Digital Watermarking

Information Hiding

Information hiding is a general term encompassing many sub disciplines.

The three main sub disciplines include :

1. Steganography – the art of concealed communication
2. Watermarking – the practice of hiding a message in a media
3. Cryptography – the practice of keeping messages secret

What is a Watermark?

A distinguishable mark or device impressed in the substance of a sheet of paper during manufacture, usually barely noticeable except when the sheet is held against strong light.

- to make forgery more difficult
- to record the manufacturer’s trademark

Digital Watermark

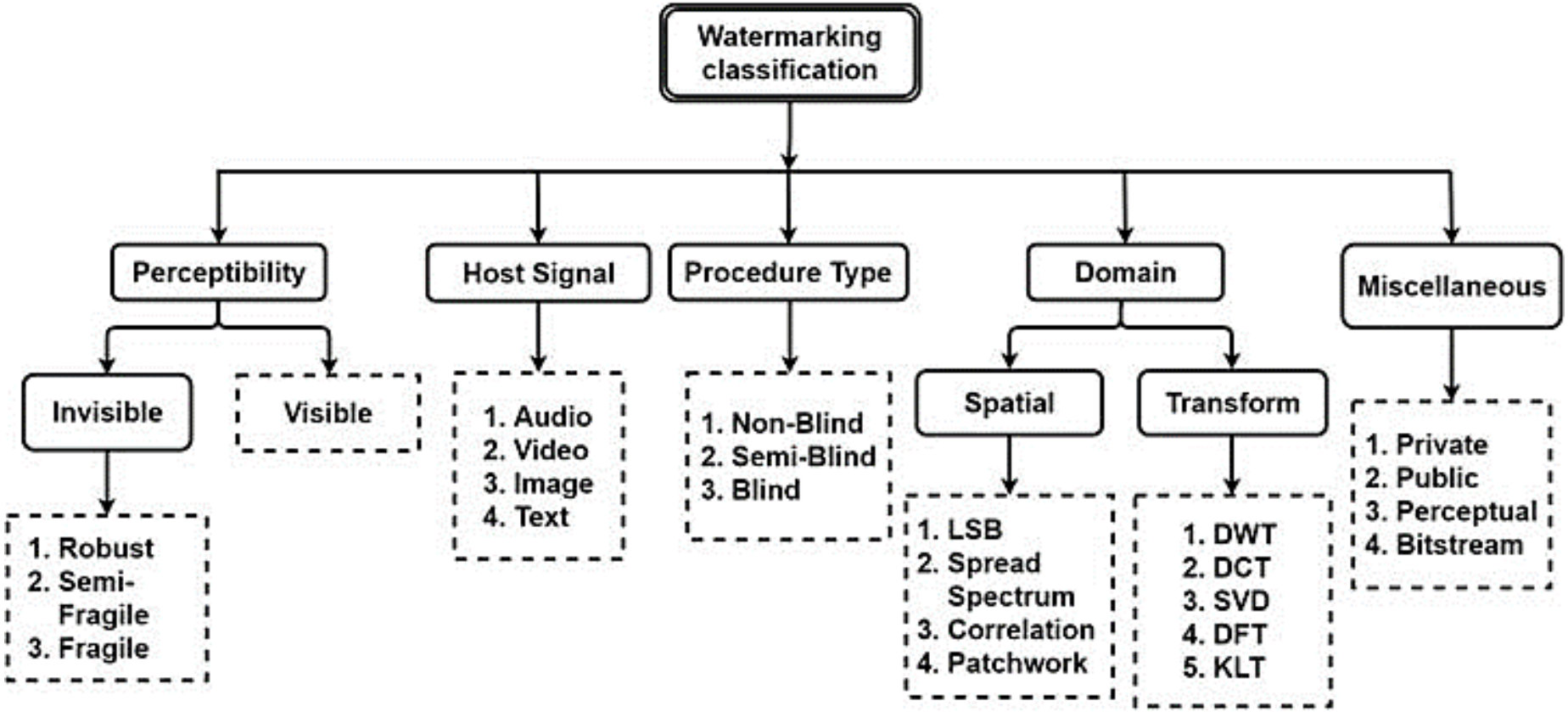
Digital watermarking is a technology for embedding various types of information in digital content. In general, information for protecting copyrights and proving the validity of data is embedded as a watermark.

