Annales UMCS Informatica AI XII, 4 (2012) 123–139
DOI: 10.2478/v10065-012-0021-3

Annales UMCS
Informatica
Lublin-Polonia
Sectio AI
http://www.annales.umcs.lublin.pl/

Security issues on digital watermarking algorithms

Wioletta Wójtowicz[1*], Marek R. Ogiela[1†]

[1] *AGH University of Science and Technology*
30 Mickiewicza Ave, 30-059 Krakow, Poland

Abstract – This paper gives a general introduction to the digital watermarking procedures and their security aspects. The first issue is to clarify unifying and differentiating properties of steganography and watermarking. Then the most important aspects of digital watermarking are reviewed by studying application, requirement and design problems. We put emphasis on the importance of digital watermark as an effective technology to protect intellectual property rights and legitimate use of digital images. In the paper we provide an overview of the most popular digital watermarking methods for still images available today. The watermarking algorithms are divided into two major categories of spatial and transform domains. Because of outstanding robustness and imperceptibility the transform domain algorithms are the mainstream of research. Popular transforms of images include the DFT (Discrete Fourier Transform) ([**1, 2, 3, 4, 5**]), DCT (Discrete Cosine Transform) ([**1, 3, 6, 5**]) and DWT (Discrete Wavelet Transform) ([**1, 3, 4, 7, 6, 5**]). In the paper we emphasize the advantageous features of DWT such as local time-frequency and multi-scale analysis, preserving the quality of host image and ensuring high robustness of watermark. Finally, we present three algorithms which are based on the combination of DWT and some other transformations like DFT ([**4**]), DCT ([**6**]) and the Arnold transform ([**7, 6**]). Finally, we discuss security requirements and possible attacks on the watermarking systems.

1 Introduction

Today after the fast development in multimedia technologies, the data transmission and transformation we are concerned with are digital, rather than analog, communication and media. But the technology development has its disadvantages such as illegal accessibility to private information for people. Therefore it is essential to have

*wwojtowi@agh.edu.pl
†mogiela@agh.edu.pl

a knowledge to be able to limit the illegal accessibility to private information. To deal with this serious problem of copyright protection and data authentication, the watermarking technique could be applied. Watermarking just like steganography is strongly connected with cryptography - secure communication used for protecting information in computer systems ([**8, 9**]). Although both watermarking and steganography describe techniques that are used to imperceptibly convey information by embedding it into the cover data, they differ from each other in many aspects.
Steganographic methods are usually not robust against modification of the data, or have only limited robustness and protect the embedded information against technical modifications that may occur during transmission and storage, like format conversion, compression, or digital-to-analog conversion.
Watermarking, as opposed to steganography, has the additional requirement of robustness against possible attacks. Robustness has strong implications in the overall design of a watermarking system. Watermarks do not always need to be hidden, as some systems use visible digital watermarks, but most of the literature has focused on imperceptible (invisible, transparent, or inaudible, depending on the context) digital watermarks which have wider applications. A popular application of watermarking is to give a proof of ownership of digital data by embedding copyright statements. It is obvious that for this application the embedded information should be robust against manipulations that may attempt to remove it. The information hidden by a watermarking system is always associated with the digital object to be protected or with its owner while steganographic systems just hide any information ([**1, 2**]).

2 Digital watermarking characteristics

The digital watermarking technique (watermark insertion or watermark embedding) is generally speaking hiding useful information in the multimedia data (cover media). The hidden information (a watermark), may be the serial number or random number sequence, copyright message, ownership identifier, text, binary or gray level image etc. After watermark embedding the original cover media become slightly modified and the modified content is called the watermarked content. In this paper we are focused just on the invisible watermarks inserted in the digital still images, especially their properties and algorithms of their insertion. It will be proven that depending on the watermarking application and purpose, different requirements arise resulting in various design issues (e.g. [**4, 7, 6**]).
For the real-world digital watermarking systems in still images, a few very general properties, shared by all proposed systems, can be identified. One of them is watermark imperceptibility which is a common requirement and independent of the application purpose. It is normally required to have a watermarked image which should perceptually be as close to the original image as possible. To evaluate the imperceptibility, generally, the Peak Signal to Noise Ratio (PSNR) can be used. The PSNR is the least mean square errors between the original and watermarked images (Fig. 4).

The additional requirements have to be taken into consideration when designing watermarking techniques ([**1**]):

- **Robustness** of the watermarked data against modifications and/or malicious attacks is one of the key requirements in watermarking. However, there are applications where it is less important than in others.
- **Watermark security** is essential because in most applications, such as copyright protection, the secrecy of embedded information needs to be assured.
- **Capacity** of the watermarking system is defined as the maximum amount of information that can be embedded in the cover work.
- **Computational cost** should be as low as possible because the watermarking method with high complex algorithms will require both more software and hardware resources.

The ultimate watermarking method should resist any kind of distortion introduced by the standard or malicious data processing. Thus, practical systems must implement a compromise between the robustness and the competing requirements like invisibility and information rate.

