Robust Image Watermarking Theories and Techniques: A Review

Hai Tao\textsuperscript{1}, Li Chongmin\textsuperscript{2}, Jasni Mohamad Zain\textsuperscript{1}, Ahmed N. Abdalla\textsuperscript{3}

\textsuperscript{1}Faculty of Computer System and Software Engineering, University Malaysia Pahang, Malaysia
\textsuperscript{2}Department of mathematics and information, Qinghai Normal University, China
\textsuperscript{3}Faculty of Electrical and Electronic Engineering, University Malaysia Pahang, Malaysia

ABSTRACT
Over the past several decades, digital information science has emerged to seek answers to the question: can any technique ensure tamper-resistance and protect the copyright of digital contents by storing, transmitting and processing information encoded in systems where digital content can easily be disseminated through communication channels? Today it is understood that the answer is yes. This paper reviews the theoretical analysis and performance investigation of representative watermarking systems in transform domains and geometric invariant regions. Digital watermarking is a technology of embedding watermark with intellectual property rights into images, videos, audios, and other multimedia data by a certain algorithm. The basic characteristics of digital watermark are imperceptibility, capacity, robustness and false positive of watermarking algorithm and security of the hiding place. Moreover, it is concluded that various attacks operators are used for the assessment of watermarking systems, which supplies an automated and fair analysis of substantial watermarking methods for chosen application areas.

Keywords: watermarking, robust, attacks, security.

1. Introduction
Because of the fast and extensive growth of network technology, digital information can be distributed with no quality loss, low cost and nearly instantaneous delivery. Protection of multimedia content has recently become an important issue because of the consumers’ insufficient cognizance of the ownership of intellectual property. Thus, over the past several decades, digital information science has emerged to seek answers to the question: can researchers ensure tamper-resistance and protect the copyright of digital contents by storing, transmitting, and processing information encoded in systems where digital content can easily be disseminated through communication channels? Today it is understood that the answer is yes, and many research groups around the world are working towards the highly ambitious technological goal of protecting the ownership of digital contents, which would dramatically protect inventions represented in digital form for being vulnerable to illegal possession, duplication and dissemination [83]. Digital watermarking [16] is the process of embedding or hiding digital information called watermark into a multimedia product, and then the embedded data can later be extracted or detected from the watermarked product, for protecting digital content copyright and ensuring tamper-resistance, which is indiscernible and hard to remove by unauthorized persons.

Digital watermarking is seen as a partial solution to the problem of securing copyright ownership [80]. Essentially, watermarking is defined as the process of embedding sideband data directly into the samples of a digital audio, image, or video signal. Sideband data is typically “extra” information that must be transmitted along with a digital signal, such as block headers or time synchronization markers. It is important to realize that a watermark is not transmitted in addition to a digital signal, but rather as an integral part of the signal samples. The value of watermarking comes from the fact that regular sideband data may be lost or modified when the digital signal is converted between formats, but the samples of the digital signal are (typically) unchanged[72].

To clarify this concept further, it is useful to consider an analogy between digital watermarks
and paper watermarks. Watermarks have traditionally been used as a form of authentication for legal documents and paper currency. A watermark is embedded within the fibers of paper when it is first constructed, and it is essentially invisible unless held up to a light or viewed at a particular angle. More importantly, a watermark is very difficult to remove without destroying the paper itself, and it is not transferred if the paper is photocopied. The goals of digital watermarking are similar; in the next section, it will be shown that digital watermarks require similar properties.

Before the concept of watermarking can be explored further, three important definitions must first be established. A host signal is a raw digital audio, image, or video signal that will be used to contain a watermark. A watermark itself is loosely defined as a set of data, usually in binary form, that will be stored or transmitted through a host signal. The watermark may be as small as a single bit, or as large as the number of samples in the host signal itself. It may be a copyright notice, a secret message, or any other information. Watermarking is the process of embedding the watermark within the host signal. Finally, a key may be necessary to embed a watermark into a host signal, and it may be needed to extract the watermark data afterwards[16].

Up to now, two traditionally-used strategies, spatial-domain [68] and transform domain [28][80] techniques have been developed for digital image watermarking. The former category is designed to insert directly a watermark into the original image by a factor, which would lead to fair-quality watermarked images. The latter approach, for taking advantage of perceptual properties, is devised to embed a watermark into the frequency-domain of the original images. These types of watermarking schemes have good performances of robustness in comparison to the most common signal processing manipulations such as JPEG compression, filtering, and addition of noise [16][14][40][49][59]. Signal processing operators are applied to watermarked images for removing the watermark or decreasing its energy so that the extracted watermark is unrecognizable or insufficient as the validate evidence. Unfortunately, the ineffectiveness of existing traditional watermarking algorithms is described by the robustness against unintentional or malicious geometric attacks [37]. Geometric attacks induce synchronization errors between the original and the extracted watermark during the detection process. In other words, the watermark still exists in the watermarked image, but its positions have been changed. Therefore, while traditional watermarking systems require the creation of a framework of the resilience to watermarked data geometrical modifications, creation and enforcement of synchronization errors correction of such frameworks is now possible. Besides facilitating more efficient copyrighted protection and robustness against desynchronization, adaptation of geometrically invariant image features can potentially offer a greater robust capacity to detect watermarks without synchronization errors, especially when applied to survive local distortions such as random bending attacks. Development of such a framework is an essential starting point for organizations that wish to improve or replace currently existing watermarking algorithm-based pixel frequency or other transform coefficients for watermark embedding, and develop a set of means to establish and maintain feature-based watermarking of geometric distortions correction.