A digital watermark is a digital signal or pattern inserted into digital content. The digital content could be a still image, an audio clip, a video clip, a text document, or some form of digital data that the creator or owner would like to protect. The main purpose of the watermark is to identify who the owner of the digital data is, but it can also identify the intended recipient.

History

- Nearly 700 years ago, Italy – indicate the paper brand and the mill that produced it
- 18th century – used as anti counterfeiting measures on money and other documents
- Term “Watermark” – introduced – end of the 18th century
- First example – similar technology – patent filed in 1954 by Emile Hembrooke – identifying music works
- In the year of 1988, Komatsu and Tominaga were the foremost to use the term digital watermarking.

Characteristics	Steganography	Watermarking
Purpose of use	Hidden data transmission	Control of integrity, authenticity, authorship protection
Protected object	Secret message	Cover image
Result of embedding	Stego image	Watermarked image
Main security threat	Steganalysis	Image distortion
Main security criterion	Resistance to steganalysis	Robustness
Imperceptibility of embedding	High	Usually high, but in some cases not required
Embedding capacity	Usually high	May be different
Need to extract embedded information	Yes	Not in all cases


Watermarking classification
Perceptibility
Host Signal
Procedure Type
Domain
Miscellaneous
Invisible
Visible
1. Audio
2. Video
3. Image
4. Text
1. Non-Blind
2. Semi-Blind
3. Blind
Spatial
Transform
1. Private
2. Public
3. Perceptual
4. Bitstream
1. Robust
2. Semi-Fragile
3. Fragile
1. LSB
2. Spread Spectrum
3. Correlation
4. Patchwork
1. DWT
2. DCT
3. SVD
4. DFT
5. KLT

Watermarking Characteristics

- **Imperceptibility**
  - The modifications caused by watermark embedding should be below the perceptible threshold. The imperceptibility means that the watermark should not be visible to the human visible system as the watermark should look like genuine.
- **Robustness**
  - The robust meaning is nothing but the strength of the watermark to obstruct the manipulations of the media like the compression, scaling and cropping. The definition of the robustness explains that the watermark can be detected after the operations on te content, the operations are like filtering, compression, color correction and geometrical changes.
  - The ability of the watermark to resist distortion introduced by standard or malicious data processing.

Watermark Characteristics

- **Security**
  - A watermark is secure if knowing the algorithms for embedding and extracting does not help unauthorized party to detect or remove the watermark
- **Complexity**
  - The complexity explains about the difficulty and the time needed for the watermark embedding.
- **Verification**
  - The verification is a process and it has a private key function or a public key function.

Watermark Characteristics

- **Payload**
  - The amount of information that can be stored in a watermark
- **Informed (non oblivious, or private)**
  - The original un watermarked cover is required to perform the extraction process
- **Blind (oblivious, or public)**
  - The original un watermarked cover is NOT required to perform the extraction process

Visible Digital Watermark

Visible digital watermarks, as the name implies, are watermarks that can be seen with the human eye. They are also called perceptible digital watermarks. Visible watermarks take the form of logos or transparent images and text placed on the digital media requiring protection. They are also very easy to create.

Visible watermarks might be easy to create, but they are ineffective. For example, a visible digital watermark placed on media like pictures or videos can be overlaid or even cropped.

