Second LACCEI International Latin American and Caribbean Conference for Engineering and Technology (LACCEI'2004)
"Challenges and Opportunities for Engineering Education, Research and Development"
2-4 June 2004, Miami, Florida, USA

Digital Watermark Survey and Classification

Mauricio Sadicoff, M.Sc.
HyperSol, LLC., Boca Raton, FL, USA

María M. Larrondo Petrie, PhD.
Florida Atlantic University, Boca Raton, FL, USA

Abstract

A digital watermark is defined as information embedded inside other chunks of information. This paper describes the various types of watermarks currently developed, their purposes and classifications due to insertion method, media type, key specifics, tamper-resistance, and more.

Keywords

Digital, watermark, survey, classification, security.

1. Introduction

A digital watermark is defined as information embedded inside other chunks of information.

The various types of watermarks can be better described by going through some of the most common purposes such as these, mostly defined in [1]:

- **Digital signatures** – The watermarks holds information identifying the owner of the content.
- **Fingerprinting** – The watermark hides information about the authorized user of the content.
- **Broadcast and publication monitoring** – The watermark is used to monitor the use of broadcasted or published information, as in the Philips Research initiative of creating VIVA (Visual Identity Verification Auditor) "to investigate and demonstrate a professional broadcast surveillance system. Broadcast material will be pre-encoded with an invisible unique watermark identifier."[2]
- **Authentication** – The watermark (or in this case also called a vapormark) is used to guarantee that the content showed is precisely the one created.
- **Copy control** – This watermark records information regarding possibilities of reproduction of the data. It will store simple rules similar to "data cannot be copied" or "data can be copied once only, and never afterwards".

- **Secret communication** – If there is no suspicion, information could be hidden on a watermark and sent without detection. Rivest [3] suggests that watermarking, as well as his proposed confidentiality method (chaffing) are completely different techniques from encryption and therefore such techniques effectively invalidate the strong efforts put on by the government to restrict encrypted communication as a national security issue. These methods can be as secure as encrypting information, if not more due to their unorthodox approach and unexpected character.
- **Captioning or visible logo** – Watermarks can be used in video transmission as an extra channel of communication, transmitting information such as captions or a visible logo somewhere on the media.

Watermark insertion and detection usually are symmetric. This means that whatever method used to insert the watermark inside a file will be inversed to extract the watermark. This is illustrated in figure 1:

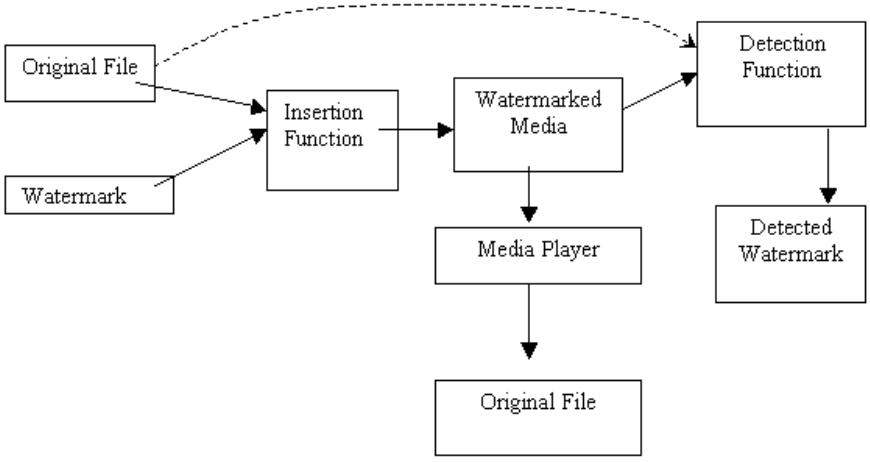

Figure 1: General Watermark insertion/detection mechanism.

Any type of watermark insertion method, if probed long enough and with enough resources, will be broken. Therefore, if the information contained in the watermark is of absolute security, none of the available watermarking methods today could guarantee its embedding and detecting safely. The problem's core then becomes how secure should the watermark be. This degree of security will vary depending on the type of watermark used and the type of attack that is probable against it. Nowadays watermark design is tailored for the application and for probable attacks.

2. Watermark Classifications

Watermarks can be classified according their various characteristics. This paper will determine the different classifications based on prior work (items 2.1 through 2.3.6) and the recognizable characteristics of various watermarks (2.3.7 and subsequent items).

2.1 Perceptibility

A typical misconception is that all watermarks are completely hidden from user perception. That is not forcibly the case. Sometimes a content generator wishes to make known that an image is copyrighted and yet does not want to interfere with the total perceptibility. As shown in [5], a watermark can be automatically inserted in an image as noise and utilize the perception masking capabilities of the human eye to make this watermark just barely visible. It does not interfere with the image and can only be

perceived if the user concentrates on the watermark. This watermark can be inserted in the image as a whole or on a non-intrusive corner, by some modification of the DCT coefficients on the JPEG algorithm.

The most interesting advantage of this method is that it can be automatically inserted, however because it blends with the image by raising or reducing the brightness of a few pixels it is not easily removed by automatic methods. Perceptible watermarks can be used in image and video media only, since it is easier to find an area in the image (or frame) where the placement of another image does not interfere with the image in itself. Audio files, specifically music recordings, are classified according to the degree of purity or low noise characteristics and thus cannot benefit from this method.

2.2 Continuous or Sampled Media

Documents are stored as either continuous (analog) or sampled (digital) media. An example of continuous media storage is a cassette tape. The information on a cassette is not digitized and does not need any sampling to be played. A .WAV file on a PC would be an example of sampled media. This file stores information about the sound to be played back in a discrete, discontinuous manner.