2.1 The most popular watermarking applications

The requirements that watermarking systems have to comply with are always based on their application. In general watermarking methods have to be robust, but different levels of required robustness can be identified. One of the simplest classification determines two basic types of watermarks: robust and fragile watermarks ([**3**]). Fragile watermarks are those that have only very limited robustness. They are applied to detect modifications of the watermarked data, rather than conveying information.
Applications of robust watermarks:

- **Copyright protection** is probably the most prominent application of watermarking today. The objective is to embed information about the source, and thus typically the copyright owner, of the data in order to prevent other parties from claiming the copyright on the data. Thus, the watermarks are used to resolve rightful ownership, and this application requires a very high level of robustness.
- **Fingerprinting** relates to watermarking applications where information such as the creator or recipient of digital data is embedded as watermarks. This is useful to monitor or trace back illegally produced copies of the data that may circulate, and is very similar to serial numbers of software products.
- A robust watermark sometimes can be used to the **access control** of some systems. Then it is desirable to have a copy and usage control mechanism to prevent illegal copy of the content or limit the number of times of copying.

Applications for fragile watermarks:

- In commercial applications, the **authentication of document** is very important. Thus, to protect document, a large number of authentication techniques can be used and one of them is the fragile watermark embedding.
- The second approach of the fragile watermark is the **authentication of evidence** in the court. The images of crime or violation can be used as evidence in the court, but the authentication of such evidence should be proved.
- **Complete authentication** checks the integrity of the entire multimedia. To authenticate the content, a simple way is to compute a signature and use this signature to perform authentication. Any change in the content will cause change of the signature, thus we can detect the modification.

2.2 Watermarking systems

A typical watermarking system is shown in Fig. 1. There are two key parts: a watermark embedding system (watermark insertion, watermark embedding) and a watermark recovery system (also called watermark extraction or watermark decoder).

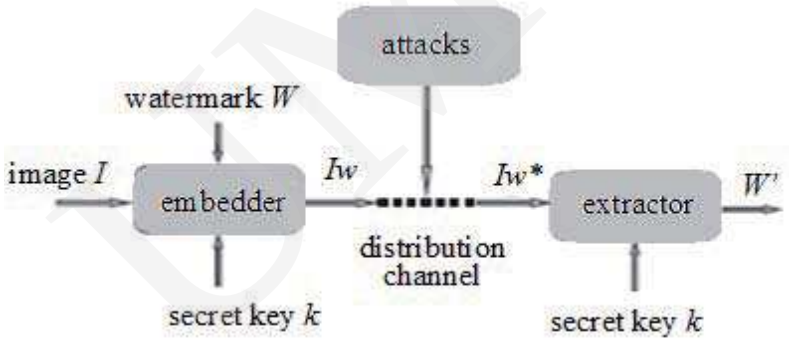

Fig. 1. Typical watermark system model.

The input to the scheme is the watermark W, the cover image I and an optional public or secret key k. The key may be used to enforce security of the system. All practical systems employ at least one key, or even a combination of several keys (i.e. Arnold transform in **[7, 6]**). The output of the watermarking embedder is the watermarked content Iw, which should be as close to I as possible in most cases. Then in the distribution channel (i.e. the Internet) the watermarked image Iw is exposed for the authorized or unauthorized attacks, which could make it modified (Iw^*). As a consequence, the inputs to the watermark extractor are the modified watermarked content Iw^*, secret key k and depending on the method, the original image or the watermark. Finally the output of the extractor is either the recovered watermark W' or some confirmation if Iw^* contains the watermark W. There are some types of watermarking schemes arising from different combinations of inputs and outputs of the

extractor ([**3**]): nonblind watermarking (systems require at least the original image, which is used as a hint to find where the watermark could be in Iw), semiblind watermarking (does not use the original data for detection), blind watermarking (it requires neither the original image I nor the embedded watermark W). To sum up, some issues

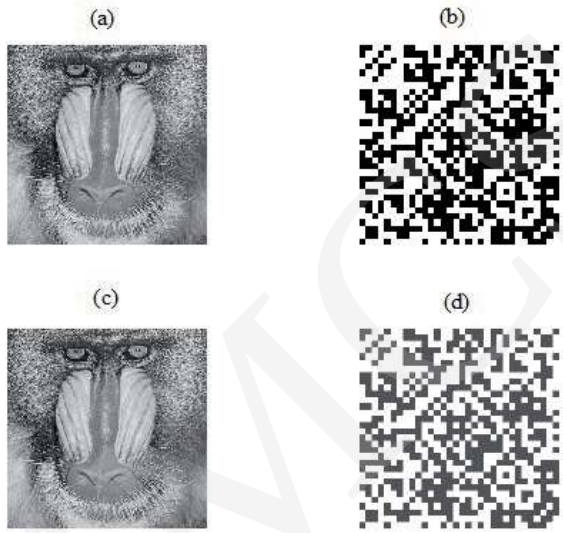

Fig. 2. Simple watermarking scheme in DWT domain, original image (a), watermark (b), watermarked image (c), extracted watermark (d).

of designing watermarking systems should be identified. Firstly if Iw is unmodified, the detected watermark W' should be exactly the same as W (Fig. 2.). Secondly, for robust watermarking if Iw is modified, W' should still match W well to give clear judgment of existence of watermark. Finally, for fragile watermarking, W' will be totally different from W after even the slight modification to Iw. What is more, W' indicates the possible tampering to Iw^* and gives information about degradation of Iw.