Invisible Watermark

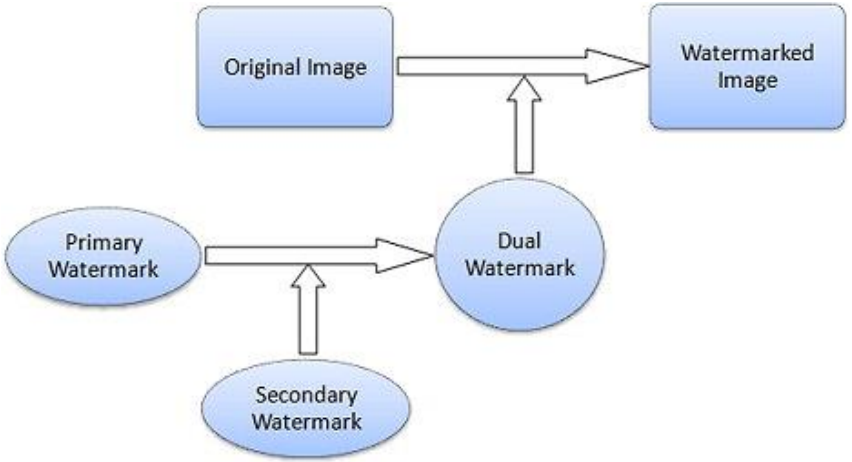
Visible Watermark

Invisible Digital Watermark

Invisible digital watermarks are watermarks that the human eye cannot see. They are usually pieces of code covertly embedded into digital media to attribute ownership and protect copyright. When invisible watermarks are used, it is highly impossible to differentiate between the watermarked media and its original. The invisible digital watermarks and the information it carries can only be processed by special software using a computer.

Dual Watermark

Dual watermark is the combination of visible and invisible watermark. An invisible watermark is used as a backup for the visible watermark.



Fragile Digital Watermark

Fragile watermarks are watermarks that lose their effectiveness when they undergo transformation. When fragile digital watermarks undergo alterations like resizing or compression, they are destroyed. Because of this, they are used to detect when any digital media has been tampered with. Eg. Tamper proofing.

Semi Fragile Digital Watermark

The semi-fragile watermarks are primarily used to certificate the integrity and authenticity of image data. The semi-fragile watermark, combining the advantages both of the robust watermark and the fragile watermark, is mainly used for fuzzy authentication of digital images. Under the premise that the image content is basically unchanged, the image is allowed to have a certain degree of distortion. Eg. Image Compression.

Robust Digital Watermarking

Robust watermarking is the watermarking algorithm that can survive not only such general operations such as compression, adding noise, filtering, A/D or D/A conversion, and so forth, but also such geometric attacks such as rotation, scaling translation, shearing, and so forth. It is often used in ownership protection. They are difficult to remove from the object they are embedded.

Blind and Non Blind Digital Watermarking

- In blind digital watermarking, none of the information of the host is used in extracting the embedded watermark.
- In non-blind digital watermarking some host information is used. Typically, the less the host information is used, the more complicated the watermarking scheme becomes.

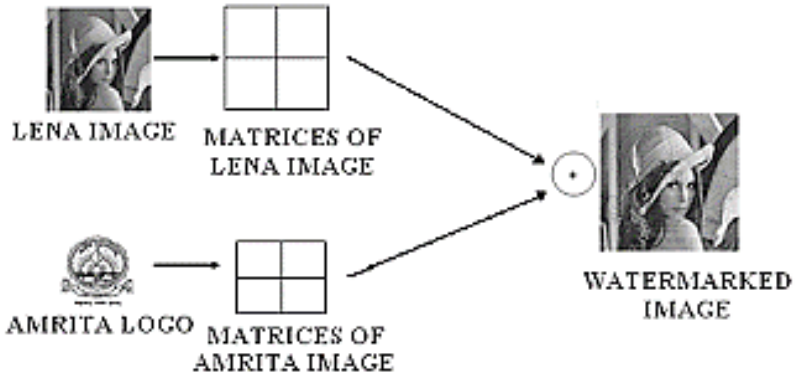
Private and Public Digital Watermarks

Private Watermark : requires at least original data to recover watermark. They are also known as ‘Secure Watermarks’.

Public Watermark : requires neither original data nor embedded watermark to recover watermark information. These can be understood and modified by anyone using certain algorithms. These are not secure.

Spatial Digital Watermarking

Spatial (domain) watermarking is a digital watermarking technique that involves embedding the digital watermark into the cover image or video pixels. In this technique, the code carrying the information is inserted into the pixels by changing its color value or intensity. The two main spatial domain watermarking forms are the Least Significant Bit and the Additive watermarking.