In the first section, the properties of the general watermarking frameworks that are exploited in the process of encoding and detecting watermarking are shortly reviewed. A survey of the key digital image watermarking algorithms and techniques is presented subsequently. The characteristics of watermarking systems are described for evaluating the performance of watermarking systems. There are five important issues that are usually considered in the most practical application; they are highlighted in the following subsections. In addition, digital watermarking is described as an efficient method for the protection of ownership rights of digital audio, image, video and other data types. It can be applied to different applications including digital signatures, fingerprinting, broadcast and publication monitoring, authentication, copy control, and secret communication. Watermarking attacks can be classified into two broad categories: destruction attacks: including image compression, image cropping, spatial filtering, among others; and synchronization attacks: including image rotation, image shifting and pixelshine deletion. The chapter lists and describes some of these conventional attacks in the following sections.

For constructing geometric invariant watermarking, four mainstream schemes are introduced by literature reviews on watermarking algorithms robust to the
geometrical distortions. Most of these efforts confine to theoretically analyzing and quantifying the effect of the global and local affine transform to the performance of the watermarking algorithms.

2. Watermarking backgrounds

With the rapid proliferation of globally-distributed computer networks technologies and popularity of multimedia systems, fashionable and economical digital recording and storage devices have made it possible to construct the platform, where it became considerably facilitated to not only acquire, represent, replicate, distribute and transmit multimedia contents in digital formats without degradation of quality, but also manipulate them. General-purpose computers and graphics editing programs provide ultimate playgrounds for an amateur who does not have any ripe experience and professional skills. He conveniently processes an image or tamper specific objects without remaining any appreciable traces, for introducing the best in your digital images, reconstructing them into anything you can imagine, and demonstrating them in extraordinary ways. In the multimedia publishing industries, researchers, scientists, and practicing engineers attach close importance to pervasive advancements because the unauthorized manipulation and the unrestricted reproduction of original digital multimedia can easily be disseminated through communication channels such as the Internet. Consequently, there is an urgent demand for effective techniques to ensure tamper-resistance and prevent pirates from causing damage to the owners of digital content. Because of possible copyright issues, the copyright and intellectual property of digital multimedia data should be protected from illegal possession, duplication and dissemination. Three complementary techniques are being introduced: encryption, steganography and watermarking [16][67][24].

Encryption [44] is a conspicuous and secure technique to converse data into a scrambled code that can be distributed and deciphered through a private or public network. Generally speaking, in both research and application fields, encryption and cryptographic algorithms serve copyright owners as an approach to protect the secure transmission of confidential multimedia data between a distributor or publisher and the purchaser of the multimedia data over public channels. Using variations of symmetrical and asymmetrical styles or forms of encrypting data, the permuted original multimedia contents are non-recognizable in appearance, unsystematic, and disorderly [82]. Although encryption algorithms can be applied to avoid illegal access to digital contents, it appears that encryption by itself is not sufficient enough to prevent an unauthorized pirate from illegally replicating multimedia content and protect multimedia data all along its lifetime. Once multimedia content has been decrypted into its original style and the protection of information is invalidated for further manipulations because there is no degradation of quality in subsequent works and no verification differences between one copy and any other derivative copy. Therefore, unauthorized replicating copy and transmission of multimedia data cannot be obstructed [83].

Steganography [54][77] represents a technique that is used to convey communicating secret data by writing hidden messages into an appropriate multimedia carrier, e.g., audio, image and video. The existence of the message is suspected except by the sender and intended receiver. Unlike cryptography techniques, the goal of steganography is to conceal the very existence of the hidden messages, together with avoiding arouse suspicion and not attracting attention to themselves. However, steganography conventionally involves associating secret point-to-point transmission and communication. Thus, steganography approaches are typically not resistant to transformation of the carriers, or hold only restricted robustness.

The definition of digital watermarking [16-17] emerges while trying to overcome the limitations of encryption and steganography in enforcement and protection of intellectual property rights. Compared to the idea of encryption, the watermark information is inserted into its original form and does not hinder users from listening to, viewing, watching, or manipulating the content. And unlike steganography, digital watermarking technologies are to establish the identity of information to avoid the unauthorized embezzlement. Generally, additional information is embedded directly into the original multimedia or host signal which is useful and valuable, and the message itself is not unnecessary to be secret.
3. Image watermarking frameworks

3.1 Image watermarking framework

Digital watermarking systems typically include two primary components: the encoder and the decoder. The inputs are the cover media data, the embedding security key, and watermarks in the watermark encoder. The encoder inserts a machine-readable code (watermark) into audio, video, and pictures with variant embedding algorithms, conceptions and schemes by modifying physical or electronic media and almost all watermarking procedures are controlled by private keys, which are assigned to the insertion and extraction procedure to extract the watermark information suitably and to warrant fundamental security. The outputs are the security key and the watermarked contents in the watermark encoder. A watermark extractor or detector involves a two-step process. Watermark retrieval is the first step that applies some scrambling algorithms to extract a sequence referred to as retrieved watermarks. Then, in the second step, the embedded watermarks are detected and extracted from a suspected signal of containing watermarks. The second step normally requires the analysis and comparison of the unreliable watermark with the original one, and the consequences could be several kinds of confidence assessment displaying the similarity between the extracted watermark and the original one.

