. The main idea of the filter is to run through the signal entry by entry, replacing each entry with the median of neighboring entries. The sliding entry-by-entry pattern of neighboring samples is called "window". For 2D signals such as images, several different window patterns are possible (such as "box" or "cross" patterns).

This kind of filtering is widely used in digital image processing because, under certain conditions, it preserves edges while removing noise. Compared to linear smoothing filters of a similar size, median filtering provides superior noise-reduction capabilities, resulting in considerably less blurring images.

Any watermark signal can be considered as a small varying signal inside the host image. The characteristics of the watermark signal can be considered equivalent to that of a noise in an image capturing equipment. Since the Median Filter is effective against many types of noise in digital images, it can potentially be harmful for the watermark as well.

Gaussian Smoothing is a yet another filter widely used in digital image processing. It aims at reducing the presence of noise in an image and shares many common properties with other smoothing processes such as Median Filtering. However, a distinct difference between Median

Filtering and Gaussian Smoothing is that the latter replaces a pixel intensity value with the weighted average of neighboring pixels. The weights are selected in accordance to 2D Gaussian distribution with particular variance. Applying a Gaussian Smoothing to an image is equivalent to convolving the image with a Gaussian function. This procedure is also known as a two-dimensional Weierstrass transform [78].

Gaussian Smoothing can be considered a low-pass filter and represents a potential threat to the watermark signal located in high-frequency domain of the host. Nevertheless, it is a common pre-processing technique, for instance, in edge detection and therefore might be an unintentional attack.

Histogram Equalization is the image processing of a different kind. While the mentioned above smoothing techniques manipulate with local characteristics of an image, Histogram Equalization is based on the statistics of the whole image. Histogram representation of pixel intensities serves for numerous spatial domain processing techniques ranging from image analysis to image enhancement. During the Histogram Equalization procedure, the pixel values are being modified in order to fit the original histogram to the desired model. As a target, different models can be used, but uniform distribution is the most common. In that case, a new (equalized) intensity of a pixel is proportional to the (corresponding) value of cumulative density function of intensities in the original image. This is an optimal fitting strategy in regards to $MSE$ criterion.

Histogram manipulation is a non-linear transform that might affect most of the sub-bands of an image. Changes in the spatial domain are even bigger. Therefore, Histogram Equalization can cause sufficient distortions of the watermark signal whatever domain is chosen. On the other hand, for a human observer the appearance of the attacked image can be better than the original.

Developments in digital image processing software made popular such a technique as sharpening [104]. Unsharp Masking is one of the most widely used procedures in that class. Conditional (or thresholded) amplification of high-frequency components is the main idea of this procedure. A high-frequency component can be extracted using a Gaussian low-pass filter. Then, it needs to be amplified in accordance with the preset amplification parameter. Finally, if the amplitude of the obtained high-frequency component is higher than the threshold, it will be added to the original image.

In practice, Unsharp Masking is used in order to enrich details of an image that might seem blurry. As one can see, the low-frequency component is not altered while the high-frequency component might be amplified. Hence, the watermark signal in a low- or high-frequency should result in either identical or amplified version of itself. However, it is impossible to predict all the settings the Unsharp Masking will be configured under: window size (and form) for a Gaussian filter might be the main source of uncertainty.

# Chapter 6

# Publications and Results

This chapter consists of six publications presenting development and progress of the research that can be characterized by a kind of SVD watermarking, from modification of orthogonal matrices to singular values; minimization of embedding distortions, from simplified approach to more advanced; watermark encoding, from scalar to multidimensional and from domain specific to domain invariant.

Publication I (PI) describes and evaluates a new SVD-based watermarking method modifying orthogonal matrices of 4x4 image blocks. This is a domain-specific watermarking and can not be utilized for the transforms other than SVD. Several embedding rules can be used to encode a watermark in $U$-matrix which is first fitted to a proposed model of an orthogonal matrix according to a relatively simple procedure to minimize embedding distortions. Compared to a method described in [105], the robustness for the proposed method was higher under the condition of JPEG compression and in some cases outperformed the reference method for more than 46%.

A more advanced multi-step procedure of minimization of embedding distortions is proposed in Publication II (PII) for watermark encoding by modification of orthogonal matrices of SVD. Two embedding rules have been proposed for watermarking which provide different robustness and transparency. A wider range of popular attacks has been applied as well as more state of the art SVD-based watermarking methods participate in the comparison. According to the results the robustness of the proposed watermarking method toward some attacks is more than 36% higher.

Publication III (PIII) represents another SVD-based watermarking method modifying orthogonal matrices that uses quite advanced procedure to minimize embedding distortions. Similarly to the previously proposed methods this one is domain-specific and can operate only with SVD. The method modifies both left and right orthonormal matrices in order to embed a bit of a watermark using Van Elfrinkhof's rotational model. A new embedding rule with adjustable parameters has been proposed for watermarking. In addition, a criterion of watermarking performance has been suggested for adaptation during embedding. The method demonstrates better robustness toward some attacks in comparison with other known blind watermarking methods.

A different scalar watermarking approach based on quantization of singular values of 4x4 blocks has been described and evaluated in Publication IV (PIV). A concept of Initial Data Loss (IDL) is introduced in order to increase robustness under high intensity AWGN. The method also exploits a new form of distribution of quantized samples which creates a distinctive feature and two criteria are proposed in order to express it numerically. Further, the criteria are utilized by a procedure for estimation of a gain factor after possible Gain Attack (GA). Compared to other well-known quantization methods, the proposed method is superior under different types of popular attacks where it demonstrates up to $10^4$ times better performance.

Publication V (PV) contains some theoretical results defining robustness under AWGN for the watermarking approach first presented in Publication IV. The quantization model is described in a formal way using first order predicate logic. For that purpose, conditions that are important for watermark embedding are discussed and formalized in several key logical expressions. Especial attention is paid to the case when the watermarked image is attacked by GA which intensity is not known to the sender and a set of constant parameters providing sufficiently good performance were proposed. Experiments were conducted in the domain of SVD, however, the watermarking method is domain invariant and can be used with other transforms.

A new approach of Distortion Compensation (DC) for the two dimensional Quincunx Lattice Quantization is described in Publication VI (PVI). The choice of a new direction of quantization is explained by the form of Voronoi cell of the lattice elements. Parameter $\alpha$ controls DC and can be adjusted to achieve a better performance in a noisy channel. An experimental evaluation of robustness under AWGN is conducted using first singular values of 8x8 blocks of natural images and compared with several other known two-dimensional Lattice as well as Scalar Quantization methods with conventional DC.

The diagram presented on Fig. 6.1 provides an overview for application for the methods detailed in the thesis. It is assumed that original image $I$ has dimension of $m \times n$. Watermarking data can always be considered as a bit sequence $\mathbf{D}_W$ which should be embedded into $I$ according to secret key $K$. Because of the requirement to be robust, a watermark bit is ought to be embedded into a block which is not smaller than $4 \times 4$ in size. This implies that small images might not be suitable to incorporate $\mathbf{D}_W$ and this condition should be checked first. Then, image is divided on $4 \times 4$ blocks (their selection might be made according to $K$, or optionally $K$ can be used to scramble $\mathbf{D}_W$) and SVD for each block is performed. Next, the choice has to be made on what component of the decomposition will be modulated in order to embed $\mathbf{D}_W$. This is the step where the distinction between the methods first appears. The methods detailed in Publications I-III embed $\mathbf{D}_W$ using orthogonal matrices, while the methods in Publications IV-VI quantize singular values (SV).

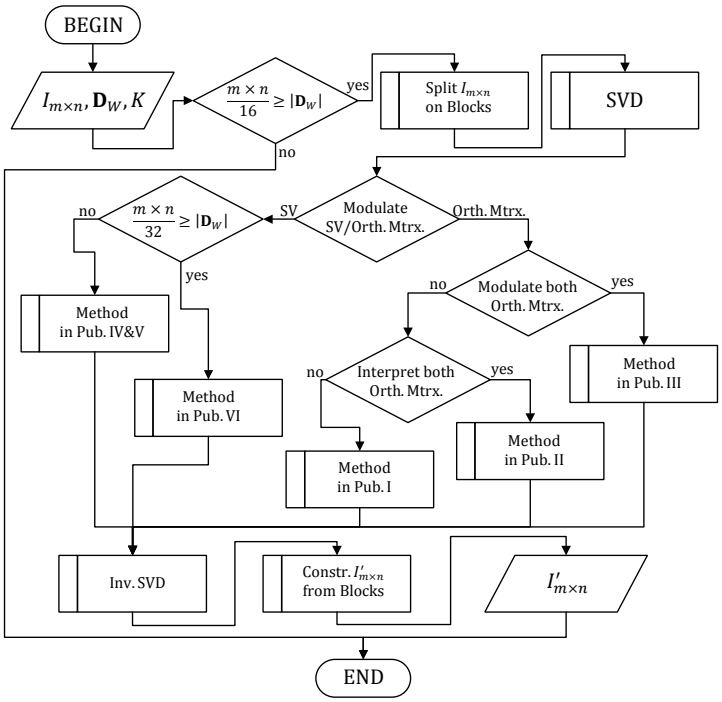



**Figure 6.1.** Application diagram for DIWmethods included in the thesis

The both left and right orthogonal matrices might be interpreted for watermark extraction (for instance, Rule 1 in PII), but not necessary need to be modified to embed data. The method in PI modifies and interprets elements of only one matrix.

In case SV are chosen to be quantized, one might decide between scalar or vector quantization approaches. The method described in PVI is an example of simple but efficient vector quantization approach where a bit of $\mathbf{D}_W$ is embedded into two samples (one sample is modified but the both are interpreted). This requires double quantity of blocks compared to scalar approach. Therefore, condition $m \times n \geq 32|\mathbf{D}_W|$ should hold in order to apply this method (8x8 block size has been mentioned in PVI, but one can use 4x4 as well).

After embedding is completed, an inverse SVD needs to be performed. In case SV were modulated (according to the methods presented in PIV, PV and PVI), original left and right orthogonal matrices of the corresponding blocks have to be used to compose watermarked 4x4 blocks. However, if orthogonal matrices were modified (according to the methods described in PI, PII and PIII), SVs of the corresponding blocks may be adjusted in order to reduce differences between original and watermarked image blocks.

Finally, the watermarked image $I'_{m\times n}$ has to be constructed from 4x4 watermarked blocks where the position of each block may be dependent on $K$. The obtained image contains watermark $\mathbf{D}_W$ and can be transmitted in a robust way over communication channel which is characterized by some type of noise like, for instance, AWGN.

## 6.1 Publication I: A Simple Model of Orthogonal Matrix for Low-Distortion Watermarking

This is the first article of the thesis that is inspired by the idea of new SVD-based watermarking presented in [105, 106]. The article is focused on the modification of the idea of manipulation with orthonormal matrix of image blocks. The goal of the paper is to confirm in experimental way that the proposed modification is indeed efficient for digital image watermarking tasks.

The idea of SVD-based watermarking presented in [105] can be described as follows. In order to embed a watermark bit in a 4x4 image block $B$ it needs to be decomposed:

$$B = \sum_i S_{i,i} \cdot U_i \cdot V_i^T \ ,$$

where $U_i$ and $V_i$ are the columns of the left and the right orthonormal matrices, respectively. The watermarking is utilized by modifying the 2-nd and the 3-rd elements in $U_1$. The modified matrix $U^*$ and original matrices $S$ and $V$ are composed back into modified block $B^*$:

$$B^* = U^* \cdot S \cdot V^T \ .$$

The motivation behind this approach is that image block component $B_1^* = S_{1,1} \cdot U_1^* \cdot V_1^T$ is the most important and robust component of $B^*$as $S_{1,1}$ is the largest singular value. If some attack occurs it will most likely be localized in $B_2^*$, $B_3^*$ or $B_4^*$ as those components are considered to represent higher frequency domain. In addition, modification of $U_1$ is preferential over traditionally manipulated $S_{1,1}$ because: a) $U_1$ is a vector and this fact provides potentially more efficient embedding as a contrast to a scalar $S_{1,1}$; b) matrix $U^*$ is normalized which makes it invariant to a popular multiplication (scaling) attack called Gain Attack (GA).

However, the mentioned papers do not address several important issues. First, the modified matrix $U^*$ is not orthonormal in general. The authors propose to modify the elements of the first column of (originally orthonormal) matrix $U$, but they do not provide any description of an approach that will adjust the rest of the columns. After composing $B^*$, there is no evidence that SVD of $B^*$ will produce $U^*$, $S$ and $V$. Instead, the result of SVD can be different which will cause errors in the extracted watermark even though no attack is applied. Second, only one embedding rule is explored. Two elements out of four are modified in $U_1$ while one might modify all the four. Moreover, inequalities of a higher order can be used as an embedding rule which can potentially improve robustness-transparency trade-off for the method under different attacks.

In Publication I, based on the idea presented in [105] we introduced two-stage watermarking with optimization. A simple model of orthogonal matrix $A$ was proposed:

$$A = \begin{bmatrix} -a & c & d & b \\ d-b & a & c & \\ b & d-c & a & \\ c & a & b-d & \end{bmatrix}.$$

On the first stage, all the four elements $a, b, c, d$ of $A$ need to be adjusted in order to approximate $U$. In the paper, this approximation is represented as optimization. For the purpose of watermarking, five new embedding rules were proposed:

$$L1_4: (-1)^b \cdot \left( \left\| \left( U^*_{1,1}, U^*_{2,1} \right) \right\|_1 - \left\| \left( U^*_{3,1}, U^*_{4,1} \right) \right\|_1 \right) \geq T;$$

$$L2_4: (-1)^b \cdot \left( \left\| \left( U^*_{1,1}, U^*_{2,1} \right) \right\|_2 - \left\| \left( U^*_{3,1}, U^*_{4,1} \right) \right\|_2 \right) \geq T;$$

$$L\infty_4: (-1)^b \cdot \left( \left\| \left( U^*_{1,1}, U^*_{2,1} \right) \right\|_\infty - \left\| \left( U^*_{3,1}, U^*_{4,1} \right) \right\|_\infty \right) \geq T;$$

$$L1_2: (-1)^b \cdot \left( \left\| U^*_{2,1} \right\|_1 - \left\| U^*_{3,1} \right\|_1 \right) \geq T;$$

$$L2_2: (-1)^b \cdot \left( \left\| U^*_{2,1} \right\|_2 - \left\| U^*_{3,1} \right\|_2 \right) \geq T.$$

Here, $T$ is the positive threshold which influences the trade-off between robustness and transparency of the watermark; $b$ is the value of the watermark bit.

Each rule has been approbated independently during the experiment. On the second stage, according to each rule, the first column of $A$ has been modified while minimizing the amount of change. As one can see, after modification of the first column, orthogonality can be preserved by updating the corresponding elements in the remaining columns. This guarantees correct extraction of the watermark. The two stages of optimization could, definitely, be combined in a single optimization task for a better result, but in that case complexity would rise significantly.

During the experiment, several grayscale images of dimension 512x512 were chosen for watermarking with 1024 bit sequence. After watermarking, the images were separately attacked by AWGN, Speckle Noise, Salt and Pepper, JPEG. Upon watermark extraction, Bit Error Rate (BER) was calculated for each embedding rule and compared with the BER of the method described in [105]. The embedding carried out (threshold $T$ was correspondingly adjusted) in a way that the quality of the watermarked images was comparable with the images watermarked according to [105]. For all the mentioned embedding rules, the proposed modification of the watermarking method demonstrated sufficiently better performance under JPEG compression attack. Rule $L2_4$ is especially advantageous in that sense and for some images demonstrates 46% improvement. For the rest types of attacks, the method does not have any sufficient advantage.

## 6.2 Publication II: An Advanced Model of Orthogonal Matrix for Watermarking by Multiplication and Multi-Step Distortion Reduction

This paper is the natural continuation of the previous one with a higher emphasis on the problem of reducing embedding distortions. In the experimental part, we have extended the comparison of watermarking performance by including some of the well-known SVD-based methods. Also, the variety of attacks on watermarked images has been increased which provides a better overlook on the efficiency of the proposed method.

As the previous paper, this paper targets the aspect of digital watermarking by modifying an orthonormal matrix of SVD of image blocks. However, there are two aspects that differ the current paper from the preceding one. First, different embedding rules were used and one of the rules takes into account the first columns of both left and right orthonormal matrices. Second, a watermark bit is embedded by multiplying one of the original orthonormal matrices with several other orthogonal matrices. As a contrast, in the previous paper the original orthogonal matrix was replaced with the approximated one.

The two following rules were proposed for watermark embedding and extraction:

$$Rule\#1:\begin{cases} Embedding:\ (-1)^{bit}\, Y'_{k,1}Y'_{k,2} = Max * min(Th, Max); \\ Extraction: bit = \left(2 + sign\left(Y'_{k,1}Y'_{k,2}\right)\right) mod\ 3. \end{cases}$$

$$Rule\#2:\begin{cases} Embedding:\ (-1)^{bit}\, Y'_{k,1} = Th; \\ Extraction: bit = \left(2 + sign\left(Y'_{k,1}\right)\right) mod\ 3. \end{cases}$$

$$Max = max\left(\left|Y_{k,1}\right|,\ \left|Y_{k,2}\right|\right),$$

$$Y_{k,1} = \boldsymbol{u}_1\boldsymbol{ref}_1 - m',$$

$$Y_{k,2} = \boldsymbol{v}_1\boldsymbol{ref}_2 - m'',$$

$$Y'_{k,1} = \boldsymbol{u}'_1\boldsymbol{ref}_1 - m',$$

$$Y'_{k,2} = \boldsymbol{v}'_1\boldsymbol{ref}_2 - m'',$$

where $Th$ stands for positive threshold; $\boldsymbol{u}_1 = (U_1)^T$, $\boldsymbol{v}_1 = (V_1)^T$, $\boldsymbol{u}'_1 = (U_1^*)^T$, $\boldsymbol{v}'_1 = (V_1^*)^T$; $\boldsymbol{ref}_1$ and $\boldsymbol{ref}_2$ are some reference vectors (columns) for the left and the right orthonormal matrices, respectively; $m'$ and $m''$ are the mean values of $\boldsymbol{u}_1\boldsymbol{ref}_1$ and $\boldsymbol{v}_1\boldsymbol{ref}_2$, respectively. The first rule depends on the first columns of the both $U^*$ and $V^*$, while the second rule takes into account only the first column of $U^*$.

A new procedure of watermarking minimizes embedding distortions. It utilizes the following property: result of multiplication of (two and more) orthogonal matrices is an

orthogonal matrix too. For a given (original) orthogonal matrix, another (modulating) orthogonal matrix can be found so that their multiplication produces a matrix with the desired first column. This column should satisfy condition (according to one of the rules) in order to embed a particular bit of the watermark. Either left- or right-isoclinic modulating matrix is calculated for this task. The resulting matrix can be multiplied further with another modulating matrix. The purpose of the second and the third multiplications is to reduce distortions introduced by the first multiplication. However, the mentioned steps should not change already modified first column of the resulting matrix. Euler-Rodrigues matrices are used on these stages. Each corresponding singular value is adjusted after each multiplication in order to minimize distortion.

For the experiment, we have chosen 16 grayscale images of dimension 512x512. The watermark had 512 bits in length. The methods proposed in [89, 92, 106] were compared under requirement that the quality of the watermarked images is nearly similar. The watermarked images were attacked with the following attacks: AWGN, Salt & Pepper, JPEG, Median Filtering, Cropping and Rotation. Compared with the other state of the art methods, the proposed method performs reasonably good. Its robustness is the best under Cropping and Rotation attacks. Also, under AWGN, Salt & Pepper and JPEG the proposed method (Rule #1) demonstrates negligibly worse robustness compared to the best achieved rate for other methods. However, it should be noticed that PSNR of the watermarked images is always slightly higher than for other methods. Unfortunately, under Median Filtering the robustness of the proposed method is significantly worse than the best score achieved according to [89], but is still better than for the methods from [92, 106].

## 6.3 Publication III: Criterion-Based Multiplicative Watermarking of Orthogonal Matrix

The article develops further the idea of SVD-based watermarking by multiplying original orthogonal matrices $U$ and $V$ with rotation orthogonal matrices defined according to Van Elfrinkhof's formulae. Here, in contrast to the previous paper, in order to satisfy an embedding rule matrices $U$ and $V$ can be multiplied with the corresponding rotation matrix only once. Hence, this process is described as a constrained minimization of embedding distortion. In addition, a criterion of watermarking efficiency is proposed that can be used for the purpose of adjustment of the threshold parameter $Th$.

The distortion $G$ introduced by watermarking 4x4 fragment $I_k$ is defined as

$$G = \|I'_k - I_k\|_2^2 .$$

Here $I'_k = U'S'(V')^T$, $U' = R_U U$, $V' = R_V V$. The orthogonal matrices $R_U$ and $R_V$ can be defined according to Van Elfrinkhof's formulae:

$$R = \begin{pmatrix} ap - bq - cr - ds & -aq - bp + cs - dr \\ bp + aq - dr + cs & -bq + ap + ds + cr \\ cp + dp + ar - bs & -cq + dp - as - br \\ dp - cq + br + as & -dq - cp - bs + ar \\ -ar - bs - cp + dq & -as + br - cq - dp \\ -br + as - dp - cq & -bs - ar - dq + cp \\ -cr + ds + ap + bq & -cs - dr + aq - bp \\ -dr - cs + bp - aq & -ds + cr + bq + ap \end{pmatrix},$$

where $a, b, c, d, p, q, r, s$ are reals and $a^2 + b^2 + c^2 + d^2 = 1$, $p^2 + q^2 + r^2 + s^2 = 1$. Further, it has been shown that $G = \|U'S'(V')^T - I_k\|_2^2 = \|US'V^T - R_U^T I_k R_V\|_2^2$. A diagonal matrix has been represented as $S' = S + \Delta S$ where $\Delta S$ is also a diagonal matrix. The embedding distortion is therefore:

$$G = \|U\Delta S V^T + I_k - R_U^T I_k R_V\|_2^2 .$$

In order to simplify the optimization we switched to a suboptimal solution by minimizing $G^* = \|I_k - R_U^T I_k R_V\|_2^2$ while showing that it is always possible (and simple) to find such $\Delta S$ that $G \leq G^*$. The embedding rule is $(-1)^{bit}(\boldsymbol{u} R_U^T Ref R_V \boldsymbol{v}^T - m) = Th$, where $Ref$ is some 4x4 matrix. The complete description of the constrained optimization task is therefore:

$$\begin{cases} G^* = \|I_k - R_U^T I_k R_V\|_2^2 \to min; \\ (-1)^{bit}(\boldsymbol{u} R_U^T Ref R_V \boldsymbol{v}^T - m) = Th. \end{cases}$$

In general, this is a non-linear Least Squares (LS) task. In order to achieve sufficiently good results, one might reduce its complexity (nonlinearity). This can be done by looking on $M = R_U^T Ref R_V$ first. If the matrices $R_U^T$, $R_V$ are left- or right-isoclinic, elements of matrix $M$ in general are not linear expressions. However, if $Ref$ is orthonormal then $M$ is as well. We might constrain $M$ to some model of orthogonal matrix like, for instance, matrix $A$ in the Publication I, which means that the constraint can be linear. Further, we might propose some model for $R_U$ (with linear expressions for its elements). Therefore, matrix $R_V$ is defined as $R_V = Ref^T R_U M$. The goal function $G^*$ is equivalent to $G^{**}$:

$$G^{**} = \|I_k M^T - R_U^T I_k Ref^T R_U\|_2^2 \to min .$$

The constraint depends on $M$ and does not depend on $R_U$. So, we can first minimize $G^{***} = \|I_k M^T - I_k Ref^T\|_2^2 \to min$ (with constraint). Then, we adjust $R_U$ with fixed $M$ in order to reduce $G^{**}$. One can see that it is always possible to obtain $G^{**} \leq G^{***}$ (according to the mentioned models of rotation matrices, it is always possible to turn $R_U$ to be identity). This is a suboptimal procedure, but the significant advantage is that orthonormal $Ref$ allows keeping it relatively simple.

In the paper, a heuristic criterion $C$ has been proposed in order to estimate the degree of efficiency of embedding for every block. The criterion combines aspects of invisibility and robustness (the trade-off is controlled by $\alpha$ and $\beta$). For the robustness aspect, the assumption is that some of the distortions can be modelled by some additive pattern which is defined by four coefficients. This might be seen as a limitation (because the dimension of the block is 4x4=16 pixels), but on the other hand the well-known fact is that many distortions are localized in a part of the spectrum (usually high frequency). Using $C$ one might adjust $Th$ during embedding for a better performance.

During the experiment, five different orthonormal matrices were considered for $Ref$. A matrix that provides minimum variance for the term $\boldsymbol{u}'Ref\boldsymbol{v}'^{T}$ was chosen in order to reduce embedding distortions. Different length watermarks were embedded in 512x512 grayscale images: 64 bit long watermark was embedded repeatedly 8 times, while 512 bit long watermark was embedded without repetition. The same attacks as in the previous paper were applied to the watermarked images. Upon watermark extraction, BER for the proposed method was usually lower under JPEG (non-redundant embedding case) and cropping attacks than that for [89, 106]. A reasonable assumption is that the further improvement of the method is possible if the size of the block is increased. However, the model of embedding (with minimization of distortions) that preserves orthogonality of the modified matrices can become prohibitively complex. This was one of the main reasons to move toward quantization technique in our subsequent watermarking experiments.

## 6.4 Publication IV: Lossy Scalar Quantization in Asymmetric Manner

This is our first paper targeting some of the problems related to watermarking using scalar Quantization Index Modulation (QIM) with Distortion Compensation (DC). In the paper, two kinds of attacks are mainly assumed: AWGN and GA. In order to increase efficiency of DC-QIM in the case of AWGN, we proposed a new model of DC under which Initial Data Loss (IDL) is possible. IDL means that a part of the watermarking data is interpreted wrongly during embedding (initially and deliberately), but no distortion is introduced to an image for this part of data. Also, an asymmetric form of the distribution of quantized samples serves as a distinctive feature which is used by the proposed procedure of recovery after GA. The robustness of the proposed watermarking scheme under the mentioned attacks is compared with the state of the art quantization approaches like canonical scalar DC-QIM and RDM [97, 107].

In the case of AWGN, it has been shown that some of the parameters of the model of canonical DC-QIM comply with the requirements of the upper capacity limit model described by Costa [96, 73]. However, such requirements like, for example, infinite size of the codebook can not be satisfied which reduces the actual capacity in practice. We try to solve this problem in a different way: a given original sample can be modified only if there is a “good enough” code-word for this purpose. The decision about encoding for every sample (micro level) is based on the estimated robustness-transparency trade-off for this sample. While transparency (embedding distortion) is easy to estimate, robustness can not be estimated analytically for DC-QIM.

Nevertheless, a good practice is to assign different levels of robustness for different code-words in the codebook (or rank them). The samples that are not quantized cause IDL. In addition, the whole codebook is re-designed (modifications on the macro level) which causes quantized samples to be distributed in a different to canonical DC-QIM way (Figure 6.2).

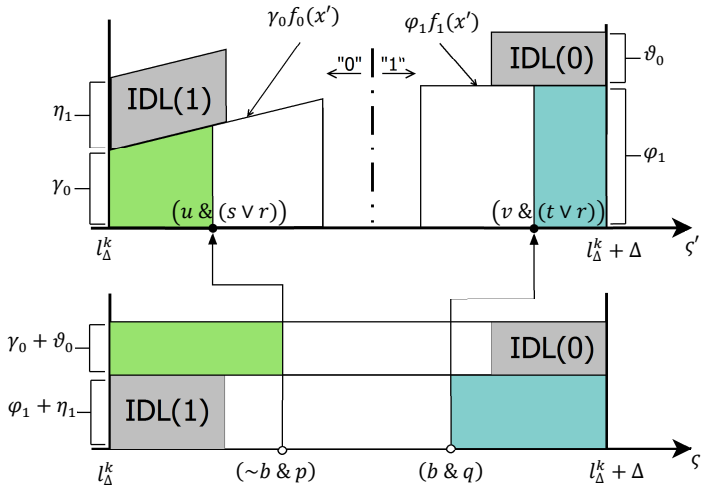



**Figure 6.2.** Illustration of the process of separation

Distributions of samples for a single embedding interval are depicted before (bottom) and after (top) quantization. The length of the interval is Δ, its left endpoint is $l_\Delta^k$. Distributions of the quantized samples coded as "0" and "1" are given by $f_0(x')$ and $f_1(x')$, respectively. Distributions of the samples that will not be modified but are labeled as "0" and "1" are denoted as IDL(0) and IDL(1), respectively. The fractions of the quantized samples that will be distributed according to $f_0(x')$, $f_1(x')$, IDL(0), IDL(1) are $\gamma_0$, $\varphi_1$, $\vartheta_0$, $\eta_1$, respectively. For further explanations of the watermarking scheme, please, refer to the original article. There are two main consequences of IDL and asymmetric distribution of the quantized samples. First, because some of the samples are not quantized, the embedding distortion is lower. In watermarking, the severity of the attack is measured by Watermark to Noise Ratio (WNR), $\mathrm{WNR} = 10\log_{10}(D/\sigma_n^2)$, $D$ is embedding distortion, $\sigma_n^2$ is the variance of AWGN. For a given value of WNR, lower $D$ implies lower $\sigma_n^2$ (less severe attack). This means that the samples that are modified (non-IDL) have more chances to be extracted correctly after an attack. Second, the asymmetric distribution is a distinctive feature and will most likely remain asymmetric even after quite a severe AWGN.

The second feature is used to recover the watermark signal after GA. For this purpose, we have proposed a procedure that utilizes one of the criteria of asymmetry of the distribution of attacked samples. As the result of GA, all the samples are multiplied with a constant that deviates from 1 very slightly, but is not known to the receiver. Therefore, extraction of the watermark with original Δ is not efficient and a new $\tilde{\Delta}''$ has to be estimated. Asymmetry of the distribution is a suitable indicator of the right estimate, but if the estimation is wrong the distribution appears nearly uniform. In order to measure asymmetry, we proposed two different criteria that demonstrate similar performance. For a range of possible estimates, the procedure retrieves

different distributions. Finally, the estimate that provides the highest value for a criterion will be chosen for $\tilde{\Delta}''$. For instance, odd central moments might serve as possible realizations for the criterion that is sensitive to asymmetry in the distribution. Despite the procedure calculates a criterion for a range of possible estimates, the asymptotic complexity is $O(n)$ because the number of possible estimates is much smaller than the number of samples $n$.

The robustness of the proposed watermarking technique was verified under AWGN and GA. The results were compared with the results of conventional DC-QIM, RDM and other popular QIM-based methods [108, 109]. For the experiment, 87 natural grayscale images with resolution 512x512 were selected. The images were split on 4x4 blocks and their first singular values were quantized. For the proposed watermarking method, we have conducted experiments with IDL technique as well as without it. After attacking, watermark message was extracted from the images and the mutual information between embedded and decoded messages was calculated. It should be noted that term "capacity" in the paper need to be replaced with mutual information between original message and the message decoded by a particular decoder. Different decoders might provide different performance. In the paper we experiment with (two options for) hard-decision decoder as this is the most practical. In one of its variants, the decision regions remain the same for any kind (and severity) of the attack. The other variant defines decision regions based on the median of the distribution of the attacked samples inside embedding interval. According to the results, our method with IDL performs up to $10^4$ times better than DC-QIM under AWGN. In addition to that, thanks to the proposed procedure of recovery it performs up to $10^3$ times better than RDM under GA.

## 6.5 Publication V: A Convenient Logical Framework for New Scalar DC-QIM

In this conference paper, we present research results following the idea of modification of DC-QIM described in the previous article. First order predicate logic is used with the aim to provide more straightforward description for the quantization model. Besides that, new analytical results were obtained for the proposed quantization model. The obtained results can be used during the experiment in order to reduce the amount of computations or to increase the precision.

First order predicate logic provides simple yet efficient descriptive and reasoning tools. We believe that the ideas of modification of DC-QIM become straightforward. Moreover, the model of quantization expressed using first order predicate logic can be used as a framework for development of new watermarking methods. A suitable example illustrating this side of our model is to substitute $f_0(x')$ and $f_1(x')$. Obviously, the arguments remain valid and the only task is to provide they are sound. The robustness aspect is one of the most important, but, as it was mentioned above, in case with DC-QIM an optimal distribution of quantized samples is not feasible to derive analytically. The expressions for $f_0(x')$ and $f_1(x')$ were proposed in our paper in a heuristic manner. Other distributions of the quantized samples might provide better robustness and our model can be used to explore this aspect.

It is assumed throughout the paper that the distribution of original samples inside embedding interval is uniform. Following this concept, it has been shown that error rate in case of AWGN

attack depends only on $\Delta/\sigma_n$ and $\boldsymbol{\theta}$, where $\boldsymbol{\theta}$ is the vector of parameters defining distribution of quantized samples. For the experimental evaluation of robustness, this result is useful because for a given $\boldsymbol{\theta}$ the error rate can be calculated using one parameter ($\Delta/\sigma_n$ instead of $\Delta$ and $\sigma_n$ substituted separately). This will reduce the number of computational cycles without sacrificing the quality of the result. Further, the result can be used for the estimation of parameters of GA. The proposed procedure of GA recovery is efficient for watermark extraction, but one might want to estimate $\sigma_n$ as well. Such an option might be in a special demand in semi-fragile or fragile watermarking applications [45, 50]. A possible solution is to include in the watermark some (relatively short) pilot sequence known to the both sender and receiver. After GA, the watermark is extracted. The error rate can be calculated using the pilot sequence. Now, if $\boldsymbol{\theta}$ is agreed between sender and receiver, the latter can estimate $\tilde{\Delta}''/\sigma_n$ and, finally, calculate $\sigma_n$ (because $\tilde{\Delta}''$ is already estimated by the procedure of GA recovery).

Achieving the highest possible robustness requires adjusting $\boldsymbol{\theta}$ for different WNR. On the other hand, strict constraint agreed between sender and receiver about $\boldsymbol{\theta}$ might cause a lower performance for the proposed watermarking scheme. We have investigated this issue and compared the results under GA for the proposed method NS-QIM, constrained version NSC-QIM and state of the art RDM (Figure 6.2).

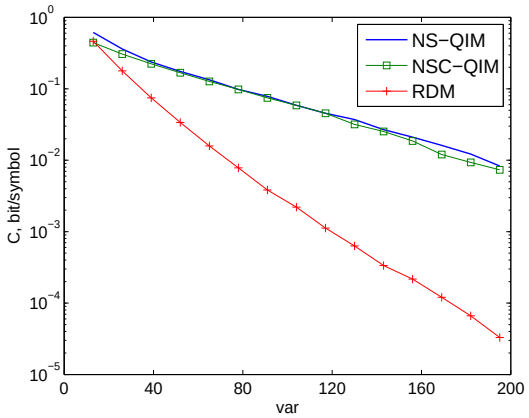



**Figure 6.3.** Information extracted under GA followed by AWGN

For the experiment, the first singular values of 4x4 blocks from 512x512 grayscale images were quantized. In order to equalize the embedding conditions for all the methods, the same Document to Watermark Ratio (DWR=28dB) was used. The usage of the term “capacity” in the paper is not an absolutely correct and needs to be replaced with “extracted information”. As one can see from Figure 6.3, it is possible to define constant parameters for $\boldsymbol{\theta}$ that sacrifice the performance very little. On the other hand, advantages of NSC-QIM over RDM are still quite considerable. The results published in this and the previous paper confirm that some aspects of watermarking can be indeed improved by introducing non-standard DC for scalar QIM. In our following research articles we investigate non-standard DC for multidimensional QIM.

## 6.6 Publication VI: Benefits of Non-Standard DC for Multidimensional QIM

A modified two-dimensional quantization approach is discusses in this article. With the aim to embed a bit of a watermark message it uses a famous Quincunx lattice. A new Distortion Compensation technique is proposed in order to improve robustness of the method. The experimental results obtained for the method are compared with the results of other popular 2D as well as scalar quantization methods.

This is our first paper considering multidimensional quantization technique. Thus, a relatively simple yet popular Quincunx lattice is utilized for watermarking. Lattices are widely used in various coding applications, their design has been discussed and their properties have been quite extensively studied in the literature [94, 98]. However, DC lattice quantization is a relatively new topic which is usually discussed in the context of digital watermarking. In case of the Quincunx lattice quantization, a bit of a watermark has to be embedded by modulation of two original samples. For this purpose, a pair of samples is seen as a point in 2D space. Before embedding, any point in the space needs to be assigned to a cell ("embedding cell" is obtained from coarse lattice) depending on a bit value ("0" or "1"). Then, in conventional DC lattice quantization, the point has to be shifted toward the center of the corresponding "embedding cell". For a correct watermark extraction, one must provide that any point from "embedding cell" is shifted inside smaller "extraction cell" obtained from fine lattice (Figure 6.4).

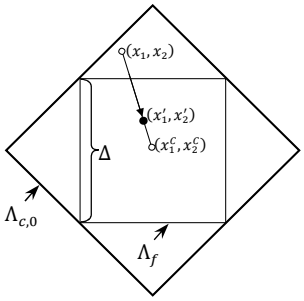



**Figure 6.4.** Conventional DC Quantization

On the figure, "embedding cell" corresponds to $\Lambda_{c,0}$ while "extraction cell" corresponds to $\Lambda_f$. The center of the both cells is $(x_1^C, x_2^C)$. Original point has coordinates $(x_1, x_2)$ and modulated point has coordinates $(x_1', x_2')$.

After a bit is embedded, the modulated pair $(x_1', x_2')$ may be attacked by AWGN. An error occurs if $(x_1', x_2')$ is shifted by AWGN outside "extraction cell". Since AWGN is a random process, this can be done in an infinite number of ways. Nevertheless there is one scenario that has the highest probability.

A new DC technique proposed in our paper minimizes the probability that corresponds to Maximum Likelihood Error Scenario (MLES). Nonetheless one can see it as a minimax strategy which is quite a popular approach for decision making in the presence of uncertainty or in complex systems [110]. Our motivation can be explained in the following way. For a given value of the compensation parameter $\alpha$, point $(x_1, x_2)$ is being shifted on distance $r_{x_1,x_2}$ so that $r_{x_1,x_2} = \sqrt{(x_1 - x_1')^2 + (x_2 - x_2')^2}$. Embedding distortion in that case depends only on $r_{x_1,x_2}^2$ and

does not depend on the direction of the shift. If $r_{x_1,x_2}$ is already decided, one would try to optimize the direction for the best robustness. For that purpose, one needs to estimate the error rate for all the points that have distance $r_{x_1,x_2}$ from $(x_1, x_2)$ and to choose $(x_1', x_2')$ that has the lowest rate for a given WNR. However, estimation of that rate for many points and under all the possible combinations of WNR, $\alpha$ is a tremendous computation task. Therefore, due to inability to estimate all the necessary rates, a simple minimax principle is utilized. It should be noted that in the case when the "extraction cell" is a hypersphere, conventional DC is a minimax strategy as well. For a non-spherical form of the cell, one may shift $(x_1, x_2)$ in the direction orthogonal to the nearest border of "extraction cell" (Figure 6.5).

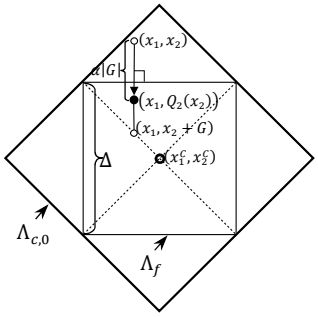



**Figure 6.5.** New MLES-based DC

The mentioned way of modulation is a minimax strategy until the nearest diagonal of "extraction cell" is reached. In our modification of Quincunx DC-QIM we implement that kind of modulation. Minimax strategy for any point on the diagonal is to shift along the diagonal in the direction to the center (that kind of modulation will be the subject of our further research). Obviously, the implemented approach requires for any $(x_1, x_2)$ modulation shift to be smaller or equal to that of the conventional DC with the same $\alpha$. Therefore, embedding distortion for our method is smaller than that for the standard DC. Also, either $x_1$ or $x_2$ is modified and never the both. Nevertheless, it should be emphasized that the proposed way of embedding is two-dimensional as the value of the modified coordinate depends on the other coordinate.

For the experimental comparison of the Modified Quincunx (MQ) and the conventional Quincunx with DC, 72 natural grayscale images of different resolution were selected. Additionally, we have examined two types of hexagonal 2D lattice watermarking methods with traditional DC [94, 97]. For each image, 8x8 blocks were formed from adjacent pixels and passed to SVD transform. From each block, only the first singular value was chosen to form the sequence of coefficients $\boldsymbol{x}$. All the watermarked images were attacked by AWGN and a simple hard-decision decoder using the described "extraction cells" was utilized after. The amount of the mutual information between embedded and decoded messages was compared for all the methods. Under WNR lower than -3dB, the advantage of MQ over all the rest methods is considerable. However, for higher WNR values its performance is the worst. A natural explanation for this is that the implementation of minimax modulation strategy (in the current realization) is limited by the diagonals of "extraction cell". Fortunately, for a better performance under any WNR one can combine MQ with the traditional DC Quincunx because the same decoding rule is used in the both cases.

# Chapter 7

# Conclusions

New methods of digital image watermarking operating in SVD domain were considered in the thesis which consists of six publications. They specifically present the entire process of the research, including image processing, watermark embedding (encoding), attacking of the watermarked images, watermark extraction (decoding) and evaluation of the results obtained. The final aim of the research is to improve the efficiency of blind digital image watermarking under different circumstances that can be described by the purpose of the watermarking as well as the type and severity of the attack.

Firstly, for image processing stage, the kind of transform is the same in all six publications. The advantages of SVD are due to relatively high robustness to most of the image processing techniques as well as imperceptibility of watermarking in that domain to human eyes. Unlike traditional 2D image transforms, like for example, DCT, the basis in SVD is not the same for different image blocks. Therefore, watermarking information can be embedded in orthogonal matrices and singular values alike. The number of singular values that interpret important visual information is fewer than that for DCT, but the number of important elements of orthogonal matrices is larger. Utilization of the latter creates interesting opportunities for vector embedding, which robustness can potentially be improved for higher dimensionality.

Secondly, watermark encoding approaches are different for each publication. The most attention in the thesis is paid to this aspect. The approaches described in the thesis can be divided into two parts depending on a constraint that should be taken into account during embedding. The first part consists of Publications I-III where the requirement of orthonormality of the modified matrix is the main constraint of embedding. The second part consists of Publications IV-VI that describe watermarking methods addressing new approaches in quantization of singular values and do not assume any specific constraints.

Publications I-III consider modification of orthogonal matrices of SVD of square image blocks which is relatively new and not a very popular approach in DIW. Nonetheless, it has some benefits such as imperceptibility, sufficiently high robustness to AWGN, JPEG and is invariant to GA. Information can be embedded in the first column of the left or right orthogonal matrices as those columns represent the most significant layer of a block.

Methods detailed in Publications I-III perform vector quantization which is defined using different embedding rules. They are based on threshold parameter $Th$ which is a contrast to other vector quantization approaches like, for instance, those based on lattice quantization. Particularly, it means that only one projection of a vector is interpreted for watermarking which might be seen as a certain limitation. Nevertheless, the mentioned approach has several advantages one of which is simplicity due to existence of only one border that separates “0” from “1” and is a contrast to periodic nature of lattices. Another advantage is that watermark robustness-transparency trade-off can be regulated by an owner without informing receiver

(consumer). Unlike lattice quantization, where the distances between the elements of lattice are important for correct extracting, threshold extracting does not require distance parameter.

Different embedding rules in Publications I-III require different encoding of a watermark. With the aim to interpret a bit of a watermark one or two orthogonal matrices can be used. The first column of a single orthogonal matrix defines watermark bit in Publication I. Several new watermarking rules were proposed where each of them has some advantages under different kinds of attacks. In Publication I, rule $L2_4$ was recommended as the most universal. Both the orthogonal matrices were used for watermark encoding/decoding in Publication II, but only one was modified at a time. Developing this idea further, the both left and right orthogonal matrices were proposed to modify in Publication III.

Another important characterization for the methods in Publications I-III is that they have specific constraints due to the requirement of orthonormality for the modified matrices. The disadvantage of the approaches previously described in the literature is inability to fulfill the latter condition [105, 106]. The mentioned reason causes distortion of the embedded watermark prior to transmission. In order to avoid this, matrix elements that do not interpret watermarking data (e.g. in the second, third and fourth columns) need to be also modified. The modification is defined according to mathematic model which is different in every publication.