3 Watermarking domains

The watermark embedding techniques should assure that insertion of the watermark modifies the cover image in a perceptually invisible manner. For the sake of imperceptibility and robustness, it is worth considering different image domains, especially that transformation of image domain could be treated as watermarking technique as well. There are two general categories: spatial domain techniques and transform domain techniques, that are described in this section.

The application purpose and desired robustness of watermark influence the design process. For example if we need a method that is resilient to the JPEG compression with

high compression factors, it is probably more efficient to employ a method working in a transform domain than to use the one that works in the spatial domain. Similarly, if the method should accommodate generalized geometrical transformations, that is rotation, scaling and shearing, an approach in the spatial domain is probably more suitable.

3.1 Spatial domain watermarking

The spatial domain is the most straight forward way to hide the information, because the watermark is directly embedded in the cover image. As a direct consequence, one of the advantages of spatial domain watermarking is simplicity of the proposed methods. Another advantage could be an exact control of the maximum difference

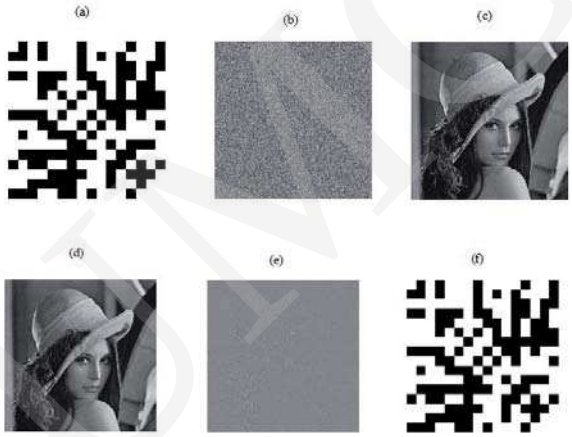

Fig. 3. Watermarking scheme with watermark as modulated pseudo random noise; watermark (a), noise modulation of watermark (b), original image (c), original image with modulated watermark (d), demodulation of watermark (e), detected watermark (f), ([**5**]).

between the original and watermarked contents as well as possibilities of designing near-lossless watermarking systems. The most common techniques, which are used for watermarking in the spatial domain, are:

- adding the pseudo random noise pattern to the intensity of image pixels - the most straightforward method, the noise signal is usually integers like $(-1, 0, 1)$, the noise is generated by a key and should not correlate with the content of the image (Fig. 3).
- Least significant bit (LSB) modification is a common method for embedding the watermark in the host data. The method consists in the manipulation of LSBs of images in a manner which is not detectable and imperceptible to human eyes. The basic idea is that the LSBs of original 8-bit grey level

image are discarded first, then the LSBs are replaced by the permuted binary watermark of the same size as the original one (Fig. 4).

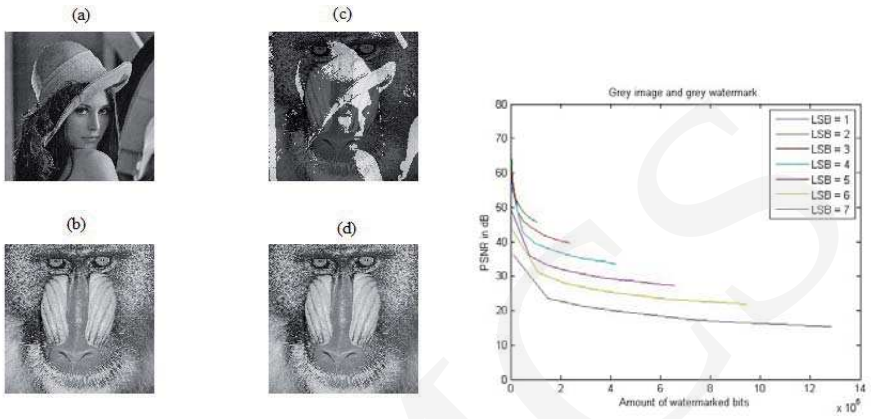

Fig. 4. LBS algorithm for grey cover image and grey watermark, original image (a), watermark (b), watermarked image (c), watermark - only the most significant bits (d) and PSNR as a function of amount of watermarked bits.

3.2 Frequency domain watermarking

Transformation of image from the spatial to the frequency domain rises both imperceptibility and robustness of watermarking schemes. The most popular transforms operating in the frequency domain are: Discrete Fourier Transform (DFT), Discrete Cosine Transform (DCT) and Discrete Wavelet Transform (DWT).