Frequency Digital Watermarking

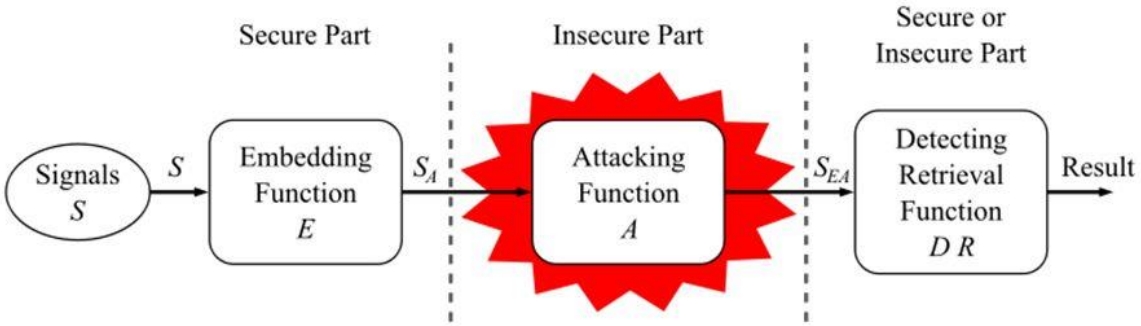
Frequency Digital Watermarks use the technique of inserting codes into specific frequency bands of a digital medium. When the frequencies of the image are separated, the watermark becomes visible. This technique is more complex than the spatial digital watermark and hence more effective.

Digital Watermarking Life phases

The information needs be embedded in the media. The signal which is embedded is the host signal and the information is called digital watermark. The process has 3 main parts:

1. **Embed –** In this part, the digital signal is embedded with the digital watermark.
2. **Attack –** The moment when the transmitted media is changed, it becomes a threat and is called an attack to the watermarking system.
3. **Protection –** The detection of the watermark from the noisy signal which might have altered media (JPEG compression, rotation, cropping, and adding noise) is called Protection.

LIFE-CYCLE PHASES



Produce watermarked signal

The marked signal is modified

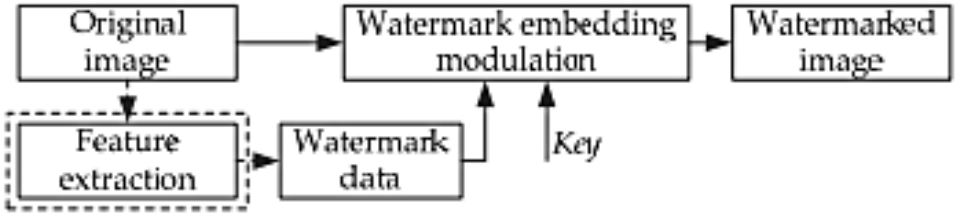
Attemp to extract watermark from signal

Watermark Generation

This function generates a suitable watermark according to the watermarking objectives in an application. In a simple data-hiding application, a watermark can be the embedding-data (e.g., message, *m*, other image data, *j*) itself (along with any side information). In an advanced application, a watermark may require to have certain properties (depending upon the watermarking objectives). For example, in a copyright protection application, a watermark may need to be 'robust' against certain processing techniques and/or attacks. Failure to consider those properties may result in technical flaws and security vulnerabilities. Although watermark generation is mainly constrained by the required properties, it starts with necessary inputs and their properties.

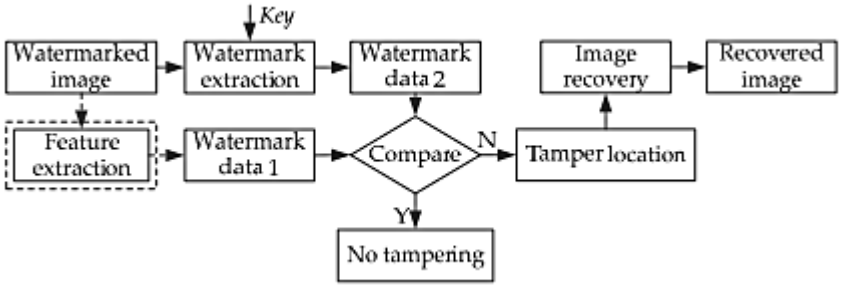
Watermark embedding

As the data-hiding component, watermark embedding function considers where and how to embed the watermark satisfying various requirements of the cover objects (here, digital images). For example, ‘perceptual similarity’ requirements (that control which pixels can be modified to what extent) of medical images may limit the embedding region. There are different domains (e.g., spatial, transform) for embedding, which are computed directly from an input image. Embedding types may also be different (e.g., invisible, invertible or reversible, blind, etc.).