3.2 Encoder

In order to combine a watermark with a digital document, for example, images, you need an image (CO), a watermark (W) that contains the watermark information, a security key(K) and an encoding algorithm (E) to create a watermarked image (CW). The encoder takes the signature and the cover document and generates the watermarked image, which is described as a function:

\[ CW = E(CO, W, K) \]  

(1)

In this case, secret or public keys and other parameters can be used to extend the watermarking encoder. The watermark is considered to be robust if it is embedded in such a way that the watermark can survive even if the watermarked data CW go through severe distortions.

3.3 Detector

The watermark detection procedure is depicted as follows:

\[ W' = e(CW, K, \ldots) \]  

(2)

where e(.) is the detection algorithm whilst C and W are the optional inputs for the detection function.

The watermark is extracted using a decoder function (e). In this case, the decoder D loads the watermarked, normal or corrupted image CW, and extracts the hidden signature W. Using nonblind and blind watermarking techniques in Figure 1 and Figure 2, the decoder D loads an additional image CO, which is often the original image, to extract the watermarking information by correlation.

4. Characteristics of watermarking systems

It is essential to define the criteria of a watermarking system for the comparison results against equivalence of group of the assessment attained by assessing the performance of other watermarking methods. Obviously, each watermarking system should have particular properties regarding the given application; therefore, there is no unique set of properties that all watermarking systems have to satisfy. Generally, there are five important issues that are usually considered in the most practical application; they are highlighted in the following subsections.
4.1 Imperceptibility

Imperceptibility is an essential condition for digital watermarking; that is, the visual similarity between the watermarked version and original one of the media element and the perceptual quality of the original signal should be transformed imperceptibly by the insertion of the watermark. There are two main reasons why it is important to keep the imperceptibility of the host media after the encoding with watermark data. Firstly, the presence or absence of a watermark cannot be distinguished from the primary purpose of the original media, if the watermarked media is so badly distorted that its value is lost. In addition, suspicious perceptible artifacts may introduce a watermark in existence, and perhaps its precise location being detected from host media. This information may provide accesses for distorting, substituting, or removing the watermark data maliciously. Therefore, the information embedded in it may no longer be available.

4.2 Robustness

One of the most commonly measured properties is that watermark signals must be reasonably resilient to various attacks and common signal processing operations in digital watermarking systems. Once some watermark signal is inserted in the original content, distortions may be applied to the signal unavoidably when the signal is encoded, decoded, and distributed across the Internet. These distortions may be designed to apply the expected distortion to the watermarked signals or compress it before transmission, and they may or may not significantly disrupt the watermarked signals. It is impossible for a watermarking system to be robust against all signal processing operations whereas the requirement is application subordinate and dependent. For the digital watermarking of images, the good watermarking method is likely to resist against noise addition, filtering processing, geometrical transformations such as scaling, translation and rotation, and also JPEG compression.

4.3 Capacity

Capacity is defined using the largest quantity of information that inserted watermarks are capable of hiding, and embedded watermarks can be extracted credibly for the purposes of authentication, and copyright safeguards. Under the condition of imperceptibility as well as the requirements of robustness, the capacity relies on the size of the original data. The more original patterns are attainable, more bits are able to be embedded. However, inserting as much watermark information as possible is a more difficult task in digital watermarking. Very often, a prerequisite for capacity relies on the practicable application used for watermarking. For the audio, the capacity would relate to the amount of inserted bits in every second that is communicated. For images, the capacity may refer to the amount of embedded bits into pixels or patterns of the images. For the video, the capacity refers to either the amount of bits in every second or the bits' amount per frame. In a word, the fewer the amount of bits of capacity included in a watermark, the larger the opportunity of it being computationally complex; fewer false positives or finer granularities and bigger capacity will enlarge the potential operations of the data inserting method and construct the verification judgment more credible.

Note: Therefore, the conditions of imperceptibility, robustness, and capability are conflicted and limited by each other. One may want to increase the watermarking strength in order to increase the robustness but these results in a more perceptible watermark. On the other hand, under the condition of imperceptibility, a watermark would have to be created with the maximum possible separation to avoid a situation where a small corruption of the watermarked image would lead to erroneous watermark detection. Similarly, one can increase the data payload by decreasing the number of samples allocated to each hidden bit but this is counterbalanced by a loss of robustness. In other words, for any watermarking scheme, it is impossible to meet these three requirements simultaneously. As a result, a good trade-off among these requirements has to be achieved.

4.4 Security

All existing watermarking algorithms which are not secure cannot be used for copyright protection, data authentication, or tracking the illegal distribution of digital content. Therefore, the watermarking algorithm is safe and robust, if the attacker, using watermarking procedures and knowledge, does not know the key used for
watermarking digital content. Thus, the hidden watermark information cannot be destroyed or damaged. In addition, the complexity of the watermark process may be safety-related because the attacker will be discouraged to search the insertion in an embedding space and long key position. Therefore, in order to improve the security of the algorithm, it can enlarge the embedded space, and increase the size of the keys split into small pieces of cover image.