3.2.1 Discrete Fourier Transform

Any digital image of the size $N_1 \times N_2$ can be treated as a matrix with N_1 rows and N_2 columns. The intensity values of image pixels, which are represented by the elements of this matrix are denoted by the two-dimensional signal $f(n_1, n_2)$, where $n_1 = 0, 1, 2, \ldots, N_1 - 1$, $n_2 = 0, 1, 2, \ldots, N_2 - 1$. For each pair of frequencies $(k_1, k_2) \in N_1 \times N_2$ we can define the Discrete Fourier Transform as follows

$$F(k_1, k_2) = \frac{1}{\sqrt{N_1 N_2}} \sum_{n_1=0}^{N_1-1} \sum_{n_2=0}^{N_2-1} f(n_1, n_2) \exp(-\frac{i2\pi n_1 k_1}{N_1} - \frac{i2\pi n_2 k_2}{N_2}). \quad (1)$$

This representation provides many interesting characteristics of the image, which could be developed in watermarking techniques. It is essential to study how geometrical transformations of the image in the spatial domain affect this image in the DFT domain. Below there are the examples of the most popular transformations: translation, scaling and rotation

$$f(n_1 + a, n_2 + b) \leftrightarrow F(k_1, k_2)\exp(-i(ak_1 + bk_2)), \tag{2}$$

$$f(n_1 a, n_2 b) \leftrightarrow F(k_1/a, k_2/b), \tag{3}$$

$$f(n_1 cos\theta - n_2 sin\theta, n_1 sin\theta + n_2 cos\theta) \leftrightarrow F(n_1 cos\theta - n_2 sin\theta, n_1 sin\theta + n_2 cos\theta). \tag{4}$$

It seems that shifting in the spatial domain is cleared like the linear shifting in the DFT phase. The scaling in the spatial domain appears to be an inverse one in the frequency domain. The rotation of image just identical to θ in the spatial domain is caused by the rotation of DFT of image to the same extent. Some of the changes like rotation and scaling affect the DFT magnitude. However, replacing the DFT transform with an invariant one like the log-polar mapping (5) (LMP), which is also called the Fourier-Mellin transform ([**3, 4**]) has some advantages

$$u = e^{\rho} cos(\theta), \qquad v = e^{\rho} sin(\theta). \tag{5}$$

After applying the DFT, which is invariant to translation, every amplitude in the DFT at the position (u, v) is projected in a new coordinate space (ρ, θ). In this new coordinate system, rotation and scaling are converted into translation. By calculating the amplitude of the DFT of the LMP, the resulting domain is invariant rotation, scaling, and translation (RST). This property is very important, especially that the watermarks in the DFT domain are robust against translations. As a consequence, construction of the transform domain, which could be insensitive to rotations and scaling is possible.

In practice in watermarking techniques, which are based on the DFT transform, DFT is used twice times. Firstly, when we change the domain of original cover image and secondly, after changing the coordinates to the log-polar ones. This transformation is more often used than the common DFT.

3.2.2 Discrete Cosine Transform

For every block (size $N \times N$) of image, the connection between the spatial domain image pixels $f(n_1, n_2)$, and the DCT domain coefficients $F(k_1, k_2)$ is

$$F(k_1, k_2) = \sum_{n_1=0}^{N-1} \sum_{n_2=0}^{N-1} C(n_1)C(n_2)f(n_1, n_2)cos(\frac{\pi(2n_1 + 1)k_1}{2N})cos(\frac{\pi(2n_2 + 1)k_2}{2N}), \tag{6}$$

where $C(n_1) = C(n_2) = \sqrt{\frac{1}{N}}$ for $k_1 = k_2 = 0$ or $C(n_1) = C(n_2) = \sqrt{\frac{2}{N}}$ for $k_1 = k_2 = 1, 2, \ldots, N - 1$. After computing DCT of image, the frequencies are divided into three main categories: low, medium and high frequencies. Due to the fact that the low frequencies are sensitive to image distortions and high frequencies are eliminated during the JPEG compression, the watermark is embedded in the medium frequencies. Embedding the watermark in these coefficients enables robustness against compression (which was impossible in the DFT domain) and other distortions ([**6**]). This method has also some disadvantages such as a lack of robustness to the geometrical distortions (rotation, scaling etc.).