Publication I explores the idea of approximation of the original orthogonal matrix of SVD using a simple model of 4x4 matrix. It guarantees that after modification of the first column (according to embedding rule), the rest of the matrix can be easily adjusted in order to become orthogonal. The positive outcome of such approach is that for some grayscale images it demonstrates 46% better performance under a JPEG attack in comparison with the method described in the literature. In Publication II, instead of approximation, an original orthogonal matrix is multiplied with a modelled orthogonal matrix. The embedding procedure includes several multiplicative steps that have different aims. Multiplication with the matrix defined according to Van Elfrinkhof's formulae is performed on the first step which aims to satisfy an embedding rule while minimizing distortions. During the subsequent steps, the multiplication is performed with matrices presented in Euler-Rodrigues form and the aim is to reduce the distortion of the first step. The process of the minimization of embedding distortions is essentially sub-optimal but its complexity is relatively low. In order to further increase watermarking performance, a criterion of watermarking efficiency is proposed in Publication III where, in addition, a different embedding rule is used. The criterion combines indices of embedding distortion and robustness with the aim to select one from among a set of watermarked blocks. For each original 4x4 block of an image, the mentioned set is created by using different values for the parameter $Th$ in the embedding rule. The usage of such heuristic criterion provides better adaptation of the watermarked image to the expected distortion or attack. The proposed watermarking method was tested under several kinds of attacks such as AWGN, Salt&Pepper, JPEG, 3x3 Median Filtering, Cropping and Rotation. The results of the experiment show that in comparison with other reference methods the proposed one provides better performance under JPEG and geometric attacks (around 40%).

Encoding methods detailed in Publications IV-VI differ from those in Publications I-III in two main aspects: visual meaning of original indices as well as distribution of modified indices.

For visual meaning, the most controversy between SV and elements of orthogonal matrix is because the latter defines a structure pattern of a layer of image block while the first defines importance of this layer. Accordingly, orthogonal matrices are normalized while SVs are not which makes GA harmful even though it might be unnoticeable to human eye. The main difference in distribution of quantized samples is due to fact that methods in Publications IV-VI utilize for encoding lattice-based quantizers which have only local influence on final distribution. Traditionally, this has caused an additional complication under GA as local features are not easy to reconstruct [111]. On the other hand, to the benefit of SV quantization is that it does not require complex minimization of embedding distortions which is described in Publications I-III. Also, characteristics of lattice-based quantizers under AWGN are well-studied and represented in the literature [73, 97].

The GA vulnerability issue was successfully addressed in Publications IV-V where a new model for Scalar Quantization with Distortion Compensation has been proposed. According to it samples were quantized in a way that the final distribution is asymmetric. This creates a distinctive feature that is used later by the procedure for GA recovery.

Robustness under AWGN has also been addressed in Publications IV-VI. Quantization Index Modulation is an approach which is well-known in digital watermarking. Nevertheless, some aspects of its robustness can still be considerably improved and, for instance, under high intensity AWGN the performance of known in the literature methods is quite low. In order to improve it, a concept of IDL was introduced in Publication IV according to which samples in pre-defined positions are not quantized in order to avoid significant embedding distortions. During the experiment, the watermarked images were attacked by AWGN and GA after which the results of the watermark extraction were compared with such state of the art methods as DC-QIM and RDM [97, 107]. Under a high-intensity AWGN the proposed IDL concept is especially beneficial that is reflected in up to $10^4$ times higher performance. In case of GA, the procedure utilizing a proposed criterion of an asymmetric distribution of quantized samples helps to recover a watermark with high accuracy which results in up to $10^3$ times higher performance than achievable by RDM.

The proposed in Publication IV quantization method has several parameters that influence robustness-transparency trade-off. For a given intensity of attack, the parameters need to be adjusted in order to provide the highest possible performance of the watermarking method. This kind of optimization should be done once only and the optimal parameters can be used further in watermarking applications in practice. However, the optimization process is quite complex and time-consuming as it is mostly based on brute-force search. One aspect of this problem is addressed in Publication V where using analytical derivations it has been shown that robustness of the proposed quantization method depends on $\Delta/\sigma_n$ and $\boldsymbol{\theta}$. Therefore, only those parameters need to be considered to find the best robustness-transparency trade-off. Additionally, a case of constant parameters $\boldsymbol{\theta}$ (that are the same for any level of noise) has been studied there and shown that compared to RDM, considerable advantage can also be achieved.

Publication VI contains description and evaluation of a new multi-dimensional quantization with DC based on Quincunx lattice. Compared to the scalar quantization case discussed in Publications IV-V, a new aspect of efficiency under AWGN arises here. This aspect is connected

with both the form of Voronoi cells used for quantization and the direction of quantization shift. Lattices and the impact of the form of their Voronoi cells on quantization (without DC) performance are well-studied [98, 94]. However, any solid theoretical study of DC applied to lattices is not yet known. In the paper, a Maximum Likelihood Error Scenario is explained for 2D Quincunx lattice. According to MLES, the probability of an error for a point inside a particular Voronoi cell depends on how close the point is to the boundary of the cell and a new DC technique is proposed that minimizes probability of MLES for every given point (in 2D space). The new technique reveals up to $10^2$ times better performance under low WNRs compared with the schemes utilizing conventional DC.

Thirdly, principles of watermark decoding and evaluation of the results were similar for all the methods detailed in Publications I-VI. Despite the fact that different embedding rules and lattice quantizers were used, in every publication a watermark was decoded according to hard decision principle. Its advantage is that decoder's complexity is low while performance is relatively good. For different methods mentioned in Publications I-III, evaluation and comparison of the performance was made under the requirement of similar quality of the watermarked images. Under that condition, the severity of different attacks was the same for the methods being compared. In those cases, the transparency of watermarking was expressed using simple measure of MSE between the original and the watermarked images or PSNR. In Publications IV-VI performance of different methods was compared under the same WNR. The performance itself was expressed using BER and mutual information between embedded and extracted watermarks in Publications I-III and Publications IV-VI, respectively.

In addition to simple distance-based objective quality measures, subjective quality assessment has also been conducted in Publication VI. In some cases the above mentioned quality measures might not represent the goal of watermarking adequately. Therefore, the decision about the quality of watermarked images has to be made by the professional whose task is to construe particular type of visual information. An example of the watermarking protection of a diagnostic image which quality has been assessed by a group of medical experts is described in Publication VI. Usually such images contain a patient's personal information and need to be protected, but a watermark should not interfere with important diagnostic information which is interpreted by medical staff.

Finally, several directions for further research and possible improvements can be outlined. Regarding watermark encoding, there is a twofold development of the methods described in Publication I-III. The first part is in increasing of the size of image blocks which will offer more elements in the first column of orthogonal matrix. Consequently, a better robustness-transparency trade-off can be achieved. The second part is in adjusting embedding rules in a way that a certain desired distribution of modified indices is provided. The reason for this is that the shape of target distribution of modified indices better reflects robustness compared to threshold-based rules. For the methods detailed in Publications IV-VI, the following opportunities can be explored. Because of the fact that DC techniques with IDL have not received enough attention among the researchers, new models utilizing it might be beneficial. For example, logical constructions used in Publication V for the description of quantization method can be utilized as a framework for further improvements which can be done by a simple substitution of new pdf of

quantized samples. In regards to new multidimensional approach proposed in Publication VI, MLES concept can be combined with a different distribution of quantization distortion. Lastly, the methods from Publications IV-VI are universal and can be implemented in different transform domains like, for instance, DCT or DWT and do not require any additional modification. At any rate, promising results were presented in the thesis which will hopefully stimulate productive discussions in the field of digital image watermarking.

# Personal Contributions

All the publications in this thesis were the work of the author of the thesis and Martti Juhola (hereafter referred to as MJ). All the ideas regarding methods for Digital Image Watermarking presented in the publications came from the author who also collected original images, implemented and conducted all the experiments. MJ commented on all the manuscripts, provided guidance and supervised the entire research project.

# Bibliography


[1] Furht, B., Kirovski, D.: Multimedia Security Handbook. CRC Press, Boca Raton (2004)

[2] Guidelines for Secure Use of Social Media by Federal Departments and Agencies. (Accessed Sep 2009) Available at: https://cio.gov/wp-content/uploads/downloads/2012/09/Guidelines_for_Secure_Use_Social_Media_v01-0.pdf

[3] Fridrich, J.: Steganography in Digital Media: Principles, Algorithms, and Applications. Cambridge University Press, New York, NY, USA (2009)

[4] Cox, I., Miller, M., Bloom, J., Fridrich, J., Kalker, T.: Digital Watermarking and Steganography (2 ed.). Morgan Kaufmann Publishers Inc., San Francisco, CA, USA (2007)

[5] Zeng, W., Yu, H., Lin , C.-Y.: Multimedia Security Technologies for Digital Rights Management. Academic Press, Burlington (2006)

[6] Tipton, H., Nozaki, M.: Information Security Management Handbook, Sixth Edition. Taylor & Francis (2012)

[7] Ponceleon, D., Nusser, S., Zbarsky, V., Cerruti, J., Nin, S.: Enabling secure distribution of digital media to SD-cards. In : Proceedings of the 14th annual ACM international conference on Multimedia (MULTIMEDIA '06), pp.495-496 (2006)

[8] Kubesch, A. S., Wicker, S.: Digital rights management: The cost to consumers [Point of View]. Proceedings of the IEEE 103(5), 726-733 (May 2015)

[9] d'Ornellas, M. C.: Applying Digital Rights Management to Complex Content Management Systems. In : Proceedings of the 11th IEEE International Conference on Computational Science and Engineering, 2008 (CSE '08)., pp.429-435 (July 2008)

[10] Borders, K., Zhao, X., Prakash, A.: Securing sensitive content in a view-only file system. In : Proceedings of the ACM workshop on Digital rights management (DRM '06), pp.27-36 (2006)

[11] Cole, E.: Hiding in Plain Sight: Steganography and the Art of Covert Communication (1 ed.). John Wiley & Sons, New York, USA (2003)

[12] Bianchi, T., Piva, A.: Secure Watermarking for Multimedia Content Protection: A Review of its Benefits and Open Issues. IEEE Signal Processing Magazine 30(2), 87-96 (Mar. 2013)

[13] Choi, J., Chun, S., Cho, J.-W.: Smart SecureGov: Mobile Government Security Framework. In : Proceedings of the 15th Annual International Conference on Digital Government Research, New York, NY, USA, pp.91-99 (2014)

[14] Hartung, F., Ramme, F.: Digital rights management and watermarking of multimedia content for m-commerce applications. Communications Magazine, IEEE 38(11), 78-84 (Nov 2000)

[15] Pan, W., Coatrieux, G., Cuppens-Boulahia, N., Cuppens, F., Roux, C.: Watermarking to Enforce Medical Image Access and Usage Control Policy. In : Proceedings of the Sixth International Conference on Signal-Image Technology and Internet-Based Systems (SITIS), pp.251-260 (Dec. 2010)

[16] Katzenbeisser, S., Petitcolas, F.: Information Hiding Techniques for Steganography and Digital Watermarking. Artech House, Norwood, USA (2000)

[17] Gupta, G., Pieprzyk, J., Wang, H.: An Attack-localizing Watermarking Scheme for Natural Language Documents. In : Proceedings of the 2006 ACM Symposium on Information, Computer and Communications Security, New York, NY, USA, pp.157-165 (2006)

[18] Low, S. H., Maxemchuk, N. F.: Performance comparison of two text marking methods. IEEE Journal on Selected Areas in Communications 16(4), 561-572 (May 1998)

[19] Low, S. H., Maxemchuk, N. F., Lapone, A. M.: Document identification for copyright protection using centroid detection. IEEE Transactions on Communications 46(3), 372-383 (Mar 1998)

[20] Cox, I., Kilian, J., Leighton, T., Shamoon, T.: Secure spread spectrum watermarking for images, audio and video. In : Proceedings of International Conference on Image Processing, 1996., vol. 3, pp.243-246 vol.3 (Sep 1996)

[21] Cvejic, N., Seppänen, T.: Spread spectrum audio watermarking using frequency hopping and attack characterization. Signal Processing 84(1), 207-213 (2004)

[22] Nishimura, A.: Audio Watermarking Based on Amplitude Modulation and Modulation Masking. In : Proceedings of the 1st International Workshop on Information Hiding and Its Criteria for Evaluation, New York, NY, USA, pp.49-55 (2014)

[23] Zolotavkin, Y., Lukichov, V., Vasyura, A.: A novel approach to the security of data hidden in multimedia objects. In : Proceedings of 42nd Annual IEEE International Carnahan Conference on Security Technology, 2008 (ICCST 2008), Prague, pp.23-28 (Oct 2008)

[24] Braudaway, G., Magerlein, K., Mintzer, F.: Protecting publicly available images with a visible image watermark. Proc. SPIE 2659, 126-133 (1996)

[25] Marvel, L. M., Boncelet, C. G. ., Retter, C. T.: Spread spectrum image steganography. IEEE Transactions on Image Processing 8(8), 1075-1083 (Aug 1999)

[26] Mathai, N. J., Kundur, D., Sheikholeslami, A.: Hardware implementation perspectives of digital video watermarking algorithms. IEEE Transactions on Signal Processing 51(4), 925-938 (Apr 2003)

[27] Kundur, D., Karthik, K.: Video fingerprinting and encryption principles for digital rights management. Proceedings of the IEEE 92(6), 918-932 (2004)

[28] Chan, P.-W., Lyu, M.: A DWT-Based Digital Video Watermarking Scheme with Error Correcting Code. In Qing, S., Gollmann, D., Zhou, J., eds. : Information and Communications Security 2836. Springer Berlin Heidelberg (2003) 202-213

[29] Cho, J.-W., Prost, R., Jung, H.-Y.: An Oblivious Watermarking for 3-D Polygonal Meshes Using Distribution of Vertex Norms. IEEE Transactions on Signal Processing 55(1), 142-155 (Jan 2007)

[30] Wu, H.-T., Cheung, Y.-M.: A Fragile Watermarking Scheme for 3D Meshes. In : Proceedings of the 7th Workshop on Multimedia and Security, New York, NY, USA, pp.117-124 (2005)

[31] Bazin, C., Le Bars, J.-M., Madelaine, J.: A Blind, Fast and Robust Method for Geographical Data Watermarking. In : Proceedings of the 2Nd ACM Symposium on Information, Computer and Communications Security, New York, NY, USA, pp.265-272 (2007)

[32] Deng, X., Xu, G., Sun, G., Man, J.: Software Watermarking Based on Dynamic Program Slicing. In : Proceedings of the International Conference on Intelligent Information Hiding and Multimedia Signal Processing, 2008 (IIHMSP '08)., pp.461-464 (Aug 2008)

[33] Cousot, P., Cousot, R.: An Abstract Interpretation-based Framework for Software Watermarking. SIGPLAN Notes 39(1), 173-185 (Jan 2004)

[34] Jinchao, X., Guosun, Z.: A Software Watermarking Algorithm Based on Stack-State Transition Graph. In : Proceedings of the 4th International Conference on Network and System Security, 2010 (NSS'10), pp.83-88 (Sept 2010)

[35] Abdel-Hamid, A. T., Tahar, S., Aboulhamid, E.: IP watermarking techniques: survey and comparison. In : Proceedings of the 3rd IEEE International Workshop on System-on-Chip for Real-Time Applications, 2003., pp.60-65 (June 2003)

[36] Cui, A., Chang, C., Tahar, S.: IP Watermarking Using Incremental Technology Mapping at Logic Synthesis Level. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems 27(9), 1565-1570 (Sept 2008)

[37] Nie, T., Kisaka, T., Toyonaga, M.: A post layout watermarking method for IP protection. In : Proceedings of IEEE International Symposium on Circuits and Systems, 2005 (ISCAS 2005)., pp.6206-6209 Vol. 6 (May 2005)

[38] Couch, I.: Digital and Analog Communication Systems 3rd edn. Prentice Hall PTR, Upper Saddle River, NJ, USA (1990)

[39] Gallager, R.: Information Theory and Reliable Communication. John Wiley & Sons, Inc., New York, NY, USA (1968)

[40] Lu, C.-S.: Multimedia Security: Steganography and Digital Watermarking Techniques for Protection of Intellectual Property. IGI Global, Hershey, USA (2005)

[41] Tsai, H.-M., Chang, L.-W.: Secure reversible visible image watermarking with authentication. Signal Processing: Image Communication 25(1), 10-17 (2010)

[42] Yang, Y., Sun, X., Yang, H., Li, C.-T.: Removable visible image watermarking algorithm in the discrete cosine transform domain. Journal of Electronic Imaging 17(3), 8-11 (2008)

[43] Zolotavkin, Y., Juhola, M.: An SVD-based Transparent Watermarking Method. In : Proceedings of the International Conference on E-Technologies and Business on the Web (EBW2013), Bangkok, pp.85-90 (May 2013)

[44] Lim, Y., Xu, C., Feng, D.: Web Based Image Authentication Using Invisible Fragile Watermark. In : Proceedings of the Pan-Sydney Area Workshop on Visual Information Processing - Volume 11, Darlinghurst, Australia, Australia, pp.31-34 (2001)

[45] Fridrich, J.: Security of fragile authentication watermarks with localization. Proceedings of SPIE 4675, Security and Watermarking of Multimedia Contents IV 4675, Vol. 4675, 691-700 (Apr. Apr. 2002)

[46] Liu, S.-H., Yao, H.-X., Gao, W., Liu, Y.-L.: An image fragile watermark scheme based on chaotic image pattern and pixel-pairs. Applied Mathematics and Computation 185(2), 869-882 (2007) Special Issue on Intelligent Computing Theory and Methodology.

[47] Barni, M., Bartolini, F., Cappellini, V., Piva, A.: Robust watermarking of still images for copyright protection. In : Proceedings of 13th International Conference on Digital Signal Processing Proceedings, pp.vol.2, 499-502 (2-4 Jul 1997)

[48] Barni, M., Bartolini, F., Furon, T.: A general framework for robust watermarking security. Signal Processing 83(10), 2069-2084 (2003)

[49] Lin, E., Podilchuk, C., Delp III, E.: Detection of image alterations using semifragile watermarks. Proc. SPIE 3971, 152-163 (2000)

[50] Altun, O. .: A Set Theoretic Framework for Watermarking and Its Application to Semifragile Tamper Detection. IEEE Transactions on Information Forensics and Security 1(4), vol.1, no.4, 479-492 (Dec. Dec. 2006)

[51] Dharwadkar, N. V. .: Non-blind watermarking scheme for color images in RGB space using DWT-SVD. In : Proceedings of International Conference on Communications and Signal Processing (ICCSP), 2011, pp.489-493 (10-12 Feb. 2011)

[52] Su, P.-C., Kuo, C.-C., Wang, H.-J.: Blind digital watermarking for cartoon and map images. Proc. SPIE 3657, 296-306 (1999)

[53] Eggers, J. J., Girod, B.: Blind watermarking applied to image authentication. In : Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '01), 2001., vol. 3, pp.1977-1980 vol.3 (2001)

[54] Chen, T., Chen, T.: A Framework for Optimal Blind Watermark Detection. In : Proceedings of the 2001 Workshop on Multimedia and Security: New Challenges, Ottawa, Ontario, Canada, pp.11--14 (2001)

[55] Akhaee, M., Sahraeian, S., Jin, C.: Blind Image Watermarking Using a Sample Projection Approach. IEEE

Transactions on Information Forensics and Security 6(3), 883-893 (Sept. 2011)

[56] Fridrich, J., Goljan, M., Du, R.: Invertible authentication. Proc. SPIE 4314, 197-208 (2001)

[57] Domingo-Ferrer, J., Sebe, F.: Invertible spread-spectrum watermarking for image authentication and multilevel access to precision-critical watermarked images. In : Proceedings of International Conference on Information Technology: Coding and Computing, pp.pp.152-157 (April 2002)

[58] Van Leest, A., Haitsma, J., Kalker, T.: On digital cinema and watermarking. Proc. SPIE 5020, 526-535 (2003)

[59] Rouvroy, G., Standaert, F.-X., Lefèbvre, F., Quisquater, J.-J., Macq, B., Legat, J.-D.: Reconfigurable Hardware Solutions for the Digital Rights Management of Digital Cinema. In : Proceedings of the 4th ACM Workshop on Digital Rights Management, New York, NY, USA, pp.40-53 (2004)

[60] Le, T., Nguyen, K., Le, H.: Literature Survey on Image Watermarking Tools, Watermark Attacks and Benchmarking Tools. Proceedings of International Conference on Advances in Multimedia 0, 67-73 (2010)

[61] Emmanuel, S., Kankanhalli, M.: A digital rights management scheme for broadcast video. Multimedia Systems 8(6), 444-458 (2003)

[62] Cho, Y., Ahn, B., Kim, J.-S., Kim, I.-Y., Kim, S.: A study for watermark methods appropriate to medical images. Journal of Digital Imaging 14(1), 184-186 (2001)

[63] Giakoumaki, A., Pavlopoulos, S., Koutsouris, D.: Secure and efficient health data management through multiple watermarking on medical images. Medical and Biological Engineering and Computing 44(8), 619-631 (2006)

[64] Jafari, M., Safavi-Naini, R., Saunders, C., Sheppard, N.: Using Digital Rights Management for Securing Data in a Medical Research Environment. In : Proceedings of the Tenth Annual ACM Workshop on Digital Rights Management, New York, NY, USA, pp.55-60 (2010)

[65] Gibbs, S.: Instagram for doctors: how Figure 1 is crowdsourcing diagnoses. (Accessed Apr. 27, 2015) Available at: http://www.theguardian.com/technology/2015/apr/27/instagram-for-doctors-figure-1-crowdsourcing-diagnoses

[66] Kleinman, Z.: 'Instagram for doctors' to be launched in Europe. (Accessed Oct. 7, 2014) Available at: http://www.bbc.com/news/technology-29521986

[67] Luoma, E., Vahtera, H.: Current and emerging requirements for digital rights management systems through examination of business networks. In : Proceedings of the 37th Annual Hawaii International Conference on System Sciences, 2004., pp.10 pp.- (Jan 2004)

[68] Memon, N., Wong, P. W.: A buyer-seller watermarking protocol. IEEE Transactions on Image Processing 10(4), 643-649 (Apr 2001)

[69] Soo Ju, H., Jeong Kim, H., Hoon Lee, D., In Lim, J.: An Anonymous Buyer-Seller Watermarking Protocol with Anonymity Control. In Lee, P., Lim, C., eds. : Information Security and Cryptology — ICISC 2002 2587. Springer Berlin Heidelberg (2003) 421-432

[70] Poh, G., Z'aba, M.: On the Security and Practicality of a Buyer Seller Watermarking Protocol for DRM. In : Proceedings of the 4th International Conference on Security of Information and Networks, New York, NY, USA, pp.251-254 (2011)

[71] Thomas, T., Emmanuel, S., Subramanyam, A. V., Kankanhalli, M. S.: Joint Watermarking Scheme for Multiparty Multilevel DRM Architecture. IEEE Transactions on Information Forensics and Security 4(4), 758-767 (Dec 2009)

[72] Bossert, M.: Channel Coding for Telecommunications 1st edn. John Wiley \& Sons, Inc., New York, NY, USA (1999)

[73] Costa, M.: Writing on dirty paper. IEEE Transactions on Information Theory 29(3), 439-441 (May 1983)

[74] Voloshynovskiy, S., Pereira, S., Pun, T., Eggers, J. J., Su, J. K.: Attacks on digital watermarks: classification,

estimation based attacks, and benchmarks. Communications Magazine, IEEE 39(8), 118-126 (Aug 2001)

[75] Avcibas, I., Sankur, B., Sayood, K.: Statistical evaluation of image quality measures. Journal of Electronic Imaging 11(2), 206-223 (2002)

[76] Sheikh, H., Bovik, A.: Image information and visual quality. IEEE Transactions on Image Processing 15(2), 430-444 (Feb. 2006)

[77] Pitas, I.: Digital image processing algorithms and applications. John Wiley \& Sons (2000)

[78] Petrou, M., Petrou, C.: Image Processing : The Fundamentals. John Wiley & Sons (2010)

[79] Lin, S., Chen, C.-F.: A robust DCT-based watermarking for copyright protection. IEEE Transactions on Consumer Electronics 46(3), 415-421 (Aug 2000)

[80] Xiao, J. .: Toward a Better Understanding of DCT Coefficients in Watermarking. In : Proceedings of Pacific-Asia Workshop on Computational Intelligence and Industrial Application, 2008. PACIIA '08., pp.vol.2, 206-209 (19-20 Dec. 2008)

[81] Lukitchov, V., Vasyura, A., Zolotavkin, Y.: Improvement of pattern hiding method of data application in JPEG-files. In : Proceedings of 42nd Annual IEEE International Carnahan Conference on Security Technology, 2008 (ICCST 2008), Prague, pp.69-75 (2008)

[82] Vasconcelos, N.: Discrete Cosine Transform. In: 161C - Digital Signal Processing II. (Accessed Oct 19, 2015) Available at: http://www.svcl.ucsd.edu/courses/ece161c/handouts/DCT.pdf

[83] Welstead, S.: Fractal and wavelet image compression techniques. SPIE Optical Engineering Press (Bellingham, WA (1999)

[84] Liu, C.-L.: Digital Image and Signal Processing. (Accessed Oct 19, 2015) Available at: http://disp.ee.ntu.edu.tw/tutorial/WaveletTutorial.pdf

[85] Benedetto, J.: Wavelets: mathematics and applications 13. CRC press (1993)

[86] Meerwald, P., Uhl, A.: Survey of wavelet-domain watermarking algorithms. Proc. SPIE 4314, 505-516 (2001)

[87] Venkateswarlu, S., Reddy, P., Raju, Y.: Watermarking for JPEG2000 Compression Standard on FPGA. In : Proceedings of the International Conference and Workshop on Emerging Trends in Technology, New York, NY, USA, pp.309-314 (2010)

[88] Ganic, E., Eskicioglu, A.: Robust DWT-SVD domain image watermarking: embedding data in all frequencies. In : Proceedings of the 2004 workshop on Multimedia and security (MM&Sec '04), pp.166-174 (2004)

[89] Li, Z., Yap, K.-H., Lei, B.-Y.: A new blind robust image watermarking scheme in SVD-DCT composite domain. In : Proceedings of the 18th IEEE International Conference on Proceedings of Image Processing (ICIP'11), 2011, pp.2757-2760 (Sept. 2011)

[90] Manjunath, M. .: A new robust semi blind watermarking using block DCT and SVD. In : Proceedings of IEEE International Conference on Advanced Communication Control and Computing Technologies (ICACCCT), 2012, pp.193-197 (23-25 Aug. 2012)

[91] Trefethen, L., Bau, D.: Numerical Linear Algebra. Cambridge University Press (1997)

[92] Gorodetski, V., Popyack, L., Samoilov, V., Skormin, V.: SVD-based Approach to Transparent Embedding Data into Digital Images. In : Proceedings of the International Workshop on Information Assurance in Computer Networks: Methods, Models, and Architectures for Network Security (MMM-ACNS '01), pp.263-274 (2001)

[93] Modaghegh, H. .: A new adjustable blind watermarking based on GA and SVD. In : International Conference on Innovations in Information Technology, 2009. IIT '09., pp.6-10 (15-17 Dec. 2009)

[94] MOULIN, P., KOETTER, R.: Data-Hiding Codes. Proceedings of the IEEE 93(12), 2083-2126 (Dec 2005)

[95] Malvar, H. S., Florencio, D. A. F.: Improved spread spectrum: a new modulation technique for robust

watermarking. IEEE Transactions on Signal Processing 51(4), 898-905 (Apr 2003)

[96] Eggers, J., Bäuml, R., Tzschoppe, R., Girod, B.: Scalar Costa Scheme for Information Embedding. IEEE Transactions on Signal Processing 51(4), 1003-1019 (APR. 2003)

[97] Chen, B., Wornell, G.: Quantization index modulation: a class of provably good methods for digital watermarking and information embedding. IEEE Transactions on Information Theory 47(4), 1423-1443 (May 2001)

[98] Zamir, R., Feder, M.: On lattice quantization noise. IEEE Transactions on Information Theory, 1152-1159 (Jul 1996)

[99] Esen, E., Alatan, A.: Data hiding using trellis coded quantization. In : Proceedings of the IEEE 12th Signal Processing and Communications Applications Conference, pp.pp.21-24 (Apr 2004)

[100] Wang, X., Zhang, X.-P.: A new implementation of trellis coded quantization based data hiding. In : Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2008), pp.pp.1689-1692 (Apr 2008)

[101] Chou, J., Pradhan, S., Ramchandran, K.: On the duality between distributed source coding and data hiding. In : Proceedings of the Thirty-Third Asilomar Conference on Signals, Systems, and Computers, pp.1503-1507 (Oct. 1999)

[102] Wharton, E., Panetta, K., Agaian, S.: Human visual system based similarity metrics. In : Proceedings of IEEE International Conference on Systems, Man and Cybernetics (SMC 2008), pp.685-690 (Oct. 2008)

[103] Chikkerur, S., Sundaram, V., Reisslein, M., Karam, L.: Objective Video Quality Assessment Methods: A Classification, Review, and Performance Comparison. IEEE Transactions on Broadcasting 57(2), 165-182 (June 2011)

[104] Song, C., Sudirman, S., Merabti, M., Llewellyn-Jones, D.: Analysis of Digital Image Watermark Attacks. In : Proceedings of Consumer Communications and Networking Conference (CCNC), pp.1-5 (9-12 Jan. 2010)

[105] Chang, C.-C., Tsai, P., Lin, C.-C.: SVD-based digital image watermarking scheme. Pattern Recognition Letters(26), 1577–1586 (2005)

[106] Tehrani, I., Ibrahim, S.: An enhanced SVD based watermarking using U matrix. In : Proceedings of 5th International Conference on Computer Sciences and Convergence Information Technology (ICCIT), pp.627-631 (2010)

[107] Pérez-González, F., Mosquera, C., Barni, M., Abrardo, A.: Rational dither modulation: a high-rate data-hiding method invariant to gain attacks. IEEE Transactions on Signal Processing 53(10), 3960-3975 (Oct. 2005)

[108] Esen, E., Alatan, A.: Forbidden Zone Data Hiding. In : IEEE International Conference on Image Processing, pp.pp.1393-1396 (Oct. 2006)

[109] Ramkumar, M., Akansu, A.: Signaling Methods for Multimedia Steganography. IEEE Transactions on Signal Processing 52(4), 1100-1111 (Apr. 2004)

[110] Du, D.-Z., Pardalos, P.: Minimax and applications. Kluwer Academic Publishers (1995)

[111] Shterev, I., Lagendijk, R.: Amplitude Scale Estimation for Quantization-Based Watermarking. IEEE Transactions on Signal Processing 54(11), 4146-4155 (Nov. 2006)

Publication I

# SVD-based Digital Image Watermarking on Approximated Orthogonal Matrix

Yevhen Zolotavkin and Martti Juhola

# SVD-Based Digital Image Watermarking on Approximated Orthogonal Matrix

Yevhen Zolotavkin, Martti Juhola
*Computer Science, School of Information Sciences, University of Tampere, Kanslerinrinne 1, Tampere, Finland*
*yevhen.zolotavkin@uta.fi, martti.juhola@sis.uta.fi*

Keywords: Digital Image Watermarking, Singular Value Decomposition, Robustness, Distortions, Transparency.

Abstract: A new watermarking method based on Singular Value Decomposition is proposed in this paper. The method uses new embedding rules to store a watermark in orthogonal matrix $U$ that is preprocessed in advance in order to fit a proposed model of orthogonal matrix. Some experiments involving common distortions for grayscale images were done in order to confirm efficiency of the proposed method. The robustness of watermark embedded by our method was higher for all the proposed rules under condition of jpeg compression and in some cases outperformed existing method for more than 46%.

## 1 INTRODUCTION

Multimedia is becoming increasingly important for human communication. In some cases the protection of multimedia from unauthorized usage is a critical requirement. Existing and widely used techniques in Digital Right Protection (DRP) do not always provide reliable defence against cybercriminals. One of the main difficulties is connected with degradation of quality of media content caused by application of DRP related tools. Indeed value of perceptual content of media is of the same importance as the question of ownership. The situation is complicated by increasing number of multimedia processing tools that do not contradict officially with DRP policy, but can introduce some specific distortions like, for example, compression. New and more sophisticated methods are needed to satisfy the requirements which complexity is growing.

One of the branches of DRP is Digital Image Watermarking (DIW). The needs of DIW could be different depending on a particular application. For example, it might be required that a watermark resists as much influence as possible (robust watermarking) (Barni, 1997), resists some kinds of influence and indicates presence of other kinds (semi-fragile watermarking) (Altun, 2006), (Pei, 2006), and just indicates (fragile) (Fridrich, 2002).

In order to increase robustness under some constraint that somehow represents invisibility (or transparency) many methods have been proposed during the last 20 years (Cox, 2007). The most successful among them are methods operating in transform domain. Widely used transforms are DFT, DCT, DWT (Fullea, 2001), (Lin, 2000). Those well-known transforms are parameterized in advance and do not depend on an image fragment being transformed. Therefore only a set of coefficients is important to represent a fragment according to a particular transform. However usually few coefficients in the set are used for watermarking.

The drawback is that number of significant coefficients of transformed fragment (and significance of some coefficients as well) could vary between different fragments (Xiao, 2008). Consequently different parts of a watermark could be embedded with non-equal robustness that worsens the total extraction rate under an assumption of some kind of distortion.

Another concern is that embedding of a watermark requires quantization of coefficients. A proper robustness-transparency trade-off for a particular application requires different quantization steps for different fragments. However information about quantization steps should be transmitted separately.

Different type of transform is provided by Singular Value Decomposition (SVD). It assures that the number of coefficients encapsulating image

fragment's features is small and constant. These coefficients form a diagonal in a matrix of singular values. However SVD is a unique transform which is different for every fragment and information about the transform is in left and right orthonormal matrices. Utilization of singular values for watermarking provides good trade-off between robustness and invisibility (Yongdong, 2005).

Though, elements of left and right orthonormal matrices could also be used for watermarking. The main complication for modification of elements of left and right orthonormal matrices is that matrices can become non-orthogonal. This considerably worsens robustness of a watermark.

The main contribution of this paper is to provide a watermarking method that modifies left orthonormal matrix in a way it remains orthonormal. Another contribution is utilization of different embedding rules that provide different robustness-transparency trade-off which improves flexibility (adjustability) of watermarking.

The rest of the paper is organized as following: a short review of relevant watermarking methods exploiting SVD is given in the Section 2; Section 3 bears our own approach which is described in detail; then, some experimental results are represented in Section 4 followed by a discussion of their importance in Section 5; finally, in Section 6 the paper is concluded by general remarks regarding relevance of our approach and its influence on future research.

## 2 SVD-BASED WATERMARKING

Watermarking methods utilizing SVD have become especially popular during the last 10 years.

This transform decomposes image fragment $I$ on two orthogonal matrices $U$ and $V$ and diagonal matrix $S$ containing singular values:

$$I = U \cdot S \cdot V^T. \quad (1)$$

Virtually any component from such decomposition can be used for watermark embedding. There are SVD-based watermarking methods that are blind (Modaghegh, 2009), semi-blind (Manjunath, 2012) and non-blind (Dharwadkar, 2011). In spite of that the classification is quite clear, some methods, for example, state they do not require for extraction any additional media except a key, but during watermarking the region of embedding is carefully chosen to optimize robustness-transparency trade-off (Singh, 2012). Evidently it is not absolutely fair to compare performance of pure blind methods with random key toward performance of such region specific methods as the latter require new key (different size) for each new image which is a lot of additional information.

Starting from the first methods modifying just the biggest singular value of decomposed image fragment (Sun, 2002), continued further by more sophisticated methods combining DCT-SVD (Lin, 2000), (Manjunath, 2012), (Quan, 2004), DWT-SVD (Dharwadkar, 2011), (Fullea, 2001), (Ganic, 2004) and methods optimizing trade-off between robustness and transparency for SVD-based watermarking (Modaghegh, 2009) only few among those approaches consider for embedding orthogonal matrices $U$ and $V$. The papers discussing blind embedding in orthogonal matrix are (Chang, 2005) (Tehrani, 2010) where watermarking methods that operate on $U$ are proposed. The difference between them is that in (Tehrani, 2010) some additional block-dependent adjustment of a threshold is done. Realizations and computational requirements for both methods are quite simple. However, their impact is not only in increased robustness compared, for example, to (Sun, 2002). The methods also could be modified in order to embed larger watermarks. The idea to switch from standard approach of modification of one singular value (as it is usually done in most SVD-based watermarking schemes) to modification of the first column in $U$ provides better adaptation to robustness-transparency requirement. The first column contains several elements that are of equal significance. Their significance is the same as it is for the biggest singular value which is clear when equation (1) is rewritten in a different form:

$$I = \sum_i S_{i,i} \cdot U_i \cdot V_i^T, \quad (2)$$

where $U_i$ and $V_i$ are corresponding columns of $U$ and $V$ respectively. Being constructed from $i$ different significance layers image fragment $I$ has scaling factor $S_{i,i}$ on each layer. Adoptive quantization of the first scaling factor is not always the best alternative for watermarking because it requires transmission of additional information about quantization steps. Therefore it would be more beneficial to modify the first layer in a more sophisticated manner that provides adaptation which purely corresponds to blind strategy. Such attempt is made by Chang (2005) and Tehrani (2010) by introducing a rule with a threshold. The rule is applied to a pair of elements in the first column of $U$

and can be used for embedding with different robustness-transparency rate for each block.

Nevertheless approaches presented by Chang (2005) and Tehrani (2010) have some disadvantages because the authors did not develop a tool to achieve orthogonality and normalization of modified matrix $U$. On the other hand SVD guarantees that during extraction of a bit of a watermark from a square block all three resulting matrices are orthogonal. Therefore matrices that were used to compose a block during embedding phase are not equal to the matrices calculated during extraction phase. This fact obviously could cause misinterpretation of a bit of a watermark. Another disadvantage of Chang's (2005) and Tehrani's (2010) approaches is that they used only one embedding rule that considers only two out of four elements in a column. Obviously there is a better way to minimize distortions of embedding if more elements are taken into account.

In order to increase the performance of SVD-based blind watermarking in $U$ domain some improvements are proposed in this paper. First we provide that modified $U$-matrix is orthonormal which improves robustness. Second we propose different embedding rules that maintain different robustness-transparency trade-off which improves flexibility. Third we minimize embedding distortions which reduces visual degradation of original image.

# 3 PROPOSED METHOD

Taking into account disadvantages of previously proposed SVD-based watermarking methods new approach is considered in this section. The improvements incorporated in our approach provide that altered $U$ matrix is orthogonal and normalized. Different embedding rules are also proposed.

Satisfying orthogonality requirement would consequently imply better robustness as all the changes introduced to the most robust part of a matrix (the first column) would not have projections on other dimensions (defined by second, third and fourth columns) except the dimension defined by that part. In order to provide this a special kind of approximation of an initial orthogonal matrix is proposed.

Another improvement considered to enhance robustness while preserving most of an original image is normalization of altered orthogonal matrix. Even in case each of original orthogonal matrices defined by SVD is normalized, embedding of a watermark according to (Chang, 2005), (Tehrani, 2010) cancels this quality. In contrast to that our embedding method assures each watermarked orthogonal matrix is normalized.

The way watermark bits are interpreted also significantly influences robustness. The only kind of matrix elements interpretation described in (Chang, 2005), (Tehrani, 2010) is the comparison of absolute values of the second and the third elements in the first column. In some cases we could greatly benefit from different ways of interpretation that take into account more elements. Our method of embedding utilizes five different embedding rules where each rule has an advantage under an assumption of some kind of distortion.

## 3.1 Approximation of Orthogonal Matrix

The approximation of an initial orthogonal matrix proposed in this paper is based on 4x4 matrix that can be described by 4 variables in different combinations. Each combination creates an entry in a set. One matrix $A$ from the possible set is represented as following:

$$A = \begin{bmatrix} -a & c & d & b \\ d-b & a & c & \\ b & d-c & a & \\ c & a & b-d & \end{bmatrix}. \quad (3)$$

This matrix is always orthogonal and under an assumption single row (or column) is normalized the whole matrix is normalized too. Similarly to widely used basis functions this matrix is described compactly (just 4 variables) but in contrast to them each separate element in a row (or column) is free from being functionally dependent on others. Such a quality makes these matrices quite suitable for accurate and computationally light approximations of original orthogonal matrices obtained after SVD of square image fragments. Moreover every matrix from the set is a distinctive pattern which could be used to assess the distortions introduced after watermark is embedded. Optionally this distinction could be used to determine during extraction which matrix from equally suitable $U$ and $V$ caries watermark's bit. The whole set of proposed orthogonal matrices and option to choose between embedding in $U$ or $V$ is necessary to achieve minimal total distortion that consists of an approximation error and a distortion caused by embedding according to some rule.

There could be several approximation strategies considering models from the proposed set of

orthogonal matrices. The main idea of embedding is to provide extraction of watermark bits from orthogonal matrices obtained after SVD with highest possible rate while preserving high enough image quality. Extraction is possible if during embedding a watermarked image fragment is composed using one diagonal matrix $S$ and two orthogonal matrices $U_w$ and $V$ (here $U_w$ is defined to store a bit).

Suppose now we are preparing (or approximating) the first orthogonal matrix $U$ for embedding, so the result is $U_w^p$, but the second orthogonal matrix $V$ remains unchanged. As we do not embed in singular values there is no need to care about the content of the diagonal matrix except the requirement that it should be diagonal. So let modified matrix of singular values be $S^*$ and possibly different from original $S$. Having the original image fragment $I$ of size 4x4 it can be written:

$$\left(U_w^p\right)^T \cdot I \cdot V = S^*. \tag{4}$$

Note that in case of such approximation strategy it is only required to satisfy twelve off-diagonal elements of $S^*$ are as small as possible (in Least Squares sense). Then after approximation is done those twelve elements should be put to zero, so approximation error causes some distortion of image fragment before the actual embedding.

Another approximation strategy is to provide both $S$ and $V$ are unchanged. In that case it is necessary to approach:

$$U_w^p \cdot S \cdot V^T = I. \tag{5}$$

This is more challenging task as it is required to match sixteen pixels as close as possible using the same model of orthogonal matrix defined by just four variables. However, this kind of approximation strategy could have some advantage in perceptual sense because singular values are preserved.

For our particular realization of watermarking method it was decided to limit watermark embedding by the first kind of approximation only. In order to show in more details the approximation with proposed orthogonal matrix let us substitute the matrix product $I \cdot V$ in (4) with 4x4 matrix $B$:

$$\left(U_w^p\right)^T \cdot B = S^*. \tag{6}$$

Now let's substitute $\left(U_w^p\right)^T$ with orthogonal matrix A in (3):

$$A \cdot B = S^*. \tag{7}$$

Matrix $S^*$ for simplicity could be transformed from 4x4 to 1x16 vector $S_v^*$ by rearranging elements of $S^*$ row by row which will lead to the following equation:

$$[a\ \ b\ \ c\ \ d] \cdot B^* = S_v^*, \tag{8}$$

where

$$B^* = \begin{bmatrix} -B_{1,1} & -B_{1,2} & -B_{1,3} & -B_{1,4} & B_{3,1} & B_{3,2} & B_{3,3} & B_{3,4} & B_{4,1} \\ B_{4,1} & B_{4,2} & B_{4,3} & B_{4,4} & -B_{2,1} & -B_{2,2} & -B_{2,3} & -B_{2,4} & B_{1,1} \\ B_{2,1} & B_{2,2} & B_{2,3} & B_{2,4} & B_{4,1} & B_{4,2} & B_{4,3} & B_{4,4} & -B_{3,1} \\ B_{3,1} & B_{3,2} & B_{3,3} & B_{3,4} & B_{1,1} & B_{1,2} & B_{1,3} & B_{1,4} & B_{2,1} \\ B_{4,2} & B_{4,3} & B_{4,4} & B_{2,1} & B_{2,2} & B_{2,3} & B_{2,4} & & \\ B_{1,2} & B_{1,3} & B_{1,4} & B_{3,1} & B_{3,2} & B_{3,3} & B_{3,4} & & \\ -B_{3,2} & -B_{3,3} & -B_{3,4} & B_{1,1} & B_{1,2} & B_{1,3} & B_{1,4} & & \\ B_{2,2} & B_{2,3} & B_{2,4} & -B_{4,1} & -B_{4,2} & -B_{4,3} & -B_{4,4} & & \end{bmatrix}.$$

Equation (8) can be simplified by ignoring 1, 6, 11 and 16 columns and elements of $B^*$ and $S_v^*$ respectively because for the current kind of approximation diagonal elements of $S^*$ are not important. By doing so we will get $B^{**}$ and zero vector $\mathbf{0}_{1\times12}$:

$$B^{**} = \begin{bmatrix} -B_{1,2} & -B_{1,3} & -B_{1,4} & B_{3,1} & B_{3,3} & B_{3,4} & B_{4,1} & B_{4,2} & B_{4,4} \\ B_{4,2} & B_{4,3} & B_{4,4} & -B_{2,1} & -B_{2,3} & -B_{2,4} & B_{1,1} & B_{1,2} & B_{1,4} \\ B_{2,2} & B_{2,3} & B_{2,4} & B_{4,1} & B_{4,3} & B_{4,4} & -B_{3,1} & -B_{3,2} & -B_{3,4} \\ B_{3,2} & B_{3,3} & B_{3,4} & B_{1,1} & B_{1,3} & B_{1,4} & B_{2,1} & B_{2,2} & B_{2,4} \\ & & & B_{2,1} & B_{2,2} & B_{2,3} & & & \\ & & & B_{3,1} & B_{3,2} & B_{3,3} & & & \\ & & & B_{1,1} & B_{1,2} & B_{1,3} & & & \\ & & & -B_{4,1} & -B_{4,2} & -B_{4,3} & & & \end{bmatrix},$$

$$[a\ \ b\ \ c\ \ d] \cdot B^{**} = \mathbf{0}_{1\times12}. \tag{9}$$

It is natural to suggest that simplest solution for (9) is $a = b = c = d = 0$, but taking into account requirement for $A$ to be normalized the solution is not as trivial:

$$\begin{cases} [a\ \ b\ \ c\ \ d] \cdot B^{**} = \mathbf{0}_{1\times12} \\ a^2 + b^2 + c^2 + d^2 = 1 \end{cases}. \tag{10}$$

Obviously such a regularized overdetermined system represents non-linear Least Squares task.