4.5 False positive

A false positive is defined as not actually containing the watermark in the process of watermark detection. It refers to the amount of false positives that it is predictable to happen in a precondition amount of runs of the detector. Likewise, the possibility can be discussed about any precondition detector run by a false positive occurring. There are two subtle distinctive ways to describe this probability, which are often confusing in some papers. The two diverge in whether the host image or the watermark is contemplated the arbitrary variable. In the first explanation, the false positive probability is the possibility that precondition a settled host image and arbitrarily chosen watermarks, the detector will state that a watermark exists in that image. The watermarks are constructed from a perturbation that is defined by the design method of a watermark construction. Conventionally, watermarks are generated either by a Gaussian sequence or by a bit-encoding algorithmic rule, unrelated to random number generating systems. In common situations, the false positive probability, depending on this first definition, is truthfully sovereign of the host image and only rely on the approach of generating a watermark. In the second definition, the false positive possibility is that randomly chosen images and preconditioned a settled watermark. The detector will retrieve that watermark in an image. The perturbation is greatly application-based determined, where the image is chosen. Medical images, natural images, music videos, graphics, and surveillance video all possess very distinctive statistics. Moreover, while these perturbations are varied from each other, also they are probable to be specific varied from the statistics method of the watermark generation systems. Hence, this second definition of false positive probabilities is absolutely distinctive from the first definition of them.

5. Classification of digital watermarking applications

5.1 Digital Watermarking for Copyright Protection

Copyright protection appears to be one of the first applications for which digital watermarking were targeted. The metadata in this case contains information about the copyright owner. It is imperceptibly embedded as a watermark in the cover work to be protected. If users of digital content (music, images, and video) have easy access to watermark detectors, they should be able to recognize and interpret the embedded watermark and identify the copyright owner of the watermarked content. An example of one commercial application created for that purpose is Digimarc Corporation’s ImageBridge Solution. The ImageBridge watermark detector is made available in a form of plug-ins for many popular image processing solutions such as Adobe PhotoShop or Corel PhotoPaint. When a user opens an image using a Digimarc-enabled application, Digimarc’s watermark detector will recognize a watermark. It will then contact a remote database using the watermark as a key to find a copyright owner and his contact information. An honest user can use that information to contact the copyright owner to request permission to use the image.

5.2 Fingerprinting

Additional data embedded by watermark in this application is used to trace the originator or recipients of a particular copy of multimedia file. For example, watermarks carrying different serial or ID numbers are embedded in different copies of multimedia information before distributing them to a large number of recipients. The algorithms implemented in fingerprinting applications need to be invisible and must also be invulnerable to intentional attacks and signal processing modifications such as lossy compression or filtering. Fingerprinting should be resistant to the collusion attack, that is, it is impossible to embed more than one ID number in the host multimedia file; otherwise, a group of users with the same image containing different fingerprints would be able to collude and validate the fingerprint or create a copy without any fingerprint.
5.3 Copy Control

Watermarks can also be used for copy prevention and control. Fragile watermarks can be used for copy control by having digital player devices detect a fragile watermark and refuse to play a music file or a video clip if no proper signature watermark is detected, preventing people from making illegal copies of copyrighted material. The main challenge that such systems face is that the whole system will only work if all player devices contain a watermark detector. Users will always choose a device that can play and record illegal copies. Actually, a copy protection mechanism that includes digital watermarking at its core is currently being considered for standardization and second generation DVD players may well include the ability to read watermarks and act based on their presence or absence [15].

The above represent a few example applications where digital watermarks could potentially be of use. In addition, there are many other applications for rights management and protection like tracking use of content, binding content to specific players, automatic billing for viewing content, broadcast monitoring, among others. From the variety of potential applications exemplified above, it is clear that a digital watermarking technique needs to satisfy a number of requirements. Since the specific requirements vary with the application, watermarking techniques need to be designed within the context of the entire system in which they are to be employed. Each application imposes different requirements and would require different types of invisible or visible watermarking schemes or a combination thereof.

6. Attack operators

The assessment of watermarking systems supplies an automated and fair analysis of substantial watermarking methods for chosen application areas. At present, numerous investigators and researchers utilize their own designed assessment systems, which does not require the capability of comparison each other. Therefore, The assessment procedure can be very complicated, and the current research is on assessment methods with unique attacks for images (for example, attainable tools Optimark [51], StirMark [65], Checkmark [10] or for particular applications like DRM [43][5][58]. A categorization of extensive watermarking attacks for evaluating the robustness is presented in [29], where the attacks are classified into protocol, geometrical, removal, and security attacks. In [71], the description is expanded by going into evaluation attacks or [27][42][9][8] present attacks to acquire knowledge about the secrets of the procedures (inserting and/or retrieval/ detection) for also evaluating the security of watermarking schemes. Furthermore, the devoting to the security and robustness estimation, for example, in [11] the imperceptibility of different watermarking methods is evaluated; the thesis illustrates some of these traditional attacks in the following sections. The authors examine other generally employed attacks for watermarking techniques in [71] [80].

6.1 Removal attacks

Removal attacks intent to accomplished removing watermarks from the host image. This classification includes lossy compression, denoising, demodulation, quantization, averaging and collusion attacks.