3.2.3 Discrete Wavelet Transform

DWT also provides excellent space for image watermarking. Contrary to DFT and DCT, DWT is a hierarchical transformation, which enables analysis of image in the spatial-frequency domain. Then watermark could be added to the coefficients of transformation, that are exposure and frequency functions. Wavelets representation of any one-dimensional signal could be generalized, therefore the wavelet multi-resolution representation of the image is proven ([**3**]). The entity of this method is decomposition of the two-dimensional signal $f(n_1, n_2)$ into the sequence of signals with the decreasing resolution. The algorithm decomposes the original image into four subimages and obviously every component has a dimension equal to a quarter of the original image dimension. The subimages could be next decomposed recursively in the same way. Finally, the representation of the image on many resolution levels could be obtained. In practice to get the wavelet decomposition, the highpass filter (H) and lowpass filter (L) are applied to the rows and columns of the image in the spatial domain. Mathematically the wavelet transformation of the image is the convolution operation on the image pixels. Decomposition could be repeated recursively and for the nth level of decomposition, the following components of image are obtained:

$$L_n = [L_{n_1} * [L_{n_2} * L_{n-1}]_{\downarrow 2,1}]_{\downarrow 1,2} \tag{7}$$

$$D_n^V = [L_{n_1} * [H_{n_2} * L_{n-1}]_{\downarrow 2,1}]_{\downarrow 1,2} \tag{8}$$

$$D_n^H = [H_{n_1} * [L_{n_2} * L_{n-1}]_{\downarrow 2,1}]_{\downarrow 1,2} \tag{9}$$

$$D_n^D = [H_{n_1} * [H_{n_2} * L_{n-1}]_{\downarrow 2,1}]_{\downarrow 1,2} \tag{10}$$

where: $*$ is the convolution operator, L_0 - the original image, $\downarrow 2,1$ - the sampling of rows, $\downarrow 1,2$ - the sampling of columns, H_{n_1}, L_{n_1} - the highpass and lowpass filtres along the rows, H_{n_2}, L_{n_2} - highpass and lowpass filtres along the columns.
The lowpass filters remove high frequency elements from the image. Thus from the image the convulsive changes in intensity of adjacent pixels are removed. Due to this, the edges on the image seem to be blurred. These filters are usually used to eliminate noise or to smooth slightly the image. The highpass filter extracts high frequency elements from the image. These filters are used to specify the details of the image or the edges of objects. The L_n component is obtained by lowpass filtering of the component L_{n-1}, thus for n equal to 1 it complies the original image. Then D_n^l, for $l = H, V, D$ are obtained by filtering L_{n-1} in appropriate directions. Thus these components include: horizontal, vertical and diagonal details of the image respectively. The three-level wavelet decomposition scheme is shown in Fig. 5. It is proven that the details of image are placed in the low frequency region (LL_n), therefore inserting the watermark in this area will damage quality of the image. On the other hand, the high frequency coefficients include less information (HH_n), so they are not robust against the compression. For these reasons, the watermark is usually inserted in the medium frequency area (LH_n and HL_n).

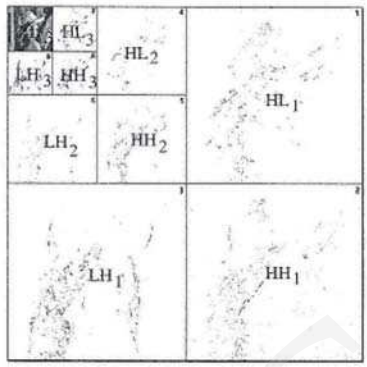

Fig. 5. Multiscale decomposition of the "Lena" image, ([**3**]).

4 A survey of current watermarking techniques in the frequency domain

The copyright of still images is currently studied in the majority of publications in the field of watermarking. Based on several recent publications related to this issue, this section discusses three actual algorithms for data embedding in digital still images. These methods proceeded from combination of wavelets and other transforms to improve robustness, imperceptibility and security of watermarking schemes.
Manoochehr, Pourghassem i Shahgholian in [**4**] suggest the method based on the combination of DWT and Fourier-Mellin Transform. The proposed algorithm is evaluated in various logos and images watermarks. The embedding procedure starts with computation DWT of the cover image and the watermark. After one-level decomposition the horizontal or vertical subband (the region of medium frequency) is selected from the cover image and from the watermark. Afterwards FMT is applied to both components and finally these subbands are combined. Using the inverse transformations IFMT and IDWT, the watermarked image can be obtained. The described algorithm

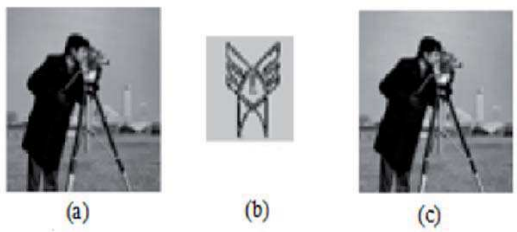

Fig. 6. Original image (a), logo (b), watermarked image (c), ([**4**]).

is tested on various families of waves and in most cases there is no difference between the original image and the watermarked one. The influences of attacks like rotation, noise, low pass filtering, brightness decrease, brightness increase, resizing are measured on the watermarked image. It seems that DWT is robust against the noise attack and FMT is robust against the geometrical transformations, therefore the robustness of watermarked image against the mist of attacks is improved by combination of these two transforms (Fig. 6).