For further embedding it is required to prepare a set of approximated orthogonal matrices where

matrix $A$ is just one possible variant for final decision.

Five embedding rules were introduced to improve robustness. Each rule is a condition that could be satisfied in different ways, so we tried to minimize distortions introduced on that step too. Thanks to simplicity of our orthogonal matrix model minimization of embedding distortions can also be done quite easily. Suppose that as a result of watermark embedding matrix $A$ has been changed and become $A^*$. Because it is required to keep $A^*$ normalized we will accept for further simplicity that there is some vector $[\Delta a, \Delta b, \Delta c, \Delta d]$ with length 1 which is orthogonal to $[a, b, c, d]$ and $A^*$ is formed from

$$a^* = \sqrt{1-n^2}\cdot a + n\cdot \Delta a,\ b^* = \sqrt{1-n^2}\cdot b + n\cdot \Delta b;$$
$$c^* = \sqrt{1-n^2}\cdot c + n\cdot \Delta c,\ d^* = \sqrt{1-n^2}\cdot d + n\cdot \Delta d;$$

where $0 \le n \le 1$. The result of extraction of a watermarked image fragment from unwatermarked will be:

$$A\cdot S^*\cdot V^T - A^*\cdot S^*\cdot V^T = (A - A^*)\cdot S^*\cdot V^T. \quad (11)$$

Matrix $A - A^*$ is orthogonal as $A^*$ has the same structure as $A$. Consequently the Sum of Square Residuals (SSR) between watermarked and unwatermarked fragments can be defined as:

$$SSR = norm^2(A - A^*)\cdot \sum_{i=1}^{4}(S^*_{i,i})^2 =$$
$$= norm^2\left(A - \left(\sqrt{1-n^2}\cdot A + n\cdot \Delta A\right)\right)\sum_{i=1}^{4}(S^*_{i,i})^2. \quad (12)$$

Here $\Delta A$ is formed from $\Delta a, \Delta b, \Delta c, \Delta d$ and is normalized. Further simplification taking into account the previously made assumptions will produce an equation:

$$SSR = 2\left(1 - \sqrt{1-n^2}\right)\cdot \sum_{i=1}^{4}(S^*_{i,i})^2. \quad (13)$$

According to (13) distortion of image fragment caused by watermark embedding in our method depends on the length of the vector added to the first column of orthogonal matrix $A$ and does not depend on a vector's orientation in contrast to the method proposed in (Chang, 2005), (Tehrani, 2010). This quality could greatly simplify procedure for minimization of watermarking distortions and enable more different embedding rules to be used. Equation (13) also provides an understanding that the same embedding amplitude could lead to different distortions in different image fragments because of influence of singular values.

## 3.2 Embedding Rules

Proposed embedding rules could be split in two groups. The first group consists of rules $L1_4$, $L2_4$ and $L\infty_4$ that utilize all the four elements of the first column of orthogonal matrix for both embedding and retrieving. The second group consists of rules $L1_2$ and $L2_2$ that utilize just two elements for retrieving, however, could change four elements for embedding because optimization takes place under normalization constraint. Further suppose we are embedding bit $b$ in $U$ with a positive non-zero threshold $T$:

$$L1_4: (-1)^b \cdot \left(\left\|\left(U^*_{1,1}, U^*_{2,1}\right)\right\|_1 - \left\|\left(U^*_{3,1}, U^*_{4,1}\right)\right\|_1\right) \ge T;$$

$$L2_4: (-1)^b \cdot \left(\left\|\left(U^*_{1,1}, U^*_{2,1}\right)\right\|_2 - \left\|\left(U^*_{3,1}, U^*_{4,1}\right)\right\|_2\right) \ge T;$$

$$L\infty_4: (-1)^b \cdot \left(\left\|\left(U^*_{1,1}, U^*_{2,1}\right)\right\|_\infty - \left\|\left(U^*_{3,1}, U^*_{4,1}\right)\right\|_\infty\right) \ge T; \quad (14)$$

$$L1_2: (-1)^b \cdot \left(\left\|U^*_{2,1}\right\|_1 - \left\|U^*_{3,1}\right\|_1\right) \ge T;$$

$$L2_2: (-1)^b \cdot \left(\left\|U^*_{2,1}\right\|_2 - \left\|U^*_{3,1}\right\|_2\right) \ge T.$$

For each embedding rule there is the same additional normalization constraint and the same goal function to minimize distortions (that is quite simple thanks to the proposed orthogonal matrix):

$$\begin{matrix}\left\|\left(U^*_{1,1}, U^*_{2,1}, U^*_{3,1}, U^*_{4,1}\right)\right\|_2 = 1 \\ \sum_{i=1}^{4}\left(U^*_{i,1} - U_{i,1}\right)^2 \to min\end{matrix}. \quad (15)$$

## 3.3 Watermarking Procedure

After embedding is done the resulting matrix $U^*$ should be composed with $S^*$ and $V^T$ which produces watermarked image fragment $I^*$. However, it is necessary to notice that $I^*$ contains real-valued pixels instead of integers. There are many possible kinds of truncation and each kind distorts orthogonal matrix $U^*$, but, for example, simple round operation is quite negligible to retrieve a bit for some

reasonable $T$ (0.02 works well for all the embedding rules). A diagram of watermark embedding is shown on Figure 1.

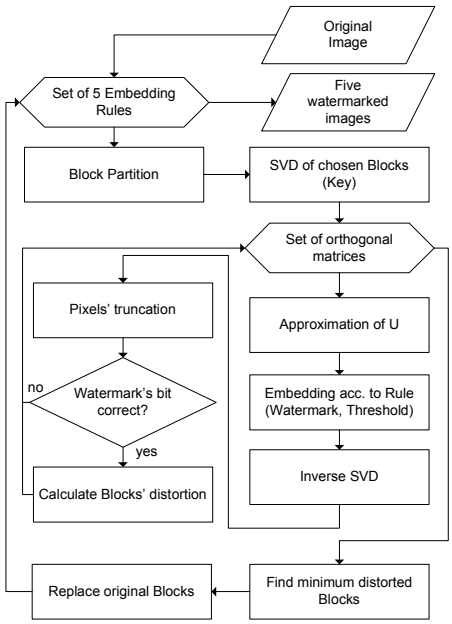



Figure 1: Watermark embedding diagram.

As it follows from the diagram the least distorted watermarked fragments are chosen in order to replace the corresponding original fragments of the image. This is thanks to availability of different orthogonal matrices in the set used for the approximation. It is necessary to notice that in the current realization we utilized constant threshold for all the blocks, but threshold adaptation can be done in the future more easily (at once, non-iteratively) compared to (Tehrani, 2010) as distortion in our method depends only on the amplitude of a vector added to the first column of $U$.

To extract a watermark from the watermarked image it is required to know the key and the rule. However in contrast to embedding the extraction threshold for each rule is zero. The extraction diagram is given on Figure 2.

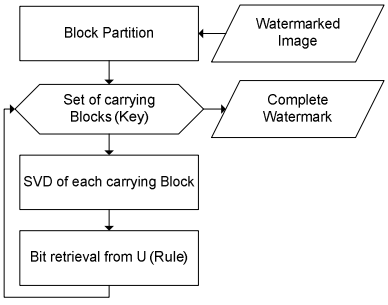



Figure 2: Watermark extraction diagram.

In our realization we also avoided embedding area to be limited only by blocks with greater complexity as defined in (Chang, 2005), (Tehrani, 2010), because due to some kind of distortion complexity (namely the number of non-zero singular values per block) could change and the person extracting a watermark could mismatch a key on different set. Another reason is that such a set has different size for different images which forces to use synchronized PRNG (Pseudorandom Number Generator, not steady key as we use) between embedder and extractor which is impractical.

# 4 EXPERIMENTAL RESULTS

In order to confirm the improvements of the proposed watermarking method some experiments took place. Each result has been compared with the result provided by the method described in (Chang, 2005) under the same circumstances. Original images, watermarking key and watermark were absolutely identical. We tried to adjust parameters so that Peak Signal to Noise Ratios (PSNRs) between each original image and the corresponding watermarked one were very close for both methods. There were four kinds of distortions used in the experiment: white Gaussian noise, speckle noise, “salt&pepper” and Jpeg compression.

Three grayscale host images with dimension 512x512 and bitdepth 8 bit were used for watermark embedding. Those images appear to be tested quite widely in papers related to image processing and are namely: livingroom.tif, mandril.tif and cameraman.tif (Figure 3-5). The choice of images for watermarking could be explained in a way that we tried to compare a performance of the proposed method on images with different amount of fine details. Here image livingroom.tif contains some areas with fine details, mandril.tif has a lot of fine details and cameraman.tif contains few details while having quite large areas with almost constant background.

The watermark for all our tests is the same and is 1024 bit long. For the better visual demonstration of each method’s robustness it has been prepared in a form of square binary 32x32 image that depicts Canadian maple leaf. Each bit of the watermark has been embedded according to the same key (generated randomly) for all the images. The key defines 4x4 image fragments used for watermarking and is 16384 bit long. Extraction is done using the same key. Without distortions extraction of the

watermark is absolutely correct for all the methods and images.

Figure 3: Original grayscale image livingroom.tif.

Taking into account that different rules were used for embedding in our method and the approximation had been done previously comparison with the method proposed in (Chang, 2005) is more complex. The only parameter influencing robustness in that method is a threshold, but embedding with the same threshold has different impact on an image when both methods are used. Therefore, the threshold for the method proposed in (Chang, 2005) has been adjusted after embedding by our method is done in a way that each in a pair of the corresponding watermarked images has the same (or very similar) PSNR.

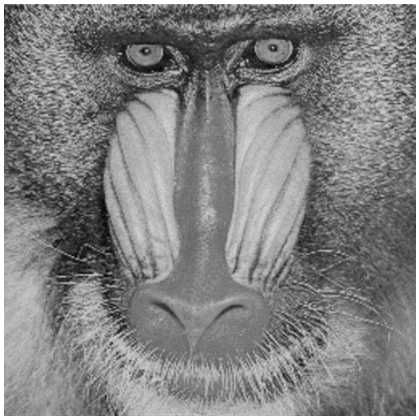


Figure 4: Original grayscale image mandril.tif.

Figure 5: Original grayscale image cameraman.tif.

Four models of distortions applied to the watermarked images in our experiment could be split in two types according to the noise nature: additive and non-additive. Distortions with additive noise are namely Gaussian and speckle. Before applying distortions to watermarked images their pixel values were scaled to match interval [0, 1]. The mean for Gaussian is 0 and the variance shown in tables is 0.0006. Speckle noise adds, to each pixel $p_{i,j}$, the term $x \cdot p_{i,j}$ where $x$ is distributed uniformly with mean 0 and variance 0.001. Distortions utilizing non-additive noise types are “salt & pepper” and lossy jpeg-compression. In our experiments we have applied 3% “salt & pepper” and 75 image quality for jpeg (Matlab realization).

An extraction with the key has been done afterwards. To compare the results we used the value 1-BER (Bit Error Rate) which indicates the fraction of correctly extracted bits of a watermark. We have placed the values 1-BER calculated according to each method, embedding rule and distortion type in separate table for each image (Tables 1-3). Each result has been averaged among 100 runs for all kinds of distortions except jpeg (as it is straightforward and does not contain random component). For better comparability each row with results from our method was neighbored to a row containing results with similar PSNR from method (Chang, 2005). For every pair of rows better indicator of robustness toward particular distortion is bolded.

Table 1: Results of watermark extraction for livingroom.tif.

| **Method, Rule** | **Gaussian, 0.0006** | **Speckle, 0.001** | **Salt & pepper, 0.03** | **Jpeg, 75** |
|---|---|---|---|---|
| $L1_4$, 46.13dB | 0.9325 | 0.9823 | 0.8451 | **0.9844** |
| Chang,46.02dB | **0.9737** | **0.9997** | **0.8986** | 0.9170 |
| $L2_4$, 49.68dB | **0.8581** | 0.9288 | 0.8333 | **0.9268** |
| Chang,49.60dB | 0.8571 | **0.9464** | **0.8952** | 0.7324 |
| $L\infty_4$, 49.93dB | **0.8797** | **0.9602** | 0.8805 | **0.9092** |
| Chang,49.83dB | 0.8410 | 0.9326 | **0.8967** | 0.7227 |
| $L1_2$, 50.22 dB | **0.8833** | **0.9660** | **0.8954** | **0.8076** |
| Chang,50.22dB | 0.8063 | 0.8961 | 0.8950 | 0.6865 |
| $L2_2$, 50.22 dB | **0.8847** | **0.9662** | **0.8975** | **0.8066** |
| Chang,50.22dB | 0.8063 | 0.8961 | 0.8950 | 0.6865 |

Table 2: Results of watermark extraction for mandril.tif.

| **Method, Rule** | **Gaussian, 0.0006** | **Speckle, 0.001** | **Salt & pepper, 0.03** | **Jpeg, 75** |
|---|---|---|---|---|
| $L1_4$, 42.37dB | 0.9681 | 0.9902 | 0.8690 | **0.9961** |
| Chang,42.29dB | **0.9976** | **1.0000** | **0.9070** | 0.9775 |
| $L2_4$, 46.12dB | 0.9026 | 0.9469 | 0.8539 | **0.9297** |
| Chang,46.11dB | **0.9648** | **0.9988** | **0.8988** | 0.8174 |
| $L\infty_4$, 46.70dB | 0.9138 | 0.9685 | 0.8837 | **0.8652** |
| Chang,46.65dB | **0.9492** | **0.9949** | **0.8981** | 0.7822 |
| $L1_2$, 47.55dB | **0.9099** | 0.9715 | **0.9000** | **0.8057** |
| Chang,47.54dB | 0.9060 | **0.9736** | 0.8979 | 0.7236 |
| $L2_2$, 47.54dB | **0.9111** | 0.9716 | 0.8978 | **0.8076** |
| Chang,47.54dB | 0.9060 | **0.9736** | **0.8979** | 0.7236 |

Table 3: Results of watermark extraction for cameraman.tif.

| **Method, Rule** | **Gaussian, 0.0006** | **Speckle, 0.001** | **Salt & pepper, 0.03** | **Jpeg, 75** |
|---|---|---|---|---|
| $L1_4$, 45.70dB | 0.8908 | 0.9745 | 0.8322 | **0.9336** |
| Chang,45.70dB | **0.9153** | **0.9932** | **0.8918** | 0.8125 |
| $L2_4$, 50.89dB | **0.8123** | **0.8933** | 0.8104 | **0.8926** |
| Chang,50.82dB | 0.7927 | 0.8876 | **0.8471** | 0.6094 |
| $L\infty_4$, 51.06dB | **0.8419** | **0.9327** | **0.8667** | **0.8467** |
| Chang,51.05dB | 0.7808 | 0.8712 | 0.8410 | 0.5840 |
| $L1_2$, 52.32dB | **0.8377** | **0.9338** | **0.8591** | **0.7227** |
| Chang,52.30dB | 0.6716 | 0.7365 | 0.8317 | 0.4922 |
| $L2_2$, 52.31dB | **0.8419** | **0.9348** | **0.8603** | **0.7217** |
| Chang,52.30dB | 0.6716 | 0.7365 | 0.8317 | 0.4922 |

Images watermarked by the proposed method are depicted in Figures 6-8. The rule $L2_4$ has been used for this particular demonstration and PSNRs are 49.68 dB, 46.12 dB and 50.89 dB for livingroom.tif, mandril.tif and cameraman.tif respectively.

Figure 6: Watermarked grayscale image livingroom.tif.

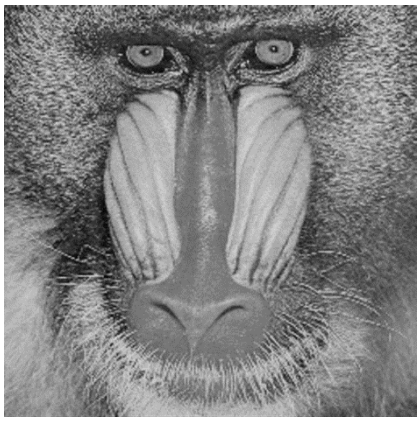
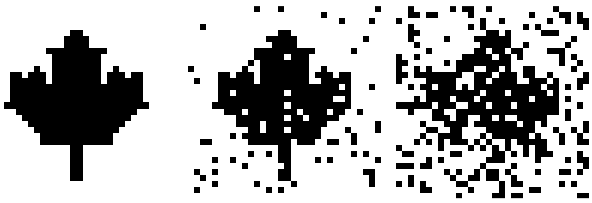


Figure 7: Watermarked grayscale image mandril.tif.

Figure 8: Watermarked grayscale image cameraman.tif.

The threshold for the method proposed in (Chang, 2005) has been adjusted so that very similar PSNR has been achieved for each watermarked image. Compression according to jpeg standard has been done then. The watermarks extracted from the watermarked image livingroom.tif by both methods are shown together with the original watermark

extracted from non-distorted watermarked image (Figure 9).



(a) (b) (c)

Figure 9: Original and distorted by jpeg compression watermarks.

The demonstrated binary images represent watermarks extracted with rates 1 (Figure 9. (a), both methods, no distortion), 0.9268 (Figure 9. (b), our method, jpeg 75), 0.7324 (Figure 9. (c), method (Chang, 2005), jpeg 75).

# 5 DISCUSSION

Comparing the rate of correct watermark extraction for our method and the method proposed in (Chang, 2005) and further developed in (Tehrani, 2010) we can state the following. Robustness demonstrated by our method against jpeg attack is much better than those demonstrated by (Chang, 2005). This is true for all the embedding rules, but to be said separately rule $L2_4$ provides the greatest improvement for all the trials with jpeg-compression: it is about 27% better on livingroom.tif, about 14% better on mandril.tif and more than 46% better on cameraman.tif.

For other types of distortions including Gaussian, speckle, "salt&pepper" noises rules $L1_2$ and $L2_2$ performed better than others: about 10% outperform (Chang, 2005) for Gaussian on livingroom.tif, just 1% better than (Chang, 2005) for Gaussian on mandril.tif, but 27% better than (Chang, 2005) for speckle on cameraman.tif. There was no considerable advantage found for "salt&pepper" noise for any rule. However rules $L1_4$ and $L2_4$ usually perform worse under conditions with Gaussian, speckle, and "salt&pepper" noises. The rule $L\infty_4$ has considerable advantage on jpeg which is close to the advantage $L2_4$ has and under conditions with Gaussian, speckle, and "salt&pepper" noises in some cases performs several percent better than (Chang, 2005) (livingroom.tif and cameraman.tif). The highest achievement for rule $L1_4$ is to be 15% better toward (Chang, 2005) under jpeg-attack for cameraman.tif, but the gaps in trials with Gaussian, speckle, and "salt&pepper" noises are sometimes too high, so, it should probably be rejected from future experiments.

It is possible to issue a short guidance for end-user that reflects better flexibility of proposed method utilizing different rules: embedding rules $L1_2$ and $L2_2$ should be used if there are comparable chances for each kind of tested distortions to occur; rule $L\infty_4$ is better to be used when chances of jpeg compression are higher; we recommend to use rule $L2_4$ in case the only kind of possible distortion is jpeg.

The threshold used in all our embedding rules was the same. On the other hand, PSNRs of the watermarked images are quite high. So, in the future we would like to experiment with different values of the threshold (probably greater) and also apply adaptation for each block as it is proposed in (Tehrani, 2010). Another direction we might wish to explore is an embedding in $U$ matrix of the blocks of greater size, but this requires a different model of orthogonal matrix to be used for approximation.

# 6 CONCLUSIONS

The watermarking method operating on $U$-domain of $SVD$ transform was proposed. Its robustness is better than those for the method proposed in (Chang, 2005). The improvements are due to optimizations done on two stages of embedding.

The first stage serves for the approximation of $U$ matrix of transformed 4x4 image blocks. The approximation was done according to the proposed model that describes orthogonal matrix analytically. This procedure allows to preserve orthogonality of $U$ matrix after watermark bit is embedded. Orthogonality of $U$-matrix improves extraction rate.

The second stage represents an embedding according to one of five proposed embedding rules. Each of the embedding rules has its own trade-off between robustness and transparency which allows to choose the best rule for particular application. A minimization of embedding distortions was done for each rule during embedding which reduces degradation of original image.

Several kinds of attacks were applied to test robustness. It was experimentally confirmed that for each kind of attack there is a different embedding rule which is more preferable than the others. However, watermarking according to each of the proposed embedding rules outperforms the method proposed in (Chang, 2005) under condition of JPEG-attack.

## ACKNOWLEDGEMENTS

The first author is thankful to Tampere Program in Information Science and Engineering for the support.

## REFERENCES

Altun, O., Sharma, G., Celik, M., Bocko, M., 2006. A Set Theoretic Framework for Watermarking and Its Application to Semifragile Tamper Detection. *IEEE Transactions on Information Forensics and Security*, vol.1, no.4, 479-492.

Barni, M., Bartolini, F., Cappellini, V., Piva, A., 1997. Robust watermarking of still images for copyright protection. In *Proceedings of 13th International Conference on Digital Signal Processing Proceedings*, IEEE, Piscataway, NJ, USA, vol.2, 499-502.

Chang, C., Tsai, P., Lin, C., 2005. SVD-based digital image watermarking scheme. *Pattern Recognition Letters*, 26, 1577–1586.

Cox, I., Miller, M., Bloom, J., Fridrich, J., Kalker, T., 2007. *Digital Watermarking and Steganography*, Morgan Kaufmann Publishers Inc.. San Francisco, CA, USA, 2nd edition.

Dharwadkar, N.V., Amberker, B.B., Gorai, A., 2011. Non-blind watermarking scheme for color images in RGB space using DWT-SVD. In *Proceedings of ICCSP'11, International Conference on Communications and Signal Processing*. IEEE, Piscataway, NJ, USA, 489-493.

Fridrich, J., 2002. Security of fragile authentication watermarks with localization. *Proceedings of SPIE 4675, Security and Watermarking of Multimedia Contents IV*, 4675, Vol. 4675, 691-700.

Fullea, E., Martinez, J. M., 2001. Robust digital image watermarking using DWT, DFT and quality based average. In *Proceedings of MULTIMEDIA'01, 9th International Conference on Multimedia.* ACM, New York, NY, USA, 489-491.

Ganic, E., Eskicioglu, A. M., 2004. Robust DWT-SVD domain image watermarking: embedding data in all frequencies. In *Proceedings of MM&Sec'04, Workshop on Multimedia and security.* ACM, New York, NY, USA, 166-174.

Lin, S.D., Chen, C. F., 2000. A robust DCT-based watermarking for copyright protection. *IEEE Transactions on Consumer Electronics*, 46, 415-421.

Manjunath, M., Siddappaji, S., 2012. A new robust semi blind watermarking using block DCT and SVD. In *Proceedings of ICACCCT'12, International Conference on Advanced Communication Control and Computing Technologies.* IEEE, Piscataway, NJ, USA, 193-197.

Modaghegh, H., Khosravi, R. H., Akbarzadeh, T., 2009. A new adjustable blind watermarking based on GA and SVD. In *Proceedings of IIT '09, International Conference on Innovations in Information Technology.* IEEE, Piscataway, NJ, USA, 6-10.

Pei S. C., Zeng Y. C., 2006. Tamper proofing and attack identification of corrupted image by using semi-fragile multiple-watermarking algorithm. In *Proceedings of the ASIACCS'06, Symposium on Information, computer and communications security.* ACM, New York, NY, USA, 166-174.

Quan, L., Qingsong, A., 2004. A combination of DCT-based and SVD-based watermarking scheme. In *Proceedings of ICSP'04, 7th International Conference on Signal Processing.* IEEE, Piscataway, NJ, USA, vol.1, 873- 876.

Singh, P., Agarwal, S., 2012. A region specific robust watermarking scheme based on singular value decomposition. In *Proceedings of SIN'12, 5th International Conference on Security of Information and Networks.* ACM, New York, NY, USA, 117-123.

Sun, R., Sun, H., Yao, T., 2002. A SVD- and quantization based semi-fragile watermarking technique for image authentication. In *Proceedings of 6th International Conference on Signal Processing.* IEEE, Piscataway, NJ, USA, vol.2, 1592- 1595.

Tehrani, I.O., Ibrahim, S., 2010. An enhanced SVD based watermarking using U matrix. In *Proceedings of ICCIT'10, 5th International Conference on Computer Sciences and Convergence Information Technology.* IEEE, Piscataway, NJ, USA, 627-631.

Xiao, J., Wang, Y., 2008. Toward a Better Understanding of DCT Coefficients in Watermarking. In *Proceedings of PACIIA'08, Pacific-Asia Workshop on Computational Intelligence and Industrial Application.* IEEE, Piscataway, NJ, USA, vol.2, 206-209.

Yongdong, W., 2005. On the security of an SVD-based ownership watermarking. *IEEE Transactions on Multimedia*, vol.7, no.4, 624- 627.

Publication II

# An SVD-based Transparent Watermarking Method

Yevhen Zolotavkin and Martti Juhola

Publication III

# A New Blind Adaptive Watermarking Method Based on Singular Value Decomposition

Yevhen Zolotavkin and Martti Juhola

# A New Blind Adaptive Watermarking Method Based on Singular Value Decomposition

Yevhen Zolotavkin
Computer Science, School of Information Sciences
University of Tampere
Tampere, Finland
yevhen.zolotavkin@uta.fi

Martti Juhola
Computer Science, School of Information Sciences
University of Tampere
Tampere, Finland
martti.juhola@sis.uta.fi

***Abstract*—A blind watermarking method on the basis of Singular Value Decomposition is proposed in this paper. Each bit of a watermark is being enclosed in 4x4 blocks. The method modifies the both left and right orthonormal matrices in order to embed a bit. A new embedding rule with adjustable parameters has been proposed for watermarking. The modification of orthonormal matrices is accomplished according to Van Elfrinkhof's rotational model. Distortions of watermark embedding are minimized. A criterion of watermarking performance has been proposed that combines robustness and transparency. An adaptation on the basis of the criterion has been employed. Popular attacks have been applied and experimental results have been represented. The proposed watermarking method demonstrates better robustness toward some attacks in comparisonwith other known blind watermarking methods.**

***Keywords-Digital Image Watermarking, Singular Value Decomposition, Robustness, Distortions, Transparency***

## I. INTRODUCTION

Security of data is a very important requirement of modern society. There are many different aspects of security that are applicable in different circumstances. One of the most important aspects is a protection of digital rights for a work produced by an author. These kinds of information security problems are addressed by Digital Image Watermarking (DIW).

To protect a digital image by the means of DIW it is necessary to enclose a digital watermark that would witness an owner[1]. Therefore there are three important characteristics for a particular watermarking method: robustness, transparency and data payload.

Robustness is an ability to withstand different kinds of attacks [2]. It is impervious to provide robustness toward all the possible attacks especially if their intensities are high. Hence this requirement is quite specific. However, mostly robustness against noise, some kinds of filtering and geometric attacks is required. The most approved index of robustness for an extracted watermark is Bit Error Rate (BER).

Transparency is an ability to preserve original image by watermarking it. There are many measures of image quality that could be applied to define transparency quantitatively[3]. Though, the most popular measure is Peak Signal to Noise Ratio (PSNR).

Data payload is a number of watermark bits embedded into an image. There might be different requirements to data payload as there might be different kind of information to witness an ownership. Nevertheless higher payload provides better protection as the watermark can be more unique. Small binary graphical logos are the most popular choice in watermarking. Sequences of randomly generated bits without visual meaning are also favored.

The original image can be modified in many different ways to embed a watermark. Original pixel values can be changed directly which is a kind of spatial transform. Modification of the Least Significant Bit is a good example of such kind of transforms[1]. Another kind of embedding is to change coefficients that have some spectral meaning which is a frequency domain transform. Some suitable examples are watermarking methods on the basis of Discrete Cosine Transform (DCT) [4] and Discrete Wavelet Transform (DWT)[5]. Robustness and transparency can be greatly influenced by the kind of transform chosen for embedding. Usually modification of some spectral coefficients is more favorable as they are more robust against noise and image processing attacks.

Singular Value Decomposition (SVD) is a unique kind of transform[6]. It separates an image fragment on several independent layers. The number of layers is much less than that, for example, for DCT. Therefore the most important layer is quite stable to various attacks.

An efficiency of watermarking also depends on a rule exploited for embedding. Each embedding rule could have several parameters that influence robustness-transparency tradeoff. Those parameters could remain constant for the whole watermarking procedure or be different (adopted) for each independent block. Usually embedding with adopted parameters provides better watermarking performance.

There are many existing SVD-based watermarking methods. The best of them provide adaptation of embedding parameters. However, additional information is usually required for extraction which limits their usage. For those few methods that do not urge transfer of additional information embedding requires modification of more coefficients in a

block. This implies that larger blocks are used and lower data payload can be maintained.

In this paper we propose new SVD-based blind watermarking method with adaptation. The method doesnot need additional information except a key to extract a watermark. It uses the both orthonormal matrices obtained by SVD of 4x4 block to embed a bit of a watermark. The proposed method provides good robustness-transparency tradeoff and high data payload.

The rest of the paper is organized asfollowing: a short review of relevant watermarking methods exploiting SVD is given in the Section II; Section III bears our own approach which is described in detail; then, some experimental results are represented in Section IV followed by a discussion of their importance in Section V; finally, in Section VI the paper is concluded by general remarks regarding relevance of our approach and its influence on future research.

## II. SVD-BASEDWATERMARKING

An image fragment $I_k$ of size $n \times n$ is being decomposed according to SVD[6] in the following way:

$$I_k = USV^T = \begin{pmatrix} U_{1,1} \cdots U_{1,n} \\ U_{2,1} \cdots U_{2,n} \\ \vdots \\ U_{n,1} \cdots U_{n,n} \end{pmatrix} \times \begin{pmatrix} S_{1,1} & 0 & \cdots & 0 \\ 0 & S_{2,2} & \cdots & 0 \\ & & \vdots & \\ 0 & 0 & \cdots & S_{n,n} \end{pmatrix} \times \begin{pmatrix} V_{1,1} \cdots V_{1,n} \\ V_{2,1} \cdots V_{2,n} \\ \vdots \\ V_{n,1} \cdots V_{n,n} \end{pmatrix}^T, \quad (1)$$

where $U$ and $V$ are some orthonormal matrices and $S$ is a diagonal matrix of singular values.

An alternative representation demonstrates that fragment $I_k$ is being decomposed on $n$ independent layers where geometry of $i$-th layer is defined by a pair of $i$-th columns (one from matrix $U$ and one from $V$) and a luminance component $S_{i,i}$:

$$I_k = \sum_i S_{i,i} U(1 \ldots n, i) V(1 \ldots n, i)^T \quad (2)$$

The luminance $S_{1,1}$ has the biggest value and mostly this value is much bigger than the othervalues $S_{i,i}$, $i > 1$. Therefore the first layer is the most substantial and provides the best robustness for watermarking.

### A. Methods Modifying Singular Values

Popular strategy for SVD-based methods that modify singular values is to quantize the biggest value of a block depending on the corresponding bit of a watermark.

The first paper introducing SVD for Digital Image Steganography and Watermarking was[7]. A blind technique with high data payload and without adaptation was proposed for color RGB images. However, the resulting robustness-transparency tradeoff was not satisfying mostly because of inability to quantize singular values of different blocks with different steps.

Another pioneering paper exploiting SVD for watermarking is[8] where noninvertible non-blind scheme was introduced. However, later in[9] it has been shown that the scheme is vulnerable to a kind of attack counterfeiting an original watermark because too much of reference information should be saved for a detector.

In the paper[10] the first in the literature DWT-SVD watermarking method was proposed. The method demonstrates good robustness and provides high data payload. However, distortion of original image is quite considerable, the method is non-blind and doesnot assume an adaptation.

The method proposed in[11] applies adoptive quantization to DWT-SVD. The method is robust against JPEG-compression. However, it is made deliberately fragile to other kinds of distortions like noise, median filtering or cropping. The information about quantization steps for all blocks should be transmitted. Another drawback is considerable degradation of original image.

Among the recent works exploiting adoptive quantization of singular values the paper[12] introduces quite robust watermarking method. However, information about quantization parameters should be transferred to extract a watermark.

One of the most robust blind watermarking schemes based on SVD-DCT transform was proposed in[13]. Two bits of a watermark are being embedded in 32x32 macro-block. An adaptation is applied to each block. The method doesnot require any additional information except a keyfor extraction.

### B. Methods ModifyingOrthonormal Matrices

In the literature there are few watermarking approaches that modify orthonormal matrices of SVD. The advantage of such kind of watermarking is that more elements are available for modification.

The paper[14] proposes a watermarking method that modifies the left orthonormal matrix of SVD. The whole image of size 512x512 is split on fragments 4x4 and SVD is applied to each of them. A bit of the watermark is embedded by modifying the second and the third elements in the first column of left orthonormal matrix. The method provides sufficient data payload and quality of watermarked images which PSNR was higher than 42 dB. However, robustness toward common distortions like JPEG-compression, Gaussian noise and cropping isnot high.

Another paper exploiting the idea of embedding a watermark in orthonormal matrix of SVD is[15]. The watermarking scheme proposed in[14] was developed further in order to improve robustness-invisibility tradeoff. Instead of embedding a bit of a watermark with constant threshold for all the blocks the authors proposed to adjust the threshold. The adjustment is done in a way that PSNR of each modified blocks is higher 42 dB whenever it is possible. The method provides considerable data payload equal to 2048 bit per image. Robustness-invisibility tradeoff is also better compared to[14]. Nevertheless its robustness is not sufficient toward, for example, JPEG-compression.

There are several shortcomings in the mentioned above two methods proposed in[14] and[15]. First modified matrices are not orthonormal which could cause an embedded bit to be lost even without influence of the third person or noise. Second none of the methods uses an adequate criterion to adapt the threshold for each block. The PSNR-based criterion and 42dB limit arenot obvious. Third both methods utilize only the left orthonormal matrix while utilization of the both could provide more elements for watermarking and improve robustness-transparency tradeoff.

## III. Proposed Watermarking Method

Proposed in this paper watermarking method modifies $U$ and $V$ that are left and right orthonormal matrices of SVD of particular image block $I_k$.

Each new watermarked image fragment $I'_k$ that carries corresponding bit is composed from two orthonormal matrices $\{U', V'\}$and a diagonal matrix of singular values:

$$I'_k = U'S'(V')^T. \quad (3)$$

Image block $I'_k$ should be decomposed by SVD again in order to extract a bit. The decomposition always returns orthogonal matrices. With the aim to assure that a bit is extracted correctly matrices $U'$ and $V'$ should be orthogonal when $I'_k$ is composed. Otherwise the matrices of the decomposition will not be the same as the matrices used to compose a watermarked block.

In order to provide orthogonality of $U'$ and $V'$ a multiplication with rotational matrix can be applied. Any rotational matrix $R$ is always orthonormal and multiplication with another orthonormal matrix, for example, $U$ will produce new orthonormal matrix. Any column of $U$ could be seen as a point and rotation according to $R$ changes coordinates of a point. This kind of modification of coordinates of a point can be used to embed a bit.

Our method embeds each bit of a watermark in a square fragment of image which size is $4 \times 4$. Only the first column of $U$ and the first column of $V$ represent a watermark bit.Transforms that are necessary for watermarking can be defined as$T_L: U \to U'$, and $T_R: V \to V'$. New watermarked matrices $\{U', V'\}$ are defined using rotation matrices $R_U$and $R_V$:

$$U' = R_U U, \quad (4)$$

$$V' = R_V V. \quad (5)$$

### A. Embedding Rule

Modified matrices $\{U', V'\}$ should satisfy some requirements necessary for proper extraction of a bit of a watermark. Those requirements can be expressed in a watermarking rule. Further we use a definition of transposed first columns of $U'$ and $V'$ respectively: $\boldsymbol{u}' = [U'(1,1), U'(2,1), U'(3,1), U'(4,1)]\boldsymbol{v}' = [V'(1,1), V'(2,1), V'(3,1), V'(4,1)]$. Two main components of a rule are reference matrix $Ref$ and a threshold $Th$. The rule is expressed as the following equation:

$$(-1)^{bit}\left(\boldsymbol{u}' Ref \boldsymbol{v}'^T - m\right) = Th, \quad (6)$$

where$m$ is a mean of the term $\boldsymbol{u}' Ref \boldsymbol{v}'^T$. Higher threshold $Th$ implies higher level of embedding distortions, but the robustness is also higher. To extract a bit of a watermark it is necessary to calculate the following expression:

$$bit = \left(2 + sign\left(\boldsymbol{u}' Ref \boldsymbol{v}'^T - m\right)\right) mod\ 3. \quad (7)$$

### B. Minimization of Embedding Distortions

While robustness of a watermark depends on the parameters of embedding rule invisibility of a watermark is a subject for minimization of some criteria as, for example, a Residual Sum of Squares (RSS) between original and altered pixel values of a block. In this subsection we presume that the proposed embedding rule is used and the both matrices $U$ and $V$ are being modified.

The proposed goal function $G$ for a watermarked fragment $I'_k$ is:

$$G = \|I'_k - I_k\|_2^2. \quad (8)$$

The goal function can be rewritten in order to include rotational matrices $R_U$and $R_V$:

$$G = \|I'_k - I_k\|_2^2 = \|U'S'(V')^T - I_k\|_2^2 = \\ = \|US'V^T - R_U^T I_k R_V\|_2^2. \quad (9)$$

If we define $S' = S + \Delta S$ where $\Delta S$ is also a diagonal matrix, expression (9) becomes:

$$G = \|U\Delta S V^T + I_k - R_U^T I_k R_V\|_2^2. \quad (10)$$

In case we further denote

$$G^* = \|I_k - R_U^T I_k R_V\|_2^2, \quad (11)$$

becomes clear that it is always possible to adjust $\Delta S$ in (10) to provide $G \le G^*$. It is possible to modify on the first stage $\{R_U, R_V\}$ with the aim to minimize $G^*$ and adjust $\Delta S$ on the second stage to minimize $G$. Such approach has its advantages and disadvantages. The advantage is that the approach is simpler because it doesnot require variables of $\Delta S$ to be taken into account and optimized on its first stage; optimization of $\Delta S$ on the second stage does not influence robustness; global minimum is easy to reach on the second stage. The disadvantage is that the solution is suboptimal in principle.

Taking into account that $\boldsymbol{u}' = (R_U \boldsymbol{u}^T)^T$and $\boldsymbol{v}' = (R_V \boldsymbol{v}^T)^T$, first stage optimization task including embedding constraint can be defined as:

$$\begin{cases} G^* = \|I_k - R_U^T I_k R_V\|_2^2 \to min; \\ (-1)^{bit}(\boldsymbol{u} R_U^T Ref R_V \boldsymbol{v}^T - m) = Th. \end{cases} \quad (12)$$

Therefore rotational matrices $R_U$ and $R_V$ should be calculated according to optimization procedure.

### C. Model for Rotations

Rotational matrix $R$ in four dimensional space can be fully described according to Van Elfrinkhof'sformulae [16]:

$$R = \begin{pmatrix} ap - bq - cr - ds & -aq - bp + cs - dr & -ar - bs - cp + dq & -as + br - cq - dp \\ bp + aq - dr + cs & -bq + ap + ds + cr & -br + as - dp - cq & -bs - ar - dq + cp \\ cp + dp + ar - bs & -cq + dp - as - br & -cr + ds + ap + bq & -cs - dr + aq - bp \\ dp - cq + br + as & -dq - cp - bs + ar & -dr - cs + bp - aq & -ds + cr + bq + ap \end{pmatrix} \quad (13)$$

where $a, b, c, d, p, q, r, s$ are reals and $a^2 + b^2 + c^2 + d^2 = 1$, $p^2 + q^2 + r^2 + s^2 = 1$.

Rotational matrix $R$ can be decomposed on matrices $\{R^L, R^R\}$ that describe left-isoclinic and right-isoclinic rotations:

$$R = R^L R^R \quad (14)$$

$$R^L = \begin{pmatrix} a & -b & -c & -d \\ b & a & -d & c \\ c & d & a & -b \\ d & -c & b & a \end{pmatrix}, \quad (15)$$

$$R^R = \begin{pmatrix} p & -q & -r & -s \\ q & p & s & -r \\ r & -s & p & q \\ s & r & -q & p \end{pmatrix}. \quad (16)$$

In general rotational matrices $\{R_U, R_V\}$ that figurate in the optimization task (12) are represented as compositions of left- and right-isoclinic rotations:

$$R_U = R_U^L R_U^R, \quad (17)$$

$$R_V = R_V^L R_V^R. \quad (18)$$

However, in some cases simpler model of rotational matrix is applicable.

In case $Ref$ is an orthonormal matrix there are two consequences: a) $Ref$ can be seen as some rotational matrix and can be decomposed on left- and right-isoclinic rotational matrices $Ref^L$and $Ref^R$:

$$Ref = Ref^L Ref^R; \quad (19)$$

b) term$R_U^T Ref R_V$ in (12) is also an orthonormal matrix and according to a) $R_U^T Ref R_V = R_U^T Ref^L Ref^R R_V$. In order to express any orthonormal matrix by term $R_U^T Ref^L Ref^R R_V$ it is enough that matrices $R_U$ and $R_V$ are left- and right-isoclinic respectively. Therefore utilization of orthonormal $Ref$ for watermarking could significantly simplify goal function $G^*$ and make embedding easier.

### D. Criterion of Watermarking Performance

In order to provide high watermarking performance it is necessary to minimize embedding distortions and to adjust robustness. It would be much easier to judge a tradeoff between robustness and transparency for each block separately. Threshold value $Th$ influences embedding distortions as well as robustness of a bit of a watermark for each particular block. Therefore several different values of $Th$ for each block could provide sufficient variety of transparency-robustness pairs. A decision about the best$Th$ for each block should be made according to some criterion.

Embedding distortions can be easily estimated according to, for example, RSS, but in order to estimate robustness we have to make some assumptions regarding distortion patterns. Those distortions are usually represented by signal processing or noise.

One possible way to check if an embedded bit is robust is to add each possible distortion pattern to a block and perform SVD to extract a bit. However, it would be computationally unreasonable. Therefore another kind of estimation of robustness is required.