6.1.1 Denoising and lossy compression attacks

This category of attacks is relatively broad and contains common image processing operators such as lossy compression, image denoising and quantization. Image denoising, also understood as filtering, is mainly related to maximum likelihood, a minimax criterion or a minimum mean square error, a maximum a posteriori probability. The resulting filtering operator is decided by the selected criteria, and also by the priors on the cover image and the watermark. Compression is a popular scheme for attacking watermarked images or audio. Two common compression schemes: lossy compression, such as VQ compression, and JPEG compression for image processing have lately been considered to have approximately the same impact on noise removal as denoising. For the attackers to remove the hidden watermarks, they may compress the watermarked images with some other VQ codebooks and decode the VQ indices to get the reconstruction. The VQ compression schemes are effective for attacking some of the existing algorithms. Both lossy compression and denoising can importantly diminish the capacity of the watermarking channel establishing the output of various substitutable channels to zero for each bit of watermark.
6.1.2 Remodulation attacks

Since lossy compression and denoising have been widely presented in the literature, with some applications of low bit rate coding and image enhancement, respectively; it is not incredible that they are also famous attack tools for the watermarking community. On the other hand, remodulation attacks are a rather fresh theory unique to the watermarking attacks. A systematic remodulation attack was first demonstrated in [32]. In this algorithm, the watermark was forecasted using subtracting from the host stego image to the median filtered version of stego image. The forecasted watermark was also truncated, high-pass filtered, and the subtraction is done from the stego image with a constant amplification parameter with2. Since the median filtering mainly takes away the noise in the high-frequency section, the low-frequency section cannot correctly estimate the value according to this filter. In the situation of a highly consonant between the amplification parameter and the estimated watermark, the attacks have the guidance to a diminishment in extensive correlation in the matched filter with decoding.

An almost equivalent attack with weighted mean forecasting was introduced in [26]. In this work, authors reported their success at removing watermarks generated by the watermarking strategy introduced in [78] and the Digimarc commercial software. In addition, in [66], a Wiener attack is presented. The presented attack comprises three steps: forecasting of the watermark according to the Wiener filter, subtracting from the stego with some strength parameter to the estimated watermark, and adding stationary Gaussian noise. In [50], the impact of the attacks is discussed from the information-theoretic point of view and it is concluded that the attack of additive white Gaussian noise can be optimal asymptotically with respect to removing the watermark when the intensity of the noise is big in comparison to the energy of the watermark.

6.1.3 Averaging and collusion attacks

In this group, other attacks are collusion attacks and statistical averaging. With respect to the collusion attacks, numerous examples of the same data are attainable, but the attacked data set is constructed by possessing only a little section of each data set and reconstructing a novel attacked data set from these sections at this time. The latter illustrates an attack where many samples of a precondition data set, each time logged in with a different secret key or distinctive watermark, are averaged to evaluate the attacked data. For example, each frame can be inserted using a different key or a different watermark into video watermarking schemes. If the amount of data set is sufficiently huge, the inserted watermark cannot be discovered anymore supposing that it will output zero mean on average. In [20], collusion and the averaging attacks are discussed in using videos and complementary countermeasures are recommended. The other kind of attack that diminishes the decoding and detection of the watermark is the mosaic attack [55]. The attack was produced in the structure of automatic copyright protection frameworks that investigate the Internet and download images for checking the existence of the watermarked images on pirate websites. The mosaic attack does not attempt to remove the watermark with some image processing approaches, but rather it leads to producing problems for the watermark detection splitting the image into the small fragments.

6.2 Geometrical attacks

Compared to the removal attacks, geometric deformation attacks do not plan for the removal inserted watermark, but for distortion of it through temporal or spatial transformations of the stego contents. The attacks are normally described as follows: the detected watermark loses synchronization with the inserted information. The most famous integrated software versions for geometrical attacks are Stirmark [56] and Unzign watermark removal software from 1999. The global attacks are scaling, rotation, translation, change of aspect ratio, and shearing a link up with a kind of extensive affine transformation. The translation /cropping and column/line removals are also merged in Stirmark. Unzign presents local pixel jittering and is extremely efficient to attack watermarking schemes in spatial domains. Stirmark presents both local and global geometric distortions. Most current watermarking approaches are robust against these attacks owing to the application of particular synchronization procedures. If the robustness of global affine
transformations is a little or a lot a resolved problem, the local random transformations integrated by Stirmark always remain an open issue almost for all methods. The random bending attacks exploit the background that the human visual system is insensitive towards local affine modifications and shifts. Thus, the locally shifted, rotated and scaled pixels are without distortions in significant visual aspects. The thesis will also discuss dedicated attacks, which intend to test the efficiencies of proposed algorithms.

6.3 Cryptographic attacks

Cryptographic attacks are quite equivalent to the attacks applied in cryptography. There are the seriously forced attacks which intend to discover secret information using the exhaustive searches. Ever since numerous watermarking systems utilize a secret key, it is greatly significant to use keys with a safe length. In addition, another attack is the so-called Oracle attacks in this category [13] and [53], which is able be applied to produce a non-watermarked image while a device of a watermark detector is attainable..

6.4 Protocol attacks

The protocol attacks intend to attack the definition of the watermarking applications. The protocol attack was introduced by [19]. They present the structure of unidirectional watermark and demonstrate that watermarks requisite for being non-invertible in applications of copyright protection. The concept of inversion comprises of the truth that attackers who have a copy of the stego contents can represent that the data also includes the attackers' watermark using subtracting to his own watermark information. The activities can produce an ambiguity's condition with respect to the authentic ownership of the contents. The prerequisite of non-inevitability on the watermarking system suggests that it should not be potential to detect or extract a watermark from non-watermarked images.