In [**6**] there was proposed the blind watermarking algorithm based on DCT and DWT. To evaluate this method the host grey scale image and the binary image as the watermark are used. To ensure the security of scheme, the watermark is three times Arnold scrambled before embedding and the number of scrambling is used as the key k_1. Algorithm starts with the three level wavelet decomposition of original image and selection of LL_3 component for further analysis. Afterwards the selected subband is divided into $8 * 8$ non overlapping blocks. Then DCT is performed on each block and the transformation coefficients are obtained. In the next step the medium frequency coefficients are selected to embed the watermark as the following method. If the values of watermark bits are equal to 0, the values of pseudorandom sequence X_t , generated by the key k_2 are added to the medium frequency coefficients. In the opposite case, the medium frequency coefficients are not changed. To obtain the watermarked image for each block, IDCT is performed to reconstruct LL_3, then IDWT is used three times to obtain the watermarked image. In order to validate the effectiveness and

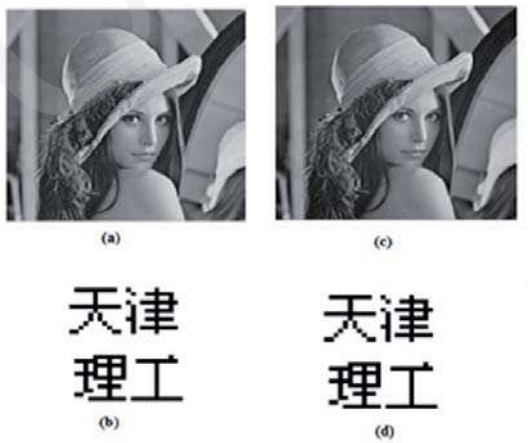

Fig. 7. Original image (a), watermark (b), watermarked image (c), extracted watermark (d), ([**6**]).

robustness of the proposed method, there are carried out a series of image attacks on the watermarked image. The performance of algorithm against the common attacks such as Gaussian noise, JPEG compression and cropping is satisfactory, but when

the watermarked image is subjected to image scaling attack, the algorithm does not perform very well (Fig. 7).

QiWei Lin, JiSheng and XuFeng Wu propose blind digital image watermarking algorithm based on the multi-strategy ([**7**]). The described method is combined with the DWT decomposition, spread spectrum technique and Arnold transformation. The multi-strategy technique and the watermark spread spectrum code technique are adopted and several copies of bilevel watermark are interlaced in different resolutions of image in the wavelet domain. Applying the distinct dimensional (f_1, f_2, f_3, f_4) Arnold transformation to the coefficient of four subband image: LH_1, HL_1 and LH_2, HL_2 we can obtain the spread spectrum code and the secret key (f_1, f_2, f_3, f_4). Since the intensity of the embedding watermark is different affecting various resolution layers of the wavelet factor, so in these resolution layers we employ the difference embedding strategy (different watermark embedding intensity factors, pseudo random sequence as a marker). Finally, the secret key, inverse Arnold transform and IDWT are used to obtain the watermarked image.

This quite complex method rises the imperceptibility and the robustness to resist various attacks. The embedding different intensity spread spectrum watermark codes obviously improve imperceptibility. On the other hand, the embedding watermark on different resolution layers of the DWT domain, using different embedding factors or different lengths of the spread spectrum code must rise the robustness of the scheme. Due to the fact that LH_2 and HL_2 are the root from LL_1, they are located in the lower frequency band and have a smaller size, so the change of their coefficients will have a greater effect on the watermarked image, therefore fewer data can be embedded in LH_2 and HL_2. On the contrary, in the LH_1 and HL_1 layers, the above mentioned issue can be well settled. As a result, many properties of the watermarking scheme

Fig. 8. Watermarked image after clipping (a), original watermark (b), extracted watermark (c), ([**7**]).

were improved. The robustness of the proposed algorithm (JPEG compression, adding noise, and etc.) as well as the imperceptibility and security of the watermarked images were improved (Fig. 8).

5 Security of watermarking systems and possible attacks

A digital watermark security refers to the inability of the unauthorized users to modify, remove, detect or estimate the watermark. The aim of an attacker is usually to eliminate, remove or degrade the effectiveness of the watermark, to disable the detector or to attack the concept of the watermarking application. An attack is considered successful if the attacker disrupts any stage of the watermarked life cycle (see Fig. 1). Thus, the content owner and the watermarking software have to ensure that each stage is secured against such manipulations.

Depending on the watermarks applications, their security properties have to fulfill different requirements. In ensuring the ability of watermarking techniques to meet these requirements, it is necessary to identify all possible risks and make some assumptions about the capabilities of the adversary as well. For example, if the adversary knows *nothing about the watermarking algorithm*, he or she must rely on general knowledge of the weaknesses from which most watermarking algorithms suffer. But in some cases, it is possible for the adversary to obtain *more than one watermarked image*. The adversary can often exploit this situation to remove watermarks, even when he or she does not know the algorithm (e.g. collusion attacks). Additionally, for the systems that require a very high level of security, it is better to assume that the adversary knows *everything about the algorithm* except one or more secret keys. Such adversary may be able to find and exploit weaknesses in the detection strategy. Finally, in some cases, we can assume that the adversary has *a watermark detector*. Then even if the adversary knows nothing about the algorithm, access to a detector gives him or her an advantage in attacking the watermark.