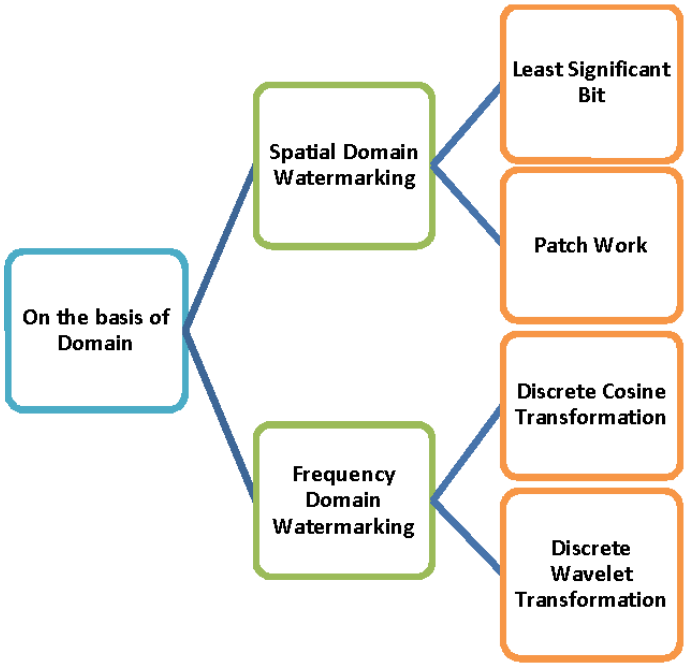


Watermark detection

In the detection process, the watermarking information in the tested image is extracted and compared with the original watermark. If the two are consistent, we consider that the image has not been tampered with; if the two are inconsistent, we consider that the image has been tampered with. Then the algorithm enters the tamper detection and recovery process. In addition, when the watermarking information is generated using image content features, in the authentication process, there is no need to provide the original watermarking information. Only the comparison between the extracted watermarking information and image content features is needed.



Digital Watermarking Algorithms



Spatial Domain Techniques

Techniques in spatial domain class generally share the following characteristics:

1. The watermark is applied in the pixel domain.
2. No transforms are applied to the host signal during watermark embedding.
3. Combination with the host signal is based on simple operations, in the pixel domain.
4. The watermark can be detected by correlating the expected pattern with the received signal.

Least Significant Bit

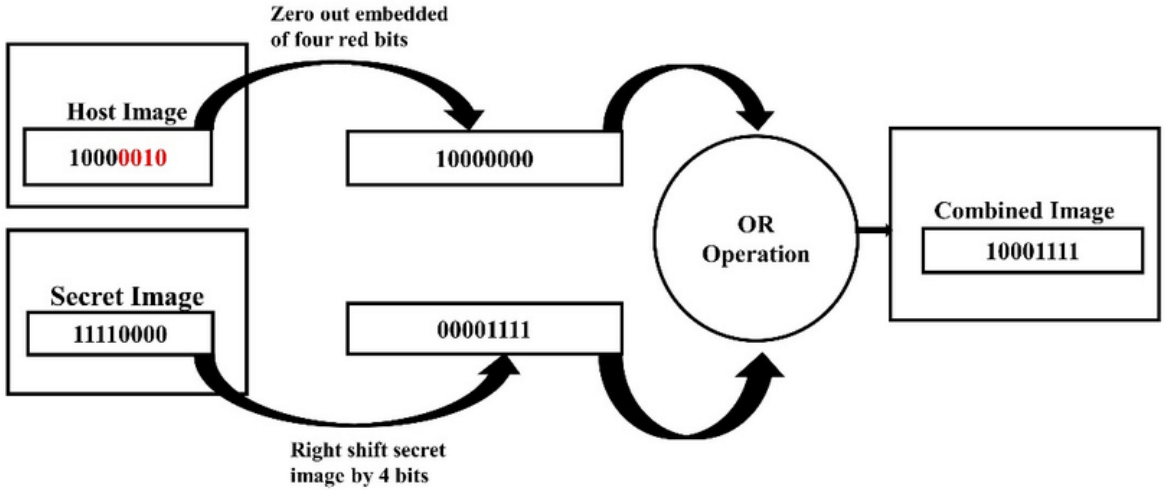
The simplest algorithm is Least Significant Bit (LSB) Insertion, in which each 8-bit pixel's least significant bit is overwritten with a bit from the watermark. Given the extraordinarily high channel capacity of using the entire cover for transmission in this method, a smaller object may be embedded multiple times. In a digital image, information can be inserted directly into every bit of image information or the more busy areas of an image can be calculated so as to hide such messages in less perceptible parts of an image. Two techniques were presented to hide data in the spatial domain of images by them. These methods were based on the pixel value's Least Significant Bit (LSB) modifications.