According to the proposed watermarking rule a bit of a watermark cannot be influenced by a distortion pattern that doesnot change the first column of orthonormal matrix. Therefore let us consider a special case of distortion that occurs when $S'$ is being changed to $S''$, each column of $U'$ and $V'$except the first is being rotated:

$$U'' = (R'_{\boldsymbol{u}'} U'^T)^T, \quad (20)$$

$$V'' = (R'_{\boldsymbol{v}'} V'^T)^T. \quad (21)$$

Such rotations are represented by rotational matrices $R_{\boldsymbol{u}'}$ and $R_{\boldsymbol{v}'}$ respectively where each matrix can be described by Euler-Rodrigues formulae [16]:

$$R' = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & a^2 + b^2 - c^2 - d^2 & 2(bc - ad) & 2(bd + ac) \\ 0 & 2(bc + ad) & a^2 - b^2 + c^2 - d^2 & 2(cd - ab) \\ 0 & 2(bd - ac) & 2(cd + ab) & a^2 - b^2 - c^2 + d^2 \end{pmatrix} \quad (22)$$

A special distortion pattern $Dis_k^*$ of watermarked image fragment $I'_k$can be expressed in that case:

$$Dis_k^* = I'_k - I''_k = U'\left(S' - R'^T_{\boldsymbol{u}'} S'' R'_{\boldsymbol{v}'}\right) V'^T. \quad (23)$$

If we further define term $\left(S' - R'^T_{\boldsymbol{u}'} S'' R'_{\boldsymbol{v}'}\right)$ as $SR^*$ it can be seen that:

$$SR^* = \begin{pmatrix} SR_{1,1} & 0 & 0 & 0 \\ 0 & SR_{2,2} & SR_{2,3} & SR_{2,4} \\ 0 & SR_{3,2} & SR_{3,3} & SR_{3,4} \\ 0 & SR_{4,2} & SR_{4,3} & SR_{4,4} \end{pmatrix}. \quad (24)$$

For general case distortion pattern for $k$-th fragment is denoted as $Dis_k$ and general $SR$ is defined:

$$SR = U'^T Dis_k V'. \quad (25)$$

The measure $\|\boldsymbol{dv}\|_2^2$ where

$$\boldsymbol{dv} = (SR_{1,2}, SR_{1,3}, SR_{1,4}, SR_{2,1}, SR_{3,1}, SR_{4,1}). \quad (26)$$

can be used as an indicator of changes in $\boldsymbol{u}'$ and $\boldsymbol{v}'$ for a single distortion pattern $Dis_k$, because no changes in $\boldsymbol{u}'$ and $\boldsymbol{v}'$ imply that $\|\boldsymbol{dv}\|_2^2 = 0$. If all the distortion patterns $\{Dis_k\}$ are taken into account then appropriate indicator of possible changes in $\boldsymbol{u}'$ and $\boldsymbol{v}'$ is $Var(\|\boldsymbol{dv}\|_2^2)$.

We further assume that random distortion pattern $Dis$ can be approximated as $Dis = \sum_{i=1}^{4} r_i \boldsymbol{dis}_i$, where $\{\boldsymbol{dis}_i\}$ is a set of four independent components, each represented as 4x4 matrix, and $\{r_i\}$ is a set of four independent normally distributed zero-mean random variables. This assumption is due to the nature of random distortion pattern for distortions caused by some popular image processing (for example JPEG).Usually such distortion patterns can be described by several high-frequency components.

The indicator of robustness for a particular pair $\{U', V'\}$ can now be defined as:

$$Var(\|\boldsymbol{dv}\|_2^2) = \sum_{i=1}^{4} \|\boldsymbol{dv}_i\|_2^4 Var(r_i^2) + \\ +4\sum_{\substack{i=2 \\ j<i}}^{4} \left[ \left(\boldsymbol{dv}_i \boldsymbol{dv}_j^T\right)^2 Var(r_i) Var(r_j) \right], \quad (27)$$

where $\boldsymbol{dv}_i = (SR_{1,2}^i, SR_{1,3}^i, SR_{1,4}^i, SR_{2,1}^i, SR_{3,1}^i, SR_{4,1}^i)$ and $SR^i = U'^T \boldsymbol{dis}_i V'$. The main advantages of the proposed indicator of robustness are that it takes into account multivariate distribution of distortion patterns and can be easily computed for any pair $\{U', V'\}$.

It is necessaryto estimate the watermarking performance in order to choose an appropriate threshold value $Th$ for a particular block. To estimate the performance we united indicators of embedding distortions and robustness in a single criterion $C$ that is determined as:

$$C = \alpha \frac{G}{\sum_{i,j=1}^{4} I_k^2(i,j)} + \beta \sqrt{\frac{Var(\|\boldsymbol{dv}\|_2^2)}{(Th * S_{1,1}')^4}}, \quad (28)$$

where $\alpha$ and $\beta$ are some positive constants defined empirically. Lower value of $C$ corresponds to better watermarking performance. Depending on requirements to the tradeoff between invisibility and robustness different values of $\alpha$ and $\beta$ can be used.

Therefore in order to provide lower value of $C$ for a particular image fragment $I'_k$ goal function $G$ should be minimized several times, each time with different value of $Th$. The value of $Th$ that provides the lowest $C$ is the best for a particular block.

### E. The Steps of Watermarking

The method of watermark embedding can be described as following:

1) Define a set of $n$ different threshold values $\{Th_j\}$, $j = 1 \dots n$, that can be used in each block to embed a bit of a watermark;
2) Split the whole image $I$ on fragments of size $4 \times 4$;
3) Select image fragments for watermark embedding according to some secret key;
4) For a particular selected image fragment $I_k$ provide that watermarked fragment $I'_{k,j}$ satisfies embedding condition (12) and $G_j$ is minimized for each $Th_j$, calculate $C_j$;
5) Replace each $I_k$ by $I'_{k,j}$ that has the lowest $C_j$.

Watermark extraction can be specified by the steps:

1) Split the whole watermarked image $I'$ on fragments of size $4 \times 4$;
2) Select image fragments for watermark extraction according to the key;
3) Apply SVD to each selected fragment $I'_k$ and obtain $\{U', V'\}$;
4) Substitute $\{U', V'\}$ in equation (7) and calculate a bit.

## IV. Experimental Results

The performance of the proposed watermarking method was compared with two different blind SVD-based methods proposed in[13] and[15]. Several tests were conducted in order to emphasize differences between the methods. First an influence of different orthonormal reference matrices $Ref$ on the level of embedding distortions was explored without adaptation. Then some results of watermarking using the proposed method with adaptation were compared with the results of the other methods. Finally, results of watermarking with increased data payload were analyzed.

### A. Different Reference Matrices

Reference matrix $Ref$ is an important component for adjusting the proposed embedding rule. In order to select a matrix that provides better watermarking performance we have compared embedding distortions for different orthonormal reference matrices under condition with no adaptation e.g., equal threshold $Th$ has been applied to all the blocks.

We presumed that all the considered orthonormal matrices-candidates provide equal robustness. Hence the level of embedding distortions is the only important characteristic that could be different for different matrices. Variance of the term $\boldsymbol{u}' Ref \boldsymbol{v}'^T$ influences embedding distortions. Lower embedding distortions correspond to a reference matrix that provides lower $Var\left(\boldsymbol{u}' Ref \boldsymbol{v}'^T\right)$.

Five orthogonal normalized matrices were proposed as candidates for $Ref$.

The first matrix has just one non-zero element in each column (row):

$$Ref\_1 = \begin{pmatrix} 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}. \quad (29)$$

The second matrix has two non-zero elements with equal absolute values in each column:

$$Ref\_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 & 0 & 0 \\ -1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & -1 \end{pmatrix}. \quad (30)$$

The third matrix has three non-zero elements with equal absolute values in each column:

$$Ref\_3 = \frac{1}{\sqrt{3}} \begin{pmatrix} 0 & 1 & 1 & 1 \\ -1 & 1 & -1 & 0 \\ -1 & 0 & 1 & -1 \\ -1 & -1 & 0 & 1 \end{pmatrix}. \quad (31)$$

All the elements of the fourth matrix have equal absolute values:

$$Ref\_4 = 0.5 \begin{pmatrix} 1 & -1 & -1 & -1 \\ -1 & 1 & -1 & -1 \\ 1 & 1 & 1 & -1 \\ 1 & 1 & -1 & 1 \end{pmatrix}. \quad (32)$$

The fifth matrix has different number of non-zero elements in different columns (rows):

$$Ref\_5 = \frac{1}{3} \begin{pmatrix} 2 & -1 & 0 & -2 \\ 0 & 0 & -3 & 0 \\ -1 & 2 & 0 & -2 \\ 2 & 2 & 0 & 1 \end{pmatrix}. \quad (33)$$

All the five orthonormal matrices were used to collect five sets of indices. Each $i$-th set contains 262144 index values $\boldsymbol{u}'Ref\_i\boldsymbol{v}'^T$ calculated for each 4x4 block from 16 different grayscale images with resolution 512x512. The parameters of the distributions of index values for each set are given in Table I.

TABLE I. PARAMETERS OF INDEX DISTRIBUTION FOR DIFFERENT REFERENCE MATRICES

| | $Ref\_1$ | $Ref\_2$ | $Ref\_3$ | $Ref\_4$ | $Ref\_5$ |
|---|---|---|---|---|---|
| **Mean** | -0.00030 | 0.00005 | 0.00086 | -0.00047 | 0.00036 |
| **Variance** | 0.0818 | 0.0523 | 0.0913 | 0.1031 | 0.0981 |

From the table it can be seen that the second reference matrix $Ref\_2$ provides that the variance of the set $\{\boldsymbol{u}'Ref\_2\boldsymbol{v}'^T\}$ is the smallest. Therefore the second reference matrix should be used in order to provide better watermarking performance.

### B. Adjustment of the Proposed Criterion

A value of criterion $C$ according to (28) depends on estimate of $Var(\|\boldsymbol{dv}\|_2^2)$ which in turn depends on a set of distortion patterns $\{Dis_k\}$. Each pattern in the set is approximated as $Dis_k = \sum_{i=1}^{4} r_{i,k}\boldsymbol{dis}_i$ .However,for each pattern its complete (exact) representation should be obtained first: $Dis'_k = \sum_{i=1}^{16} r_{i,k}\boldsymbol{dis}_i$. Hence there are two important stages: collect distortion patterns $\{Dis'_k\}$ ;define the most important components $\{\boldsymbol{dis}_i\}, i = 1 \dots 4$.

In our tests all the 16 test images were split on blocks 4x4 which produced set$\{I_k\}$. Compression according to JPEG with quality factor 50 and 3x3 median filtering have been applied in turn to each of 16 test images. Therefore 32 distorted images were obtained. Each distorted image was again split on blocks 4x4 which produced set $\{I''_{k,g}\}$ , where $g = 1$ corresponds to JPEG compression and $g = 2$corresponds to median filtering. For each distorted block $I''_{k,g}$ distortion pattern $Dis'_{k,g}$ has been computed:

$$Dis'_{k,g} = I''_{k,g} - I_k. \quad (34)$$

Four the most important components $\{\boldsymbol{dis}_i\}$, $i = 1 \dots 4$ were defined using Principal Component Analysis (PCA) from the collection of $\{Dis'_{k,g}\}$, where $k = 1 \dots 524288$. For that purpose each distortion pattern $Dis'_{k,g}$ has been represented as a point in 16-dimensional space. First four eigenvectors (with the highest eigenvalues) returned by PCA have been obtained in a form of 1x16 vectors. Each vector has been rearranged to corresponding 4x4 (matrix) component and a set $\{\boldsymbol{dis}_i\}$has been formed.

The set $\{Th_j\}$ for the adaptation was $\{0.002, 0.003, 0.004, 0.005, 0.006\}$, which means adaptation procedure required 5 iterations for each block.

### C. Watermarking Results

The methods proposed in[13] and[15] provide quite different robustness-transparency tradeoffs and different data payloads. In order to make comparison fair the same watermark bit sequence consisted of 64 bits was used for all the methods. Each bit was embedded by each method redundantly (8 times) in randomly chosen blocks. Positions of chosen blocks were the same for the proposed method and the method of Tehrani. Four grayscale images of size 512x512 were selected for comparison of the methods. The chosen images were Lena, Baboon, Cameraman and Livingroom. All the attacks mentioned in the experiment were simulated by StirMark Benchmark 4. For all the methods Bit Error Rates (BERs) of a watermark extracted from each image were calculated. For Gaussian Noise (GN) and Salt&Pepper attacks the rates were averaged over 100 runs. The results are represented in Table II.

The payload in the particular test was just 64 bits which is very low. The proposed method and the method of Tehrani[15] use blocks of the same size 4x4 to embed a bit of a watermark. Therefore maximum payload of the both methods for 512x512 image is 16384 bit per image. However, for Li's method [13] the size of a macro-block is 32x32 and 2 bits are being embedded in each, which limits maximum payload by just 512 bit per image.

In order to compare the methods under a condition with the highest common payload the embedding redundancy for each method was adjusted in a different way. Then the methods were tested using the same 4 grayscale images. The

method of Li has been used for embedding of 512 bit long sequence without redundancy. The method of Tehrani has been used for embedding of the same sequence with redundancy 6. The proposed method has been used for embedding with redundancy 8. The results are represented in Table III.

TABLE II. RESULTS OF 64-BIT WATERMARK EXTRACTION

| Image, Method, PSNR | GN, PSNR=35 dB | Salt&Pepper, 3% | JPEG, Q=50 | 3x3 Median Filter | Cropping, 75% | Rotation, 0.25° |
|---|---|---|---|---|---|---|
| Lena, Tehrani, 52.76dB | 5.58 | 6.35 | 4.68 | 9.38 | 10.35 | 9.38 |
| Lena, Li, 42.58dB | **0** | **1.83** | **0** | **0** | 13.20 | **4.68** |
| Lena, proposed, 53.07dB | 5.42 | 6.07 | 3.13 | 4.68 | **9.81** | 7.81 |
| Baboon, Tehrani, 50.32dB | 3.34 | 4.92 | 3.13 | 7.81 | **9.39** | 7.81 |
| Baboon, Li, 41.95dB | **0** | **1.59** | **0** | **0** | 14.46 | **4.68** |
| Baboon, proposed, 51.50dB | 3.81 | 5.47 | 1.56 | 4.68 | 11.60 | 9.38 |
| Cameraman, Tehrani, 52.80 dB | 6.28 | 6.23 | 3.13 | 9.38 | 11.67 | 10.94 |
| Cameraman, Li, 41.75dB | **0** | **2.08** | **0** | **0** | 12.53 | **4.68** |
| Cameraman, proposed, 53.32dB | 6.14 | 6.62 | 1.56 | 6.25 | **11.03** | 7.81 |
| Livingroom, Tehrani, 50.59dB | 3.51 | 5.82 | 3.13 | 10.94 | 12.72 | 9.38 |
| Livingroom, Li, 42.26dB | **0** | **1.13** | **0** | **0** | 15.09 | **4.68** |
| Livingroom, proposed, 51.36dB | 4.12 | 7.36 | 3.13 | 6.25 | **11.83** | 10.94 |

TABLE III. RESULTS OF 512-BIT WATERMARK EXTRACTION

| Image, Method, PSNR | GN, PSNR=35 dB | Salt & Pepper, 3% | JPEG, Q=50 | 3x3 Median Filter | Cropping, 75% | Rotation, 0.25° |
|---|---|---|---|---|---|---|
| Lena, Tehrani, 43.82dB | 6.85 | 7.37 | 5.66 | 10.94 | 12.47 | 10.16 |
| Lena, Li, 42.39dB | **1.98** | **5.15** | **2.73** | **1.37** | 51.37 | 19.92 |
| Lena, proposed, 44.07dB | 5.56 | 6.07 | 3.32 | 4.88 | **9.75** | **8.00** |
| Baboon, Tehrani, 43.19dB | 4.35 | 5.96 | 4.69 | 11.91 | 12.25 | 9.38 |
| Baboon, Li, 41.87dB | **2.13** | **4.84** | 2.15 | **1.56** | 45.03 | 20.31 |
| Baboon, proposed, 43.42dB | 3.81 | 5.47 | **1.76** | 4.69 | **11.60** | **9.18** |
| Cameraman, Tehrani, 43.91dB | 7.21 | 6.85 | 4.30 | 10.35 | 12.03 | 11.33 |
| Cameraman, Li, 41.85dB | **2.25** | 6.78 | 2.54 | **1.95** | 48.52 | 21.09 |
| Cameraman, proposed, 44.12dB | 6.39 | **6.52** | **1.56** | 6.64 | **10.92** | **8.20** |
| Livingroom, Tehrani, 42.95dB | 4.31 | 6.92 | 5.47 | 11.72 | 13.82 | **10.35** |
| Livingroom, Li, 42.39dB | **2.32** | **6.56** | **1.95** | **2.34** | 50.26 | 21.88 |
| Livingroom, proposed, 43.86dB | 4.15 | 7.34 | 3.32 | 6.05 | **11.71** | 10.74 |

Even for increased data payload quality of images watermarked by the proposed method remains quite acceptable and is definitely the best among all the watermarked images in the test (Table III). It can be seen from Fig.1 that watermarked Lena image looks quite pure (PSNR=44.07dB).

Figure 1. Watermarked Lena image with PSNR=44.07dB.

## V. DISCUSSION

The proposed method is blind and only blind methods[13] and [15] were selected to compare the performance. The reason why other well-known SVD-based non-blind or semi-blind methods were rejected from the comparison is that they require additional information to be transferred.

From the watermarking results with low data payload (64 bit) it can be seen that there is no single method which performs better compared to others for all the kinds of common distortions. However, for non-geometrical attacks the method proposed by Li demonstrates extremely high robustness. For cropping attack the proposed and Tehrani's methods perform better because smaller blocks can be better spread in an image and smaller blocks are less likely to be cropped either. The quality of the images watermarked by the proposed and Tehrani's methods is much higher compared to Li's method. The proposed method provides slightly better quality of the watermarked images compared to Tehrani's method and its robustness toward JPEG and median filtering is better.

From the watermarking results with increased data payload (512 bit) it can be seen as previously that there is no absolute favorite. Each method embeds a watermark with different redundancy and the proposed method dominates in more positions while still providing the best quality. The method proposed by Li fully dominates in Gaussian noise and median filtering attacks even without redundant embedding. However, its performance in geometric attacks (cropping and rotation) is much worse. Another concern is that quality of images watermarked by Li's method is the worst. Because of the embedding with different redundancy the quality of images

watermarked by the proposed and Tehrani's methods is comparable. Nevertheless the robustness of the proposed method is better than that of Tehrani's method except two kinds of distortion for Livingroom image. The advantage of the proposed method over Tehrani's method is especially high for JPEG and median filtering attacks and for some images BER is around 6% lower.

The proposed method and the method of Tehrani use the same SVD transform to embed a bit of a watermark in 4x4 block. Considerable advantage of the proposed method compared to the method of Tehrani in case of JPEG and median filtering attacks is mostly due to minimization of embedding distortions and proper adjustment of $Th$ for each block.

Robustness of Li's method toward most kinds of attacks is very high even without redundant embedding. It is quite obvious that the ability to withstand noise and filtering attacks is better in case a bit of a watermark is embedded in larger block. The method proposed by Li uses 32x32 macro-block to embed 2 bit. In contrast to that the proposed method uses 4x4 blocks to embed a bit.

Popular image processing techniques usually process areas that are far larger than 4x4. A good example is JPEG-compression that process blocks 8x8. Therefore the result of JPEG-attack for a particular block 4x4 depends not only on that block, but also on some neighboring pixels, which makes a prediction of changes quite difficult based only on 4x4 block. Similar observation can be made regarding median filtering that uses adjacent pixels as well.

Nevertheless robustness-transparency tradeoff can be sufficient in case a bit is embedded in 4x4 block. In some instances compromise is required between robustness and data payload or between robustness and image quality. Quality of images watermarked by Li's method is usually around 42 dB which could be not enough for some demanding applications. Maximum payload provided by Li's method is 512 bit per 512x512 image which is only a half of required payload to embed 32x32 logo. Therefore redundant embedding is impossible for that method even for quite moderate payload which implies lower robustness toward some geometric attacks.

## VI. CONCLUSIONS

New blind watermarking method based on SVD is proposed in this paper. The method embeds a bit of a watermark by modifying the first columns of the both orthonormal matrices of a transformed 4x4 image block. Multiple improvements implemented in respect to existing methods are: the both orthonormal matrices are used, model of rotations in 4D space is applied to modify orthonormal matrices, embedding distortions are minimized, the criterion of watermarking performance is proposed for adaptation.

Utilization of the both orthonormal matrices maintains better watermarking performance. Modification of the both matrices introduces lower embedding distortions compared to an approach that modifies just one. On the other hand such embedding is less affected by common image processing techniques.

Application of rotational model assures that the result of modification of orthonormal matrices is a matrix which is also orthonormal. The application guaranties that the result of the decomposition matches matrices used to compose a fragment. This is a considerable advantage over existing approaches.Rotational matrices are being adjusted in order to minimize goal function.Constraints necessary to embed a bit of a watermarkare taken into account during the minimization procedure. Minimization of embedding distortions improves transparency of watermarked images without affecting robustness.

The proposed criterion of watermarking performance takes into account embedding distortions as well as robustness of a bit of a watermark for each particular block. Considered adaptation procedure on the basis of the proposed criterion choosesappropriate threshold value for each block. This reduces embedding distortions while keeps substantial robustness. Lower level of distortions enables embedding with higher redundancy which considerably increases total robustness of a watermark.

As the result of the proposed improvements a watermark extracted (for example from Lena image) after median filtering attack has up to 6% lower BER compared to the method proposed in [15].On the other hand quality of watermarked images is higher. For some geometric attacks BER has been reduced more than 40% compared to the method proposed in [13].

## ACKNOWLEDGMENT

The first author is thankful to Tampere Program in Information Science and Engineering for the support.

## REFERENCES

[1] I. Cox, M. Miller, J. Bloom, J. Fridrich and T. Kalker, Digital Watermarking and Steganography (2 ed.), San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2007.

[2] C. Song, S. Sudirman, M. Merabti and D. Llewellyn-Jones, "Analysis of Digital Image Watermark Attacks," in *Proceedings of Consumer Communications and Networking Conference (CCNC)*, 9-12 Jan. 2010.

[3] H. R. Sheikh and A. Bovik, "Image information and visual quality," *IEEE Transactions on Image Processing,* vol. 15, no. 2, pp. 430-444, Feb. 2006.

[4] S. Lin and C.-F. Chen, "A robust DCT-based watermarking for copyright protection," *IEEE Transactions on Consumer Electronics,* vol. 46, no. 3, pp. 415-421, Aug 2000.

[5] E. Fullea and J. Martinez, "Robust digital image watermarking using DWT, DFT and quality based average," in *Proceedings of the ninth ACM international conference on Multimedia (MULTIMEDIA '01)*, 2001.

[6] L. Trefethen and D. Bau, Numerical Linear Algebra, Cambridge University Press, 1997.

[7] V. Gorodetski, L. Popyack, V. Samoilov and V. Skormin, "SVD-based Approach to Transparent Embedding Data into Digital Images," in *Proceedings of the International Workshop on Information Assurance in Computer Networks: Methods, Models, and Architectures for Network Security (MMM-ACNS '01)*, 2001.

[8] R. Liu and T. Tan, "An SVD-based watermarking scheme for protecting rightful ownership," *IEEE Transactions on Multimedia,* vol. 4, no. 1, pp. 121-128, Mar 2002.

[9] X.-P. Zhang and K. Li, "Comments on "An SVD-based watermarking scheme for protecting rightful Ownership"," *IEEE Transactions on Multimedia,* vol. 7, no. 3, pp. 593-594, June 2005.

[10] E. Ganic and A. Eskicioglu, "Robust DWT-SVD domain image watermarking: embedding data in all frequencies," in *Proceedings of the 2004 workshop on Multimedia and security (MM&Sec '04)*, 2004.

[11] P. Bao and X. Ma, "Image adaptive watermarking using wavelet domain singular value decomposition," *IEEE Transactions on Circuits and Systems for Video Technology,* vol. 15, no. 1, pp. 96-102, Jan. 2005.

[12] H.-C. Wu, R.-J. Jang and Y.-C. Liu, "A robust watermarking scheme based on singular value decomposition and quantization technique," in *Proceedings of International Computer Symposium (ICS)*, Dec. 2010.

[13] Z. Li, K.-H. Yap and B.-Y. Lei, "A new blind robust image watermarking scheme in SVD-DCT composite domain," in *Proceedings of 18th IEEE International Conference on Image Processing (ICIP),* Sept. 2011.

[14] C.-C. Chang, P. Tsai and C.-C. Lin, "SVD-based digital image watermarking scheme," *Pattern Recognition Letters,* no. 26, p. 1577–1586, 2005.

[15] I. O. Tehrani and S. Ibrahim, "An enhanced SVD based watermarking using U matrix," in *Proceedings of 5th International Conference on Computer Sciences and Convergence Information Technology (ICCIT)*, 2010.

[16] H. P. Manning, Non-Euclidean Geometry, HardPress, 2012.

# An SVD-based Transparent Watermarking Method

Yevhen Zolotavkin
Computer Science, School of Information Sciences
University of Tampere
Tampere, Finland
yevhen.zolotavkin@uta.fi

Martti Juhola
Computer Science, School of Information Sciences
University of Tampere
Tampere, Finland
martti.juhola@sis.uta.fi

***Abstract*—A new watermarking method on the basis of Singular Value Decomposition is proposed in this paper. The method is blind and modifies one of the orthonormal matrices of the decomposition of 4x4 block to enclose a bit of a watermark. A procedure for minimization of embedding distortions is considered. Two embedding rules have been proposed for watermarking which provide different robustness and transparency. Popular attacks have been applied and experimental results have been illustrated. According to the results the robustness of the proposed watermarking method toward some attacks is more than 36% better compared to other known blind watermarking methods.**

***Keywords—Watermarking, Singular Value Decomposition, Robustness, Transparency***

## I. INTRODUCTION

Protection of digital rights is one of the most important tasks of cyber security. These kinds of problems are addressed by Digital Image Watermarking (DIW), where a digital image is being protected by enclosing a digital watermark. Three important characteristics for a particular watermarking method are robustness, transparency and data payload. The tradeoff between robustness and transparency depends on a kind of transform which provides coefficients for modification as well as embedding rule that interprets values of coefficients.

In most cases robustness against noise, some kinds of filtering and geometric attacks is required [1]. The most approved index of robustness is Bit Error Rate (BER) of a watermark upon extraction. Transparency is an ability to preserve original image by watermarking it and many measures could be used to define it quantitatively [2]. Though, the most popular measure is Peak Signal to Noise Ratio (PSNR). Data payload is a number of watermark bits embedded into an image.

The original image can be modified in many different ways to embed a watermark. Modification of the Least Significant Bit is a simplest example of watermarking in spatial domain [3]. Some suitable examples of frequency domain watermarking are the methods on the basis of Discrete Cosine Transform (DCT) [4] and Discrete Wavelet Transform (DWT) [5]. Usually modification of some spectral coefficients is more favorable as it is more robust against noise and image processing attacks. Singular Value Decomposition (SVD) is a unique kind of transform [6]. It separates an image fragment on several independent layers which number is much less than that, for example, for DCT. Therefore the most important layer is quite stable to various attacks.

An efficiency of watermarking also depends on a rule exploited for embedding. Each embedding rule could have several parameters that influence robustness-transparency tradeoff.

In this paper we propose new SVD-based blind watermarking method which minimizes embedding distortions. It uses an orthonormal matrix obtained by SVD of 4x4 block to embed a bit of a watermark. The proposed method provides good robustness-transparency tradeoff and high data payload.

The rest of the paper is organized as following: a short review of relevant watermarking methods exploiting SVD is given in the Section II; Section III bears our own approach which is described in detail; then, some experimental results are represented in Section IV followed by a discussion of their importance in Section V; finally, Section VI concludes the paper.

## II. SVD-BASED WATERMARKING

### A. SVD Transform

An image fragment $I_k$ of size $n \times n$ is being decomposed according to SVD [6] in the following way:

$$I_k = USV^T = \begin{pmatrix} U_{1,1} \cdots U_{1,n} \\ U_{2,1} \cdots U_{2,n} \\ \vdots \\ U_{n,1} \cdots U_{n,n} \end{pmatrix} \times \times \begin{pmatrix} S_{1,1} & 0 & \cdots & 0 \\ 0 & S_{2,2} & \cdots & 0 \\ & & \vdots & \\ 0 & 0 & \cdots & S_{n,n} \end{pmatrix} \times \begin{pmatrix} V_{1,1} \cdots V_{1,n} \\ V_{2,1} \cdots V_{2,n} \\ \vdots \\ V_{n,1} \cdots V_{n,n} \end{pmatrix}^T, \quad (1)$$

where $U$ and $V$ are some orthonormal matrices and $S$ is a diagonal matrix of singular values.

An alternative representation demonstrates that fragment $I_k$ is being decomposed on $n$ independent layers where geometry of $i$-th layer is defined by a pair of $i$-th columns (one from matrix $U$ and one from $V$) and a luminance component $S_{i,i}$:

$$I_k = \sum_i S_{i,i} U(1 \dots n, i) V(1 \dots n, i)^T \quad (2)$$

The luminance $S_{1,1}$ has the biggest value and for most fragments of natural images this value is much bigger than the

This paper was supported by Tampere Program in Information Science and Engineering (TISE).

other. Therefore the first layer is the most substantial and provides the best robustness for watermarking.

### B. Review of SVD-based Watermarking Methods

Popular strategy for SVD-based methods that modify singular values is to quantize the biggest value of a block depending on the corresponding bit of a watermark. The first SVD-based blind method for DIW was proposed in [7]. It features high data payload and uses color RGB images. Another pioneering noninvertible non-blind scheme exploiting SVD was introduced in [8]. However, later in [9] it has been shown that the scheme is vulnerable for the counterfeit attack and is invertible. In the paper [10] the first in the literature DWT-SVD non-blind watermarking method was proposed which demonstrates good robustness and high data payload. The downside is that embedding distortions are quite considerable. The semi-fragile method proposed in [11] applies adaptive quantization to DWT-SVD. The disadvantage is that additional information should be transmitted and the method causes considerable degradation of original image. Recently a quite robust method with adaptive quantization of SV has appeared in [12], but additional information is required for extraction. One of the most robust blind watermarking schemes based on SVD-DCT transform was proposed in [13].

Modification of orthogonal matrices is a very rear approach of SVD watermarking. The advantage of such kind of watermarking is that more elements are available for modification. The paper [14] proposes a blind watermarking method that modifies the second and the third elements in the first column of the left orthogonal matrix of SVD of a 4x4 block. The quality of watermarked images is quite high, but robustness toward common distortions like JPEG-compression, Gaussian noise and cropping is not sufficient. Another paper developing further previous approach is [15] where authors proposed to adjust a threshold in a way that PSNR of each modified blocks is higher than 42 dB whenever it is possible. As a result robustness-invisibility tradeoff is better than for the method proposed in [14].

There are several shortcomings in the latter two methods. First, an orthonormal matrix where two elements are modified becomes non-orthonormal which could cause an embedded bit to be lost even without influence of the third person or noise. Second, only two elements of the left orthogonal matrix are used while utilization of all the four elements of the first column could improve robustness-transparency trade-off.

## III. Proposed Watermarking Method

The method described below resolves mentioned shortcomings of the methods proposed in [14] and [15] and minimizes embedding distortions. Further we consider orthonormal matrices of SVD of particular image block $I_k$ which size is $4 \times 4$. A watermark embedding rule should be used to embed and extract a bit. The watermarking method proposed in this paper modifies one of the orthonormal matrices which is either $U$ or $V$ to embed a bit of a watermark.

### A. Watermark Embedding Task

Let us assume that matrix $U$ is being modified: $U \to U'$. Singular values can be modified as well: $S \to S'$. Different embedding rules can be used, but in order to provide high robustness we suggest that the first column $\boldsymbol{u}'^{T}_1$ of the modified orthonormal matrix $U'$ interprets a bit, $\boldsymbol{u}'_1 = \left[U'_{1,1}, U'_{2,1}, U'_{3,1}, U'_{4,1}\right]$. Optionally the first column $\boldsymbol{v}^{T}_1$, $\boldsymbol{v}_1 = \left[V_{1,1}, V_{2,1}, V_{3,1}, V_{4,1}\right]$, of the other matrix $V$ can also be used by embedding rule to interpret a bit, however, only one orthonormal matrix is being modified. Each new watermarked image fragment $I'_k$ that carries corresponding bit is composed from two orthonormal matrices $\{U', V\}$ and a diagonal matrix of singular values $S'$:

$$I'_k = U'S'(V)^T. \tag{3}$$

Matrix $U'$ should be orthonormal in order to provide proper decoding of a bit of a watermark. Otherwise the result of SVD of $I'_k$ might be different and a left orthonormal matrix $U''$ might be obtained such that $U' \neq U''$. In that case a bit of a watermark might be inverted.

We propose to modify matrix $U$ according to the model of rotations in 4D space, which assures that the resulting matrix $U'$ is orthonormal:

$$U' = R_U U \tag{4}$$

Rotational matrix $R_U$ in four dimensional space can be fully described according to Van Elfrinkhof's formulae [16]:

$$R_U = R^L_U R^R_U, \tag{5}$$

$$R^L_U = \begin{pmatrix} a & -b & -c & -d \\ b & a & -d & c \\ c & d & a & -b \\ d & -c & b & a \end{pmatrix}, \tag{6}$$

$$R^R_U = \begin{pmatrix} p & -q & -r & -s \\ q & p & s & -r \\ r & -s & p & q \\ s & r & -q & p \end{pmatrix}, \tag{7}$$

where $a, b, c, d, p, q, r, s$ are reals and $a^2 + b^2 + c^2 + d^2 = 1$, $p^2 + q^2 + r^2 + s^2 = 1$. Rotational matrices $R^L_U$ and $R^R_U$ are left-isoclinic and right-isoclinic respectively [16]. According to the mentioned rotational model any orthonormal matrix $U$ can be transformed to any other orthonormal matrix $U'$. Thus, using (4) and (5) we can express modified block $I'_k$ in general case as following:

$$I'_k = R^L_U R^R_U U S'(V)^T. \tag{8}$$

In order to provide good transparency a watermark should be embedded with minimal distortions. The embedding distortion for the particular block is:

$$G = \|I'_k - I_k\|_2^2 = \|R^L_U R^R_U U S'(V)^T - I_k\|_2^2. \tag{9}$$

Parameters $R^L_U$, $R^R_U$ and $S'$ should be adjusted to minimize goal function $G$ taking into account constraints that are given by an embedding rule. This is obviously non-linear Least Squares (LS) task which solving would require considerable computational costs and a global minimum is not guaranteed.

### B. Simplified Solution for Embedding Task

With the aim to simplify embedding and make it more sufficient in computational sense we propose to alter the first column of orthogonal matrix $\boldsymbol{u}^T_1 \to \boldsymbol{u}'^T_1$ independently from the other columns. The rest columns of orthogonal matrix $U'$ should be adjusted after.

The first layer $S'_{1,1}\boldsymbol{u}'^T_1\boldsymbol{v}_1$ is the most essential as it interprets watermark's bit and the first singular value is the largest which corresponds to the highest visual importance. Hence, the task of the first phase of embedding is to satisfy a condition given by a rule and to minimize the term $\left\|I_k - S'_{1,1}\boldsymbol{u}'^T_1\boldsymbol{v}_1\right\|_2^2$. In order to make it simple we propose: a) to rotate in 4D space vector $\boldsymbol{u}_1^T$ on a minimal angle to satisfy an embedding rule, that is $\boldsymbol{u}_1^T \to \boldsymbol{u}'^T_1$; b) to adjust $S_{1,1} \to S'_{1,1}$ and minimize $\left\|I_k - S'_{1,1}\boldsymbol{u}'^T_1\boldsymbol{v}_1\right\|_2^2$. The action a) depends on a rule, but we will show further that in our case it is simple. The action b) is ordinary Least Squares (LS) task.

The rest columns of matrix $U'$ and the rest singular values in $S'$ should be defined to minimize embedding distortions $\|I_k - I'_k\|_2^2$ during the second phase, but $\boldsymbol{u}'^T_1$ and $S'_{1,1}$ have to remain unchanged. For that purpose we have proposed a special procedure that guarantees orthonormality of the resulting $U'$ and requires ordinary LS solver on each step.

In the proposed procedure each column of $U'$ starting from the second is being defined on a separate step. A special rotational matrix is used on each step that does not change columns defined on previous steps.

The steps of the proposed procedure are:

1. Calculate rotation matrix $R^L$ that transforms $\boldsymbol{u}_1^T$ to $\boldsymbol{u}'^T_1$;
2. Calculate new orthonormal matrix $U'^{(1)} = R^L U$;
3. Calculate rotation matrix $R''_{d=0}$ that does not change the first column of $U'^{(1)}$ but minimizes $\left\|I_k - \sum_{g=1}^{2} S'_{g,g}\boldsymbol{u}'^T_g\boldsymbol{v}_g\right\|_2^2$ by modifying the second column $\boldsymbol{u}'^T_2$, find $S'_{2,2}$;
4. Calculate new orthonormal matrix $U'^{(2)} = U'^{(1)}R''_{d=0}$;
5. Calculate rotation matrix $R'_{c=d=0}$ that does not change the first two columns of $U'^{(2)}$ but minimizes $\left\|I_k - \sum_{g=1}^{3} S'_{g,g}\boldsymbol{u}'^T_g\boldsymbol{v}_g\right\|_2^2$ by modifying the third column $\boldsymbol{u}'^T_3$, find $S'_{3,3}$;
6. Calculate final orthonormal matrix $U' = U'^{(2)}R'_{c=d=0}$, find $S'_{4,4}$ and compose $I'_k$.

As can be seen from the steps the final matrix $U'$ can be expressed in terms of original $U$ and calculated rotational matrices: $U' = R^L U R''_{d=0} R'_{c=d=0}$. More detailed description of each step is provided further.

Left- or right-isoclinic rotation model can be used on step 1. According to left-isoclinic rotation model:

$$R^L\boldsymbol{u}_1^T = \begin{pmatrix} a & -b & -c & -d \\ b & a & -d & c \\ c & d & a & -b \\ d & -c & b & a \end{pmatrix} \times \begin{pmatrix} U_{1,1} \\ U_{2,1} \\ U_{3,1} \\ U_{4,1} \end{pmatrix} = \boldsymbol{u}'^T_1. \quad (10)$$

Parameters $a, b, c, d$ of $R^L$ are calculated as following:

$$\begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} = \begin{pmatrix} U_{1,1} & -U_{2,1} & -U_{3,1} & -U_{4,1} \\ U_{2,1} & U_{1,1} & U_{4,1} & -U_{3,1} \\ U_{3,1} & -U_{4,1} & U_{1,1} & U_{2,1} \\ U_{4,1} & U_{3,1} & -U_{2,1} & U_{4,1} \end{pmatrix}^T \boldsymbol{u}'^T_1. \quad (11)$$

New matrix $U'^{(1)}$ can be calculated on step 2:

$$U'^{(1)} = R^L U. \quad (12)$$

A rotation matrix that does not change the first column of $U'^{(1)}$ should be calculated on the next step 3. A suitable model for that kind of changes is Euler-Rodrigues rotation matrix [16] with parameter $d = 0$:

$$R'_{d=0} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & a^2+b^2-c^2 & 2bc & 2ac \\ 0 & 2bc & a^2-b^2+c^2 & -2ab \\ 0 & -2ac & 2ab & a^2-b^2-c^2 \end{pmatrix}. \quad (13)$$

The matrix can also be represented in a different way using elements $x = a^2+b^2-c^2$, $y = 2bc$, $z = -2ac$, which provides $x^2+y^2+z^2 = 1$. The result is:

$$R''_{d=0} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & x & y & -z \\ 0 & y & \frac{z^2-y^2+(1-x)^2}{2(1-x)} & \frac{yz}{1-x} \\ 0 & z & \frac{-yz}{1-x} & \frac{z^2-y^2-(1-x)^2}{2(1-x)} \end{pmatrix}. \quad (14)$$

Matrix $U'^{(2)}$ can be defined as:

$$U'^{(2)} = U'^{(1)}R''_{d=0}. \quad (15)$$

Therefore the second column of $U'^{(2)}$ is defined as:

$$\boldsymbol{u}'^T_2 = U'^{(1)}\begin{pmatrix} 0 \\ x \\ y \\ z \end{pmatrix}. \quad (16)$$

In order to find parameters $x, y, z$ it is necessary to solve:

$$\left\|I_k - S'_{1,1}\boldsymbol{u}'^T_1\boldsymbol{v}_1 - S'_{2,2}\boldsymbol{u}'^T_2\boldsymbol{v}_2\right\|_2^2 \to min. \quad (17)$$

Here everything except the term $S'_{2,2}\boldsymbol{u}'^T_2$ is known.

After parameters $S'_{2,2}$, $x$, $y$, $z$ are found matrix $U'^{(2)}$ can be determined on step 4 according to (15).

A rotation matrix that does not change the first two columns of $U'^{(2)}$ should be calculated on the next step 5. The model for that kind of rotation is even more simplified Euler-Rodrigues matrix with parameters $c = d = 0$, $a^2 - b^2 = x$, $2ab = y$:

$$R'_{c=d=0} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & x & -y \\ 0 & 0 & y & x \end{pmatrix}. \quad (18)$$

Matrix $U'$ is defined as

$$U' = U'^{(2)}R'_{c=d=0}. \quad (19)$$

Therefore the third column of $U'$ is defined as:

$$\boldsymbol{u}'^{T}_{3} = U'^{(2)} \begin{pmatrix} 0 \\ 0 \\ x \\ y \end{pmatrix}. \quad (20)$$

The third column should satisfy:

$$\left\| I_k - S'_{1,1}\boldsymbol{u}'^{T}_{1}\boldsymbol{v}_1 - S'_{2,2}\boldsymbol{u}'^{T}_{2}\boldsymbol{v}_2 - S'_{3,3}\boldsymbol{u}'^{T}_{3}\boldsymbol{v}_3 \right\|^2_2 \rightarrow min, \quad (21)$$

where everything except the term $S'_{3,3}\boldsymbol{u}'^{T}_{3}$ is known.

After parameters $S'_{3,3}$, $x$, $y$ are found final matrix $U'$ can be defined on step 6 according to (19). The last remained undefined parameter $S'_{4,4}$ should minimize the following:

$$\left\| I_k - S'_{1,1}\boldsymbol{u}'^{T}_{1}\boldsymbol{v}_1 - S'_{2,2}\boldsymbol{u}'^{T}_{2}\boldsymbol{v}_2 - \right.$$
$$\left. -S'_{3,3}\boldsymbol{u}'^{T}_{3}\boldsymbol{v}_3 - S'_{4,4}\boldsymbol{u}'^{T}_{4}\boldsymbol{v}_4 \right\|^2_2 \rightarrow min \quad (22)$$

After $S'_{4,4}$ is found watermarked image fragment $I'_k$ can be composed. Every optimization step in the described procedure requires just ordinary LS solver which does not consume much computational sources.

### C. Steps of Watermarking

The proposed method of watermark embedding can be described as following:

1) Split the whole image $I$ on fragments of size $4 \times 4$;
2) Select image fragments for watermark embedding according to some secret key, assign to each fragment a bit of a watermark;
3) Perform SVD of a particular selected image fragment $I_k$. Check if a corresponding bit of a watermark is interpreted correctly according to a watermarking rule. In case interpretation is correct set $I'_k \leftarrow I_k$ and proceed to 6), else proceed to 4);
4) Depending on the rule modify the first column of either $U$ or $V$ to satisfy the rule, calculate $S'_{1,1}$;
5) Apply the described simplified procedure of embedding to the rest columns of the chosen matrix ($U$ or $V$) and the rest singular values in $S$. Compose watermarked image fragment $I'_k$;
6) Substitute original fragment $I_k$ by $I'_k$. If $k+1$ is less or equal to the length of a watermark set $k \leftarrow k+1$ and proceed to 3), else finalize embedding.

Watermark extraction can be specified by the steps:

1) Split the whole watermarked image $I'$ on fragments of size $4 \times 4$;
2) Select image fragments for watermark extraction according to the key;
3) Apply SVD to the selected fragment $I'_k$ and obtain $\{U', V'\}$;
4) Interpret a bit according to a watermarking rule using $\{U', V'\}$. If $k+1$ is less or equal to the length of a watermark set $k \leftarrow k+1$ and proceed to 3), else finalize extraction and output a watermark.

## IV. Experimental Results

### A. Embedding Rules

Orthogonal matrices should satisfy some requirements that can be expressed in a rule. This is necessary for proper embedding and extraction of a bit of a watermark. Main components of a rule are $4 \times 1$ reference vectors $\boldsymbol{ref}_1$ and $\boldsymbol{ref}_2$ and a scalar threshold $Th$. Two different embedding rules are proposed.

The first rule considers the both orthogonal matrices of SVD of a block $I_k$. Matrix $U$ or matrix $V$ should be modified to embed a bit and the choice depends on properties of a block. However, the both matrices are important for extraction as the person who extracts does not have additional information about embedding conditions. We further use notation $\{U', V'\}$ because during extraction the both matrices should be treated equally.

Let us use the following notations. For original block $I_k$ we define two indexes: $Y_{k,1} = \boldsymbol{u}_1\boldsymbol{ref}_1 - m'$, $Y_{k,2} = \boldsymbol{v}_1\boldsymbol{ref}_2 - m''$. For modified block $I'_k$ we define another two indexes: $Y'_{k,1} = \boldsymbol{u}'_1\boldsymbol{ref}_1 - m'$, $Y'_{k,2} = \boldsymbol{v}'_1\boldsymbol{ref}_2 - m''$. Here $m'$ and $m''$ are corresponding mean values. The signs $sign(Y'_{k,1})$ and $sign(Y'_{k,2})$ should be computed in order to extract a bit. Sign pairs $\{-,-\}$ and $\{+,+\}$ are interpreted as 0. Sign pairs $\{-,+\}$ and $\{+,-\}$ are interpreted as 1.