An attack can be described as any processing that circumvents the intended purpose of the watermarking technique for a given application. According to this, watermarking attacks include normal processing operations, like image compression, and unintentional destruction of the watermark. These distortions are limited to those that do not produce excessive degradations, since otherwise the transformed object would be unusable. Researchers recognize many types of possible attacks on watermarking schemes, each of them exploits a different stage of the watermarking process. Three broad categories of unauthorized actions are described as follows ([**2, 3**]).

(1) **Unauthorized embedding** There are some cases when the adversary composes and embeds a watermark on his or her own.
- copy attack - the aim of this attack is to copy a watermark from one carrier image to another;
- ambiguity attack - create the appearance that a watermark has been embedded in an image when in fact, no such embedding takes place (the adversary can use this attack to claim ownership of a distributed image);
- overmarking - the operation when the watermark is embedded in an already marked carrier image.

(2) **Unauthorized detection** In most applications we are primarily concerned with preventing people from actually decoding what a watermark says. But there are some applications, where the adversary might be satisfied with simply knowing whether or not a watermark is present and does not need to know what message the watermark encodes.
- collusion attacks - the attackers can estimate the original image if they have the same work with different watermarks or estimate the watermark if they have different works with the same watermark;
- oracle attacks - the attackers are allowed to detect watermarks (if they have access to a watermark detector), but not remove them.

(3) **Unauthorized removal** Security against unauthorized removal is required in all watermarking applications. A watermark is considered removed from an image if it is rendered undetectable. The removal of watermarks represents the most obvious form of attacking a watermark. For a watermark to be secure against unauthorized removal, it must be robust to any process that maintains the fidelity of the image. This process may be a normal process, like compression, in which case we require that a secure watermark be robust. However, it may also be a process unlikely to occur during the normal processing of the image. The most common categories of such pathological distortions are described in the following.
- signal processing operations - the robustness of the watermark against a wide range of filtering and processing operations is formulated as a necessary feature of the watermarking technology. Robustness against common signal operations such as the addition of noise or localized signal distortions is often achieved by using the spread-spectrum signalling techniques in the design spread spectrum algorithms. Spreading the watermark energy over a large spectrum minimizes the spectral density and this makes such trivial attacks impractical;
- linear filtering and noise removal - linear filtering also can be used by the adversary in an attempt to remove a watermark, e.g. a watermark with significant energy in the high frequencies might be degraded by the application of a low-pass filter;
- synchronization attacks - the examples of simple synchronization distortions include rotation, scaling, and translation for images. More complex distortions include shearing, horizontal reflection, and column or line removal in images. Even more complex distortions are possible, such as nonlinear warping of images.
- scrambling attacks - the type of scrambling can be a simple sample permutation or a more sophisticated pseudo-random scrambling of the pixels values. The degree of necessary scrambling depends on the detection strategy. The only constraint is that the scrambling is invertible or near invertible. The scrambling is simply the subdivision of the image into

subimages, and the descrambling is accomplished by the web browser itself.

The performance of attacks depends also on the type of the cover image and the embedded watermark. The most popular robust watermarks are noise watermark (e.g. Fig. 2), logo watermark (e.g. Fig. 6) and message watermark (which is included in the text).

For the developer of the watermarking algorithm, it is essential to begin with identifying the security requirements and judging the usability of the watermarking technology. The variety of cover image type, watermark style, embedding technique and application of the system carries the attacks performance over. The experimental results of the algorithms taken from [**4**] and [**6**], which were presented in section 4, could set an example of possible scenarios. In [**4**], where the logo watermark was embedded in the gray scale image, the method was based on the FMT and DWT fusion, whereas in [**6**] forthe bi-level image as the watermark and the gray scale cover image, the DCT and DWT combination algorithm was used. To demonstrate the robustness of the proposed algorithms, the effects of image distortions on the normalized correlation between the original and extracted watermark (NCC, [**4**]) were investigated. In both cases the proposed methods have major advantage over their traditional versions.

Attack	The lena test image		
	PSNR	*NC(traditional method)*	*NC(the proposed method)*
No Attack	37.29	0.9951	1
Gaussian Noise(10%)	21.53	0.9473	0.9684
Salt and Pepper Noise(10%)	20.36	0.8193	0.9856
JPEG(QF=75)	32.74	0.9902	0.9963
Cropping(S=20%)	15.32	0.9551	0.9712
Rotation(1° in clockwise)	29.98	0.5098	0.9213
Scaling(S=50%)	18.59	0.8350	0.5176

Attack	FMT+DFT	FMT+DWT
Rotate30	0.8576	**0.9912**
Rotate90	0.8224	**0.9913**
5x5 Gaussian Filtering (var=2)	0.9638	**0.9913**
Gaussian noise (var=2)	0.9554	**0.9908**
Histogram Equalization	0.9554	**0.9915**
Increase Brightness (+0.5)	0.9546	**0.9917**
Decrease Brightness(-0.5)	0.8430	**0.9917**
Resize (0.5)	1.0000	**0.9915**
Median filtering (3x3)	0.9988	**0.9914**
Contrast	0.9561	**0.9912**

Fig. 9. Results of the robustness against various attacks for the FMT–DWT([**4**]) and DCT-DWT algorithms, ([**6**]).