Least Significant Bit

Steps of Least Significant Bit (LSB) :

1. To convert RGB image to gray scale image.
2. Making double precision for the image.
3. Shifting most significant bits to low significant bits of watermark image.
4. Make least significant bit of host image to zero.
5. To add shifted version (of step 3) of the watermarked image to modified (of step 4) host image.

Least Significant Bit



Least Significant Bit

FEATURES OF LEAST SIGNIFICANT BIT:

1. The technique is easy to implement.
2. It is simple to understand.
3. The result of this technique is a stenographic- image which contains hidden data yet to appear.

Least Significant Bit

Merits :

- It is very easy to implement and understand
- The image quality is less tampered.

Demerits :

- It lacks robustness
- Vulnerable to noise and cropping, scaling.

Frequency Domain Watermarking

Transform domain techniques usually achieve better performance since the perceptual characteristics of images can be better utilized and the spread spectrum principles used in secure communications can be easily incorporated. Typically, transform domain systems perform the watermarking process independent of compression, although these two processes share some common features. Transform domain techniques embed watermarks with visually recognizable patterns in the images as a set of independent and identical distributed sequences drawn from a Gaussian distribution into the perceptually most significant frequency components of an image

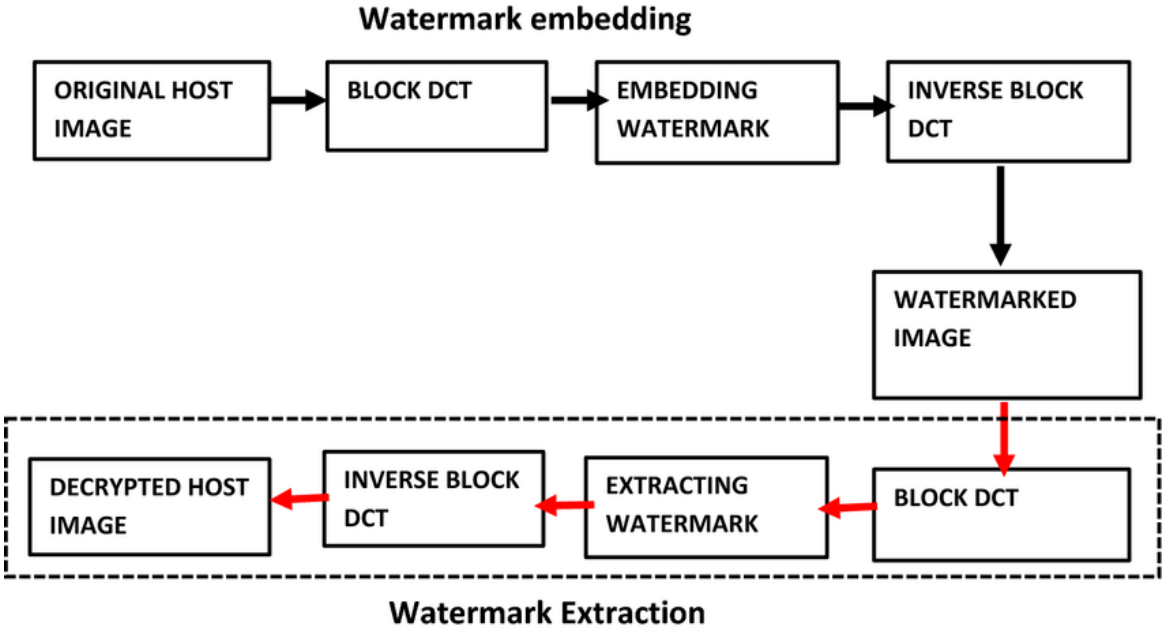
Discrete Cosine Transform Watermarking

In DCT the image is divided into different frequency band, and then embedding watermarking into the middle frequency bands of an image. DCT represents data in terms of frequency space rather than an amplitude space. DCT based watermarking techniques are robust compared to spatial domain techniques.

Steps in DCT Watermarking Algorithm (Block Based).