6.5 Image Shifting and Line Deletion

The attackers may change the watermarked image nearby vertically and horizontally, or remove an entire line of pixels, to distort the watermark information delivered. For the embedding watermarks in the VQ or DCT domains, image shifting may result in the algorithm of extracting the watermark to miss the resynchronization of the watermarked images. How to acquire an acceptable quality in the watermarked image and to preserve the capability for recovering the embedded watermark with the image shifting scheme is another topic for robust watermarking. Although the aforementioned categorization makes it potential to have an understandable segregation between the different kinds of attacks, it is unavoidable to the reminder that a malicious attacker usually uses not only a single attack, but rather a combination of various attacks at the moment. Such a probability is forecasted in the Stirmark benchmark, where all geometric transformations are practically accompanied by the attack of lossy compression.

7. Transform domains

Spatial domain watermarking is attractive because it provides a better intuition on how to attain optimal tradeoff between robustness, capacity and imperceptibility. Thus, coming up with public spatial domain algorithms which survive a broad range of manipulations is an important issue. However, the most serious problem of spatial domains is the weakness of robustness. Therefore, watermarking schemes in spatial domains usually are used singly. In electronics, control systems engineering, and statistics, frequency domain is a term used to describe the domain for analysis of mathematical functions or signals with respect to frequency, rather than time. A frequency-domain representation can also include information on the phase shift that must be applied to each sinusoid in order to be able to recombine the frequency components to recover the original signal. In this section, the concept of Singular Value Decomposition and Discrete Wavelet Transform are introduced for decomposing images.

7.1 Singular Value Decomposition

The singular value decomposition (SVD) of a matrix with real or complex entries is one of the fundamental tools of mathematics. This type of algorithms has proven to be robust in watermarking systems. It was given detailed properties and other applications for SVD in [34]. In this section, we summarize the definitions of SVD and the SVD-based watermarking scheme.

Although SVD works for any $N \times M$ matrix, and without loss of the generality, our discussion will be
limited in the $N \times M$ matrix with real entries. It is noted that the SVD applies more generally to complex-valued rectangular matrices, while we restrict our discussion to real-valued, square matrices. The singular value decomposition of $A$ is represented by,

$$A = U\Sigma V^T$$  \hspace{1cm} (3)

Where $U \in R^{N \times N}$ are the unitary matrix, and $\Sigma \in R^{N \times N}$ is a diagonal matrix and the superscript $T$ denotes matrix transposition. The diagonal elements of $\Sigma$, denoted by $\sigma_i$ are called the singular values of $A$ and these are assumed to be arranged in decreasing order $\sigma_i \geq \sigma_{i+1}$. The columns of $U$ denoted by $U_i$'s are called the left singular vectors while the columns of $V$ denoted by $V_i$'s are called the right singular vectors of $A$. It is easy to see that $\sigma_i$, $V_i$ and $U_i$ satisfy:

$$AV_i = \sigma_i U_i$$  \hspace{1cm} (4)

$$U_i^T A = \sigma_i V_i^T$$  \hspace{1cm} (5)

In [42], the watermarking scheme is fundamentally flawed algorithm, this is because one attacker can always claim that this watermark was the embedded one and he can claim ownership of the watermarked image, using the singular vectors of any fake watermark in the detection stage. For correcting fault scheme, a SVD-based watermarking algorithm, which explores the optimal scaling factors of watermark embedding, is presented.

7.2 Discrete Wavelet Transform

Multiresolution image representations using Discrete Wavelet Transform (DWT) have received wide range of attention in the recent years. It is a fast pyramidal algorithm and an efficient mathematical tool that decomposes an image into hierarchical subbands. Each sub-band is logarithmically spaced in the frequency domain. DWT separates an image into a lower resolution and labels the resulting subimages. LL (the approximation) which is the coarse overall shape, covers the low-frequency components that contain most of the energy in the image and LH (horizontal details), HL (vertical details) and HH (diagonal details) which represent higher-frequency detailed information have the finer scale wavelet coefficients according to the filters used to generate the sub-image. The wavelet components are then used to obtain the next coarse overall shape by further iterating LL in this process and we get the details (LL1, LH1, HL1, and HH1). The size of the input data reduces gradually. The size at each succeeding octave is one-fourth the size of the previous one. This process is repeated several times until the desired final scale is obtained as shown in Figure 3. The Figure shows the sketch of the decomposition of an image in three resolution levels through DWT. In the analysis of reconstruction, the inverse DWT is performed by a similar structure with the corresponding synthesis subimages, consisting of an inverse the approximation coefficients and the detail coefficients. This feature makes the watermark more robust in comparison with spatial methods against various distortions. The applications of DWT are efficient due to two important reasons. One: it has the irregular distribution of the inverse transform value in a pyramidal wavelet domain. Two: it has the low linear complexity which requires lower computational cost $O(n)$ when compared with the computational cost consumed by Fourier and cosine transforms $O(n \log(n))$. 