If an embedding rule is not satisfied for an original block $I_k$ and there is a need of modification of one of orthogonal matrices we propose to check inequality $|Y_{k,1}| \leq |Y_{k,2}|$. Matrix $U$ should be modified if the inequality is true. Otherwise matrix $V$ should be modified. Such guidance reduces embedding distortions as the level of total distortion for $I'_k$ depends mostly on distortion of the first column. A column that provides smaller absolute value of dot product with reference vector requires lower distortion to change the sign of the product.

In order to maintain good robustness-transparency tradeoff we consider two parameters in the rule that limit embedding distortions. The first parameter is empirically defined positive real threshold $Th$, which is necessary to regulate robustness-transparency tradeoff. The second parameter $Max$ depends on the properties of a block: $Max = max(|Y_{k,1}|, |Y_{k,2}|)$. The first embedding rule is:

$$Rule\#1: \begin{cases} Embedding: (-1)^{bit} Y'_{k,1}Y'_{k,2} = Max * min(Th, Max); \\ Extraction: bit = \left(2 + sign(Y'_{k,1}Y'_{k,2})\right) mod\ 3. \end{cases}$$

The second rule interprets the first column of $U$ only and only $Th$ limits embedding distortions:

$$Rule\#2: \begin{cases} Embedding: (-1)^{bit} Y'_{k,1} = Th; \\ Extraction: bit = \left(2 + sign(Y'_{k,1})\right) mod\ 3. \end{cases}$$

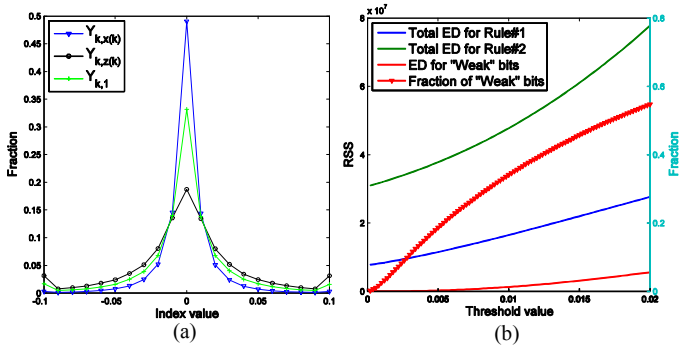
For the both proposed rules the task of adjustment of the first column of corresponding matrix is easy. Let us assume that the $Rule\#1$ is used and for a particular block $I_k$ the rule is not satisfied and $|Y_{k,1}| \leq |Y_{k,2}|$. Thus, it is necessary to find $\boldsymbol{u}'^{T}_{1}$ which belongs to the plane determined by $\boldsymbol{u}^{T}_{1}, \boldsymbol{ref}_1$ and the origin. This new column $\boldsymbol{u}'^{T}_{1}$ should satisfy $\boldsymbol{u}'_1\boldsymbol{ref}_1 = (-1)^{bit} min(Th, |Y_{k,2}|) + m'$.

### B. Characteristics of Embedding Rules

In order to analyze watermarking properties of the proposed embedding rules we first compared histograms of corresponding indices that are being interpreted by rules. Sixteen grayscale images with resolution 512x512 were used, which provided 262144 different blocks of size 4x4. The

following sets of indices were considered: $\{Y_{k,1}\}$, $\{Y_{k,x_k}\}$, $\{Y_{k,z_k}\}$, where $k = 1 \dots 262144$, $x_k = arg\min(|Y_{k,1}|, |Y_{k,2}|)$, $z_k = arg\max(|Y_{k,1}|, |Y_{k,2}|)$. The sets $\{Y_{k,x_k}\}$ and $\{Y_{k,1}\}$ are potentially being modified if $Rule\#1$ or $Rule\#2$ are used respectively. The variances of set $\{Y_{k,x_k}\}$ and set $\{Y_{k,1}\}$ are good estimates to compare embedding distortions for each rule. Higher variance of set $\{Y_{k,z_k}\}$ corresponds to higher variance of $Max$ and less equal ability to withstand disturbances by different blocks if $Rule\#1$ is used. The following reference vectors were used by the rules: $\boldsymbol{ref}_1^T = \boldsymbol{ref}_2^T = (0.5 \quad -0.5 \quad 0.5 \quad -0.5)$.

The histograms of all the mentioned above sets are shown on Fig.1a.




(a) (b)

Fig. 1. a) Histograms of indices $Y_{k,1}$, $Y_{k,x_k}$, $Y_{k,z_k}$; b) Graphs of embedding distortions for $Rule\#1$ and $Rule\#2$.

It can be seen that $Var\{Y_{k,z_k}\} > Var\{Y_{k,1}\} > Var\{Y_{k,x_k}\}$. From comparison $Var\{Y_{k,1}\} > Var\{Y_{k,x_k}\}$ one could derive that the $Rule\#1$ causes lower embedding distortions. Furthermore the higher $Var\{Y_{k,z_k}\}$ the more advantageous $Rule\#1$ compared to $Rule\#2$ in terms of transparency and less advantageous in terms of robustness.

We have used 16 test images to embed the same watermark using the both rules. The dependencies between different values of $Th$ and Residual Sum of Squares (RSS) for each rule are represented graphically on Fig.1b. Two main graphs to compare transparency are Embedding Distortions (ED) for $Rule\#1$ and $Rule\#2$. Another two graphs ED for "weak" bits and Fraction of "weak" bits provide details about robustness of a watermark embedded using $Rule\#1$. The definition of "weak" bit is used for a bit embedded using $Rule\#1$ in a block where $Max < Th$. The Fraction of such "weak" bits shows how big the disadvantage in robustness is compared to $Rule\#2$. Additionally the graph ED for "weak" bits demonstrates how ED is distributed between "weak" and "typical" bits embedded with $Max \geq Th$. For example, for $Th = 0.02$ cumulative ED for more than a half of all the blocks is less than 20% of total ED. This is quite unequal distribution of ED which worsens robustness in general when compared to $Rule\#2$. On the other hand $Rule\#2$ introduces much higher distortions during embedding.

### C. Watermarking Results

The proposed watermarking method was used with the both embedding rules. Sixteen different grayscale images of size 512x512 were used for watermarking. The watermarking performance was compared with the performance of other blind SVD-based methods proposed by Tehrani [15], Li [13] and Gorodetski [7] (adopted for grayscale images). Selected methods provide quite different robustness-transparency tradeoffs and different data payloads. With the aim to make comparison fair we used the same watermark for all the methods and all the images. We also tried to adjust methods in a way that quality of the watermarked images is comparable. For that purpose different methods performed embedding with different redundancy. The watermark payload was 512 bits per image as this is the maximum payload of Li method. The redundancy rates were: 12 for the proposed method and $Rule\#1$, 8 for the proposed method and $Rule\#2$, 6 for Tehrani's method, 1 for Li's method, 6 for Gorodetski's method. All the watermarked images undergo attacks simulated by StirMark Benchmark 4. For all the methods Bit Error Rates (BERs) of extracted watermarks were averaged among all the 16 watermarked images. The results are represented in Table I.

TABLE I. RESULTS OF WATERMARK EXTRACTION

| **Method, PSNR** | **GN, PSNR 35dB** | **Salt & Pepper, 3%** | **JPEG, Q=50** | **3x3 Median Filter** | **Cropping, 75%** | **Rotation, 0.25** |
|---|---|---|---|---|---|---|
| Tehrani, 43.47dB | 5.42 | 6.55 | 5.04 | 11.32 | 8.92 | 10.32 |
| Li, 42.13dB | **2.19** | 5.11 | **2.35** | **1.83** | 38.12 | 20.81 |
| Gorodetski, 42.85dB | 4.89 | **4.41** | 7.57 | 10.58 | 8.83 | 9.42 |
| Proposed, *Rule*1, 44.13dB | 3.74 | 4.61 | 3.16 | 6.24 | **1.62** | **8.30** |
| Proposed, *Rule*2, 43.87dB | 5.35 | 6.73 | 3.79 | 8.61 | 5.15 | 9.53 |

In our experiment we set $Th = 0.002$ for both proposed rules. The quality of images watermarked according to the proposed method and $Rule\#1$ remains quite acceptable. This can be witnessed by Fig.2 where the watermarked Livingroom image is depicted.

Fig. 2. Watermarked Livingroom image with PSNR=44.29dB.

## V. DISCUSSION

The proposed watermarking method provides good robustness-transparency tradeoff in conjunction with both proposed embedding rules. Considerable advantages are

achieved in comparison with the methods proposed by Gorodetski and Tehrani. Another point is that our method maintains much higher robustness toward geometrical attacks in comparison with the method proposed by Li.

High robustness for the proposed method and the both proposed rules in comparison with the methods proposed by Gorodetski and Tehrani is mainly due to higher redundancy of embedding. Higher redundancy is possible thanks to low distortions of embedding of a bit in a single 4x4 block. Embedding according to $Rule\#1$ provides better results in terms of the both robustness and transparency in comparison with that for $Rule\#2$. It can be explained by recalling that $Rule\#1$ causes much lower distortions while fraction of "weak" bits for threshold $Th = 0.002$ is also small (Fig. 1b).

Comparison with the method proposed by Li leads to different conclusions depending on a kind of attack. For signal processing and noise attacks (except Salt&Pepper) the method of Li demonstrates better robustness than the proposed method with $Rule\#1$. On the other hand robustness toward geometric attacks is much worse for Li's method. This is because such factors as smaller blocks and higher redundancy are more beneficial in such kind of attacks (especially cropping). Another advantage of the proposed method is considerably better transparency of the watermarked images.

The proposed method does not provide the best robustness in all the cases of different attacks. However, its robustness is sufficient and embedding distortions are low. Maximum data payload is the same as for the method of Tehrani and Gorodetski and is equal to 16384 bits per 512x512 image, which is quite large. All the mentioned characteristics make the proposed method favorable in most of watermarking applications.

## VI. CONCLUSIONS

The watermarking method proposed in this paper is blind. It uses SVD of a 4x4 image block in order to embed a bit of a watermark by modifying an orthonormal matrix which is either left or right. The modification is done according to one of the proposed embedding rules and the procedure for minimization of embedding distortions is considered. Redundant embedding is applied with the aim to maintain good robustness-transparency tradeoff.

Modification of one of the orthonormal matrices of SVD of 4x4 image block is done according to the model of rotations in four dimensional space. Optimization tasks are being solved with the aim to minimize distortions. The modification is split on two phases that simplifies each separate optimization task and reduces computation load.

Two different embedding rules were proposed which are $Rule\#1$ and $Rule\#2$. The $Rule\#1$ interprets both orthonormal matrices of SVD, while the $Rule\#2$ interprets only the left one. Each rule can be used in conjunction with the proposed method which leads to different embedding distortions and robustness.

Redundant embedding is beneficial in a combination with the proposed watermarking method as the latter introduces low distortions, but maintains considerable payload. Utilization of different redundancy rates for various applications yields to different robustness-transparency tradeoffs.

The efficiency of the proposed method was confirmed experimentally. For some kinds of popular attacks BER of watermark extraction is 36% lower compared to other known methods.

## REFERENCES


[1] C. Song, S. Sudirman, M. Merabti and D. Llewellyn-Jones, "Analysis of Digital Image Watermark Attacks," in *Proceedings of Consumer Communications and Networking Conference (CCNC)*, 9-12 Jan. 2010.

[2] H. R. Sheikh and A. Bovik, "Image information and visual quality," *IEEE Transactions on Image Processing,* vol. 15, no. 2, pp. 430-444, Feb. 2006.

[3] I. Cox, M. Miller, J. Bloom, J. Fridrich and T. Kalker, Digital Watermarking and Steganography (2 ed.), San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2007.

[4] S. Lin and C.-F. Chen, "A robust DCT-based watermarking for copyright protection," *IEEE Transactions on Consumer Electronics,* vol. 46, no. 3, pp. 415-421, Aug 2000.

[5] E. Fullea and J. Martinez, "Robust digital image watermarking using DWT, DFT and quality based average," in *Proceedings of the ninth ACM international conference on Multimedia (MULTIMEDIA '01)*, 2001.

[6] L. Trefethen and D. Bau, Numerical Linear Algebra, Cambridge University Press, 1997.

[7] V. Gorodetski, L. Popyack, V. Samoilov and V. Skormin, "SVD-based Approach to Transparent Embedding Data into Digital Images," in *Proceedings of the International Workshop on Information Assurance in Computer Networks: Methods, Models, and Architectures for Network Security (MMM-ACNS '01)*, 2001.

[8] R. Liu and T. Tan, "An SVD-based watermarking scheme for protecting rightful ownership," *IEEE Transactions on Multimedia,* vol. 4, no. 1, pp. 121-128, Mar 2002.

[9] X.-P. Zhang and K. Li, "Comments on "An SVD-based watermarking scheme for protecting rightful Ownership"," *IEEE Transactions on Multimedia,* vol. 7, no. 3, pp. 593-594, June 2005.

[10] E. Ganic and A. Eskicioglu, "Robust DWT-SVD domain image watermarking: embedding data in all frequencies," in *Proceedings of the 2004 workshop on Multimedia and security (MM&Sec '04)*, 2004.

[11] P. Bao and X. Ma, "Image adaptive watermarking using wavelet domain singular value decomposition," *IEEE Transactions on Circuits and Systems for Video Technology,* vol. 15, no. 1, pp. 96-102, Jan. 2005.

[12] H.-C. Wu, R.-J. Jang and Y.-C. Liu, "A robust watermarking scheme based on singular value decomposition and quantization technique," in *Proceedings of International Computer Symposium (ICS)*, Dec. 2010.

[13] Z. Li, K.-H. Yap and B.-Y. Lei, "A new blind robust image watermarking scheme in SVD-DCT composite domain," in *Proceedings of 18th IEEE International Conference on Image Processing (ICIP)*, Sept. 2011.

[14] C.-C. Chang, P. Tsai and C.-C. Lin, "SVD-based digital image watermarking scheme," *Pattern Recognition Letters,* no. 26, p. 1577–1586, 2005.

[15] I. O. Tehrani and S. Ibrahim, "An enhanced SVD based watermarking using U matrix," in *Proceedings of 5th International Conference on Computer Sciences and Convergence Information Technology (ICCIT)*, 2010.

[16] H. P. Manning, Geometry of Four Dimensions, Dover Publications, 1956.

Publication IV

# A New QIM-Based Watermarking Method Robust to Gain Attack

Yevhen Zolotavkin and Martti Juhola

Hindawi Publishing Corporation
International Journal of Digital Multimedia Broadcasting
Volume 2014, Article ID 910808, 14 pages
http://dx.doi.org/10.1155/2014/910808

*Research Article*

# A New QIM-Based Watermarking Method Robust to Gain Attack

**Yevhen Zolotavkin and Martti Juhola**

*Research Center for Information and Systems, School of Information Sciences, University of Tampere, Kanslerinrinne 1, 33014 Tampere, Finland*

Correspondence should be addressed to Yevhen Zolotavkin; zhzolot@suremail.us

Received 26 May 2014; Accepted 22 August 2014; Published 11 September 2014

Academic Editor: Harald Kosch



We propose a new watermarking method based on quantization index modulation. A concept of initial data loss is introduced in order to increase capacity of the watermarking channel under high intensity additive white Gaussian noise. According to the concept some samples in predefined positions are ignored even though this produces errors in the initial stage of watermark embedding. The proposed method also exploits a new form of distribution of quantized samples where samples that interpret "0" and "1" have differently shaped probability density functions. Compared to well-known watermarking schemes, this provides an increase of capacity under noise attack and introduces a distinctive feature. Two criteria are proposed that express the feature numerically. The criteria are utilized by a procedure for estimation of a gain factor after possible gain attack. Several state-of-the-art quantization-based watermarking methods were used for comparison on a set of natural grayscale images. The superiority of the proposed method has been confirmed for different types of popular attacks.

## 1. Introduction

Digital media have a great impact on many aspects of modern society. Some aspects assume that we deal with audio-visual data that relates to a person or an organization. Information about the relation quite often should be preserved. Watermarking approach is to insert the information in the media itself [1]. However, in that case the watermark might be intentional or not altered by the third party. In order to avoid an alteration the watermark needs to be robust [2]. Other characteristics except robustness may also be important. Watermark invisibility and payload are among them. Invisibility is important to assure that the quality of the media does not degrade significantly as a result of watermarking [3]. High data payload might be needed in some applications in order to define many aspects of ownership.

In the field of digital image watermarking (DIW) digital images are used as a media (or host). DIW incorporates many different techniques and one of the most popular among them is quantization index modulation (QIM). Methods that belong to QIM are widely used in blind watermarking where neither original media nor watermark is known to the receiver [4]. For the purpose of evaluating robustness the watermarked image is being attacked and additive white Gaussian noise (AWGN) is the most popular condition for that. Theoretical limit of the channel capacity which is achievable by QIM under AWGN was first derived in [5].

In most cases quantization is implemented to some coefficients rather than to signal samples. In order to obtain coefficients a transform is applied to a host signal. It has been shown that some transforms provide coefficients that are more robust to popular image processing algorithms like JPEG, geometric modifications, and so forth [6, 7].

It is assumed that during quantization each of the original coefficient values belongs to one of equally spaced intervals. Further, inside each interval coefficients to interpret "0" and "1" are selected. The task of quantization is to separate coefficients that represent different bits inside each interval. The separation efficiency influences robustness and invisibility. The result of the separation can be characterized by the size of original interval, distribution of separated samples, and the distortion incurred by the separation.

However, all the known implementations of QIM are far from achieving the capacity limit under AWGN. The simplest scalar realization of QIM is to replace all the coefficient values from a certain interval by a single value equal either to the left or right endpoint depending on a bit of a watermark [8].

Hence, the distribution of quantized samples that represent both "0" and "1" is degenerate (or Dirac). Nevertheless, the capacity of the simplest QIM (further referred to as QIM without "simplest") is less than 10% of the limit value for the condition when noise and watermark energies are equal. More advanced realization of DC-QIM is to replace each coefficient value from an original interval by a corresponding value taken from one out of two disjoint intervals that are subsets of the original one [9]. A parameter $0.5(1 - \alpha)$ is to control the size of these intervals relatively to the original. The distribution for "0" and "1" in that case is uniform. Parameter $\alpha$ is being adjusted depending on noise level in order to maximize capacity. Method DC-QIM is widely used and provides the highest capacity under AWGN among known practical realizations. However, considering AWGN attack only, the most evident gap under high noise intensity is caused by low capacity in comparison with the theoretical limit.

Some other modifications of QIM have emerged over the past years. Forbidden zone data hiding (FZDH) modifies only a fraction (controlled by $\alpha$) of coefficient values in each interval of original values [10]. Despite the fact that FZDH performs slightly worse than DC-QIM the paper represents a promising idea on how to reduce embedding distortions. Another idea was proposed by the authors of Thresholded Constellation Modulation (TCM) that use two different quantization rules to modify coefficients inside the original interval [11]. Each rule is applied only to samples from a particular subinterval and $\beta$ defines their endpoints. The value of a shift is different for any value from a subinterval according to the first quantization rule. The second rule is to provide an equal shift to all the values from another subinterval. There are two shift directions in order to embed "0" and "1."

The main advantage of the techniques based on QIM with different kinds of compensation [9–11] is a considerable robustness against AWGN. The limitation is that synchronization is required to extract a watermark. Even minor distortion of a different kind can make embedded information unreadable. The simplest realization of such kind of distortion is a gain attack (GA) which performs a constant scaling of the whole sequence of watermarked coefficients. The scaling factor might be close to 1 and cause very little visual distortion, but it is unknown to the receiver which causes asynchronous extraction. Usually GA is followed by AWGN that complicates retrieval of the watermark [12]. Vulnerability to GA causes one of the most critical gaps for practical implementation of QIM-based methods with compensation.

Many different approaches have been developed to improve robustness against GA of QIM related methods [13]. Most approaches can be classified into two groups where the main idea of the first group is to estimate the unknown factor while the idea of second is to quantize coefficients that are invariant to scaling of original signal.

Estimation of the scaling factor requires modelling. Some feature that is unique for the watermarked and attacked signal might be described by a model [14]. The scaling factor may be included in the model and to be a subject to optimization. An obvious complication is that a process of feature selection is not straightforward. In some cases the feature is created instead of being selected and some permanent data agreed upon between the sender and the receiver is a suitable example. However, such compulsory agreement limits practical implementation of watermarking method. Other possible limitations are low model accuracy or computationally heavy optimization.

For instance, a kind of GA and a constant offset attack followed by AWGN are assumed in [12]. The solution proposed there is to embed a pilot signal and to use Fourier analysis to estimate the gain factor and the offset. However, an obvious disadvantage of the solution is that the precision of estimated parameters is low even for quite long pilot sequence.

The method of recovery after GA and AWGN is proposed in [15]. It uses information about dither sequence and applies maximum likelihood (ML) procedure to estimate the scaling factor. The estimation is based on a model of watermarked and attacked signal. Information about statistical characteristics of original host signal should be either known or guessed in order to define the model. Another limitation of the approach is that it requires exact information about embedding parameter $\alpha$. Computational complexity of the approach is high.

As an opposite for estimation, invariance to GA, in general, requires more complex transform of original signal (e.g., nonlinear,) to obtain coefficients. It is necessary to modify coefficients to embed a watermark. However, a model to estimate distortions of the host is more complex in that case. Distortions should be controlled which limits the choice of the kind of QIM to one that adds less complexity to a model of distortion. This, for example, might result in reducing the number of adjustable parameters of QIM. This is one of the reasons why invariant to GA approaches are more vulnerable to AWGN compared to DC-QIM.

Rational dither modulation (RDM) is one of the most popular watermarking methods invariant to GA [16]. For a particular coefficient, a ratio that depends on a norm of other coefficients is being quantized instead of a coefficient itself. The simplest QIM scheme is utilized in order to quantize the ratio and the performance of RDM under AWGN (without GA) is close to the simplest QIM. Other recent blind watermarking methods robust to GA are proposed in [17–23]. Nevertheless, for GA invariant methods the gap is caused by the reduced capacity under AWGN.

In this paper, we propose our own scalar QIM-based watermarking approach that is beneficial in several aspects. The approach addresses the mentioned gaps in the literature: it both delivers higher capacity under AWGN and recovers after GA. In order to do this, host signal coefficients are separated in a way that the resulting distributions for coefficients that interpret "0" and "1" are different. This distinctive feature is used by a simple yet efficient procedure for estimation of a scaling factor under GA. A concept of initial data loss (IDL) is introduced in order to increase watermark channel capacity under low watermark to noise ratios (WNRs). According to IDL, some fraction of wrong watermark bits is accepted during embedding procedure.

The rest of the paper is organized as follows. In Section 2 we describe our quantization model using formal logic approach and derive some constraints on the parameters of the model. In Section 3 some important watermarking characteristics of the model are evaluated analytically while the following Section 4 contains description for the procedure of recovery after GA as well as experimental results obtained under popular attacks. In Section 5 we discuss in detail experimental conditions and compare the performance of the proposed method with the performance of well-known methods. Section 6 concludes the paper and outlines possible directions for improvement.

## 2. Quantization Model

In this section we define a new model of quantization. First, it is necessary to show that according to our model the separation of original coefficients is possible and we can embed information. Formal logic approach is used to define dependencies between several conditions that are important for the separation of original coefficients. Separation argument (SA) represents the model in a compact form yet has a clear structure which is sufficient to reason the intuition behind the dependencies. Second, it is necessary to assure conditions when SA is sound.

*2.1. Formalization of SA.* Symbol $\Sigma$ will be used to denote a random variable whose domain is the space of original coefficients of a host. A particular realization of $\Sigma$ will be denoted by $\varsigma$. We will further describe our model for values $\varsigma$ that are in some interval of size $\Delta$. More specifically we will refer to an interval with integer index $k$ whose left endpoint is $l_\Delta^k$. Such an interval is referred to further as embedding interval. For any $\varsigma \in [l_\Delta^k, l_\Delta^k + \Delta]$ we define $x = \varsigma - l_\Delta^k$ and $X$ will be used to denote a random variable which represents $x$. The length $\Delta$ is selected in a way that an appropriate document to watermark ratio (DWR) is guaranteed after the separation. We also assume that $\Delta$ is small enough to derive that the distribution of $X$ is uniform. A random variable that represents separated coefficients inside $k$th interval is denoted by $X'$ and its realization is denoted by $x'$. Correspondingly, a random variable that represents separated coefficients on the whole real number line is denoted by $\Sigma'$ and its realization is denoted by $\varsigma'$. Each pair of an original $x$ and corresponding quantized $x'$ belong to the same $k$th embedding interval so that an absolute shift is never larger than $\Delta$.

Let us denote a watermark bit by $b$. Truncated pdfs $f_0(x')$ and $f_1(x')$ are used to describe the distribution of $X'$ and should be defined prior to quantization. Parameters $\eta_1$ and $\vartheta_0$ represent fractions of IDL for $b = 1$ and $b = 0$, respectively. Parameters $\varphi_1$ and $\gamma_0$ represent fractions of the samples where original values $x$ are to be modified by a quantizer for $b = 1$ and $b = 0$, respectively. It is therefore assumed that the fraction of zeros in a watermark data is $\gamma_0 + \vartheta_0$ and fraction of ones is $\varphi_1 + \eta_1$. Condition $\gamma_0 + \vartheta_0 + \varphi_1 + \eta_1 = 1$ always holds. The result of the separation in the $k$th embedding interval depends on $b$, $\eta_1$, $\vartheta_0$, $\varphi_1$, $\gamma_0$, $f_0(x')$, and $f_1(x')$. In other words $x'$ is defined by quantizer $Q_\Delta^k[\cdot]$ that has the mentioned parameters:

$$x' = Q_\Delta^k\left[x, b, \eta_1, \vartheta_0, \varphi_1, \gamma_0, f_0, f_1\right]. \tag{1}$$

We will use SA to describe the quantizer $Q_\Delta^k[\cdot]$. Each of logical atoms $p, q, r, s, t, u$, and $v$ represents some condition which is either true or false:

$$p \mid x \le \frac{\Delta\gamma_0}{\gamma_0 + \vartheta_0}, \tag{2}$$

$$q \mid x \ge \frac{\Delta\eta_1}{\varphi_1 + \eta_1}, \tag{3}$$

$$r \mid x' = x, \tag{4}$$

$$s \mid x' < x, \tag{5}$$

$$t \mid x' > x, \tag{6}$$

$$u \mid x\frac{\gamma_0 + \vartheta_0}{\Delta} = \gamma_0 \int_0^{x'} f_0\left(x'\right) dx', \tag{7}$$

$$v \mid (\Delta - x)\frac{\varphi_1 + \eta_1}{\Delta} = -\varphi_1 \int_\Delta^{x'} f_1\left(x'\right) dx'. \tag{8}$$

For example, $(\sim b\&p)$ is true if and only if "$b = 0$" and $x$ is not classified for IDL. We formalize SA in the following way:

$$\begin{gathered}\left((\sim b\&p) \supset (u\&(s \vee r))\right),\\ \left((b\&q) \supset (v\&(t \vee r))\right),\\ \left(((\sim b\&\sim p) \vee (b\&\sim q)) \supset r\right)\\ \vDash ((u\&(s \vee r)) \vee (v\&(t \vee r)) \vee r).\end{gathered} \tag{9}$$

It can be seen that SA is valid. The conclusion of SA states that the separation of coefficient values inside $k$th embedding interval is possible which means that the proposed model is suitable for information embedding. Furthermore, each premise represents an important dependency between input and output of the quantizer $Q_\Delta^k[\cdot]$ and we require that each premise is indeed true. Hence, it is necessary to enforce soundness for SA.

The intuition behind SA can be explained in the following way. Initially samples with labels "$b = 0$" and "$b = 1$" are not separated in the dimension of $x$ inside the mentioned $k$th embedding interval. In order to separate them we shift those with "$b = 0$" to the left and those with "$b = 1$" to the right. If so, shift to the right for "$b = 0$" or shift to the left for "$b = 1$" is not acceptable because it would introduce distortion and on the other hand worsen separation between "0" and "1." Therefore for $\sim b$ formula $(s \vee r)$ is true and for $b$ formula $(t \vee r)$ is true.

Another consideration is that for any two $x_i \le x_j$ with the same bit value we infer that quantization in a way that $x_i' \le x_j'$ implies less distortion than if $x_i' > x_j'$. Saving the order we preserve cumulative distribution in respect to the order. Quantized samples $x'$ that interpret "0" are distributed according to pdf $f_0(x')$; samples $x'$ that interpret "1" are distributed according to pdf $f_1(x')$. Therefore $u$ or $v$ is true if $(\sim b\&p)$ or $(b\&q)$ is true, respectively.

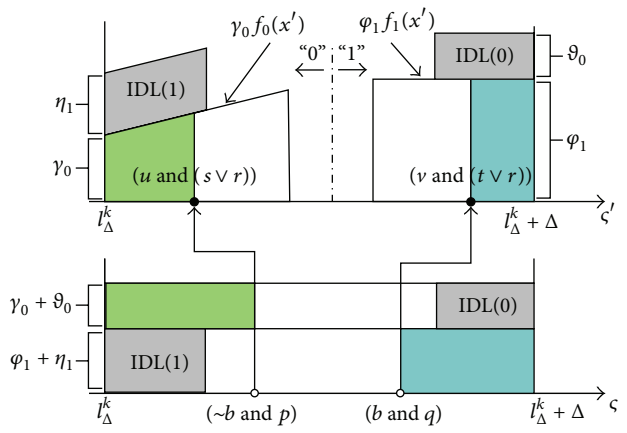



Figure 1: Illustration of the process of separation.

And, lastly, the condition for IDL is $((\sim b \,\&\sim p) \vee (b \,\&\sim q))$ and it is the case when $x$ is not modified and therefore $r$.

An illustration of an example where SA is sound is given in Figure 1. Two positions of original values are shown on the lower part of Figure 1. Condition $(\sim b \& p)$ is satisfied for the first original value and condition $(b \& q)$ is satisfied for the second. Two positions of the modified values are shown on the upper part of Figure 1. After the separation the modified values satisfy conditions $(u \& (s \vee r))$ and $(v \& (t \vee r))$, respectively. The areas of green segments on the lower and the upper parts of Figure 1 are equal. The areas of blue segments are also equal. As it can be seen on the upper part of Figure 1, the distribution of separated coefficients in $k$th embedding interval depends on $\Delta, \eta_1, \vartheta_0, \varphi_1, \gamma_0, f_0(x')$, and $f_1(x')$.

Parameters of the pdfs $f_0(x')$ and $f_1(x')$ need to be specified in order to prove soundness for the whole range of $x$ in the $k$th interval. In addition formulas (7) and (8) need to be rearranged in order to express $x'$ in a suitable way for the quantization form.

We propose such $f_0(x')$ and $f_1(x')$ that in general there is no line of symmetry which can separate them inside embedding interval. This feature will provide easier recovery after GA. It is necessary to emphasize that the proposed functions $f_0(x')$ and $f_1(x')$ only describe distributions for fractions $\gamma_0$ and $\varphi_1$, respectively (e.g., without taking into account fractions of IDL). We introduce parameters $\alpha$, $\beta$, and $\tau$ to define both pdfs $f_0(x')$ and $f_1(x')$, where $0 \le \alpha \le \beta \le 1$, as shown in Figure 2(a). As can be seen the density is zero in the subinterval $(\Delta(\beta - \alpha), \Delta\beta)$ which separates "0" from "1." In Figure 2(b) we can see the distribution of the quantized coefficients outside $k$th embedding interval as well.

Namely, the proposed truncated pdfs are a linear function and a constant:

$$f_0\left(x'\right) = \begin{cases} cx' + \tau, & \text{if } x' \in \left[0, \Delta\left(\beta - \alpha\right)\right], \\ 0, & \text{otherwise;} \end{cases} \tag{10}$$

$$f_1\left(x'\right) = \begin{cases} g, & \text{if } x' \in \left[\Delta\beta, \Delta\right], \\ 0, & \text{otherwise.} \end{cases} \tag{11}$$

The samples that belong to IDL fraction are distributed according to pdfs $\mathrm{IDL}_0(x')$ and $\mathrm{IDL}_1(x')$:

$$\begin{aligned} \mathrm{IDL}_0\left(x'\right) &= \begin{cases} \dfrac{\gamma_0 + \vartheta_0}{\Delta\vartheta_0}, & \text{if } x' \in \left[\dfrac{\Delta\gamma_0}{\gamma_0 + \vartheta_0}, \Delta\right], \\ 0, & \text{otherwise;} \end{cases} \\ \mathrm{IDL}_1\left(x'\right) &= \begin{cases} \dfrac{\varphi_1 + \eta_1}{\Delta\eta_1}, & \text{if } x' \in \left[0, \dfrac{\Delta\eta_1}{\varphi_1 + \eta_1}\right], \\ 0, & \text{otherwise.} \end{cases} \end{aligned} \tag{12}$$

*2.2. Soundness Conditions for SA.* The soundness of SA is guaranteed if it is possible to satisfy $(u \& (s \vee r))$ or $(v \& (t \vee r))$ when $(\sim b \& p)$ or $(b \& q)$ is true, respectively. The requirement to satisfy $(u \& (s \vee r))$ or $(v \& (t \vee r))$ imposes some constraints on $\alpha, \beta, c, g, \tau, \gamma_0, \varphi_1, \eta_1, \vartheta_0$, and $\Delta$. Let us find those constraints.

We start from defining parameters of $f_0(x')$ and $f_1(x')$ using property of pdf:

$$\int_0^{(\beta-\alpha)\Delta} f_0\left(x'\right) dx' = c\frac{\left(\beta - \alpha\right)^2\Delta^2}{2} + \tau\Delta\left(\beta - \alpha\right) = 1, \tag{13}$$

$$\int_{\beta\Delta}^{\Delta} f_1\left(x'\right) dx' = g\Delta\left(1 - \beta\right) = 1. \tag{14}$$

It is easy to derive from (14) that

$$g = \frac{1}{\Delta\left(1 - \beta\right)}. \tag{15}$$

According to (4), (5), and (7) condition $(u \& (s \vee r))$ is satisfied if and only if for all $x'$

$$x'\frac{\gamma_0 + \vartheta_0}{\Delta} \le \gamma_0 \int_0^{x'} f_0\left(x'\right) dx'. \tag{16}$$

Using (10) and the fact $x' \ge 0$ we can derive

$$\tau \ge \frac{\gamma_0 + \vartheta_0}{\Delta\gamma_0} - c\frac{x'}{2}. \tag{17}$$

The latter inequality should be true for all $x' \in [0, \Delta(\beta - \alpha)]$ which means

$$\tau \ge \max_{x' \in [0, \Delta(\beta-\alpha)]} \left(\frac{\gamma_0 + \vartheta_0}{\Delta\gamma_0} - c\frac{x'}{2}\right). \tag{18}$$

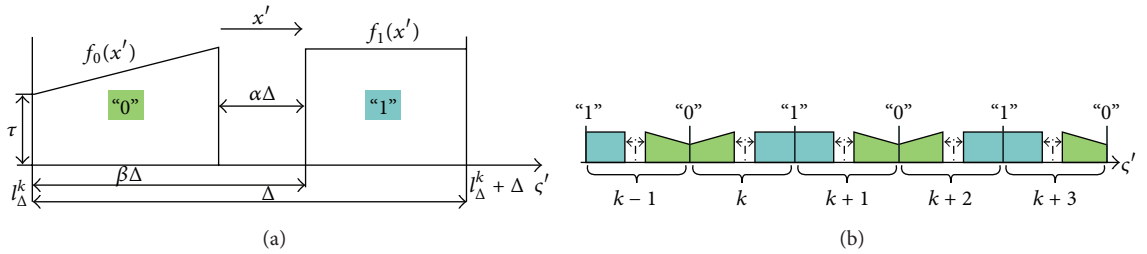
For our particular application we chose $c \ge 0$; therefore

$$\tau \ge \frac{\gamma_0 + \vartheta_0}{\Delta\gamma_0}, \tag{19}$$

and we are using the value $\tau = (\gamma_0 + \vartheta_0)/(\Delta\gamma_0)$ in our method.

Using (13) we can conclude that

$$c = 2\frac{\gamma_0 - \left(\gamma_0 + \vartheta_0\right)\left(\beta - \alpha\right)}{\gamma_0\left(\beta - \alpha\right)^2\Delta^2}. \tag{20}$$




(a) (b)

FIGURE 2: Distribution of the quantized coefficients: (a) inside $k$th embedding interval; (b) in five consecutive intervals.

Functions $f_0(x')$ and $f_1(x')$ can be fully defined now. Let us find dependencies that connect $\alpha$ and $\beta$ with $\gamma_0, \varphi_1, \eta_1$, and $\vartheta_0$. Taking into account that in our realization $c \geq 0$ we can derive from (20) that

$$\beta - \alpha \leq \frac{\gamma_0}{\gamma_0 + \vartheta_0}. \tag{21}$$

According to (4), (6), and (8) condition $(v\&(t \vee r))$ is satisfied if and only if

$$\frac{\varphi_1 + \eta_1}{\Delta} \leq g\varphi_1. \tag{22}$$

Using (15) we find that

$$\beta \geq \frac{\eta_1}{\varphi_1 + \eta_1}. \tag{23}$$

In the experiment section of the paper the goal is to find the highest capacity for a given WNR. Different values of the parameters need to be checked for that purpose. Preserving (15) and (19)–(21), (23) would guarantee soundness of SA and avoidance of using parameters' combinations that are not efficient for watermarking. This can reduce required computations.

*2.3. Embedding Equations.* For the proposed pdfs we can now define $x'$ as a function of $x$, which is the main task of the quantizer $Q_\Delta^k[\cdot]$. Let us consider conditions $(\sim b\&p)$, $(b\&q)$ separately as it is never the case when both conditions are true. We will denote $x'$ in case of $(\sim b\&p)$ by $\dot{x}'$, but in case of $(b\&q)$ the notation $\ddot{x}'$ will be used.

From (7), (10), and $\tau = (\gamma_0 + \vartheta_0)/(\Delta\gamma_0)$ it is clear that

$$0.5c\dot{x}'^2 + \tau\dot{x}' = \tau x. \tag{24}$$

Taking into account that $\dot{x}' \geq 0$ we derive

$$\dot{x}' = \frac{\sqrt{\tau^2 + 2c\tau x}}{c} - \frac{\tau}{c}. \tag{25}$$

From (8), (11), and (15) we can find that

$$\ddot{x}' = Bx + \Delta(1 - B), \qquad B = \frac{(1-\beta)(\varphi_1 + \eta_1)}{\varphi_1}. \tag{26}$$

According to (26), the values of quantized coefficients are linearly dependent on original values while according to (25) the dependency is nonlinear. Different character of dependency between quantized and original values for "0" and "1" is one of the key features of our approach. This differentiates the proposed watermarking method from the methods previously described in the literature [10–12].

## 3. Characteristics of Quantization Model

The model was proposed in the previous section. It was shown that it is suitable for coefficient separation and the conditions necessary for soundness of SA were defined. In this section we focus on efficiency of separation. The main characteristic that can be estimated analytically is the watermark channel capacity under AWGN. It is required to calculate such characteristic for different WNRs. First, we express WNR in terms of parameters of the quantization scheme. Second, we express error rates in terms of parameters of the quantization scheme. This makes it possible to include WNR in the expression for error rates (and capacity).

*3.1. Estimation of Quantization Distortions.* The variance $\sigma_n^2$ is the only parameter of AWGN attack and WNR is defined as

$$\mathrm{WNR} = 10\log_{10}\left(\frac{D}{\sigma_n^2}\right), \tag{27}$$

where $D$ is a watermark energy. Alternatively, $D$ can be seen as a distortion of a host signal, induced by the quantization. Let us define $D$.

For the matter of convenience of the experiment it is better to use a single parameter (control parameter) that can be adjusted in order to provide the desired value of $D$. While defining $D$ we choose $\Delta$ to be the control parameter and collect it in the expression for $D$. The total distortion $D$ is a sum of distortions $D_0$ and $D_1$ caused by two types of shifts that are $x \to \dot{x}'$ and $x \to \ddot{x}'$, respectively. The first distortion component $D_0$ is defined as

$$D_0 = \gamma_0 \int_0^{\Delta(\beta-\alpha)} f_0(x')\left(x' - \frac{1}{\tau}\int_0^{x'} f_0(x')\,dx'\right)^2 dx'. \tag{28}$$

Proceeding further and using (10) we can derive that

$$D_0 = \gamma_0 \int_0^{\Delta(\beta-\alpha)} (cx' + \tau)\frac{c^2 x'^4}{4\tau^2}dx'. \tag{29}$$

However, it is clear from (19)-(20) that both parameters $c$ and $\tau$ depend on $\Delta$. In order to collect $\Delta$ we introduce two independent of $\Delta$ parameters $\widehat{c} = c\Delta^2$ and $\widehat{\tau} = \tau\Delta$. This brings us to

$$D_0 = \Delta^2 Q_0,$$
$$Q_0 = \gamma_0 \left( \frac{\widehat{c}^3}{24\widehat{\tau}^2}(\beta-\alpha)^6 + \frac{\widehat{c}^2}{20\widehat{\tau}}(\beta-\alpha)^5 \right). \tag{30}$$

The second distortion component $D_1$ is defined as

$$D_1 = \varphi_1 \times \int_{\beta\Delta}^{\Delta} f_1\left(x'\right) \times \left( x' - \left( \frac{\varphi_1\Delta}{\eta_1+\varphi_1} \int_{\beta\Delta}^{x'} f_1\left(x'\right) dx' + \frac{\eta_1\Delta}{\eta_1+\varphi_1} \right) \right)^2 dx'. \tag{31}$$

Using (11), (15), and integrating in (31) we obtain

$$D_1 = \Delta^2 Q_1,$$
$$Q_1 = \varphi_1 \frac{\left((\eta_1+\varphi_1)(1-\beta)-\varphi_1\right)^2}{3(\eta_1+\varphi_1)^2}. \tag{32}$$

The total quantization distortion $D$ can be expressed in terms of $\Delta, Q_0$, and $Q_1$:

$$D = \Delta^2 \left(Q_0 + Q_1\right). \tag{33}$$

For any combination of $\sigma_n^2$, WNR, $\alpha, \beta, \eta_1, \vartheta_0, \gamma_0$, and $\varphi_1$ the required value of $\Delta$ is defined using (27) and (33) as

$$\Delta = \sqrt{\frac{\sigma_n^2 10^{0.1*\mathrm{WNR}}}{Q_0+Q_1}}. \tag{34}$$

*3.2. Estimation of Error Rates.* Bit error rate (BER) and channel capacity can be calculated without simulation of watermark embedding procedure. It is important that the kind of threshold used to distinguish between "0" and "1" is suitable for analytic estimations. Further we assume that the position of the threshold remains permanent after watermark is embedded and does not depend on attack parameters. In Figure 2(b) the position of the threshold is Th for intervals numbered $k + 2m$, $m \in \mathbb{Z}$. For the intervals numbered $k + 2m + 1$ the position of the threshold is $\Delta - \mathrm{Th}$.