1. Divide the image into non-overlapping blocks of 8×8
2. Apply forward DCT to each of these blocks
3. Apply some block selection criteria (e.g. HVS)
4. Apply coefficient selection criteria
5. Embed watermark
6. Apply inverse DCT transform on each block

Discrete Cosine Transform Watermarking



Discrete Cosine Transform Watermarking

Merits :

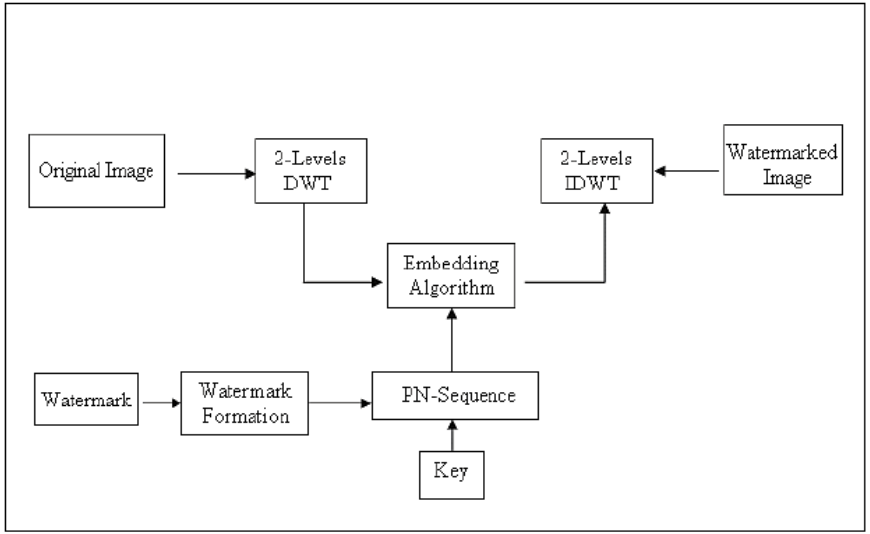
- Here the watermark is embedded directly into the coefficients of the middle value of frequency, hence the visibility of image will not be affected and the watermark will not be removed by any attack.

Demerits :

- Certain higher frequency components tend to be suppressed during the quantization step.

Discrete Wavelet transform Watermarking

In DWT a signal is split into two parts, usually high frequencies and low frequencies. The low frequency part of the signal is again split into two parts of high and low frequencies. The process can then be repeated to computes multiple “scale” wavelet decomposition



Discrete Wavelet transform Watermarking

Merits :

- Allows good localization in spatial and frequency domain .

Demerits :

- Computing cost may be higher.
- Longer compression time.
- Blur near edges of images

Difference between Spatial and Frequency Domain

Factors	Spatial Domain	Frequency Domain
Implementation Cost	Low	High
Robustness	Fragile	More Robust
Perceptual quality	High Control	Low Control
Complexity	Low	High
Time	Less	More
Capacity	High	Low
Example of Application	Mainly Authentication	Copyrights

Applications

1. **Copyright Protection**
   - Most prominent application
   - Embed information about the owner to prevent other from claiming copyright
   - Require very high level of robustness

2. **Copy Protection**
   - Embed watermark to disallow unauthorized copying of the cover
   - For example, a complaint DVD player will not playback or copy data that carry a “copy never” watermark

Applications

3. **Content Authentication**
   - Embed a watermark to detect modifications to the cover
   - The watermark in this case has low robustness, “fragile”
4. **Transaction Tracking**
   - Embed a watermark to convey information about the legal recipient of the cover
   - This is useful to monitor or trace back illegally produced copies of the cover
   - This is usually referred to as “fingerprinting”

Applications

5. **Broadcast Monitoring**
   - Embed a watermark in the cover and use automatic monitoring to verify whether cover was broadcasted as agreed.

6. **Tamper Proofing**
   - To find out if data was tampered. Eg. Change meaning of images
   - Hidden Watermarks track change in meaning
   - Issues: Accuracy of detection

7. **Quality Assessment**
   - Degradation of Visual Quality

Security

Security property of watermarking schemes as a whole may be far from easy to conceptualize. Two main possible reasons are :

(i) application-dependent properties and

(ii) the confusion between security and robustness requirements.

In practice, different image applications may require different levels of security. Some applications do not need to be secure at all since there is no ultimate benefit in circumvention of watermarking objectives. For example, where a watermark is used only to add value in which they are embedded rather than to restrict uses for some device control applications. Therefore, these types of watermarks do not need to be secure against any hostile attacks, although they still need to be robust against common processing techniques used in those applications. Although the requirements for robustness and security properties of a watermarking scheme may overlap, they need to be considered separately. For security properties, in contrast to robustness, all possible attacks that an adversary may attempt with in a particular scenario are to be studied.