![Figure 3. (a): three-level wavelet transform of the 'Lena' image](image-url)
8. Geometric invariant watermarking

Watermarking algorithms robust to the geometrical distortions have been the focus of research [81]. Several approaches to this problem include exhaustive search, synchronization pattern/template, invariant domain and implicit synchronization using image features are widely used. Most of the proposed geometrical transform invariant algorithms are actually only RST invariant. Also the systematic analysis of the watermarking algorithm performance under geometrical distortion has begun to draw great attention. Most of these efforts confine to theoretically analyzing and quantifying the effect of the global affine transform to the performance of the watermarking algorithms. Local distortions are more and more regarded as a necessary test scenario in benchmarking software. However, it is difficult to theoretically analyze its effect on watermark detection due to its complexity.

8.1 Exhaustive Watermark Search

Exhaustive search technique [31] is one of commonly used candidate approaches to coping with the watermark desynchronization problem. This approach performs the process of watermark detection over a training sequence containing each pilot geometrical inverse deformation. The watermark is recovered by searching each hypothetical distortion parameter. Obviously, it seems to be difficult to efficiently evaluate deformation parameters for the absence of the original image [1][3]. Also, the high computational complexity of performing the method is intractable. It is feasible that exploring spaces are precluded in a reasonable subset of parameters. In this way, the limitation of the search space’s size constrains essential degradation of perceptual quality in watermarked images. However, the restriction of search space suffers from unacceptable effects [36] that it can increase more errors of the synchronizer outcome, and the dramatically proliferated false positive probability because of the geometrical deformation and the interpolation errors [4]. In [6] and [45], the stochastic analysis is used as the further discussion. It is shown that the implementation of random bending attacks in Stimark enlarges the searching space and increases computation complicatedly for exhaustive search detector.

8.2 Template-based approach and self-synchronizing

Recently, one of the most straightforward solutions to cope with desynchronization attacks is composed of inserting templates along with the watermark or embedding a periodic watermark pattern into the image. These additional templates [52][64] are used as artificially embedded references for purposes of resynchronization. There is insensitive information carried in them. The template-based approach performs the watermark retrieval process by asserting the presence of a watermark and estimating and compensating the severe geometric transformation of the watermarked image for accomplishing resynchronization patterns [41]. Therefore, the registration pattern is identified and offered resilience to geometric attacks [7].

The template-based algorithms introduced in this section all are relevant to adaptive determination in the strength of the templates for constraining the search space. The contraction of searching space results in being rather susceptible to the credibility of system. Meanwhile, the template is restricted to the number of synchronization points, which tends to be the counterpart of unaccepted false positive probabilities in the parameter estimation of the applied affine transformation. Also the process of the successful watermark detection relies on the precise detection of the template because the detection error tolerance is unacceptable for the inaccuracy of the detected position. It causes the unwarranted synchronization for shrinking the search space.
Although an amount of progress for utilizing template-based watermarking techniques, these ignore the perceptual similarity between the original and watermarked image. The insertion of the watermark and the template should take the embedding position and strength into account carefully. And besides, the algorithms compromise the data payload of the watermark for keeping the restricted fidelity of the original image after the embedding process, since the embedded template tries to decrease the number of embedded information that consist of the template to guarantee synchronization and avoid potential visible artifacts.

In addition, security aspect has been paid little regard, because these algorithms are particularly vulnerable to template removal attacks [23][25][7]. If these templates are provided with characteristic features of independence from the host image, then the specific characteristics can be exploited to destroy the synchronization pattern. In another word, the templates are used to be inserted as pseudorandom noise patterns and some specific operations, for example filtering operators, can filter potentially out the local maxima. Then, the applied templates are eliminated easily by the malicious attack and the applied various geometric transformations are unable to be recognized during the template detection process. In brief, the major limitation is that this kind of approach has a severe influence on the resistance of affine transformations confronting with threats and risks of the template attack.

Self-synchronizing watermarks [29][21] are susceptible to removal or estimation attacks in much the same way as template-based methods, because an attacker can use knowledge about the watermark’s periodic tiling to remove it. For example, attackers can easily remove peaks in the autocorrelation function or the frequency domain by filtering in the spatial or the frequency domain, thus rendering the watermark vulnerable to subsequent geometric attacks.

8.3 Invariant Domain

Another solution consists in embedding the watermark in a geometrical invariant subspace. In [12], it suggested using histogram specification to hide a watermark invariant to geometrical distortions. In [60] and [38], they proposed a watermarking scheme based on the Fourier–Mellin transform [39]. A rotation, scale and translation invariant domain is obtained using the log-polar mapping (LPM) and the Fourier transform invariance properties to translations. In the resulting log-polar map, rotations and scaling come down to translations. In practice, this solution can be implemented for simple affine transformations, but it is inapplicable as soon as the image undergoes local geometrical deformations. Moreover, problems of approximation due to the discrete nature of the images, plus the reduction of the embedding space make the watermark weakly resistant to low-pass filtering and lossy compression. Alghoniemy and Tewfik [2] present another approach where the watermarking space is defined as a canonical, normalized space based on the geometric image moments. These moments are used to transform the image into a form that is independent to rotation, scale, and horizontal/vertical reflection. The watermark is embedded in this space, and then the inverse transformation is applied to obtain the final watermarked image. During detection, the moments are calculated again and used to estimate the normalization parameters. Once the image is normalized, the watermark can be detected. Rather than embedding the watermark in an invariant subspace, Solachidis and Pitas propose creating a self-similar watermark, and to embed it then into the DFT domain[63]. Thus, their method is robust to translation, since it does not affect the DFT magnitude. Since the watermark is made up of identical sectors, the detection is possible even after a degree rotation. The self-similar properties of the watermark also allows for the reduction of the number of different frequency sampling steps where the detection should be performed when the image has been cropped and scaled.