The absolute value of quantized sample in any interval is $\varsigma'$. We use $\varsigma_n'$ for a sample that is distorted by noise. Hence, $\varsigma_n'$ interprets "0" or "1" depending on belonging to $\mathbf{Z}$ or $\mathbf{O}$, respectively:

$$\mathbf{Z} = \bigcup_{m=-\infty}^{\infty} \left[ 2\Delta m + l_\Delta^k - \mathrm{Th}, 2\Delta m + l_\Delta^k + \mathrm{Th} \right), \tag{35}$$

$$\mathbf{O} = \bigcup_{m=-\infty}^{\infty} \left[ 2\Delta m + l_\Delta^k + \mathrm{Th}, 2\Delta(m+1) + l_\Delta^k - \mathrm{Th} \right). \tag{36}$$

There are two cases when errors occur in non-IDL samples. An error in "0" is incurred by a noise if and only if the both following conditions are true:

$$\left(\varsigma' \in \mathbf{Z}\right), \qquad \left(\varsigma_n' \in \mathbf{O}\right). \tag{37}$$

An error in "1" occurs if and only if the following is true:

$$\left(\varsigma' \in \mathbf{O}\right), \qquad \left(\varsigma_n' \in \mathbf{Z}\right). \tag{38}$$

Two cases when errors occur in IDL samples can be presented with the following conditions for "0" and "1," respectively:

$$\left(\varsigma' \in \mathbf{O}\right), \qquad \left(\varsigma_n' \in \mathbf{O}\right), \tag{39}$$

$$\left(\varsigma' \in \mathbf{Z}\right), \qquad \left(\varsigma_n' \in \mathbf{Z}\right). \tag{40}$$

The pdf of AWGN with variance $\sigma_n^2$ can be represented in terms of $\varsigma'$ and $\varsigma_n'$ as $f_{\mathcal{N}}[\varsigma_n' - \varsigma', 0, \sigma_n]$. In general we can estimate error rates for an interval with any integer index $k + m$. For that purpose we use generalized notations $\dot{f}_0(\varsigma')$, $\dot{f}_1(\varsigma')$, $\mathrm{I\dot{D}L}_0(\varsigma')$, and $\mathrm{I\dot{D}L}_1(\varsigma')$ for pdfs of quantized samples in any interval. For example, for even $m$ pdf $\dot{f}_0(\varsigma') = f_0[\varsigma' - (l_\Delta^k + \Delta m)]$; for odd $m$ pdf $\dot{f}_0(\varsigma') = f_0[l_\Delta^k + \Delta(m+1) - \varsigma']$. We denote $k+m$ interval by $I_{k+m} = [l_\Delta^k + \Delta m, l_\Delta^k + \Delta(m+1)]$. Then, the error rates for quantized samples in $I_{k+m}$ can be defined as

$$\begin{aligned}
\mathrm{BER}_0 &= \frac{\gamma_0}{\gamma_0+\vartheta_0} \int_{\mathbf{O}} \int_{I_{k+m}} \dot{f}_0\left(\varsigma'\right) f_{\mathcal{N}}\left[\varsigma_n' - \varsigma', 0, \sigma_n\right] d\varsigma' d\varsigma_n' \\
&\quad + \frac{\vartheta_0}{\gamma_0+\vartheta_0} \int_{\mathbf{O}} \int_{I_{k+m}} \mathrm{I\dot{D}L}_0\left(\varsigma'\right) \times f_{\mathcal{N}}\left[\varsigma_n' - \varsigma', 0, \sigma_n\right] d\varsigma' d\varsigma_n', \\
\mathrm{BER}_1 &= \frac{\varphi_1}{\varphi_1+\eta_1} \int_{\mathbf{Z}} \int_{I_{k+m}} \dot{f}_1\left(\varsigma'\right) f_{\mathcal{N}}\left[\varsigma_n' - \varsigma', 0, \sigma_n\right] d\varsigma' d\varsigma_n' \\
&\quad + \frac{\eta_1}{\varphi_1+\eta_1} \int_{\mathbf{Z}} \int_{I_{k+m}} \mathrm{I\dot{D}L}_1\left(\varsigma'\right) \times f_{\mathcal{N}}\left[\varsigma_n' - \varsigma', 0, \sigma_n\right] d\varsigma' d\varsigma_n'.
\end{aligned} \tag{41}$$

Now we can show that $\mathrm{BER}_0$ and $\mathrm{BER}_1$ can be calculated according to (41) for any chosen interval. For that purpose it is enough to demonstrate that any component in (41) remains the same for every interval. For example, we state that

$$\begin{aligned}
&\int_{\mathbf{O}} \int_{I_{k+m}} \dot{f}_0\left(\varsigma'\right) f_{\mathcal{N}}\left[\varsigma_n' - \varsigma', 0, \sigma_n\right] d\varsigma' d\varsigma_n' \\
&= \int_{\mathbf{O}} \int_0^{\Delta} f_0\left(x'\right) f_{\mathcal{N}}\left[\varsigma_n' - \left(l_\Delta^k + x'\right), 0, \sigma_n\right] dx' d\varsigma_n',
\end{aligned} \tag{42}$$

for any $m$.

Let us first assume $m = 2n$, $n \in \mathbb{Z}$. Then, $\varsigma' = x' + l_\Delta^k + 2\Delta n$, $\dot{f}_0(\varsigma') = f_0(x')$. However, it is also clear from (36) that $\mathbf{O} + 2\Delta n = \mathbf{O}$. Hence,

$$\begin{aligned}
&\int_{\mathbf{O}} \int_{I_{k+2n}} \dot{f}_0\left(\varsigma'\right) f_{\mathcal{N}}\left[\varsigma_n' - \varsigma', 0, \sigma_n\right] d\varsigma' d\varsigma_n' \\
&= \int_{\mathbf{O}+2n\Delta} \int_0^{\Delta} f_0\left(x'\right) \\
&\qquad \times f_{\mathcal{N}}\left[\left(\varsigma_n' - 2n\Delta\right)\right. \\
&\qquad\qquad \left. - \left(l_\Delta^k + x'\right), 0, \sigma_n\right] dx' d\left\{\varsigma_n' - 2n\Delta\right\},
\end{aligned} \tag{43}$$

and we prove the statement.

Now let us assume $m = 2n + 1$, $n \in \mathbb{Z}$. Then, $\varsigma' = (x' - \Delta) + l_\Delta^k + 2\Delta(n+1)$, $\dot{f}_0(\varsigma') = f_0(\Delta - x')$. For the matter of convenience we accept that $l_\Delta^k + j\Delta = 0$ for some $j \in \mathbb{Z}$. Therefore $f_{\mathcal{N}}[\varsigma_n' - \varsigma', 0, \sigma_n] = f_{\mathcal{N}}[(\varsigma_n' - 2\Delta(n+1-j)) - (-l_\Delta^k + (x' - \Delta)), 0, \sigma_n]$. Also $-(\mathbf{O} + 2\Delta(n+1-j)) = \mathbf{O}$. The property of pdf of AWGN provides that $f_{\mathcal{N}}[y, 0, \sigma_n] = f_{\mathcal{N}}[-y, 0, \sigma_n]$ and, consequently,

$$\begin{aligned}
&f_{\mathcal{N}}\left[\left(\varsigma_n' - 2\Delta\left(n+1-j\right)\right)\right. \\
&\qquad \left. - \left(-l_\Delta^k + \left(x' - \Delta\right)\right), 0, \sigma_n\right] \\
&= f_{\mathcal{N}}\left[-\left(\varsigma_n' - 2\Delta\left(n+1-j\right)\right)\right. \\
&\qquad\qquad \left. - \left(l_\Delta^k + \left(\Delta - x'\right)\right), 0, \sigma_n\right].
\end{aligned} \tag{44}$$

Using the latest equation we derive that

$$\begin{aligned}
&\int_{\mathbf{O}} \int_{I_{k+2n+1}} \dot{f}_0\left(\varsigma'\right) f_{\mathcal{N}}\left[\varsigma_n' - \varsigma', 0, \sigma_n\right] d\varsigma' d\varsigma_n' \\
&= \int_{-(\mathbf{O}+2\Delta(n+1-j))} \int_0^{\Delta} f_0\left(\Delta - x'\right) \\
&\qquad \times f_{\mathcal{N}}\left[-\left(\varsigma_n' - 2\Delta\left(n+1-j\right)\right)\right. \\
&\qquad\qquad \left. - \left(l_\Delta^k + \left(\Delta - x'\right)\right), 0, \sigma_n\right] \\
&\qquad \times d\left\{\Delta - x'\right\} \\
&\qquad \times d\left\{-\left(\varsigma_n' - 2\Delta\left(n+1-j\right)\right)\right\},
\end{aligned} \tag{45}$$

and we prove the statement.

## 4. Experimental Results

In this section we describe conditions, procedure, and results of two different kinds of experiments based on analytic estimation of capacity as well as simulations. The preferred index of attack severity is WNR (indexes $\sigma_n$ and quality of JPEG compression are also used). For a given set of embedding parameters, the error rates and capacity are estimated differently using different models suitable for each kind of experiment. However, for both kinds of experiment, the maximum capacity for a given level of attack severity is found by using brute force search in the space of all adjustable parameters.

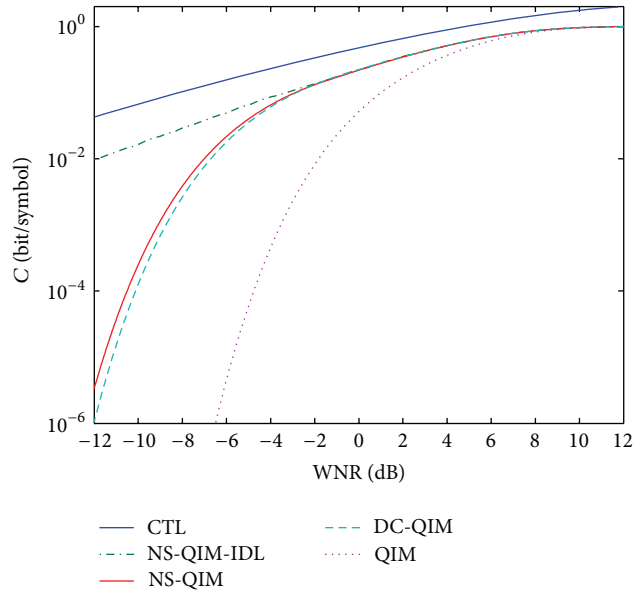



Figure 3: Analytic-based estimation of capacity under AWGN.

*4.1. Analytic Estimation of Watermarking Performance under AWGN.* In this subsection of our experiment we use $\sigma_n = 1$. Parameters $\alpha, \beta, \eta_1, \gamma_0, \vartheta_0$, and $\varphi_1$ are subjects to constraints (21), (23), $\eta_1 + \gamma_0 = 0.5$, and $\vartheta_0 + \varphi_1 = 0.5$ and the simulations are repeated for each new value of WNR. Then, the length of embedding interval $\Delta$ is calculated according to (34). Error rates are calculated according to (41).

We use two variants of the proposed quantization scheme with adjustable parameters: nonsymmetric QIM (NS-QIM) and nonsymmetric QIM with IDL (NS-QIM-IDL). Such a decision can be explained by a consideration that IDL is acceptable for some application, but other applications may require all the watermark data to be embedded correctly.

In Figure 3 the plots for channel capacity toward WNR are shown for two variants of the proposed method as well as DC-QIM and QIM [9]. The permanent thresholding $\mathrm{Th} = \Delta(\beta - 0.5\alpha)$ is applied to NS-QIM and NS-QIM-IDL. As a reference, Costa theoretical limit (CTL) [5] is plotted in Figure 3:

$$\mathrm{CTL} = \frac{1}{2}\log_2\left(1 + 10^{0.1*\mathrm{WNR}}\right). \tag{46}$$

Capacity is calculated analytically according to the description provided in the literature for DC-QIM and QIM.

During the estimation, the subsets $\widetilde{\mathbf{Z}} \subset \mathbf{Z}$ and $\widetilde{\mathbf{O}} \subset \mathbf{O}$ were used instead of $\mathbf{Z}$ and $\mathbf{O}$:

$$\widetilde{\mathbf{Z}} = \bigcup_{m=-100}^{100} \left[2\Delta m + l_\Delta^k - \mathrm{Th}, 2\Delta m + l_\Delta^k + \mathrm{Th}\right),$$
$$\widetilde{\mathbf{O}} = \bigcup_{m=-100}^{100} \left[2\Delta m + l_\Delta^k + \mathrm{Th}, 2\Delta\,(m+1) + l_\Delta^k - \mathrm{Th}\right). \quad (47)$$

Therefore, for such estimation we assume that quantized coefficients from the $k$th interval after AWGN are distributed only inside $[-200\Delta + l_\Delta^k - \mathrm{Th}, 202\Delta + l_\Delta^k - \mathrm{Th})$. The assumption is a compromise between computational complexity and the fidelity of the result.

As can be seen from Figure 3 both variants of the proposed method perform better than DC-QIM for WNR values less than −2 dB and, obviously, much higher capacity provided by DC-QIM-IDL is compared to the other methods in that range. Taking into account that DC-QIM provides the highest capacity under AWGN compared to the other known in the literature methods [12, 19], newly proposed method DC-QIM-IDL fills an important gap. Reasonably, the demonstrated superiority is mostly due to IDL.

*4.2. Watermarking Performance in Simulation Based Experiments without GA.* The advantage of analytic estimation of error rates according to (41) is that the stage of watermark embedding can be omitted and host signal is not required. The practical limitation of the approach is that $\widetilde{\mathbf{Z}}$ and $\widetilde{\mathbf{O}}$ are just subsets of $\mathbf{Z}$ and $\mathbf{O}$, respectively. Other disadvantages are that estimation might become even more complex in case the threshold position is optimized depending on the level of noise; only rates for AWGN can be estimated, but there are other kinds of popular attacks [24]. Therefore in this subsection we will also simulate watermarking experiments using real host signals.

*4.2.1. Conditions for Watermark Embedding and Extraction.* In case of experiments with real signals the parameters of the proposed watermarking scheme must satisfy some other constraints instead of (34). However, constraints (21), (23), $\eta_1 + \gamma_0 = 0.5$, and $\vartheta_0 + \varphi_1 = 0.5$ remain the same as in the analytic based experiment.

Some lower limit of DWR has to be satisfied for watermarked host, which assures acceptable visual quality. DWR is calculated according to

$$\mathrm{DWR} = 10\log_{10}\left(\frac{\sigma_H^2}{D}\right), \quad (48)$$

where $\sigma_H^2$ is the variance of the host.

Therefore, using (33) the equation for $\Delta$ in that case is

$$\Delta = \frac{\sigma_H}{\sqrt{(Q_0 + Q_1)\,10^{0.1\mathrm{DWR}}}}. \quad (49)$$

In contrast to analytic based experiment, $\sigma_n$ should be adjusted for different severity of the attack and is defined as

$$\sigma_n^2 = \frac{\sigma_H^2}{10^{0.1(\mathrm{DWR}+\mathrm{WNR})}}. \quad (50)$$

After watermark is embedded and AWGN with $\sigma_n^2$ is introduced we perform extraction and calculate channel capacity.

A variant NSC-QIM with constant (nonadjustable) parameters is also used in some experiments. The intention to adjust the parameters in order to maximize capacity is natural. However, maximization requires information about WNR to be known before watermark embedding and transmission. In some application areas level of noise (or severity of an attack) might change over time or remain unknown. Therefore watermark should be embedded with some constant set of parameters depending on expected WNR.

Different positions of the threshold can be used to extract a watermark. An optimal position of the threshold is not obvious. Placing the threshold in the middle of the interval might be inefficient because the distribution of quantized samples inside embedding interval is nonsymmetric. Two kinds of thresholding are proposed: permanent and nonpermanent. The permanent position is $\mathrm{Th} = \Delta(\beta - 0.5\alpha)$ for the intervals with numbers $k + 2m$, $m \in \mathbb{Z}$. The name "permanent" is because Th cannot be changed after embedding. Its position depends only on $\alpha$, $\beta$, and $\Delta$ and does not depend on the parameters of attack.

The nonpermanent position of Th is the median of the distribution inside each interval. Nonpermanent position may depend on the type and severity of a noise. The advantage of nonpermanent Th is that extraction of a watermark can be done without information about $\alpha$ and $\beta$.

*4.2.2. Watermarking Performance for AWGN and JPEG Attacks without GA.* The performance of the proposed method was evaluated using real host signals. For that purpose we used 87 natural grayscale images with resolution 512 × 512. Each bit of a watermark was embedded by quantizing the first singular value of SVD of 4 × 4 block. This kind of transform is quite popular in digital image watermarking and the chosen block size provides a good tradeoff between watermark data payload and robustness [7, 25]. The value of DWR was 28 dB. An attack of AWGN was then applied to each watermarked image. The resulting capacity toward noise variance is plotted for different methods in Figure 4.

It can be seen that the resulting capacity after AWGN attack is the highest for NS-QIM. The other two methods whose performance is quite close to NS-QIM are DC-QIM and FZDH. Compared to DC-QIM the advantage is more obvious for higher variance. However, for moderate variance the advantage is more obvious compared to FZDH.

Methods QIM and RDM do not have parameters that can be adjusted to different variance. Under some circumstances adjustment is not feasible for NS-QIM as well. We have chosen constant parameters $\alpha = 0.05$ and $\beta = 0.35$ for NSC-QIM in order to provide a fair comparison with QIM and RDM. The plots for NSC-QIM, QIM and RDM are marked

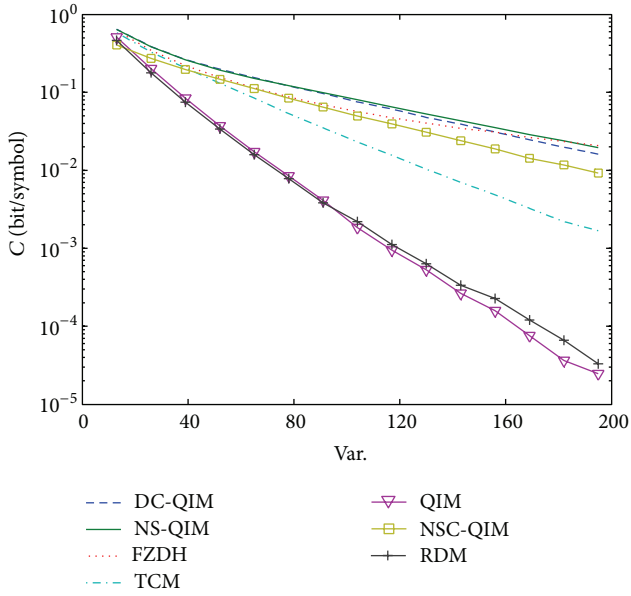



Figure 4: Capacity under AWGN for natural grayscale images.

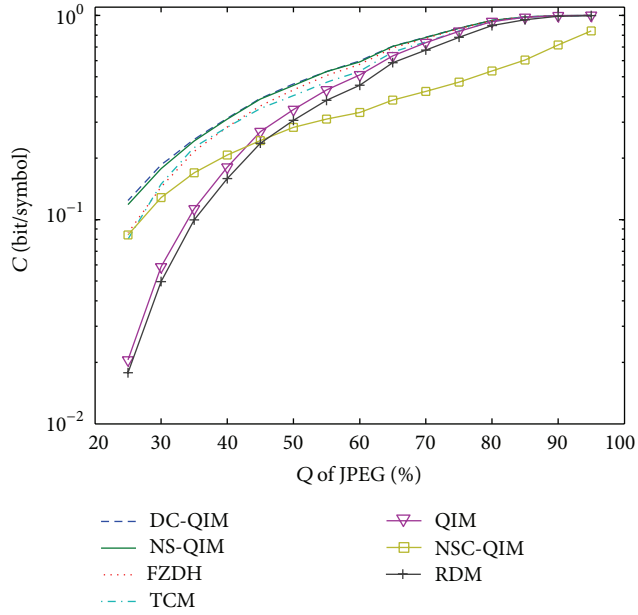



Figure 5: Capacity under JPEG for natural grayscale images.

by squares, triangles, and crosses, respectively, in Figure 4. As can be seen, NSC-QIM performs considerably better than QIM and RDM and the advantage is especially noticeable for higher noise variance.

Other image processing techniques except additive noise are able to destroy a watermark and one of them is JPEG compression which is quite popular. The capacity of the proposed watermarking method was also compared with other methods and the procedure of embedding was the same as in AWGN case. However, this time JPEG compression with different levels of quality was considered as an attack. The results are plotted in Figure 5.

According to the plots in Figure 5, the performance of NS-QIM in general is very close to that of DC-QIM but is slightly worse for low $Q$ factor. The methods FZDH and TCM provide lower capacity than NS-QIM and DC-QIM but in general are quite close to them. The worst performance is demonstrated by QIM and RDM and the disadvantage is especially noticeable for low Q. For NSC-QIM with $\alpha = 0.05$ and $\beta = 0.35$ the performance is considerably better than that for QIM and RDM under low Q but is worse for higher quality of JPEG compression.

*4.3. Procedure for GA Recovery.* It has been demonstrated that for some popular types of attack the performance of NS-QIM is comparable or better than that of DC-QIM. The mentioned DC-QIM is considered to be one of the best quantization methods for watermarking, but it is extremely vulnerable to GA. On the other hand the performance of RDM is not as good under AWGN and JPEG attacks and is comparable to that of QIM. In this subsection, we propose a procedure for GA recovery in order to fill an important gap in the literature and introduce a watermarking method that provides high efficiency under AWGN as well as GA. The procedure utilizes features that are unique for the proposed approach and have not been discussed in the field of watermarking before.

We are proposing several criteria that will be used by the procedure to provide robustness against GA for NS-QIM. The criteria exploit nonsymmetric distribution inside embedding interval and help to recover a watermarked signal after the attack. It is presumed that a constant gain factor is applied to the watermarked signal (followed by AWGN) and the task is either to estimate the factor or the resulting length of embedding interval.

Let us denote the actual gain factor by $\lambda$ and our guess about it by $\lambda'$. The length of the embedding interval (which is optimal for watermark extraction) is modified as a result of GA and is denoted by $\widetilde{\Delta} = \lambda\Delta$. Our guess about $\widetilde{\Delta}$ is $\widetilde{\Delta}' = \lambda'\Delta$.

The core of the procedure of recovery after GA is the following. For each particular value $\widetilde{\Delta}'$, noisy quantized samples $\varsigma_n'$ are being projected on a single embedding interval:

$$x_n' = \begin{cases} \varsigma_n' \bmod \widetilde{\Delta}', & \text{if } \left\lfloor \dfrac{\varsigma_n' - l_\Delta^k}{\widetilde{\Delta}'} \right\rfloor \bmod 2 = 0, \\ \widetilde{\Delta}' - \left(\varsigma_n' \bmod \widetilde{\Delta}'\right), & \text{otherwise.} \end{cases} \tag{51}$$

One of the following criteria is being applied to the random variable $X_n' \in [0, \widetilde{\Delta}']$:

$$C_1\left(\widetilde{\Delta}'\right) = \left| \frac{\operatorname{median}\left(X_n'\right)}{\widetilde{\Delta}'} - 0.5 \right|,$$
$$C_2\left(\widetilde{\Delta}'\right) = \left| \frac{E\left(\left[X_n'\right]^w\right)}{\left[\widetilde{\Delta}'\right]^w} \right|, \quad w = 2m+1,\ m \in \mathbb{N}. \tag{52}$$

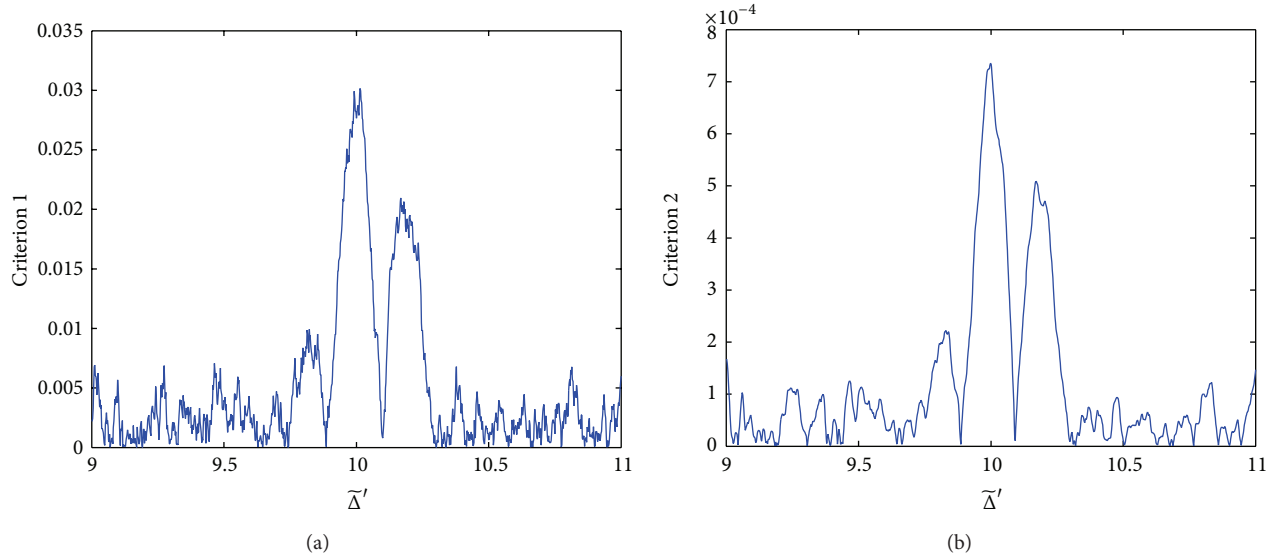



FIGURE 6: Plots of criteria values toward guessed length of embedding interval: (a) criterion $C_1$; (b) criterion $C_2$.

The value of $\widetilde{\Delta}''$ that maximizes one of the proposed criteria should be chosen as the best estimate of $\widetilde{\Delta}$:

$$\widetilde{\Delta}'' = \arg\max_{\widetilde{\Delta}'} C_{1,2}\left(\widetilde{\Delta}'\right). \tag{53}$$

The intuition behind the proposed procedure of recovery from GA is the following. The variance of the coefficients of the host signal is much larger than the length of embedding interval. Embedding intervals are placed next to each other without gaps and even small error in estimation of $\widetilde{\Delta}$ results in considerable mismatch between positions of samples inside corresponding embedding intervals. In other words, wrong assumption about $\widetilde{\Delta}$ makes distribution of $X'_n$ very close to uniform. However, in case $\widetilde{\Delta}'$ is close to $\widetilde{\Delta}$ the distribution of $X'_n$ demonstrates asymmetry because the distribution of quantized samples inside embedding interval (before GA is introduced) is indeed asymmetric. Hence, criteria $C_1$ and $C_2$ are just measures of asymmetry. The main advantage of the procedure is simplicity and low computational demand.

Experimental results demonstrate high level of accuracy of the proposed procedure of recovery after GA. Grayscale image Lena.tif with dimension $512 \times 512$ was used as a host signal for that purpose. A random watermark sequence was embedded into the largest singular values of SVD of $4 \times 4$ blocks using NS-QIM with $\alpha = 0.05$ and $\beta = 0.35$. The AWGN attack was applied after the embedding so that WNR $= -5$ dB. The length of embedding interval was 10. However we use notation $\widetilde{\Delta} = 10$ because the value is not known to the receiver and during watermark extraction the proposed recovery procedure was used. The interval of initial guess was $\widetilde{\Delta} \pm 10\%$ so that $\widetilde{\Delta}' \in [9, 11]$. Such an initial guess reflects real needs for recovery after GA because a gain factor that is outside the range 0.9~1.1 causes considerable visual distortions in most cases. The initial guess interval was split by equally spaced 1000 steps and for each step the recovery procedure was applied. The plots for values of $C_1$ and $C_2$, $w = 5$, toward guessed values of $\widetilde{\Delta}$, are shown in Figures 6(a) and 6(b), respectively.

Despite the fact that for the same $\Delta$ the difference between values of $C_1$ and $C_2$ is huge, the shapes of the plots are similar. The criteria reach their maximum at 10.042 and 9.998 for $C_1$ and $C_2$, respectively, which are quite precise estimates of the actual $\Delta$ used during watermark embedding.

*4.4. Performance for AWGN and JPEG Attacks with GA.* The embedding constraints for the current experiment are the same as described in Section 4.2.1. Among the quantization methods used for comparison the only method robust to GA is RDM. Therefore, only RDM was used as a reference to NS-QIM and NSC-QIM under GA followed by AWGN and JPEG attacks, respectively. The exact information about $\Delta$ was not used for extraction in NS-QIM and NSC-QIM cases which is equivalent to GA with unknown scaling factor.

The watermark embedding domain was the same as in previous tests: first singular values of SVD of $4 \times 4$ blocks from $512 \times 512$ grayscale images were quantized, DWR $= 28$ dB. In case of RDM, the quantized value of a particular coefficient is based on the information about the last 100 previous coefficients. For NSC-QIM the parameters of embedding were $\alpha = 0.05$ and $\beta = 0.35$. For both AWGN and JPEG attacks the same as previously ranges of parameters were used.

However, during watermark extraction no information except initial guess interval $\Delta \pm 10\%$ was used in NS-QIM and NSC-QIM cases. Criterion $C_1$ was used for the estimation of actual $\Delta$. Nonpermanent thresholding was applied to both modifications of the proposed watermarking method. In contrast to that RDM does use the exact information about quantization step. The resulting capacity toward AWGN variance is plotted for each method in Figure 7.

It can be seen from Figure 7 that both NS-QIM and NSC-QIM outperform RDM. The advantage of the proposed method is more evident for larger variance of the noise.

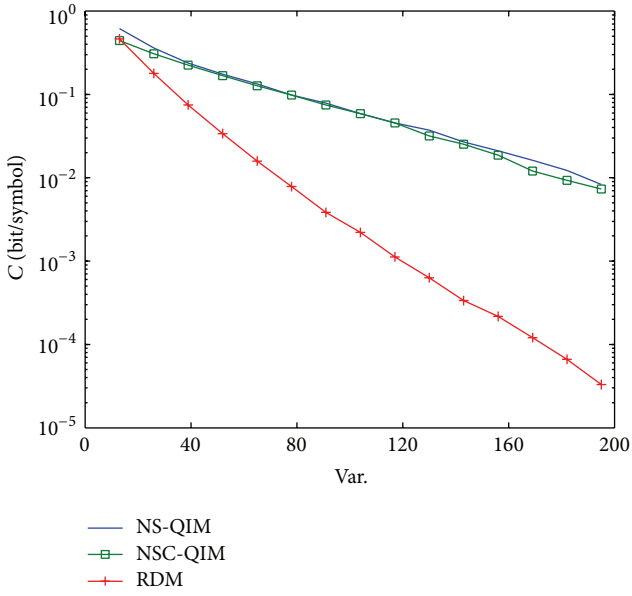



Figure 7: Capacity under GA followed by AWGN.

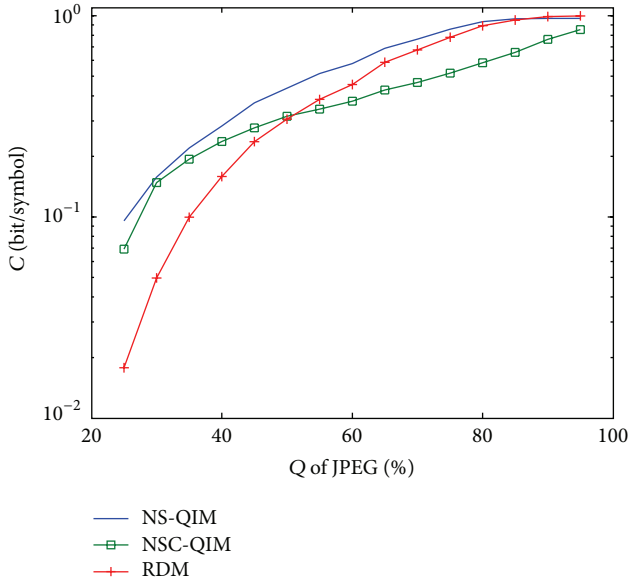



Figure 8: Capacity under GA followed by JPEG compression.

The capacity plots for NS-QIM, NSC-QIM, and RDM in case of JPEG attack are shown in Figure 8.

From Figure 8 we can conclude that both modifications of the proposed watermarking method supply higher capacity than RDM when $Q < 50\%$. However, only NS-QIM outperforms RDM in case $Q > 50\%$ and NSC-QIM performs worse than RDM for that range.

## 5. Discussion

In the experiment section we have estimated the capacity of the proposed method in both analytical and empirical ways. Following both ways we can witness that the proposed method provides higher capacity compared to the other reference methods. In this section we are to discuss in more detail measures of watermarking efficiency, conditions of the experiments, and the reasons of superiority of NS-QIM-IDL.

Channel capacity $C$ is one of the most important measures for watermarking as it indicates the maximum amount of the information that can be transmitted by a single embedded symbol [1, 12]. However, some authors in their original papers refer to error rates instead [13, 16, 19–21]. It can be demonstrated that calculations of $C$ using error rates are straightforward [26]. Capacity can be calculated according to the following expression:

$$\begin{aligned} C = \max_{p_{\text{em}}(\sim b)} \Bigg[ & p(\sim b, b) \log_2 \left( \frac{p(\sim b, b)}{p_{\text{em}}(\sim b)\, p_{\text{ex}}(b)} \right) \\ & + p(b, \sim b) \log_2 \left( \frac{p(b, \sim b)}{p_{\text{em}}(b)\, p_{\text{ex}}(\sim b)} \right) \\ & + p(\sim b, \sim b) \log_2 \left( \frac{p(\sim b, \sim b)}{p_{\text{em}}(\sim b)\, p_{\text{ex}}(\sim b)} \right) \\ & + p(b, b) \log_2 \left( \frac{p(b, b)}{p_{\text{em}}(b)\, p_{\text{ex}}(b)} \right) \Bigg], \end{aligned} \tag{54}$$

where, for instance, $p(\sim b, b)$ denotes joint probability of embedding symbol $\sim b$ and extracting symbol $b$; $p_{\text{em}}(b)$ and $p_{\text{ex}}(b)$ denote probabilities of embedding and extracting of symbol $b$. Probabilities of extracting a particular symbol can be calculated using joint probabilities:

$$\begin{aligned} p_{\text{ex}}(b) &= p(\sim b, b) + p(b, b), \\ p_{\text{ex}}(\sim b) &= p(b, \sim b) + p(\sim b, \sim b). \end{aligned} \tag{55}$$

Joint probabilities can be expressed using $p_{\text{em}}(\cdot)$ and error rates:

$$\begin{aligned} p(\sim b, b) &= p_{\text{em}}(\sim b)\, \text{BER}_{\sim b}, \\ p(b, \sim b) &= p_{\text{em}}(b)\, \text{BER}_b, \\ p(\sim b, \sim b) &= p_{\text{em}}(\sim b)\left(1 - \text{BER}_{\sim b}\right), \\ p(b, b) &= p_{\text{em}}(b)\left(1 - \text{BER}_b\right). \end{aligned} \tag{56}$$

Embedding probabilities for the methods proposed in this paper are

$$\begin{aligned} p_{\text{em}}(\sim b) &= \gamma_0 + \vartheta_0, \\ p_{\text{em}}(b) &= \eta_1 + \varphi_1. \end{aligned} \tag{57}$$

As a contrast to the watermarking approach proposed in this paper, the QIM-based methods known in the literature assume equal embedding probabilities and provide equal error rates for "0" and "1" [12, 19]. For all the mentioned in the experimental section methods (QIM, DC-QIM, RDA, FZDH, TCM, and the proposed methods) the results were collected under equal conditions of each kind of attack. In

order to compare efficiency of the proposed methods with some other state-of-the-art papers in watermarking [13, 21], their channel capacity can be calculated based on the data provided in those papers. From (54)–(56) we derive that QIM-based watermarking which has been presented in the literature capacity is

$$C = 1 + \mathrm{BER}\log_2(\mathrm{BER}) + (1 - \mathrm{BER})\log_2(1 - \mathrm{BER}). \quad (58)$$

The largest singular values of SVD of $4 \times 4$ blocks were used by all the methods for watermark embedding in the empirical estimations of capacity. Such a domain is a natural choice for many watermarking applications because it provides a good tradeoff between robustness, invisibility, and data payload [7, 27, 28]. Commonly, the largest singular values are being quantized [25]. The robustness of a watermark embedded in the domain can be explained by a consideration that the largest singular values have a great importance. For example, compared to a set of the coefficients of discrete cosine transform (DCT) the set of singular values has more compact representation for the same size of a segment of an image [29]. At the same time the block size of $4 \times 4$ is enough to avoid some visible artefacts and this guarantees invisibility under DWR = 28 dB. The data payload of 1 bit per 16 pixels is sufficient for inclusion of important copyright information and for image size $512 \times 512$ provides capacity of 2 kB.

Among the reference (and state of the art) methods used for comparison no one performs better than the proposed watermarking methods simultaneously under both AWGN and GA. Hence, the proposed methods fill the gap existing in watermarking literature. This is thanks to several new advancements used for embedding and extraction of a watermark.

In the case when AWGN is applied at the absence of GA the benefit is caused mostly by IDL and the kind of thresholding during watermark extraction. From Figure 3 it can be noticed that even without IDL variant NS-QIM delivers slightly higher capacity under low WNRs compared to DC-QIM. However, the capacity rises dramatically for low WNRs if we switch to NS-QIM-IDL. It is remarkable that the form of capacity plot in the latter case does not inherit the steepness demonstrated by the other methods. Instead, the plot shape is similar to CTL but is placed at a lower position. The explanation of such phenomena is in the quantization process. According to IDL we refuse to modify samples whose quantization brings the highest embedding distortion. In case these samples are quantized they are placed closer to the threshold which separates "0" and "1." Therefore the information interpreted by these samples is the most likely to be lost under low WNRs. Predicting the loss of information we might accept that fact and introduce IDL instead. It is a kind of "accumulation" of embedding distortion which can be "spent" on making the rest of embedded information more robust. Another unique feature is the proposed way of nonpermanent thresholding. In contrast to the permanent thresholding the information about $\alpha, \beta$ is not required for watermark extraction. Hence, during embedding these parameters can be adjusted to deliver higher capacity even in case there is no way to communicate new parameters to the receiver.

The proposed method is in advantageous position compared to RDM in the case when GA is used to attack the watermarked image. As one of its stages, GA assumes AWGN and this explains superiority of NS-QIM over RDM in general. The success of recovery is due to easy and efficient procedure that utilizes a unique feature introduced by the proposed methods. The feature is created during quantization and is a result of different quantization rules for "0" and "1."

The proposed estimation of scaling factor in this paper has some advantages compared to other known retrieving procedures. For instance, a model of a host is used in [15] to estimate the scaling factor. In contrast to that we exploit the unique asymmetric feature of the proposed quantization approach and this feature is not dependent on a host. The only important assumption about the host is that its variance is much larger than the size of embedding interval. As soon as this holds the estimation is not dependent on the model of the host which is a contrast to [15]. Also, our recovery procedure does not use any additional information except interval guess for $\widetilde{\Delta}$, which can be given roughly. These improvements imply more efficient retrieval after GA which in addition requires fewer samples.

The nonpermanent thresholding was proposed with the aim to avoid transmitting any additional information to the receiver. For example, different size of embedding interval $\Delta$ and different parameters $\alpha, \beta$ can be used to watermark different images. Nevertheless, a watermark can be extracted in case the recovery procedure and nonpermanent thresholding are used. Such feature might be beneficial in adaptation to the conditions that change.

In the paper we do not consider a constant offset attack. In some other papers like [12, 14, 19] it is assumed to be applied in conjunction with GA. Further modifications of the proposed recovery procedure are needed to cope with it. Also, another criterion that exploits different features compared $C_1$ and $C_2$ might be useful for that task. Apart from this goal we would like to experiment with other concepts of IDL. For example, it might be reasonable to allow for those samples to be shifted during quantization procedure. Such shifts may increase chances for those samples to be interpreted correctly after an attack is applied.

## 6. Conclusions

The new watermarking method based on scalar QIM has been proposed. It provides higher capacity under different kinds of attacks compared to other existing methods. The proposed NS-QIM-IDL method is the most beneficial in case of GA and AWGN. The advantages of the method are due to its unique approach to watermark embedding as well as a new procedure of recovery and extraction.

The main features of the unique approach to watermark embedding are a new kind of distribution of quantized samples and IDL. In general there is no line of symmetry inside embedding interval for the new distribution of quantized samples. This feature is used to recover a watermark after GA. The feature of IDL can reduce distortions introduced to a host signal which are caused by watermarking. This is done by letting some watermark bits to be interpreted

incorrectly at the initial phase of embedding and before any attack occurs. The proposed IDL is extremely beneficial for low WNRs under AWGN attack.

The new procedure of recovery after GA exploits the nonsymmetric distribution of quantized samples. One out of two different criteria might be chosen to serve as a goal function for the procedure. The criteria behave in a similar way despite the differences in realization. It has been demonstrated experimentally that the proposed recovery procedure estimates the original length of embedding interval with deviation of 0.02% even in case when WNR is quite low. Nonpermanent thresholding was proposed in order to avoid transmitting additional information to the site where watermark extraction is done. The technique is simple and establishes the threshold in the position of the median of the distribution inside embedding interval.

The mentioned advancements implied considerable performance improvement. Under conditions of AWGN and JPEG attacks (at the absence of GA) the capacity of the proposed method is at the same or higher level compared to DC-QIM. The most advantageous application of NS-QIM-IDL is under AWGN for WNRs around −12 dB where it performs up to $10^4$ times better than DC-QIM. Under the condition of GA followed by high level of AWGN the performance of the proposed method is up to $10^3$ times higher than that of RDM. For the case when GA is followed by JPEG with $Q = 25\%$ the capacity of the proposed method is up to 10 times higher than that of RDM. Superiority of the proposed methods under AWGN as well as GA allows narrowing the gap between watermarking performances achievable in theory and in practice.

## Conflict of Interests

The authors declare that there is no conflict of interests regarding to the publication of this paper.

## References


[1] I. Cox, M. Miller, J. Bloom, J. Fridrich, and T. Kalker, *Digital Watermarking and Steganography*, Morgan Kaufmann, San Francisco, Calif, USA, 2nd edition, 2007.

[2] M. Barni, F. Bartolini, V. Cappellini, and A. Piva, "Robust watermarking of still images for copyright protection," in *Proceedings of the 13th International Conference on Digital Signal Processing (DSP '97)*, vol. 2, pp. 499–502, Santorini, Greece, July 1997.

[3] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Transactions on Image Processing*, vol. 15, no. 2, pp. 430–444, 2006.

[4] T. Chen, "A framework for optimal blind watermark detection," in *Proceedings of the 2001 Workshop on Multimedia and Security: New Challenges*, pp. 11–14, Ottawa, Canada, 2001.

[5] M. H. M. Costa, "Writing on dirty paper," *IEEE Transactions on Information Theory*, vol. 29, no. 3, pp. 439–441, 1983.

[6] E. Ganic and A. M. Eskicioglu, "Robust DWT-SVD domain image watermarking: embedding data in all frequencies," in *Proceedings of the Multimedia and Security Workshop (MM & Sec '04)*, pp. 166–174, September 2004.

[7] K. Loukhaoukha, "Image watermarking algorithm based on multiobjective ant colony optimization and singular value decomposition in wavelet domain," *Journal of Optimization*, vol. 2013, Article ID 921270, 10 pages, 2013.

[8] B. Chen and G. Wornell, "Dither modulation: a new approach to digital watermarking and information embedding," in *Security and Watermarking of Multimedia Contents*, vol. 3657 of *Proceedings of SPIE*, pp. 342–353, April 1999.

[9] B. Chen and G. W. Wornell, "Quantization index modulation: a class of provably good methods for digital watermarking and information embedding," *IEEE Transactions on Information Theory*, vol. 47, no. 4, pp. 1423–1443, 2001.

[10] E. Esen and A. Alatan, "Forbidden zone data hiding," in *Proceedings of the IEEE International Conference on Image Processing*, pp. 1393–1396, October 2006.

[11] M. Ramkumar and A. N. Akansu, "Signalling methods for multimedia steganography," *IEEE Transactions on Signal Processing*, vol. 52, no. 4, pp. 1100–1111, 2004.

[12] J. J. Eggers, R. Bäuml, R. Tzschoppe, and B. Girod, "Scalar Costa scheme for information embedding," *IEEE Transactions on Signal Processing*, vol. 51, no. 4, pp. 1003–1019, 2003.

[13] J. Oostveen, T. Kalker, and M. Staring, "Adaptive quantization watermarking," in *Security, Steganography, and Watermarking of Multimedia*, Proceedings of SPIE, pp. 296–303, San Jose, Calif, USA, January 2004.

[14] X. Kang, J. Huang, and W. Zeng, "Improving robustness of quantization-based image watermarking via adaptive receiver," *IEEE Transactions on Multimedia*, vol. 10, no. 6, pp. 953–959, 2008.

[15] I. D. Shterev and R. L. Lagendijk, "Amplitude scale estimation for quantization-based watermarking," *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 4146–4155, 2006.

[16] F. Pérez-González, C. Mosquera, M. Barni, and A. Abrardo, "Rational dither modulation: a high-rate data-hiding method invariant to gain attacks," *IEEE Transactions on Signal Processing*, vol. 53, no. 10, pp. 3960–3975, 2005.

[17] F. Ourique, V. Licks, R. Jordan, and F. Pérez-González, "Angle QIM: a novel watermark embedding scheme robust against amplitude scaling distortions," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '05)*, pp. 797–800, March 2005.

[18] M. Zareian and H. Tohidypour, "Robust quantisation index modulation-based approach for image watermarking," *IET Image Processing*, vol. 7, no. 5, pp. 432–441, 2013.

[19] X. Zhu and J. Ding, "Performance analysis and improvement of dither modulation under the composite attacks," *Eurasip Journal on Advances in Signal Processing*, vol. 2012, no. 1, article 53, 2012.

[20] M. A. Akhaee, S. M. E. Sahraeian, and C. Jin, "Blind image watermarking using a sample projection approach," *IEEE Transactions on Information Forensics and Security*, vol. 6, no. 3, pp. 883–893, 2011.

[21] N. K. Kalantari and S. M. Ahadi, "A logarithmic quantization index modulation for perceptually better data hiding," *IEEE Transactions on Image Processing*, vol. 19, no. 6, pp. 1504–1517, 2010.

[22] E. Nezhadarya, J. Wang, and R. K. Ward, "A new data hiding method using angle quantization index modulation in gradient domain," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '11)*, pp. 2440–2443, Prague, Czech Republic, May 2011.