Attacks on the watermarking security

In the watermarking context, an attack can be roughly defined as any malicious attempt to perform unauthorized embedding, removal, or detection of a (valid or invalid) watermark. An adversary that makes such attempts can be of different capabilities (e.g., can have different inputs, and access to the watermarking functions). In practice, it is quite reasonable to assume the capabilities of expected adversaries in modelling attacks.

One categorization of the wide class of existing attacks contains four classes of attacks:

1. Removal Attacks
2. Geometric Attacks
3. Cryptographic Attacks
4. Protocol Attacks

Removal Attacks

Removal attacks aim at the complete removal of the watermark information from the watermarked data without cracking the security of the watermarking algorithm (e.g., without the key used for watermark embedding). That is, no processing, even prohibitively complex, can recover the watermark information from the attacked data. This category includes denoising, quantization (e.g., for compression), remodulation, and collusion attacks. Not all of these methods always come close to their goal of complete watermark removal, but they may nevertheless damage the watermark information significantly.

Removal Attacks

- **Statistical attacks**: Statistical attacks treat the watermark as signal noise which can be statistically modeled and removed. Statistical attacks do not need to explicitly model the watermark, its identification and removal may occur as part of another technique such as image filtering (de-noising), remodulation, or the application of a lossy compression algorithm.
- **Collusion attacks:** are applicable when many copies of a given data set, each signed with a key or different watermark, can be obtained by an attacker or a group of attackers. In such a case, a successful attack can be achieved by averaging all copies or taking only small parts from each different copy.

Geometric Attacks

Geometrical attacks aim to obscure the watermark, making it difficult to detect. This may be achieved by distorting the image and watermark data, degrading the watermark detector's ability to synchronize with the watermark and resulting in watermark detection failures; essentially the watermark hidden in the noise becomes lost in the noise. These changes are enough to break a watermark, but in such a way that the human eye or brain cannot see the change in the image.

Geometrical attacks may change global media parameters, such as rotation, aspect ratio and shearing or cropping of the image. Local changes include localized averaging, swapping or removal of pixels, slight color variations, introduction of additional noise, and whatever may introduce visible change to the image, but will confuse a given watermark protocol.

Cryptographic Attacks

Cryptographic attacks aim at cracking the security methods in watermarking schemes and thus finding a way to remove the embedded watermark information or to embed misleading watermarks. One such technique is brute-force search for the embedded secret information. Another attack in this category is the Oracle attack, which can be used to create a non-watermarked signal when a watermark detector device is available. Practically, application of these attacks is restricted due to their high computational complexity.

Protocol Attacks

"Attacking the concept of the watermarking application."

That is to say that rather than remove or distort the watermark we may interfere with its intended application, for example by re-watermarking an image with a second owner, and so confusing the media's original ownership.

Protocol Attacks

1. **Mosaic attack**:

The mosaic attack is of quite general application, and could be tailored to specific watermarking protocols used by particular media vendors.

A web-crawler, paired with a content distribution mechanism, form an automated online piracy detection system. The web-crawler scans the Internet for images and checks for the distributor's mechanism which, if detected, can be recorded and checked for licensing issues.

The attack technique is to chop an image into small tiles, which can be imperceptibly rendered by a browser to look the same as a single image. The issue here for the watermark being not only the geometrical attack (cropping) but also that the remaining image may be too small to hide a meaningful watermark inside, such as a single pixel.

Protocol Attacks

2. **Copy Attack:**

This pertains to the forgery of watermarks. “The goal of the attack it to copy a watermark from stego data to the target data without having any specific knowledge about the watermarking technology” and “the goal of the attack is not to destroy the embedded watermark, but jeopardize the application for which digital watermarks are used”. Therefore, copy attacks are relevant in situations where watermarks (or fingerprints) are used to prove the authenticity or origin of an image rather than to trace copyright ownership. As an analogy consider public/private key signing. If a bank’s SSL certificate could be copy-attacked, then transmissions could be signed without ever needing their private key. This is prevented by hashing in PKI, but media stenography works under fundamentally different constraints.