8.4 Feature Watermarking

Feature-based watermarking has raised a number of available algorithms over the last few years. For solving watermark synchronization problems, feature-region detection is the preferred strategy to resist against local geometric distortions. Generally speaking, content-based synchronization watermarking schemes follow the same basic process: detected feature points are localized at the local maxima while non-maxima suppression that eliminates pixels that are not local maxima, and the final set of features is determined by analysis of
threshold. Afterwards, extracted feature points are applied to identify regions for watermark insertion in the host image. At the receiver side, the feature points are detectible without synchronization error. The observation of feature-based synchronization has resulted in various algorithms known as region-based watermarking [22][35][36][61][62][73][74][75]. The feature points-based approach is a technique using localized watermarking algorithms. It discovers the watermark using stable feature points of images, where the watermark is independently inserted into the corresponding each local region [29]. Hence the feature-based process can be invariant to local geometrical deformations so that it is an encouraging approach to solve the robustness against geometrical deformations in the watermarking scheme with blind detection. In [7], it extracted feature points of the host image using Harris detector and produced a Delaunay tessellation on the group of stable feature points. Then, the watermark was inserted in each triangle of the tessellation. But the Harris feature points were not invariant to scaling [46] so that the approach is not resistant to the attack of scaling.

In [69], it produced a feature extraction process named Mexican Hat wavelet scale interaction. Image normalization was individually exploited to unoverlapping image circles with the firm radius and the center at the obtained feature points. In each image circle, two 32 × 32 blocks were selected for embedding watermark. In spite of the fact that the system displays experimentally the robustness to the majority of attacks, it is attackable to scaling transforms because these transform can cause the alterations of content for two blocks in the image circles with a firm radius. In [35], it presented the Harris feature points from the image with the normalized scale and inserted the watermark within each disk region in the center at the localized stallest Harris points and firm radius. The presented method can be resistant to global geometrical transformations, including scaling, rotation and moderate translation. Nevertheless, the presented scheme cannot be robust against cropping because the normalization of scale may be sensitive to cropping in nature. Recently, many researchers have turned interest orientation to the resynchronization of watermark using the scale invariant feature detectors relied on the scale-space theory in the pattern recognition fields, such as Harris-Laplace, scale invariant feature transform (SIFT) [33]. In [33], the SIFT feature is exploited to produce the circular patches as the inserting modules. The rectangular watermark is transformed to be a polar-mapped watermark, and inversely polar-mapped to determine inserting modules before watermark insertion. The watermark correspondence detection is achieved by the circular convolution. The invariant watermark of rotation is obtained by the translation characteristic of the polar-mapped circular pads. In [33], Harris-Laplace approach is applied to extract the scale invariant feature points relied on the scale selection theory using Harris corner points. The watermark is inserted after affine the normalization according to the local characteristic scale at each feature point. The characteristic scale is defined by the scale, where the normalized scale-space representation of an image achieves a maximum value, and the characteristic orientation relied on the angle of the dominant axis of an image.

In [22], a robust watermarking approach was proposed combining Tchebichef moments and the local circular regions (LCRs). LCRs are shaped by Harris-Laplace detector, and Tchebichef moments are applied to acquire the global characteristics of the LCRs. In [62], the affine invariant point detector [46] was discussed to detect feature points. For a chosen feature point, an elliptical feature region (viz., affine covariant region) is used and formed for inserting in spatial domain. Before embedding, the watermark, it is geometrically transformed into an elliptical pattern according to the shape of the region. This method has a modest robustness and provides a potential idea for resistance against complicated geometric distortions.

9. Conclusion

With the rapid proliferation of globally-distributed computer network technologies and popularity of multimedia systems, fashionable and economical digital recording and storage devices have made it possible to construct the platform, which has considerably make it easy to not only acquire, represent, replicate, distribute, and transmit multimedia contents in digital formats without degradation of quality, but also to manipulate them. However, the transmission of information over various networks is often unsafe. This lack of security can be critical depending on the nature of the transmitted media. This issue gave rise to digital watermarking, a research field which deals with the process of embedding information into
digital data in an inconspicuous manner. This in turn enables information security for various multimedia applications including copyright protection. In this paper, different watermarking techniques have been reviewed and analyzed. This is based on image processing in spatial and transform domain. Different techniques using singular value decomposition and discrete wavelet transform in transform domain have been reviewed. Additionally, the analysis of these techniques has been represented in the form of tables considering different factors of watermarking like embedding imperceptibility, capacity, security, robustness and false positive. It can be concluded that various attacks techniques are used to the assessment of watermarking systems, which supplies an automated and fair analysis of substantial watermarking methods for chosen application areas. Furthermore, watermarking algorithms robust against the geometrical distortions have been the focus of research. Several approaches to this problem, including exhaustive search, synchronization pattern/template, invariant domain and implicit synchronization using image features, are widely used.

References