[23] M. Zareian and A. Daneshkhah, "Adaptive angle quantization index modulation for robust image watermarking," in *Proceedings of the IEEE Global Communications Conference (GLOBECOM '12)*, pp. 881–884, Anaheim, Calif, USA, December 2012.

[24] C. Song, S. Sudirman, M. Merabti, and D. Llewellyn-Jones, "Analysis of digital image watermark attacks," in *Proceeding of the 7th IEEE Consumer Communications and Networking Conference (CCNC '10)*, pp. 1–5, Las Vegas, Nev, USA, January 2010.

[25] V. Gorodetski, L. Popyack, V. Samoilov, and V. Skormin, "SVD-based approach to transparent embedding data into digital images," in *Proceedings of the International Workshop on Information Assurance in Computer Networks: Methods, Models, and Architectures for Network Security (MMM-ACNS '01)*, pp. 263–274, 2001.

[26] R. Gallager, *Information Theory and Reliable Communication*, John Wiley & Sons, New York, NY, USA, 1968.

[27] Y. Zolotavkin and M. Juhola, "A new blind adaptive watermarking method based on singular value decomposition," in *Proceedings of the International Conference on Sensor Network Security Technology and Privacy Communication System (SNS and PCS '13)*, pp. 184–192, Nangang, China, March 2013.

[28] Y. Zolotavkin and M. Juhola, "SVD-based digital image watermarking on approximated orthogonal matrix," in *Proceedings of the 10th International Conference on Security and Cryptography (SECRYPT '13)*, pp. 321–330, July 2013.

[29] X. Jun and W. Ying, "Toward a better understanding of DCT coefficients in watermarking," in *Proceedings of The Pacific-Asia Workshop on Computational Intelligence and Industrial Application (PACIIA '08)*, vol. 2, pp. 206–209, Wuhan, China, December 2008.

Publication V

# Quantization Based Watermarking Approach with Gain Attack Recovery

Yevhen Zolotavkin and Martti Juhola

# Quantization Based Watermarking Approach with Gain Attack Recovery

Yevhen Zolotavkin
Computer Science, School of Information Sciences
University of Tampere
Tampere, Finland
yevhen.zolotavkin@uta.fi

Martti Juhola
Computer Science, School of Information Sciences
University of Tampere
Tampere, Finland
martti.juhola@sis.uta.fi

***Abstract*— A new Quantization Index Modulation -based watermarking approach is proposed in this paper. With the aim to increase capacity of the watermarking channel with noise we propose Initial Data Loss during quantization for some samples in pre-defined positions. Also, the proposed approach exploits a new form of distribution of quantized samples where samples that interpret "0" and "1" have differently shaped probability density functions. This creates a distinctive feature which is expressed numerically using one out of two proposed criteria. The criteria are utilized by a procedure for recovery after possible Gain Attack. Several state of the art quantization-based watermarking methods were used for comparison on a set of natural grayscale images. The superiority of the proposed method has been confirmed for different types of popular attacks.**

***Keywords*— *Quantization Index Modulation; Digital Image Watermarking; Gain Attack, Additive White Gaussian Noise; Initial Data Loss***

## I. INTRODUCTION

Multimedia plays important role in communications. In some cases ownership of multimedia data is important and needs to be protected. Digital images form considerable fraction of popular multimedia content. The task of Digital Image Watermarking (DIW) is to protect the digital rights of an owner and a watermark is being inserted in an image for that purpose. The watermark needs to be robust [1] as well as invisible [2].

Quantization Index Modulation (QIM) is one of the popular and efficient techniques in DIW. Methods that belong to QIM are widely used in blind watermarking where neither original media nor watermark is known to the receiver [3]. Usually robustness is evaluated by attacking a watermarked image with Additive White Gaussian Noise (AWGN). The limit of the channel capacity which is achievable by QIM under AWGN was first derived in [4]. Nevertheless, all the known on practice implementations of QIM are far from achieving the capacity limit.

Many different QIM-related approaches are known. The simplest scalar realization of QIM is to replace all the coefficient values from a certain interval by a single value equal either to the left or right endpoint depending on a bit of a watermark [5]. The obvious disadvantage is that the capacity of the simplest QIM for high intensity of noise is much lower than the limit. The idea behind more advanced realization of DC-QIM is to replace each coefficient value from an original interval by a corresponding value taken from one out of two disjoint intervals that are inside of the original one [6]. A parameter $0.5(1-\alpha)$, $0 \le \alpha \le 1$, is to control the size of each subinterval relatively to the original. The distribution for quantized samples "0" and "1" is uniform in that case. Parameter $\alpha$ is being adjusted depending on noise level in order to maximize capacity.

Forbidden Zone Data Hiding (FZDH) modifies only a fraction (controlled by $\alpha$) of coefficient values in each interval of original values [7]. Another idea was proposed by the authors of Thresholded Constellation Modulation (TCM) that use two different quantization rules to modify coefficients inside the original interval [8].

The main advantage of the techniques based on QIM is robustness against AWGN, but the limitation is that synchronization is required to extract a watermark. Gain Attack (GA) performs a constant scaling of the whole sequence of watermarked coefficients. The scaling factor might be close to 1 and cause very little visual distortion, but it is unknown to the receiver which causes asynchronous extraction. Usually GA is followed by AWGN that complicates retrieval of the watermark [9].

Numerous different approaches have been developed to improve robustness against GA of QIM related methods [10]. Most approaches can be classified into two groups where the main idea of the first group is to estimate the unknown factor [11] while the idea of second is to quantize coefficients that are invariant to scaling of original signal.

For instance, a kind of GA and a constant offset attack followed by AWGN is assumed in [9]. The solution proposed there is to embed a pilot signal and to use Fourier analysis to estimate the gain factor and the offset. Another method of recovery after GA and AWGN is proposed in [12]. It uses information about dither sequence and applies Maximum Likelihood (ML) procedure to estimate the scaling factor.

Invariant to GA watermarking requires more complex transform of original signal (nonlinear, for instance) to obtain coefficients. Rational Dither Modulation (RDM) is one of the most popular watermarking methods invariant to GA [13]. For a particular coefficient, a ratio that depends on a norm of other coefficients is being quantized instead of a coefficient itself. The simplest QIM scheme is utilized in order to quantize the ratio and the performance of RDM under AWGN (without GA) is close to the simplest QIM. Other recent blind water-

978-1-4799-5409-4/14/$31.00 ©2014 IEEE

marking methods robust to GA are proposed in [14], [15], [16].

In this paper, a new scalar QIM-based watermarking approach is proposed. The approach delivers high capacity under AWGN and GA. One of the several improvements is Initial Data Loss (IDL) which accepts some fraction of wrong watermark bits during watermark embedding.

The rest of the paper is organized as following. Our quantization model is described in Section II using formal logic approach. The routine of analytic-based estimation of robustness under AWGN is discussed in Section III. Next, Section IV contains description for the procedure of recovery after GA as well as experimental results obtained under popular attacks. In Section V we discuss in details experimental conditions and compare the performance of the proposed method with the performance of well-known methods. Section VI concludes the paper.

## II. Quantization Approach

A new quantization approach is described in this section. Conditions that are important for watermark embedding are discussed first and logical expressions are used to formalize them. Further, we combine all the conditions in a single logical expression and define parameters of quantization model that provides the expression is true.

### A. Formalization of Embedding Approach

Symbol $\Sigma$ will be used to denote a random variable which domain is the space of original coefficients of a host. A particular realization of $\Sigma$ will be denoted as $\varsigma$. We will further describe our model for values $\varsigma$ that are in some $k$-th interval of size $\Delta$ and its left endpoint is $l_\Delta^k$. Such an interval is referred further as embedding interval. For any $\varsigma \in [l_\Delta^k, l_\Delta^k + \Delta]$ we define $x = \varsigma - l_\Delta^k$ and $X$ will be used to denote a random variable which represents $x$. The value of $\Delta$ should be small enough so that the distribution of $X$ can be considered to be uniform. A random variable that represents separated coefficients inside $k$-th interval is denoted as $X'$ and its realization is denoted as $x'$. Correspondingly, a random variable that represents separated coefficients on the whole real number line is denoted as $\Sigma'$ and its realization is denoted as $\varsigma'$. Each pair of an original $x$ and corresponding quantized $x'$ belong to the same $k$-th embedding interval so that an absolute shift is never larger than $\Delta$.

We denote a watermark bit as $b$. Parameters $\eta_1$ and $\vartheta_0$ represent fractions of IDL for $b = 1$ and $b = 0$ respectively. Parameters $\varphi_1$ and $\gamma_0$ represent fractions of the samples where original values $x$ are to be modified (non-IDL) by a quantizer. The fraction of zeros in a watermark data is $\gamma_0 + \vartheta_0$, fraction of ones is $\varphi_1 + \eta_1$. Fractions of "0" and "1" are considered to be distributed uniformly inside $k$-th embedding interval. Condition $\gamma_0 + \vartheta_0 + \varphi_1 + \eta_1 = 1$ always holds.

Each sample with value $x$ inside $k$-th embedding interval has index $i \in \mathbb{N}$ according to its order in the host sequence. During watermarking a bit is assigned to each index $i$. We will use first order predicate logic to describe our approach. Two-place predicate $E$ is to denote correspondence between some index and the value of coefficient. For example, $Eix$ is true if a coefficient with order $i$ has value $x$. One-place predicate $B$ is to denote bit value assigned to a coefficient with particular index. For instance, $Bi$ is true if bit $b = 1$ is assigned to a coefficient with index $i$ and $\sim Bi$ is true if $b = 0$.

The coefficients inside $k$-th embedding interval should be separated in order to embed watermarking data. Further we assume that coefficients interpreting "0" are shifted to the left while those interpreting "1" are shifted to the right. Only coefficients that are not classified as IDL are to be modified. Non-IDL fraction $\gamma_0$ is considered to be the left most fraction of "0"-coefficients and non-IDL fraction $\varphi_1$ is considered to be the right most fraction of "1"-coefficients. In other words, for those $i$ and $x$ that satisfy $(Eix\&(x \le L_1)\&\sim Bi)$, $L_1 = \Delta\gamma_0/(\gamma_0 + \vartheta_0)$, modified value $x'$ is to the left from $x$. For those $i$ and $x$ that satisfy $(Eix\&(x \ge L_2)\&Bi)$, $L_2 = \Delta\eta_1/(\varphi_1 + \eta_1)$, modified value $x'$ is to the right from $x$. For this scenario, it is not acceptable to shift to the right if $(Eix\&(x \le L_1)\&\sim Bi)$ holds or shift to the left if $(Eix\&(x \ge L_2)\&Bi)$ holds because it would always introduce some distortion, but, on the other hand, this would worsen separation between "0" and "1". We will use two-place predicates $U$ and $V$ to denote that all our requirements to the replacement $x$ by $x'$ is satisfied. Now, we can formalize our first necessary requirement:

$$(\forall x)(\forall x')\left(\left(Uxx' \supset (x \ge x')\right)\&\left(Vxx' \supset (x \le x')\right)\right). \quad (1)$$

It is also required that quantized samples $x'$ are distributed according to some desired pdf. The distribution of quantized samples inside $k$-th embedding interval can be represented using two pairs of pdfs. For non-IDL samples the first pair consists of two truncated pdfs $f_0(x')$ and $f_1(x')$ when $b = 0$ and $b = 1$ respectively. For IDL samples the second pair consists of two truncated pdfs $IDL_0(x')$ and $IDL_1(x')$. The left endpoint of $f_0(x')$ and $IDL_1(x')$ is 0 while the right endpoint of $f_1(x')$ and $IDL_0(x')$ is $\Delta$.

For any two $x_i \le x_j$ with the same value of label $b$ we infer that quantization in a way that $x_i' \le x_j'$ implies less distortion than in case when $x_i' > x_j'$. Therefore, in a discrete model of distribution of quantized samples we can state that, for instance, the number of original non-IDL samples $x$ with label $b = 0$ for which $x \le x_i$ in $k$-th interval should be equal to the number of modified samples $x'$ with $b = 0$, where $x' \le x_i'$. Hence, as a result of switching from discrete to continuous model of distribution, our second pair of necessary requirements can be expressed as:

$$(\forall x)(\forall x')\left(Uxx' \supset \left(x\frac{\gamma_0+\vartheta_0}{\Delta} = \gamma_0 \int_0^{x'} f_0(x')dx'\right)\right), \quad (2)$$

$$(\forall x)(\forall x')\left(Vxx' \supset \left((\Delta - x)\frac{\varphi_1+\eta_1}{\Delta} = -\varphi_1 \int_\Delta^{x'} f_1(x')dx'\right)\right). \quad (3)$$

We set only two requirements so that they are both necessary and sufficient to define true values for predicates $U$ and $V$. Consequently, we can conclude:

$$(\forall x)(\forall x')\left(Uxx' \equiv \left(\left(x\frac{\gamma_0 + \vartheta_0}{\Delta} = \gamma_0 \int_0^{x'} f_0(x')dx'\right)\& \right.\right.$$
$$\left.\left.\&(x \ge x')\right)\right), \quad (4)$$

$$(\forall x)(\forall x')\left(Vxx' \equiv \left(\left((\Delta - x)\frac{\varphi_1+\eta_1}{\Delta} = -\varphi_1\int_{\Delta}^{x'} f_1(x')dx'\right)\&\right.\right.$$
$$\left.\left.\&(x \le x')\right)\right). \quad (5)$$

Finally, we enforce formula $F1$ to be true for our quantization approach:

$$F1 \equiv (\forall i)(\forall x)\left(\left((Eix\&(x \le L_1)\&{\sim}Bi) \supset (\exists x')Uxx'\right)\&\right.$$
$$\left.\&\left((Eix\&(x \ge L_2)\&Bi) \supset (\exists x')Vxx'\right)\right). \quad (6)$$

We require that parameters of the quantizer are independent on watermarking data and samples order inside host sequence. Formula $F2$ will be used to express this requirement. It represents condition that should be satisfied by parameters estimates:

$$F2 \equiv \left(\left((\forall x)\left((x \le L_1) \supset (\exists x')Uxx'\right)\right)\&\right.$$
$$\left.\&\left((\forall x)\left((x \ge L_2) \supset (\exists x')Vxx'\right)\right)\right). \quad (7)$$

It can be seen that

$$F2 \vdash F1 \quad (8)$$

and, hence, $F2$ represents condition which is sufficient for the quantizer to operate properly.

### B. Definition of Embedding Parameters

Parameters of the quantizer should satisfy condition (7). The goal of the paper is to estimate watermarking capacity for a given type of distortion and its intensity. However, $F2$ can be satisfied in many different ways. Therefore values of parameters should be optimized in order to provide the highest capacity. Condition (7) implies some constraints and we are going to solve this optimization task by brute force approach. With the aim to reduce computations, values of some parameters are constrained according to our own considerations. Some other parameters are constrained in accordance to the condition (7) and values of the parameters chosen in the first place. We chose parameters $\eta_1, \vartheta_0, \varphi_1, \gamma_0$ to be constrained in the first place according to our considerations. We will further derive constraints for other parameters of the quantizer. The result of the quantization in the $k$-th embedding interval depends on $b$, $\eta_1$, $\vartheta_0$, $\varphi_1$, $\gamma_0$, $f_0(x')$, $f_1(x')$. A short-hand notation for quantizer is $Q_\Delta^k[\cdot]$ where the output is $x'$:

$$x' = Q_\Delta^k[x, b, \eta_1, \vartheta_0, \varphi_1, \gamma_0, f_0(x'), f_1(x')]. \quad (9)$$

Constraints should be derived for parameters that specify $f_0(x')$ and $f_1(x')$. The pdfs that we propose are depicted on Fig.1 (a). On Fig. 1 (b) we can see the distribution of the quantized coefficients outside $k$-th embedding interval as well. Inside $k$-th interval the right endpoint for $f_0(x')$ and the left endpoint for $f_1(x')$ can be expressed using parameters $\alpha, \beta$, where $0 \le \alpha \le \beta \le 1$. Hence, according to Fig.1 (a) the pdfs can be defined as following:

$$f_0(x') = \begin{cases} cx' + \tau, & if\ x' \in [0,\ \Delta(\beta-\alpha)] \\ 0, & otherwise; \end{cases}, \quad (10)$$

$$f_1(x') = \begin{cases} g, & if\ x' \in [\Delta\beta,\ \Delta] \\ 0, & otherwise. \end{cases}. \quad (11)$$

Pdfs $IDL_0(x')$ and $IDL_1(x')$ are not depicted on Fig.1 (a) because their form and position is straightforward and only depends on parameters $\eta_1, \vartheta_0, \varphi_1, \gamma_0$:

$$IDL_0(x') = \begin{cases} \frac{\gamma_0+\vartheta_0}{\Delta\vartheta_0}, & if\ x' \in \left[\frac{\Delta\gamma_0}{\gamma_0+\vartheta_0},\ \Delta\right] \\ 0, & otherwise; \end{cases}$$

$$IDL_1(x') = \begin{cases} \frac{\varphi_1+\eta_1}{\Delta\eta_1}, & if\ x' \in \left[0,\ \frac{\Delta\eta_1}{\varphi_1+\eta_1}\right] \\ 0, & otherwise; \end{cases}$$

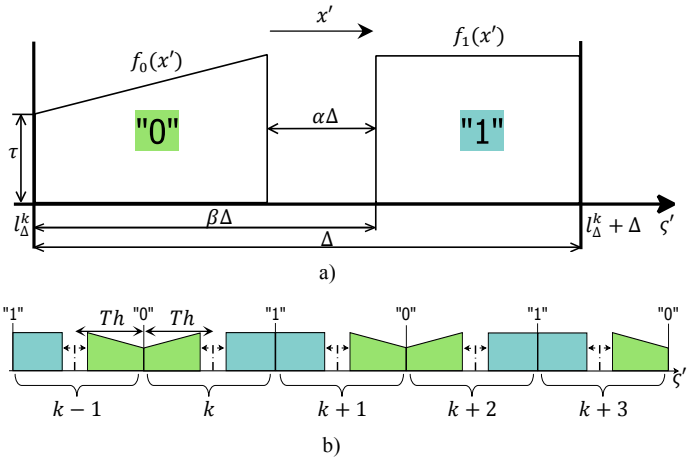



Fig. 1. Distribution of the quantized coefficients: a) Inside $k$-th embedding interval; b) In five consecutive intervals.

In accordance to (7) we will define exact values of $g, \tau, c$ in terms of $\alpha, \beta, \gamma_0, \varphi_1, \eta_1, \vartheta_0, \Delta$; constraints for $\alpha, \beta$ will be expressed in terms of $\gamma_0, \varphi_1, \eta_1, \vartheta_0$ after.

Using property of pdf and substituting expressions (10) and (11) for $f_0(x')$ and $f_1(x')$ respectively, the following can be derived:

$$\int_0^{(\beta-\alpha)\Delta} f_0(x')dx' = c\frac{(\beta-\alpha)^2\Delta^2}{2} + \tau\Delta(\beta-\alpha) = 1, \quad (12)$$

$$\int_{\beta\Delta}^{\Delta} f_1(x')dx' = g\Delta(1-\beta) = 1. \quad (13)$$

It is easy to follow from (13) that

$$g = \frac{1}{\Delta(1-\beta)}. \quad (14)$$

According to (4)

$$x'\frac{\gamma_0+\vartheta_0}{\Delta} \le \gamma_0\int_0^{x'} f_0(x')dx'. \quad (15)$$

Substituting expression for $f_0(x')$ from (10) to (15) and using the fact $x' \ge 0$, we can derive:

$$\tau \ge \frac{\gamma_0+\vartheta_0}{\Delta\gamma_0} - c\frac{x'}{2}. \quad (16)$$

Inequality (16) should be true for all $x' \in [0,\ \Delta(\beta-\alpha)]$ which means

$$\tau \ge max_{x' \in [0,\ \Delta(\beta-\alpha)]}\left(\frac{\gamma_0+\vartheta_0}{\Delta\gamma_0} - c\frac{x'}{2}\right). \quad (17)$$

In the paper we describe implementation where $c \ge 0$, therefore

$$\tau \ge \frac{\gamma_0+\vartheta_0}{\Delta\gamma_0}. \quad (18)$$

We set $\tau = (\gamma_0 + \vartheta_0)/(\Delta\gamma_0)$ in our method. Substituting this expression for $\tau$ in (12), the following can be derived:

$$c = 2\frac{\gamma_0-(\gamma_0+\vartheta_0)(\beta-\alpha)}{\gamma_0(\beta-\alpha)^2\Delta^2}. \quad (19)$$

Constraints for $\alpha$ and $\beta$ can now be expressed using parameters $\gamma_0, \varphi_1, \eta_1, \vartheta_0$. In our realization we set $c \geq 0$; from (19) it is obvious that the denominator of the right part of the equation is never negative, consequently, the numerator as well. Hence, it follows that

$$\beta - \alpha \leq \frac{\gamma_0}{\gamma_0+\vartheta_0}. \quad (20)$$

Substituting expression for $f_1(x')$ from (11) into (5), we find

$$\frac{\varphi_1+\eta_1}{\Delta} \leq g\varphi_1. \quad (21)$$

Then, using (14) we conclude

$$\beta \geq \frac{\eta_1}{\varphi_1+\eta_1}. \quad (22)$$

During brute force optimization, constrained parameters need to be set first: $\gamma_0, \varphi_1, \eta_1, \vartheta_0$ are chosen, then $\alpha$ and $\beta$ that satisfy (20) and (22) are selected. Lastly, exact values $g, \tau, c$ need to be calculated.

### C. Quantization equations

Expression for $F2$ contains predicates $Uxx'$ and $Vxx'$. However, according to (4) and (5), it is not technically possible to enforce existential quantifier $\exists$ on $x'$ for $Uxx'$ and $Vxx'$ as there is no evidence that such $x'$ exists. In order to make $F2$ verifiably true, quantizer $Q_\Delta^k[\cdot]$ should utilize models where $x'$ is expressed explicitly. We call these models quantization equations and denote them $x' = e_0(x)$ and $x' = e_1(x)$, where $e_0(\cdot)$, $e_1(\cdot)$ denote some functions. Further, for convenience we use formulas $F2_0$ and $F2_1$, where $F2 \equiv F2_0 \& F2_1$:

$$F2_0 \equiv (\forall x)\big((x \leq L_1) \supset (\exists x')Uxx'\big), \quad (23)$$

$$F2_1 \equiv (\forall x)\big((x \geq L_2) \supset (\exists x')Vxx'\big). \quad (24)$$

The following argument is valid and all its premises (25-26) and the conclusion refer to the process of embedding "0":

$$(\forall x)(\forall x')\Big(\big(x' = e_0(x)\big) \equiv Uxx'\Big), \quad (25)$$

$$(\forall x)\Big((x \leq L_1) \supset (\exists x')\big(x' = e_0(x)\big)\Big), \quad (26)$$

$$\therefore F2_0.$$

The next argument is also valid and refers to the process of embedding "1"

$$(\forall x)(\forall x')\Big(\big(x' = e_1(x)\big) \equiv Vxx'\Big), \quad (27)$$

$$(\forall x)\Big((x \geq L_2) \supset (\exists x')\big(x' = e_1(x)\big)\Big), \quad (28)$$

$$\therefore F2_1.$$

The goal of this subsection is to define $e_0(x)$ and $e_1(x)$ so that the both mentioned arguments are sound.

In order to satisfy (25), we derive expression $x' = e_0(x)$ for $Uxx'$ from (4), e.g:

$$\begin{cases} x\frac{\gamma_0+\vartheta_0}{\Delta} = \gamma_0\int_0^{x'} f_0(x')dx', \\ x \geq x'. \end{cases} \quad (29)$$

Using expression for $\tau$, the equation from (29) can be rewritten as

$$0.5cx'^2 + \tau x' = \tau x. \quad (30)$$

Solving (30) for $x'$ we define function $x' = e_0(x)$:

$$e_0(x) = \frac{\sqrt{\tau^2+2c\tau x}}{c} - \frac{\tau}{c}, \quad (31)$$

which obviously satisfies the inequality in (29). Condition (26) is also easy to enforce as function $e_0(x)$ exists on $[0, L_1]$. Therefore, the first argument is sound.

With the aim to satisfy (27), we derive expression $x' = e_1(x)$ for $Vxx'$ from (5), e.g:

$$\begin{cases} (\Delta - x)\frac{\varphi_1+\eta_1}{\Delta} = -\varphi_1\int_\Delta^{x'} f_1(x')dx', \\ x \leq x'. \end{cases} \quad (32)$$

Solving equation in (32) for $x'$ we define function $x' = e_1(x)$:

$$e_1(x) = \frac{(1-\beta)(\varphi_1+\eta_1)}{\varphi_1}x + \frac{\varphi_1-(1-\beta)(\varphi_1+\eta_1)}{\varphi_1}\Delta, \quad (33)$$

which obviously satisfies the inequality in (32). Condition (28) is also easy to enforce as function $e_1(x)$ exists on $[L_2, \Delta]$. Hence, the second argument is also sound.

## III. Robustness under AWGN

The main characteristic for any watermarking scheme is robustness. In the case of AWGN attack, robustness depends on the attack severity, which is represented by $\sigma$. In addition to that, robustness of our scheme depends on the values of the parameters used during watermark embedding, e.g. $\Delta$ and set $\boldsymbol{\theta} = \{\gamma_0, \varphi_1, \eta_1, \vartheta_0, \alpha, \beta\}$**.** Hence, information about $\sigma$, $\Delta$ and $\boldsymbol{\theta}$ is enough to estimate robustness. In this section, we derive a stronger statement that information about $\Delta/\sigma$, $\boldsymbol{\theta}$ is sufficient to perform analytic estimation of error rates for our watermarking scheme. Finally, we will demonstrate how error rates can be expressed using $WNR$ and $\boldsymbol{\theta}$.

### A. Estimation of error rates

Parameter $Th$ will denote the distance between the left endpoint of $k$-th embedding interval and the position of the threshold (dashed lines on Fig.1 (b)) that separates "0" from "1". For any coefficient that is labeled "0", an error occurs if it is being shifted by noise to one of the intervals that are interpreted as "1". We will enumerate with index $j \in \mathbb{Z}$ intervals that interpret bit values "0" and "1". Index $j$ is positive for those intervals that are to the right from the $k$-th embedding interval and is negative to the left. For instance, the $j$-th interval of "ones" is $\left[l_\Delta^k + 2j\Delta + Th, l_\Delta^k + (2j+1)\Delta - Th\right]$. There are IDL and non-IDL coefficients with label "0". In order to calculate error rates we introduce functions $\mathcal{H}_0^1(\cdot)$, $\mathcal{H}_0^2(\cdot)$, $\mathrm{C}_0^1(\cdot)$, $\mathrm{C}_0^2(\cdot)$, $\mathcal{H}_1^1(\cdot)$, $\mathcal{H}_1^2(\cdot)$, $\mathrm{C}_1^1(\cdot)$, $\mathrm{C}_1^2(\cdot)$. For the $k$-th embedding interval, the expected fraction of non-IDL coefficients with label "0" that remain after attack in interval $\left[-\infty, l_\Delta^k + 2j\Delta + Th\right]$ is $0.5 + 0.5\mathcal{H}_0^1(\sigma, \Delta, Th, j, \boldsymbol{\theta})$. Expected fraction that remains in $\left[-\infty, l_\Delta^k + (2j+1)\Delta - Th\right]$ is $0.5 + 0.5\mathcal{H}_0^2(\sigma, \Delta, Th, j, \boldsymbol{\theta})$. The fraction that enters to the $j$-th interval of "ones" is $0.5\big(\mathcal{H}_0^2(\sigma, \Delta, Th, j, \boldsymbol{\theta}) - \mathcal{H}_0^1(\sigma, \Delta, Th, j, \boldsymbol{\theta})\big)$. The expected fraction of IDL coefficients with label "0" that enters to the $j$-th interval of "ones" is $0.5\big(\mathrm{C}_0^2(\sigma, \Delta, Th, j, \boldsymbol{\theta}) - \mathrm{C}_0^1(\sigma, \Delta, Th, j, \boldsymbol{\theta})\big)$. For "0" samples, fraction of non-IDL is $\frac{\gamma_0}{\gamma_0+\vartheta_0}$ and fraction of IDL is $\frac{\vartheta_0}{\gamma_0+\vartheta_0}$. The expression for BER for "0" samples is:

$$BER_0 = \frac{0.5\gamma_0}{\gamma_0+\vartheta_0}\sum_{j=-\infty}^{\infty}\Big(\mathcal{H}_0^2(\sigma, \Delta, Th, j, \boldsymbol{\theta}) - \mathcal{H}_0^1(\sigma, \Delta, Th, j, \boldsymbol{\theta})\Big) +$$

$$+\frac{0.5\vartheta_0}{\gamma_0+\vartheta_0}\sum_{j=-\infty}^{\infty}\left(\mathrm{C}_0^2(\sigma,\Delta,Th,j,\boldsymbol{\theta})-\mathrm{C}_0^1(\sigma,\Delta,Th,j,\boldsymbol{\theta})\right). \quad (34)$$

Functions $\mathcal{H}_0^1(\cdot)$, $\mathcal{H}_0^2(\cdot)$, $\mathrm{C}_0^1(\cdot)$, $\mathrm{C}_0^2(\cdot)$ can be expressed using error function $erf(\cdot)$:

$$\mathcal{H}_0^1(\sigma,\Delta,Th,j,\boldsymbol{\theta})=\int_0^{\Delta(\beta-\alpha)} f_0(x')erf\left(\frac{Th+2\Delta j-x'}{\sigma\sqrt{2}}\right)dx\,, \quad (35)$$

$$\mathcal{H}_0^2(\sigma,\Delta,Th,j,\boldsymbol{\theta})=\int_0^{\Delta(\beta-\alpha)} f_0(x')erf\left(\frac{2\Delta(j+1)-Th-x'}{\sigma\sqrt{2}}\right)dx', \quad (36)$$

$$\mathrm{C}_0^1(\sigma,\Delta,Th,j,\boldsymbol{\theta})=\int_{\frac{\Delta\gamma_0}{\gamma_0+\vartheta_0}}^{\Delta} IDL_0(x')erf\left(\frac{Th+2\Delta j-x'}{\sigma\sqrt{2}}\right)dx'\,, \quad (37)$$

$$\mathrm{C}_0^1(\sigma,\Delta,Th,j,\boldsymbol{\theta})=\int_{\frac{\Delta\gamma_0}{\gamma_0+\vartheta_0}}^{\Delta} IDL_0(x')erf\left(\frac{2\Delta(j+1)-Th-x'}{\sigma\sqrt{2}}\right)dx'. \quad (38)$$

The $j$-th interval of "zeros" is $\left[l_\Delta^k+2j\Delta-Th, l_\Delta^k+2j\Delta+Th\right]$. The fraction of the quantized coefficients in $k$-th embedding interval that are labeled "1" and enter into $j$-th interval of "zeros" is $0.5\left(\mathcal{H}_1^2(\sigma,\Delta,Th,j,\boldsymbol{\theta})-\mathcal{H}_1^1(\sigma,\Delta,Th,j,\boldsymbol{\theta})\right)$ for non-IDL samples. For IDL samples the fraction is $0.5\left(\mathrm{C}_1^2(\sigma,\Delta,Th,j,\boldsymbol{\theta})-\mathrm{C}_1^1(\sigma,\Delta,Th,j,\boldsymbol{\theta})\right)$. Hence,

$$BER_1=\frac{0.5\varphi_1}{\varphi_1+\eta_1}\sum_{j=-\infty}^{\infty}\left(\mathcal{H}_1^2(\sigma,\Delta,Th,j,\boldsymbol{\theta})-\mathcal{H}_1^1(\sigma,\Delta,Th,j,\boldsymbol{\theta})\right)+$$
$$+\frac{0.5\eta_1}{\varphi_1+\eta_1}\sum_{j=-\infty}^{\infty}\left(\mathrm{C}_1^2(\sigma,\Delta,Th,j,\boldsymbol{\theta})-\mathrm{C}_1^1(\sigma,\Delta,Th,j,\boldsymbol{\theta})\right). \quad (39)$$

Functions $\mathcal{H}_1^1(\cdot)$, $\mathcal{H}_1^2(\cdot)$, $\mathrm{C}_1^1(\cdot)$, $\mathrm{C}_1^2(\cdot)$ can be expressed as following:

$$\mathcal{H}_1^1(\sigma,\Delta,Th,j,\boldsymbol{\theta})=\int_{\Delta\beta}^{\Delta} f_1(x')erf\left(\frac{2\Delta j-Th-x'}{\sigma\sqrt{2}}\right)dx'\,, \quad (40)$$

$$\mathcal{H}_1^2(\sigma,\Delta,Th,j,\boldsymbol{\theta})=\int_{\Delta\beta}^{\Delta} f_1(x')erf\left(\frac{2\Delta j+Th-x'}{\sigma\sqrt{2}}\right)dx', \quad (41)$$

$$\mathrm{C}_1^1(\sigma,\Delta,Th,j,\boldsymbol{\theta})=\int_0^{\frac{\Delta\eta_1}{\varphi_1+\eta_1}} IDL_1(x')erf\left(\frac{2\Delta j-Th-x'}{\sigma\sqrt{2}}\right)dx'\,, \quad (42)$$

$$\mathrm{C}_1^2(\sigma,\Delta,Th,j,\boldsymbol{\theta})=\int_0^{\frac{\Delta\eta_1}{\varphi_1+\eta_1}} IDL_1(x')erf\left(\frac{2\Delta j+Th-x'}{\sigma\sqrt{2}}\right)dx'\,. \quad (43)$$

For the rest of this section we assume that optimal watermark extraction using a threshold $Th$ requires only information about $\boldsymbol{\theta}$ and $Th=\mu(\boldsymbol{\theta})\Delta$, where $\mu(\boldsymbol{\theta})$ is a ratio that depends on $\boldsymbol{\theta}$. For example, for the most of the known QIM-based watermarking schemes $\mu$ is 0.5, [7], [9]. It is possible to demonstrate that error rates depend only on $\Delta/\sigma$, $\boldsymbol{\theta}$. It is enough to show that this is true for each expression (35-38) and (40-43). For further elimination of $\Delta$ and $\sigma$ we introduce new variables $\hat{c}=c\Delta^2$ and $\hat{\tau}=\tau\Delta$, where term $\Delta$ cancels out. For example, (35) can be rewritten:

$$\mathcal{H}_0^1(\sigma,\Delta,Th,j,\boldsymbol{\theta})=\int_0^{\Delta(\beta-\alpha)}\left(\frac{\hat{c}x'}{\Delta^2}+\frac{\hat{\tau}}{\Delta}\right)erf\left(\frac{Th+2\Delta j-x'}{\sigma\sqrt{2}}\right)dx'=$$
$$=\int_0^{(\beta-\alpha)}\left(\hat{c}\frac{x'}{\Delta}+\hat{\tau}\right)erf\left(\frac{\mu(\boldsymbol{\theta})\Delta+2\Delta j-\Delta x'/\Delta}{\sigma\sqrt{2}}\right)d\left\{x'/_\Delta\right\}. \quad (44)$$

Now, we denote $x'/\Delta$ as $\hat{x}'$ and continue:

$$\mathcal{H}_0^1(\sigma,\Delta,Th,j,\boldsymbol{\theta})=\int_0^{(\beta-\alpha)}(\hat{c}\hat{x}'+\hat{\tau})erf\left(\frac{\Delta}{\sigma}\cdot\frac{\mu(\boldsymbol{\theta})+2j+\hat{x}'}{\sqrt{2}}\right)d\hat{x}'=$$
$$=\hat{\mathcal{H}}_0^1(\Delta/\sigma,j,\boldsymbol{\theta})\,. \quad (45)$$

Further we will express $\Delta/\sigma$ in terms of $WNR$ and $\boldsymbol{\theta}$ so that it is possible to define some $\hat{\mathcal{H}}_0^1(WNR,j,\boldsymbol{\theta})$ and use it instead of $\hat{\mathcal{H}}_0^1(\Delta/\sigma,j,\boldsymbol{\theta})$.

### B. Estimation of $\Delta/\boldsymbol{\sigma}$

Measure $WNR$ is widely used in watermarking. It expresses relation between watermark and noise energies and in AWGN case is

$$WNR=10\,log_{10}\left(\frac{D}{\sigma^2}\right), \quad (46)$$

where $D$ is the energy of the watermark. Plot of watermark capacity in respect to $WNR$ is one of the characteristics that are the most meaningful for practical implementation [3], [9]. Therefore it is important to be able to express error rates using $WNR$ and the set of embedding parameters $\boldsymbol{\theta}$. For this purpose we express $\Delta/\sigma$ using $WNR$ and $\boldsymbol{\theta}$.

Parameter $D$ in (46) can be seen as a distortion of a host signal, caused by the quantization. We will define $D$ and factor it in a form $\Delta^2 Q$, where $Q$ depends only on $\boldsymbol{\theta}$. Distortion $D$ is caused by quantization of non-IDL samples and is a sum of distortions $D_0$ and $D_1$ caused by quantization of samples with label "0" and "1" respectively. The first distortion component $D_0$ is defined as

$$D_0=\gamma_0\int_0^{\Delta(\beta-\alpha)} f_0(x')\left(x'-\frac{1}{\tau}\int_0^{x'} f_0(x')dx'\right)^2 dx'\,. \quad (47)$$

According to (10) we can derive that

$$D_0=\gamma_0\int_0^{\Delta(\beta-\alpha)}(cx'+\tau)\frac{c^2x'^4}{4\tau^2}dx'.$$

Using variables $\hat{c}$ and $\hat{\tau}$ it is possible to factor $D_0$ in the form

$$D_0=\Delta^2 Q_0\,, \quad (48)$$

where

$$Q_0=\gamma_0\left(\frac{\hat{c}^3}{24\hat{\tau}^2}(\beta-\alpha)^6+\frac{\hat{c}^2}{20\hat{\tau}}(\beta-\alpha)^5\right). \quad (49)$$

The second distortion component $D_1$ is defined as

$$D_1=\varphi_1\int_{\beta\Delta}^{\Delta} f_1(x')\left(x'-\left(\frac{\varphi_1\Delta}{\eta_1+\varphi_1}\int_{\beta\Delta}^{x'} f_1(x')dx'+\frac{\eta_1\Delta}{\eta_1+\varphi_1}\right)\right)^2 dx'. \quad (50)$$

Substituting (11), (14) and integrating in (50) we obtain

$$D_1=\Delta^2 Q_1\,, \quad (51)$$

$$Q_1=\varphi_1\frac{\left((\eta_1+\varphi_1)(1-\beta)-\varphi_1\right)^2}{3(\eta_1+\varphi_1)^2}\,. \quad (52)$$

Now, factorization in the form $D=\Delta^2 Q$ can be performed using $Q=Q_0+Q_1$. Hence, according to (46) $\Delta/\sigma$ can be expressed in the following way:

$$\Delta/\sigma=\sqrt{\frac{10^{0.1*WNR}}{Q_0+Q_1}}\,. \quad (53)$$

## IV. Experimental results

Conditions and results of two different kinds of experiments are described in this section. Analytic estimations of capacity were used in experiments assuming AWGN attack. Simulations on a set of real images were used during experiments assuming GA. In the first case, the obtained results are compared with the results of QIM and DC-QIM. In the second case the results are compared with the results of RDM.

### A. Watermarking Performance under AWGN

In addition to (20) and (22) parameters $\alpha,\beta,\eta_1,\gamma_0,\vartheta_0,\varphi_1$ are subjects to constraints, $\eta_1+\gamma_0=0.5$, $\vartheta_0+\varphi_1=0.5$ during the experiment. Two variants of the proposed quantization scheme with adjustable parameters are used: Non-Symmetric QIM (NS-QIM) and Non-Symmetric QIM with IDL (NS-QIM-IDL). Here, NS-QIM is a subject to additional constraints $\eta_1=0$, $\vartheta_0=0$ (in some cases IDL is not acceptable on practice).

The plots for channel capacity toward WNR are shown on Fig. 2 for NS-QIM, NS-QIM-IDL, DC-QIM and QIM [6]. The kind of thresholding applied to NS-QIM and NS-QIM-IDL is

$\mu(\mathbf{\theta}) = \beta - 0.5\alpha$. As a reference, Costa Theoretical Limit (CTL) [4] is plotted on Fig. 2:

$$CTL = \frac{1}{2} log_2(1 + 10^{0.1*WNR}) . \quad (54)$$

Capacity is calculated analytically according to the description provided in the literature for DC-QIM and QIM (by QIM we refer to the simplest realization e.g. [5]). Only the integers from $[-100, 100]$ were used as a set for $j$ in (34) and (39) instead of the whole set $\mathbb{Z}$. Such a choice is a compromise between computational complexity and the fidelity of the result.

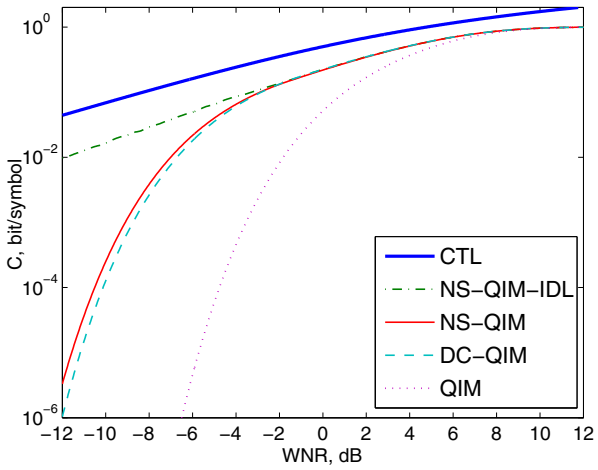



Fig. 2. Analytic-based estimation of capacity under AWGN

From Fig. 2 it can be seen that the both variants of the proposed method perform better than DC-QIM for WNR values less than -2dB. From the comparison of NS-QIM-IDL and NS-QIM it is clear that the first mentioned method is much more beneficial under low WNRs. Obviously, the demonstrated superiority is due to IDL only.

### B. *Watermarking performance under GA*

The superiority of NS-QIM over DC-QIM under AWGN attack has been demonstrated. In this subsection we explore performance of NS-QIM under GA. For comparison, RDM is chosen instead of DC-QIM as it is known to be vulnerable to GA. First, for NS-QIM we introduce procedure that aims to recover a watermark after GA. Second, we describe conditions and the results of the simulations based on real images and assuming watermark embedding followed by GA.

*1)* Procedure for GA recovery

The original length of embedding interval $\Delta$ is altered by unknown gain factor $\lambda$ and the resulting length is $\tilde{\Delta} = \lambda\Delta$. Additionally, AWGN attack is applied. We propose the procedure for GA recovery that is based on a criterion which tends to have higher values for the right length $\tilde{\Delta}$ of embedding interval. Two different criteria $C_1$, $C_2$ are introduced and exploit the unique feature of the distribution of the samples quantized according to NS-QIM. The procedure itself represents a brute force approach that substitutes guessed values $\tilde{\Delta}'$ of the length of embedding interval into a criterion and selects the one that maximizes it:

$$\tilde{\Delta}'' = arg\, max_{\tilde{\Delta}'}\, C_{1,2}(\tilde{\Delta}') , \quad (55)$$

where $\tilde{\Delta}''$ is the final estimate.

In order to calculate criteria for each particular value $\tilde{\Delta}'$, noisy quantized samples $\varsigma'_n$ are being projected on a single embedding interval:

$$x'_n = \begin{cases} \varsigma'_n \bmod \tilde{\Delta}', \ if \ \left\lfloor \frac{\varsigma'_n - l^k_\Delta}{\tilde{\Delta}'} \right\rfloor \bmod 2 = 0, \\ \tilde{\Delta}' - (\varsigma'_n \bmod \tilde{\Delta}'), \ otherwise . \end{cases} \quad (56)$$

We propose two criteria that can be applied to the random variable $X'_n \in [0, \tilde{\Delta}']$:

$$C_1(\tilde{\Delta}') = \left| \frac{median(X'_n)}{\tilde{\Delta}'} - 0.5 \right|, \quad (57)$$

$$C_2(\tilde{\Delta}') = \left| \frac{E([X'_n]^w)}{[\tilde{\Delta}']^w} \right|, w = 2m + 1, m \in \mathbb{N} . \quad (58)$$

The values of the both criteria are low if the assumption about $\tilde{\Delta}$ is wrong because in that case the distribution of $X'_n$ is very close to uniform (subscript "n" means affected by noise). However, in case $\tilde{\Delta}'$ is close to $\tilde{\Delta}$ the distribution of $X'_n$ demonstrates asymmetry. This is because the distribution of quantized samples inside embedding interval (before GA is introduced) is indeed asymmetric. Despite the procedure applies brute force, it is simple and the computational demand is low. For example, $10^3$ values from the interval $[\tilde{\Delta}'_{min}, \tilde{\Delta}'_{max}]$ are enough for recovery with high accuracy.