The results from Fig. 9 are elaborated in right papers, but with regard to them and, in fact, to many others mentioned in this and previous sections on watermarking issues, we will draw some conclusions. Having knowledge of the underlying algorithm enables the attacker to design an attack, specific for an algorithm or a class of algorithms by finding and exploiting their weaknesses. This, in turn, enables experts to examine and validate the techniques or to publish potential security flaws. To follow the considerations of attacks on watermarking systems, a few general rules should be pointed out to:

(1) The influences of most common degradation attacks like additive noise and cropping can be minimized by spreading the watermark over the dimensions where this attack takes place. The transform domain techniques are useful in

this case to spread the information all over the image so that, if possible, any remaining part could include enough information for the watermark recovery.

(2) The experiment results show that for the group of filtering and smoothing attacks, the robustness of watermark could be ensured due to embedding watermarks mid-range frequencies in the transform domain because these regions are less affected by such operations than others.

(3) Despite the fact that compression is generally an unintentional attack, very often appears in multimedia applications and its effects should be avoided as well. It is usually advisable to perform the watermark insertion task in the same domain where the compression takes place. For instance, the DCT domain image watermarking is more robust to the JPEG compression than the spatial-domain watermarking.

(4) Regarding the rotation and scaling attacks, it is proposed to use some rotation and scaling-invariant transforms (such as the Fourier-Mellin Transform, [**4**]) but one should be aware that this transformation dramatically reduces the capacity for message hiding.

6 Conclusions

This paper discusses the most important watermarking security issues. Based on several key points of watermarking in still images, a wide range of techniques has been presented. In general, invisible watermarks serve as tools for digital copyright protection, but it is not the only usage. As presented, there are various general scenarios to which watermarking may be applied. Due to the variety of aimed applications, the watermarks may have different properties, but in most cases the trade-off between imperceptibility and robustness must be gained. The great part of the paper was devoted to the analysis of two watermarking domains: the spatial domain and the frequency domain. Some practical examples of algorithms are also presented. Due to the numerous advantages of watermarking in the frequency domain, the most popular transformations: DFT, DCT and DWT were marked. Then three actual and effective watermarking algorithms were presented. Each of them arose as a combination of DWT and other techniques to improve robustness and imperceptibility of the proposed methods. Finally, the security of designed schemes and the types of possible attacks were also discussed. In many cases, the problems of unauthorized embedding and detection are directly analogous to the problems studied in cryptography. Although the analogy between the watermarking and the cryptography is useful to counter unauthorized embedding and decoding, its utility is not clear in the case of unauthorized removal. Thus, when the adversary is neither authorized to embed nor to decode, we turn to spread the spectrum techniques, rather than the cryptographic ones, to prevent unauthorized removal.

Acknowledgements

This work was supported by the AGH Univeristy of Science and Technology statutory research Grant No. 11.11.120.612

References

[1] Arnold M., Schmucker, M., Wolthusen S.D., Techniques and Applications of Digital Watermarking and Content Protection, Artech House (2003).

[2] Cox I. J., Miller M. L. et al., Digital watermarking and steganography, Morgan Kaufmann Publishers (2008).

[3] Katzenbeisser, S., Petitcolas, F.A., Information Hiding Techniques for Steganography and Digital Watermarking, Artech House (2000).

[4] Manoochehri, M., Pourghassem, H., Shahgholian, G., A Novel Synthetic Image Watermarking Algorithm Based on Discrete Wavelet Transform and Fourier-Mellin Transform, IEEE 3rd International Conference on Communication Software and Networks (ICCSN) (2011): 265.

[5] Zieliński T.P., Cyfrowe przetwarzanie sygnałów, Wydawnictwa Komunikacji i Łączności (2005).

[6] Wang Huai-bin, Yang Hong-liang, Wang Chun-dong, Wang Shao-ming, A new Watermarking Algorithm Based on DCT and DWT Fusion, International Conference on Electrical and Control Engineering (ICECE) (2010): 2614.

[7] QiWei Lin, JiSheng, XuFeng Wu, A New DWT and Multi-Strategy Watermark Embedding Algorithm, IEEE International Conference on Anti-Counterfeiting, Security and Identification (ASID) (2011): 57.

[8] Kołodziejczyk M., Ogiela M. R., Applying of security mechanisms to middle and high layers of OSI/ISO network model, Theoretical and Applied Informatics 24 (1) (2012): 95.

[9] Ogiela, M. R., Systemy utajania informacji - od algorytmów do kryptosystemów szyfrujących, Wydawnictwa AGH (2003).