*2)* Simulation results

First singular values of Singular Value Decomposition (SVD) of 4x4 blocks from 87 grayscale images with resolution 512x512 were used as coefficients for watermark embedding [17], $\gamma_0 = \varphi_1 = 0.5$. During quantization, $DWR = 28$ dB (Document to Watermark Ratio) was satisfied, where

$$DWR = 10\, log_{10}\left(\frac{\sigma_H^2}{D}\right)$$

and $\sigma_H^2$ is the variance of the host. For the proposed method, original value of $\Delta$ was not known during extraction phase which is equivalent to GA. In case of RDM, the quantized value of a particular coefficient based on the information about the last 100 previous coefficients.

For each new value $\sigma$, a brute force optimization of $\alpha$ and $\beta$ has been made in case of NS-QIM. Apart from NS-QIM, another modification NSC-QIM was introduced with constant values of parameters $\alpha = 0.05$ and $\beta = 0.35$. On practice, actual $\sigma$ might not be known during watermark embedding and constant version of NS-QIM (NSC-QIM) is addressing this case.

For NS-QIM, parameters $\alpha$ and $\beta$ are necessary for watermark extraction in case thersholding is defined in a way that $\mu(\mathbf{\theta}) = \beta - 0.5\alpha$. Such a requirement for additional information can cause a considerable limitation for the scheme on practice. Median thresholding $Th = median(X'_n)$ was used in order to avoid this.

Criterion $C_1$ was used for the estimation of actual $\Delta$. During watermark extraction, no information except initial guess interval with $\tilde{\Delta}'_{min} = 0.9\Delta$, $\tilde{\Delta}'_{max} = 1.1\Delta$, was used in NS-QIM and NSC-QIM cases. In contrast to that, RDM does use the exact information about quantization step. The resulting capacity toward AWGN variance is plotted for each method on Fig. 3.

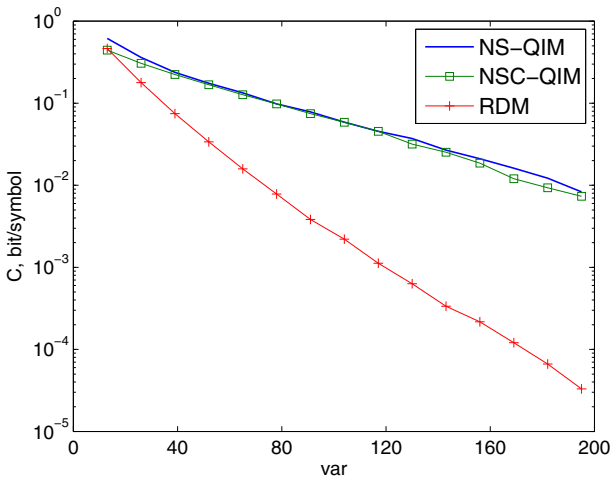



Fig. 3. Capacity under GA followed by AWGN

From Fig. 3 it can be seen that the both NS-QIM and NSC-QIM outperform RDM. For larger variance of the noise the advantage of the proposed method is more evident.

## V. Discussion

Analytical and empirical approaches were used to estimate capacity of the proposed method under AWGN and GA respectively. According to the results, the proposed method provides higher capacity compared to the other reference methods and in this section we will discuss in more details the reasons of its superiority.

In the case when only AWGN is applied, the benefit is caused by a new form of distribution of quantized samples and IDL. It is remarkable that in NS-QIM-IDL case the form of capacity plot does not inherit the steepness demonstrated by the other methods. Instead, the plot shape is similar to CTL, but is placed at a lower position. The explanation of such phenomena is in the quantization process. According to IDL we refuse to modify samples which quantization brings the highest embedding distortion. In case these samples are quantized they are placed closer to the threshold which separates "0" and "1". Therefore the information interpreted by these samples is the most likely to be lost under low WNRs. Predicting the loss of information we might accept that fact and introduce IDL instead. It is a kind of "redistribution" of embedding distortions in order to make the rest of embedded information more robust.

The proposed NS-QIM and its modification NSC-QIM perform much better than RDM under GA. This is due to the introduced procedure for estimation of scaling factor. The proposed estimation approach is more advantageous compared to other presented in the literature. For instance, a model of a host is used in [12] to estimate the scaling factor which complicates estimation and reduces its precision.

Another unique feature is the proposed median thresholding which does not require the information about $\alpha, \beta$ for watermark extraction. Therefore, during embedding these parameters can be adjusted to deliver higher capacity even in case there is no way to communicate new parameters to the receiver. Also, our recovery procedure does not use any additional information except interval $\left[\tilde{\Delta}'_{min}, \tilde{\Delta}'_{max}\right]$ that can be set roughly. These improvements imply more efficient retrieval after GA which in addition requires fewer samples.

## VI. Conclusions

In this paper, the new watermarking method based on scalar QIM has been proposed. Compared to other existing methods, it provides higher capacity under different kinds of attacks. The advantages of the method are due to its unique approach to watermark embedding as well as a new procedure of recovery and extraction.

The introduced watermark embedding approach is based on a new kind of distribution of quantized samples and IDL. There is no line of symmetry inside embedding interval for the new distribution of quantized samples. This feature is used to recover a watermark after GA. On the other hand, IDL can reduce embedding distortions introduced to a host signal. This is done by letting some watermark bits to be interpreted incorrectly at the initial phase of embedding and before any attack occurs. The proposed IDL is extremely beneficial for low WNRs under AWGN attack.

The non-symmetric distribution of quantized samples is exploited by the new procedure of recovery after GA. One out of two different criteria might be chosen to serve as a goal function for the procedure.

The mentioned advancements implied considerable performance improvement. Under conditions of AWGN the capacity of the proposed method is at the same or higher level compared to DC-QIM. Application of NS-QIM-IDL is the most advantageous under AWGN for WNRs close to -12dB where it performs up to $10^4$ times better than DC-QIM. The performance of the proposed method is up to $10^3$ times higher than that of RDM under GA followed by high level of AWGN.

## References

[1] M. Barni, F. Bartolini, V. Cappellini and A. Piva, "Robust watermarking of still images for copyright protection," in Proceedings of 13th International Conference on Digital Signal Processing Proceedings, 2-4 Jul 1997.

[2] H. R. Sheikh and A. Bovik, "Image information and visual quality," *Image Processing, IEEE Transactions on,* vol. 15, no. 2, pp. 430-444, Feb. 2006.

[3] I. Cox, M. Miller, J. Bloom, J. Fridrich and T. Kalker, Digital Watermarking and Steganography (2 ed.), San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2007.

[4] M. Costa, "Writing on dirty paper," *IEEE Transactions on Information Theory,* vol. vol.29, no. no.3, pp. 439-441, May 1983.

[5] B. Chen and G. W. Wornell, "Dither modulation: a new approach to digital watermarking and information embedding," in *Proceedings of SPIE 3657, Security and Watermarking of Multimedia Contents*, Apr. 1999.

[6] B. Chen and G. W. Wornell, "Quantization index modulation: a class of provably good methods for digital watermarking and information embedding," *IEEE Transactions on Information Theory,* vol. vol.47, no. no.4, pp. 1423-1443, May 2001.

[7] E. Esen and A. A. Alatan, "Forbidden Zone Data Hiding," in *IEEE International Conference on Image Processing*, Oct. 2006.

[8] M. Ramkumar and A. . N. Akansu, "Signaling Methods for Multimedia Steganography," *IEEE Transactions on Signal Processing,* vol. vol.52, no. no.4, pp. 1100-1111, Apr. 2004.

[9] J. J. Eggers, R. Bäuml, R. Tzschoppe and B. Girod, "Scalar Costa Scheme for Information Embedding," *IEEE Transactions on Signal Processing,* vol. vol. 51, no. no. 4, pp. 1003-1019, APR.

2003.

[10] J. Oostveen, T. Kalker and M. Staring, "Adaptive quantization watermarking," in *Security, Steganography, and Watermarking of Multimedia*, San Jose, California, USA, Jan. 2004.

[11] X. Kang, J. Huang and W. Zeng, "Improving Robustness of Quantization-Based Image Watermarking via Adaptive Receiver," *IEEE Transactions on Multimedia,* vol. vol.10, no. no.6, pp. 953-959, Oct. 2008.

[12] I. D. Shterev and R. L. Lagendijk, "Amplitude Scale Estimation for Quantization-Based Watermarking," *IEEE Transactions on Signal Processing,* vol. vol.54, no. no.11, pp. pp.4146-4155, Nov. 2006.

[13] F. Pérez-González, C. Mosquera, M. Barni and A. Abrardo, "Rational dither modulation: a high-rate data-hiding method invariant to gain attacks," *IEEE Transactions on Signal Processing,* vol. vol.53, no. no.10, pp. 3960-3975, Oct. 2005.

[14] F. Ourique, V. Licks, R. Jordan and F. Perez-Gonzalez, "Angle QIM: a novel watermark embedding scheme robust against amplitude scaling distortions," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005. (ICASSP '05)*, March 2005.

[15] M. Zareian and H. R. Tohidypour, "Robust quantisation index modulation-based approach for image watermarking," *IET Image Processing,* vol. Vol. 7, no. Iss. 5, pp. 432-441, 2013.

[16] X. Zhu and J. Ding, "Performance analysis and improvement of dither modulation under the composite attacks," *EURASIP Journal on Advances in Signal Processing,* vol. 1, no. 53, 2012.

[17] Y. Zolotavkin and M. Juhola, "A new blind adaptive watermarking method based on singular value decomposition," in *International Conference on Sensor Network Security Technology and Privacy Communication System (SNS & PCS), 2013*, May 2013.

Publication VI

# A New Two-Dimensional Quantization Method for Digital Image Watermarking

Yevhen Zolotavkin and Martti Juhola

# A New Two-Dimensional Quantization Method for Digital Image Watermarking

Yevhen Zolotavkin, Martti Juhola

Research Center for Information and Systems, School of Information Sciences, University of Tampere, Kanslerinrinne 1, 33014 Tampere, Finland

**zhzolot@suremail.us, martti.juhola@sis.uta.fi**

*Abstract*— **We propose a new technique of Distortion Compensation (DC) for the two dimensional Quincunx Lattice Quantization that is used in watermarking. The choice of a new direction of quantization is explained by the form of Voronoi cell of the lattice elements. Parameter $\alpha$ controls DC and can be adjusted to maximize the mutual information between embedded and detected message in a noisy channel. For a special case of initial distribution of original samples, quantization distortion is estimated and expressed analytically. Experimental evaluation of robustness under Additive White Gaussian Noise (AWGN) is conducted using natural images and compared with several other known two-dimensional Lattice as well as Scalar Quantization methods with conventional DC. Criterion of Watermark to Noise Ratio (WNR) was used for comparison. Among the most important findings related to the Modified Quincunx are: manipulating only a half of samples required in contrast to the other two-dimensional methods; considerably increased mutual information under low WNR; identical to the conventional Quincunx procedure of watermark extraction (decoding).**

***Keywords***— **watermarking, quantization, lattice, quincunx.**

## I. INTRODUCTION

In telecommunication networks, ownership rights protection is one of the most important aspects of multimedia usage. Many approaches applicable to Digital Image Watermarking (DIW) have been developed for the last several decades [1]. The main characteristic for a method in DIW is a trade-off between watermarking distortion and robustness to a specific attack. Image transform and watermark encoding technique are the most important factors that define the mentioned trade-off under a specific attack.

The choice of a suitable image transform is dictated by the definition of watermarking distortion and the kind of attack. Degradation of image quality can be measured according to Peak Signal to Noise Ratio (PSNR), Human Visual System (HVS) etc. [2]. The selection of the measure can rely on the intended application of the watermarked image, availability of the original, computational complexity etc. For example, watermarking in the domain of Singular Value Decomposition (SVD) is considered to have less noticeable impact in comparison with image spatial domain for the same PSNR limit [3].

Large variety of possible attacks can be considered in DIW. Among them there are additive or multiplicative kinds of noise, lossy compression techniques, gamma-correction, geometric manipulation, etc. [4] [5]. Some transforms might be suitable even for cases with multiple attacks scenarios. For instance, in case of JPEG-compression attack as well as additive noise, watermarking in the domain of Discrete Cosine Transform (DCT) might provide a sufficient robustness [6].

Encoding is another important stage of watermarking which implementation depends on relative watermark length, kind of noise, information available to a decoder, computational sources etc. Quantization Index Modulation (QIM) is a popular technique that provides sufficient robustness-distortion rate under AWGN [7]. Distortion Compensated QIM (DC-QIM) is a generalization of QIM that has an additional parameter to optimize the rate [8] [9]. For a better performance, DC-QIM of a higher dimensionality needs to be used and Nested Lattices (NL) is a suitable framework for that [10]. One of the main limitations of NL practical realizations is shape imperfection that implies non-optimal robustness-distortion rate [11]. Trellis Coded Quantization (TCQ) is another successful multidimensional approach [12, 13]. However, efficient ways of applying DC in TCQ case are not known.

In this paper, we propose a new DC technique that improves robustness-distortion rate of the two-dimensional Quincunx Lattice [11]. According to our approach, quantization is allowed only in the direction which is opposite to Maximum Likelihood Error Scenario (MLES). Therefore all the shifts are orthogonal to the closest grid of a fine lattice. Parameter $\alpha$ defines intensity of quantization shifts and its value can be optimized depending on AWGN level. The performance of the Modified Quincunx (MQ) is compared with conventional DC Quincunx as well as with other known two-dimensional lattice quantization methods.

The paper has the following structure. Detailed description of conventional Quincunx scheme with DC as well as the proposed modification is given in Section 2. Section 3 represents the stages of experimental evaluation of the Modified Quincunx and the results for other methods used for comparison. Discussion of characteristics of the proposed method is given in Section 4. The findings are concluded in Section 5.

## II. TWO-DIMENSIONAL QUINCUNX QUANTIZATION

Two-dimensional quantization is considered in this paper and symbol $X \in \mathbb{R}^2$ will be used to denote a random variable which domain is the space of pairs of original coefficients of a host. Using the pair $(x_1, x_2)$ we will denote a particular realization of X (random variables $x_1$ and $x_2$ are assumed to be independently distributed).

### *A. Conventional Quincunx Quantization with DC*

One of the most popular 2D-quantization approaches is based on Quincunx scheme [11]. A bit is embedded by modifying a pair of samples $(x_1, x_2)$. Two different coarse lattices $\Lambda_{c,0}$ and $\Lambda_{c,1}$ are used for embedding of either "0" or "1". For each element of a lattice we consider a square Voronoi cell and use the name "quantization cell" which size is controlled by parameter $\Delta$ (Figure 2). Corresponding (overlapping) grids containing the cells are depicted by solid and dashed lines, respectively (Figure 1).

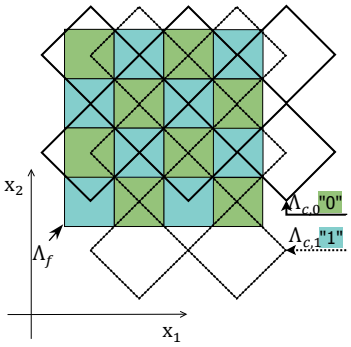



**Figure 1.** Quincunx lattice quantization and extraction cells for

For all paired original samples in a certain quantization cell the corresponding quantized samples are in the same cell too. For any point in $\mathbb{R}^2$ defined by the pair $(x_1, x_2)$ and labeled with bit $b$, the center (e.g. lattice element) of the quantization cell that the point belongs to is denoted as $(x_1^C, x_2^C)$:

$$x_1^C = F_1 + F_2 \,, \tag{1}$$

$$x_2^C = F_1 - F_2 \,, \tag{2}$$

where

$$F_1 = \Delta\left(\left\lfloor \frac{x_1+x_2}{2\Delta} + 0.5b \right\rfloor - 0.5b + 0.5\right), \tag{3}$$

$$F_2 = \Delta\left(\left\lfloor \frac{x_1-x_2}{2\Delta} + 0.5b \right\rfloor - 0.5b + 0.5\right). \tag{4}$$

The process of quantization according to DC-QIM is represented on Figure 2. Particular realization of quantized random variable $X' \in \mathbb{R}^2$ is denoted as $(x_1', x_2')$. DC-QIM defines $(x_1', x_2')$ in the following way:

$$x_1' = \alpha x_1^C + (1-\alpha)x_1 \,, \tag{5}$$

$$x_2' = \alpha x_2^C + (1-\alpha)x_2 \,, \tag{6}$$

where $\alpha \in [0,1]$.

Then, watermarked image should be transmitted and the watermark need to be extracted. A grid of fine lattice $\Lambda_f$ is constructed by smaller square Voronoi cells. The grid is used with the aim to extract information (Figure 1). Here, a half of the cells interpret "0" and the other half interpret "1" (Voronoi regions are filled with green and blue, respectively). For a cell with a particular bit label, we use name "extracting cell" and consider that it is inside the corresponding quantization cell.

An attack (or distortion) may occur during transmission. Therefore, for an extracted bit we use notation $\dot{b}$ instead of notation $b$ that was used for embedding. Therefore, the hard decision detector that we use is given as:

$$\dot{b} = 1 - mod\left(\left\lfloor \frac{\dot{x}_1'}{\Delta} - 0.5 \right\rfloor + \left\lfloor \frac{\dot{x}_2'}{\Delta} - 0.5 \right\rfloor, 2\right). \tag{7}$$

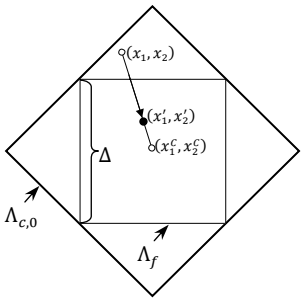



**Figure 2.** Conventional DC Quantization

The optimality (under AWGN) of the detector defined in (7) depends on the distribution of original samples inside quantization cell [8]. However it is the simplest one among known in the literature and provides reasonably good performance under an assumption that the distribution is close to uniform.

Even in case of no attack, in order to assure absolutely correct extraction of a watermark, all the elements $(x_1, x_2)$ need to be placed inside the corresponding extraction cell. However, from Figure 2 it is clear that the quantization shift in general is not orthogonal to the closest boundary of the extraction cell.

### *B. Modified Quincunx Quantization with DC*

The proposed idea of modification of 2D Quincunx can be explained using MLES concept. Under AWGN, a possible bit error is the result of the shift $(x_1', x_2') \xrightarrow{AWGN} (\dot{x}_1', \dot{x}_2')$ that can be directed randomly. Nevertheless, the smallest shift required to produce an error is the one directed normally to the nearest grid of $\Lambda_f$. According to AWGN model, the smallest possible shift has the highest probability and we use definition of MLES for that particular error case. Hence, for each particular pair of $(x_1, x_2)$, quantization in the direction opposite to MLES reduces probability of MLES with the minimal quantization distortion (Figure 3).

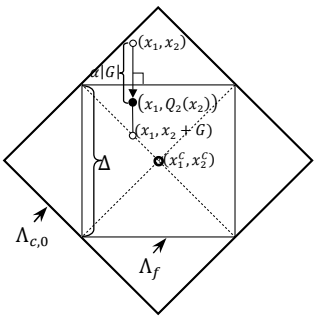



**Figure 3.** New MLES-based DC

As it can be seen from Figure 3, the shift is orthogonal to the nearest boundary of extraction cell and in case $\alpha = 1$ element $(x_1', x_2')$ reaches diagonal of the cell (not the center as the contrast to conventional DC-QIM). Elements that reach a diagonal can not be moved further. Otherwise, in order to counteract MLES efficiently, those elements would need to be moved along the diagonal toward the center. The quantization shift is denoted as $G$:

$$G = |x_1 - x_1^C| - |x_2 - x_2^C| \,. \tag{8}$$

Coordinates of a quantized element are defined as

$$(x_1', x_2') = \begin{cases} (x_1, x_2), & if\ G = 0; \\ (x_1, Q_2(x_2)), & if\ G < 0; \\ (Q_1(x_1), x_2), & otherwise; \end{cases} \quad (9)$$

where

$$Q_1(x_1) = x_1 - \alpha[sign(x_1 - x_1^C)]G\ , \quad (10)$$

$$Q_2(x_2) = x_2 + \alpha[sign(x_2 - x_2^C)]G\ . \quad (11)$$

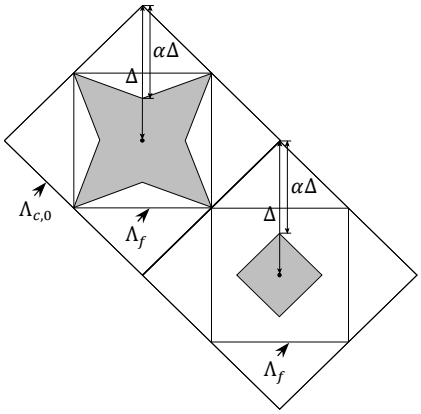



**Figure 4.** Distribution of quantized samples for MQ (top) and Conventional Quincunx (bottom)

Forms of distribution of samples quantized according to MQ and Conventional Quincunx can be compared on Figure 4 top and bottom quantization cells, respectively.

### *C. Watermarking Properties of Modified Quincunx*

The steps of watermarking procedure are:
- apply a transform to the original image $I$ and obtain a sequence of coefficients $\boldsymbol{x}$;
- choose (and synchronize with the receiver) a watermarking key $K$;
- using $K$, select pairs $\{(x_1, x_2)^1, (x_1, x_2)^2, \dots, (x_1, x_2)^N\}$ of coefficients;
- obtain pairs of quantized coefficients $\{(x_1', x_2')^1, (x_1', x_2')^2, \dots, (x_1', x_2')^N\}$ according to (9);
- replace original coefficients by corresponding coefficients from $\{(x_1', x_2')^1, (x_1', x_2')^2, \dots, (x_1', x_2')^N\}$ in accordance to $K$ and obtain $\boldsymbol{x}'$;
- apply the reverse transform to $\boldsymbol{x}'$ which will result in a watermarked image $I'$.

One of the most important characteristics for any watermarking method is embedding distortions. For the estimation of the distortions induced by the proposed quantization we assume that Δ is small enough to consider that the distribution of X inside any quantization cell is uniform. Important difference of the proposed approach compared to DC-QIM is that only one sample out of $(x_1, x_2)$ is being modified. The estimate of quantization distortions is denoted as $D$ and according to (10-11) should be expressed as

$$D = E\left((x_1 - Q_1(x_1))^2\right) = E\left((x_2 - Q_2(x_2))^2\right) = \alpha^2 E(G^2). \quad (12)$$

Referring to (8), the term $E(G^2)$ can be rewritten as

$$E(G^2) = E((x_1 - x_1^C)^2) + E((x_2 - x_2^C)^2) - 2E(|(x_1 - x_1^C)(x_2 - x_2^C)|)\ . \quad (13)$$

The form of quantization cell is symmetric and therefore

$$E((x_1 - x_1^C)^2) = E((x_2 - x_2^C)^2)\ . \quad (14)$$

Also, appealing to the symmetry we can calculate all the estimates from, for example, the first quadrant of the quantization cell. Quantization distortion in II, III and IV quadrants are the same. The following holds in the I-st quadrant

$$E(|(x_1 - x_1^C)(x_2 - x_2^C)|) = E\big((x_1 - x_1^C)(x_2 - x_2^C)\big). \quad (15)$$

Now, let us calculate (14) and (15). For simplicity we replace $x_1 - x_1^C$ with $\epsilon_1$ (also, $x_2 - x_2^C = \epsilon_2$). Because we consider only the first quadrant, parameters $\epsilon_1, \epsilon_2$ expressing quantization noise are positive there. We derive that marginal pdf

$$f(\epsilon_1) = \int_0^{\Delta - \epsilon_1} f(\epsilon_1, \epsilon_2) d\epsilon_2, \quad (16)$$

where constant joint pdf $f(\epsilon_1, \epsilon_2)$ is defined as

$$f(\epsilon_1, \epsilon_2) = \frac{1}{\rho},\ \rho = \int_0^{\Delta} d\epsilon_1 \int_0^{\Delta - \epsilon_1} d\epsilon_2 = \frac{\Delta^2}{2}\ . \quad (17)$$

Hence,

$$E(\epsilon_1^2) = \int_0^{\Delta} \epsilon_1^2 f(\epsilon_1) d\epsilon_1 = \frac{1}{\rho}\int_0^{\Delta} \epsilon_1^2 d\epsilon_1 \int_0^{\Delta - \epsilon_1} d\epsilon_2 = \frac{\Delta^4}{12\rho}\ . \quad (18)$$

Next,

$$E(\epsilon_1\epsilon_2) = \int_0^{\Delta}\int_0^{\Delta - \epsilon_1} \epsilon_1\epsilon_2 f(\epsilon_1, \epsilon_2) d\epsilon_2 d\epsilon_1 = \frac{1}{\rho}\int_0^{\Delta} \epsilon_1 d\epsilon_1 \int_0^{\Delta - \epsilon_1} \epsilon_2 d\epsilon_2 = \frac{\Delta^4}{24\rho}\ . \quad (19)$$

And, finally we calculate $D$ according to (13) and substitute the estimate for $\rho$:

$$D = 2\alpha^2[E(\epsilon_1^2) - E(\epsilon_1\epsilon_2)] = \alpha^2\frac{\Delta^2}{6}\ . \quad (20)$$

Calculated distortion $D$ for the proposed quantization method can be compared with other known scalar and 2D quantization methods mentioned in [9, 11] (Table 1). From Table 1 it can be seen that for the same pair $(\alpha, \Delta)$ the proposed approach provides the smallest embedding distortion.

**TABLE 1.** QUANTIZATION DISTORTION ESTIMATES. DISTRIBUTION OF ORIGINAL SAMPLES IS UNIFORM

| Quantization Method | Distortion |
|---|---|
| DC-QIM Scalar (1 bit/1 sample) | $\alpha^2\frac{\Delta^2}{3}$ |
| DC-QIM Qincunx (1 bit/2 samples) | $\alpha^2\frac{\Delta^2}{3}$ |
| **Modified Quincunx** (1 bit/2 samples) | $\alpha^2\frac{\Delta^2}{6}$ |
| DC-QIM Hexagonal #1 (1 trit/2 samples) | $\alpha^2\frac{5\Delta^2}{12}$ |
| DC-QIM Hexagonal #2 (2 bit/2 samples) | $\alpha^2\frac{5\Delta^2}{9}$ |

The watermarked image $I'$ should be transmitted via some channel where an attack might occur. Therefore, the receiver has access to possibly distorted watermarked image $\dot{I}'$ that he/she needs to extract the watermark from. The steps of extracting procedure are:
- apply a transform to the distorted image $\dot{I}'$ and obtain a sequence of coefficients $\dot{\boldsymbol{x}}'$;

- using $K$, select pairs $\{(\dot{x}_1', \dot{x}_2')^1, (\dot{x}_1', \dot{x}_2')^2, \dots, (\dot{x}_1', \dot{x}_2')^N\}$ of coefficients;
- from each pair $(\dot{x}_1', \dot{x}_2')^i$ extract a watermark bit $\dot{b}^i$ according to (7).

It is important that the procedure of watermark extraction for the proposed MQ scheme is the same as in the case with conventional DC-QIM Quincunx. Hence, any of the Quincunx embedding schemes can be used (without notifying the receiver) which increases the efficiency of watermarking under different types of distortions.

## III. Experimental Evaluation of the Method

In this section we experimentally evaluate robustness of the proposed quantization scheme. Further in the text, the term "extracted information" means maximized (over $\alpha$) mutual information between messages that are embedded and detected using a pair of samples. We use the hard decision detector that is described by equation (7). The robustness is estimated under consideration that AWGN attack is applied to the watermarked data.

The severity of AWGN attack is measured by WNR:

$$WNR = 10\, log_{10}\left(\frac{D}{\sigma^2}\right), \quad (21)$$

where $\sigma^2$ is the variance of AWGN. Substituting the expression for $D$ for the MQ we derive that $WNR \approx 20\, log_{10}(\alpha\Delta/\sigma) - 7.782$. The related question of watermarking performance was examined in details in [14]. Simplified settings assuming uniform distribution of original samples inside quantization interval are discussed in [15]. It was shown there that, in scalar quantization case the extracted information depends only on parameters $\alpha$ and $\Delta/\sigma$. Therefore, during the experiment, for a given pair $(WNR, \alpha)$ the required $\Delta/\sigma$ is calculated as:

$$\Delta/\sigma = \frac{1}{\alpha} 10^{\left(\frac{WNR+7.782}{20}\right)}. \quad (22)$$

For the purpose of the experiment, 72 natural grayscale images of different resolution were selected. For each image, 8x8 blocks were formed from adjacent pixels and passed to Singular Value Decomposition (SVD) transform. The choice of SVD is due to its popularity and quite moderate visual impact [16]. First singular values were chosen as the coefficients that form $\boldsymbol{x}$. A key $K$ was used to select pairs $\{(x_1, x_2)^1, (x_1, x_2)^2, \dots, (x_1, x_2)^N\}$ from $\boldsymbol{x}$. After corresponding grids with quantization cells are formed, uniform and independent 2D distribution inside cells has been confirmed according to the goodness of fit criterion. Therefore, previously derived analytical estimate for quantization distortion $D$ can be considered as precise in the case with natural signal as well.

In order to index WNR correctly, in the domain where the quantization takes place, AWGN attack was applied directly to the set $\boldsymbol{x}'$ of quantized coefficients (not the composed image $I'$). Bit Error Rate (BER) was calculated upon extraction of watermark data from $\dot{\boldsymbol{x}}'$. The diagram of the experimental approbation of the MQ is given on Figure 5.

On the diagram, original pairs of coefficients, quantized, and corrupted by AWGN are denoted as $\mathbf{X}, \mathbf{X}', \dot{\mathbf{X}}'$ respectively. Original and extracted watermark data is denoted as $\mathbf{Dt}$ and $\dot{\mathbf{Dt}}$. Procedure of embedding is denoted as $Em$ and requires $\mathbf{X}, \mathbf{Dt}, \Delta, \alpha$ as arguments. Attack is denoted as $AWGN$ and extraction is $Ex$.

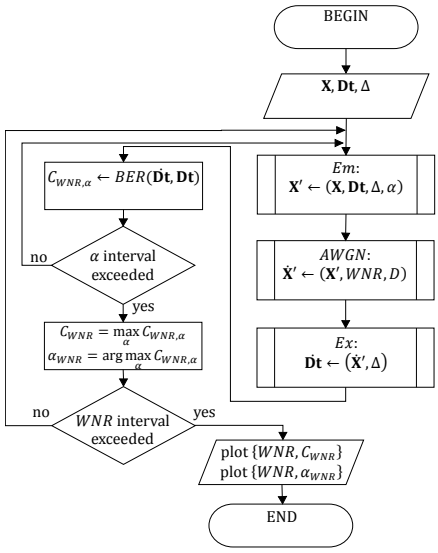



**Figure 5.** Diagram of the experiment

The final result of all iterations with the same WNR index is $\alpha_{WNR}$ that maximizes extracted information under attack with that WNR value. The whole $\alpha$-diapason is searched in order to find $\alpha_{WNR}$ for each WNR. The final result of the entire experimental approbation is the plot of the best $\alpha_{WNR}$ values drawn toward WNR (Figure 6).

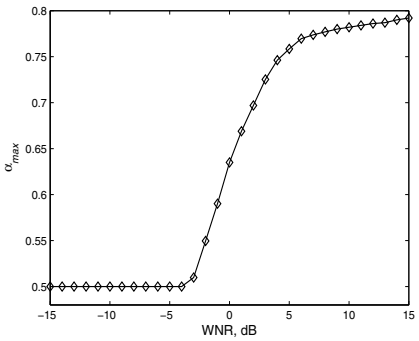



**Figure 6.** Optimal values for $\alpha$

It can be seen that optimal $\alpha$ values are increasing with the increase of WNR. It might be not clear on the stage of watermark embedding if the attack will occur in the channel during transmission. Therefore the settings of the experiment were so that in case of no distortion the extracted watermark does not contain errors. For that purpose it is required that $\alpha \geq 0.5$ (Figure 6).

The extracted information plot toward WNR is shown on Figure 7. Extracted information index $C$ has been calculated during the experiment using BER between $\mathbf{Dt}$ and $\dot{\mathbf{Dt}}$:

$$C = 1 + BER \log_2 BER + (1 - BER) \log_2 (1 - BER). \quad (23)$$

Information plots for other popular two-dimensional quantization methods are depicted on Figure 7. Comparing the performance of the MQ with the other known techniques we

can witness considerable advantage of the proposed technique under $WNR \leq -3\text{dB}$.

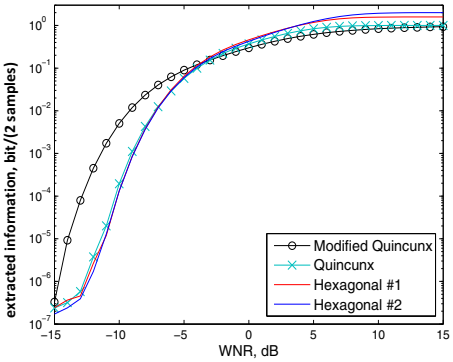



**Figure 7.** Performance for different Quantization Methods

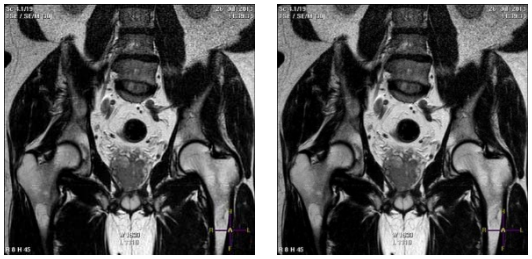
## IV. DISCUSSION

We will discuss in this section some of the most important watermarking characteristics of the Modified Quincunx.

Well-known Quincunx lattice was used which cell form is far from being hyperspherical and, therefore can not be considered optimal for Dither Modulation embedding. Nevertheless new DC technique improves watermarking properties. One of the most vivid advantages of the proposed method can be seen from Figure 7 for low WNRs. On the other hand it is clear that for higher WNRs the method's performance is the worst among the selected methods. Obviously, the explanation to this phenomenon can be found in the form of the distribution of the quantized samples. For instance, in case $\alpha = 0.5$ the distribution of the quantized samples is uniform and is inside extraction cell. Consistency and low quantization distortion make the method efficient under low WNRs. Higher WNRs require larger $\alpha$. In case $\alpha = 1$, the quantized samples are located only on the diagonals of the extraction cell. The distribution is spread far away from the center which is the main reason for poor performance in contrast to other methods. Luckily, the extraction procedures for the MQ and conventional Quincunx are the same and the advantages of the both methods can be combined. For instance, the solution might be to use MQ for $WNR \leq -3\text{dB}$ and to use Conventional Quincunx for $WNR > -3\text{dB}$.

In some cases, quantization distortion might have a crucial impact on a watermarked image. Fine Art Photography or medical imaging are good examples. In the latter case, a patient might want to assure protection for his/her personal data and limit its misuse or uncontrolled copying and distribution. However, an impact of watermarking can cause quality deterioration which results in incorrect diagnosis and subsequent treatment. In order to avoid this, watermarking energy should be low.

Unfortunately, low-energy watermarking faces some complications in digital images. One of such complications is due to discretization of pixel values of watermarked images [17]. It might cause significant BER for low-energy watermarking even prior any attack in the transmission channel. An evident advantage of the proposed in this paper quantization scheme is that only one sample out of $(x_1, x_2)$ needs to be modified during watermark embedding. Therefore, only one sample might suffer from discretization. Additionally, the estimate for $D$ for MQ is the lowest in Table 1 which means that there is higher capability to increase $\Delta$ compared to other DC-QIM based approaches. The developed technique has been tested in collaboration with medical experts. An image of prostate cancer case (Figure 8) was watermarked with Document to Watermark Ratio (DWR) equal 30dB. According to the final assessment, as a result of a double-blind review, made by medical expert group, the interpreted diagnostic information for original and watermarked images is identical. Hence, the fidelity of the watermarked images can be considered as acceptable for that particular medical application.



a) b)

**Figure 8.** Prostate cancer case: a) Original diagnostic image; b) Watermarked image

For the both MQ and Scalar DC-QIM (SDC-QIM) only one sample needs to be modified to embed 1 bit. We compared the efficiency of the methods SDC-QIM, Quincunx and MQ. With the aim to make comparison more versatile we consider three possible settings for AWGN attack on SDC-QIM. Nevertheless, settings for Quincunx and MQ are the same in each of the comparing scenarios: $N$ -bit watermark is embedded in $2N$ samples and AWGN is applied to the quantized $2N$ samples.

Scenario 1 for SDC-QIM is: $N$ samples are used for watermark embedding and AWGN attack energy is spread on $N$ quantized samples. Scenario 2: $N$ samples are used for watermark embedding but AWGN attack energy is spread on all the $2N$ samples. Scenario 3: the same $N$-bit watermark is embedded two times in $2N$ samples and AWGN attack energy is spread on all the $2N$ quantized samples. Extracted information vs. WNR plots for all the mentioned cases are shown on Figure 9.

Scenario 2 can be seen as the most beneficial for scalar quantization as in that case SDC-QIM outperforms Quincunx and MQ on all the WNR diapason. However, according to Scenario 2 fraudulent tampering (as a result of an attack) of only $N$ quantized samples out of total $2N$ samples can potentially be detected. As a contrast to that, Quincunx and MQ have potential to detect fraudulent tampering in all the $2N$ samples. This feature can be especially valuable for a number of semi-fragile and fragile watermarking applications where image integrity is important [18] [19] [20].

According to Scenario 3, tampering of all the $2N$ samples can potentially be detected using SDC-QIM. From Figure 9 it can be seen that SDC-QIM outperforms Quincunx on almost all the WNR diapason but performs worse than MQ for $WNR \leq -4\text{dB}$. Finally, SDC-QIM demonstrates the worst performance under Scenario 1 (except $WNR \geq 5\text{dB}$ where it slightly outperforms MQ).

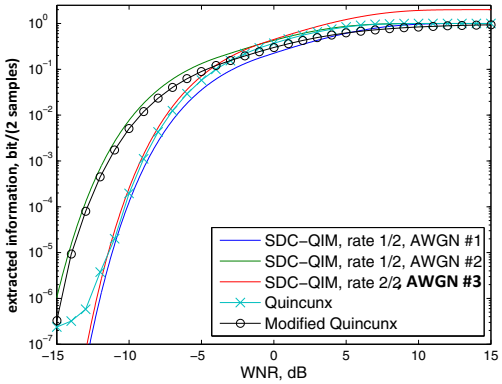



**Figure 9.** Information plots for Scalar and two-dimensional Quantization Methods

## V. Conclusions

A modified two-dimensional Quincunx-based quantization approach is proposed in this paper. Even though well-known lattice is used, a new Distortion Compensated technique results in a sufficient improvement of watermarking characteristics. In comparison to other two-dimensional approaches Modified Quincunx provides higher extracted information under intense AWGN. The demonstrated improvement is mainly due to using MLES concept that offers a new form for the distribution of quantized samples.

For a given pair of original samples, the direction of the most probable error-causing disturbance is defined according to MLES concept. The proposed DC quantization is performed in the opposite direction in order to reduce probability of MLES. Due to Quincunx-form cells, diagonals serve the natural limit for the quantization which results in a new form of the distribution of quantized samples. Quantization distortion and the form of the distribution depend on parameter $\alpha$ that value has been optimized for different levels of WNR. The procedure of watermark extraction for the Modified and conventional Quincunx is the same that allows interchangeable utilization of the methods without informing the receiver.

The efficiency of the Modified Quincunx has been estimated using natural grayscale images. According to the experimental data, the modification considerably increases extracted information (up to $10^2$ times) compared to other 2D quantization methods under low WNRs. Low quantization distortion reduces a negative impact on the perceptual quality which may contribute to a number of possible practical implementations of the method. Additionally, some properties of the method have been discussed and compared with scalar DC-QIM under several possible attack scenarios. Notably, in most cases the modified method is more advantageous than the scalar one.

Our further efforts will be to put forward an improved MLES-based DC quantization technique as well as to increase its dimensionality by utilizing, for instance, the structure of some well-known codes.

## Acknowledgment

The paper was supported by University of Tampere. Additional gratitude for provided diagnostic images and post-watermarking quality assessment is to Dr. Oleksandr Iatsyna and other members of collaborating oncological expert group from Cyber Clinic Spizhenko, Kiev, Ukraine.

## References

[1] I. Cox, M. Miller, J. Bloom, J. Fridrich and T. Kalker, Digital Watermarking and Steganography (2 ed.), San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2007.

[2] S. Chikkerur, V. Sundaram, M. Reisslein and L. J. Karam, "Objective Video Quality Assessment Methods: A Classification, Review, and Performance Comparison," *IEEE Transactions on Broadcasting,* vol. 57, no. 2, pp. 165-182, June 2011.

[3] V. Gorodetski, L. Popyack, V. Samoilov and V. Skormin, "SVD-based Approach to Transparent Embedding Data into Digital Images," in *Proceedings of the International Workshop on Information Assurance in Computer Networks: Methods, Models, and Architectures for Network Security (MMM-ACNS '01)*, 2001.

[4] B. Macq, J. Dittmann and E. J. Delp, "Benchmarking of image watermarking algorithms for digital rights management," *Proceedings of the IEEE,* vol. 92, no. 6, pp. 971-984, June 2004.

[5] H. C. Nguyen and S. Katzenbeisser, "Detection of Copy-move Forgery in Digital Images Using Radon Transformation and Phase Correlation," in *Proceedings of the Eighth International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP)*, Jul. 2012.

[6] J. Dittmann, M. Stabenau and R. Steinmetz, "Robust MPEG video watermarking technologies," in *Proceedings of the sixth ACM international conference on Multimedia (MULTIMEDIA '98)*, 1998.

[7] B. Chen and G. W. Wornell, "Quantization index modulation: a class of provably good methods for digital watermarking and information embedding," *IEEE Transactions on Information Theory,* vol. 47, no. 4, pp. 1423-1443, May 2001.

[8] P. Comesaña, F. Pérez-González and F. Balado, "On distortion-compensated dither modulation data-hiding with repetition coding," *IEEE Transactions on Signal Processing,* vol. 54, no. 2, pp. 585-600, Feb. 2006.

[9] J. J. Eggers, R. Bäuml, R. Tzschoppe and B. Girod, "Scalar Costa Scheme for Information Embedding," *IEEE Transactions on Signal Processing,* vol. 51, no. 4, pp. 1003-1019, APR. 2003.

[10] R. Zamir and M. Feder, "On lattice quantization noise," *IEEE Transactions on Information Theory,* pp. 1152-1159, Jul 1996.

[11] P. Moulin and R. Koetter, "Data-Hiding Codes," *Proceedings of the IEEE,* vol. 93, no. 12, pp. 2083-2126, Dec 2005.

[12] X. Wang and X.-P. Zhang, "A new implementation of trellis coded quantization based data hiding," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2008)*, Apr 2008.

[13] E. Esen and A. Alatan, "Data hiding using trellis coded quantization," in *Proceedings of the IEEE 12th Signal Processing and Communications Applications Conference*, Apr 2004.

[14] L. Pérez-Freire, F. Pérez-González and S. Voloshynovskiy, "An accurate analysis of scalar quantization-based data hiding," *IEEE Transactions on Information Forensics and Security,* vol. 1, no. 1, pp. 80-86, March 2006.

[15] Y. Zolotavkin and M. Juhola, "Quantization Based Watermarking Approach with Gain Attack Recovery," in *Proceedings of the IEEE International Conference on Digital Image Computing: Techniques and Applications (DICTA'14)*, Nov 2014.

[16] Y. Zolotavkin and M. Juhola, "A new blind adaptive watermarking method based on singular value decomposition," in *International Conference on Sensor Network Security Technology and Privacy Communication System (SNS & PCS), 2013*, May 2013.

[17] J. Fridrich, Steganography in Digital Media: Principles, Algorithms, and Applications, New York, USA: Cambridge University Press, 2010.

[18] J. Domingo-Ferrer and F. Sebe, "Invertible spread-spectrum watermarking for image authentication and multilevel access to precision-critical watermarked images," in *Proceedings of International Conference on Information Technology: Coding and Computing*, April 2002.

[19] W. Pan, G. Coatrieux, N. Cuppens-Boulahia, F. Cuppens and C. Roux, "Watermarking to Enforce Medical Image Access and Usage Control Policy," in *Proceedings of the Sixth International Conference on Signal-Image Technology and Internet-Based Systems (SITIS)*, Dec. 2010.

[20] S. C. Pei and Y. C. Zeng, "Tamper proofing and attack identification of corrupted image by using semi-fragile multiple-watermarking algorithm," in *Proceedings of the 2006 ACM Symposium on Information, computer and communications security (ASIACCS '06)*, 2006.