The type of media is important in that it determines the watermarks' ability to survive a D/A or A/D conversion. If a digital watermark is able to survive the a conversion to analog form and back to digital, it must have an analog counterpart. This analog counterpart is simply the analog representation of the digital watermark, which is possible for most watermarking systems. The exception are systems based in least significant bit encoding that simply need bits in order to exist.

2.3 Temporal or Static Media

Watermarks are directly related to the media where they are inserted into, since they intend to be hidden in the media and therefore should follow some of its characteristics. The relation with time is fundamental in that it also determines the amount of data involved. If the information does not change in time, the file will hold less data. An image will invariably hold less information than a video sequence at a similar resolution. Simply put, as watermarks intend to be hidden, more data in a file means more places to hide.

The variation with time also imposes some limitations. Some methods perfectly fit for still media, such as random modifications on the frequency domain, cannot be used safely on video or any time-varying media since it can be easily estimated through statistical averaging. If the modifications are not random but are the same throughout all frames, they can still be estimated through statistical analysis.

2.4 Linear or Nonlinear Insertion

Linear insertion means the final watermarked image relates to the original image through a linear equation, such as equation 1:

$$\mu^{*} = T^{-1}\left(diag(\alpha)T\mu + \beta\right) \qquad (1)$$

Where $<\alpha, \beta>$ represent two column vectors for a previously calculated watermark. They are applied to the column vector μ, which represents a finite media object. Even if μ is a movie or a sound file, it can be concatenated into one column. T is any invertible transformation matrix including discrete Fourier or cosine transforms. The watermarked object is represented by μ^*.

The linearity comes from the watermark being represented by the two column vectors <α, β>, where α is the multiplicative and β is the additive factor. A special case occurs in the so-called additive systems, when α is the identity column vector $[1, 1, 1, \ldots, 1]^T$ and the equation can be reduced to $\mu^* = \mu + T^{-1}\beta$ since $T \bullet T^{-1}$ yields the identity matrix. In this special case, the original image is only transformed by changes intensities of selected regions.

2.5 Perceptual and Statistical Encoding

The encoding of watermarks exists due to their intrinsic necessity of being included in a file without much distortion and surviving common compression techniques. The best way to achieve this goal is to encode the watermark as small changes within the perceived area of the media file such as slight changes of brightness in areas of an image, which do not noticeably degrade the image's quality and yet cannot be removed automatically without some perceived loss of data. Some watermarks are encoded using "perceptual holes" or the parts of the media object that are not usually perceptible to humans, like high frequencies in images. The main problem with this perceptual approach stands in that lossy compression attacks these same holes to reduce the size of an image and with compression techniques evolution, there is no guarantee that a today's perception hole will still exist tomorrow. This type of encoding is called perceptual since the watermark is encoded inside perceptible data.

In addition to perceptual encoding, there are also statistical encoding methods, which are characterized by one or more measures applied over either the entire document or a part of it. Statistical encoding is what allows some watermarks to be called spread-spectrum, since they are encoded statistically throughout the frequency spectrum.

2.6 Document Dependency

A watermark either depends on the document where it is being inserted or is independent of the document. Document dependency is a design choice, rather than a methodology, since it does not imply any specific encoding or rather the way the media is encoded. For instance, scaling DCT coefficients can be either dependent of the document (if we scale the largest 10% of the coefficients) or independent (if we scale the first 10% coefficients calculated). Though not deterministic as to the quality of encoding or security of the process, document dependency must be maintained in the decoding process for successful decoding.

2.7 Type of Key for Decoding

There are three types of decoding based on the type of key used for this purpose. The decoding will need a key and the original data, a secret key only, or a non-secret key only. These methods are discussed in table 1.

Table 1: Types of key for encoding

Type of Decoding	Advantages	Disadvantages
Key and original	Robust against most simple forms of attack and general file manipulation, such as scaling and rotating the media.	Extremely vulnerable to counterfeit original attacks, needs the original image intact for decoding.
Secret key only	Does not need the original document for decoding, needs secret key and method only. Cannot be decoded without the secret key. Key can be made as difficult to decode as allowed by law.	In order for one to decode the image, one needs to know the secret key and the decoding process. Therefore, an evil attacker monitoring the decoding process could get all the information needed to remove it.
Non-secret key only	All the advantages of secret key and the fact that the having the non-secret key and the decoding method does not imply having enough information to mimic or destroy the watermark.	If used as traditional non-secret key encryption, transforming data in an arbitrary way, it may cause changes in the original media data, thus making the whole process invalid.

2.8 Fidelity

Most of the watermark applications intend to insert external information on a file without disturbing the original file's perception. However, because perception holes are disappearing due to compression, watermarks lately have been modifying visible properties on images. Fidelity of a watermark is the characteristic that defines how close the watermarked file is from the original data. It can be determined by statistically averaging, throughout many files, the percentage of data that remains the same when a watermark is inserted using the same method and parameters.

2.9 Robustness

Media files might be subject to many forms of distortions or alterations before they reach the final user, especially if an attacker decides to try to remove a watermark by performing a variety of transformations. In a typical operation, an image file might suffer contrast or brightness variations to enhance a specific area, which in itself is not illegal most of the times but might erase a watermark whose insertion and deletion depends on color or brightness coefficients.

Robustness is a characteristic that depends on the type of distortion and does not apply to intentional attacks to the document but rather typical transformations. The number of distortions is infinite, since any transformation applied on a file can be considered a distortion, as long as the watermarking method is undisclosed. In addition, robustness can be determined based on two questions:

1. Can the watermark still be detected after a transformation (distortion)?
2. If it can be detected, is it still the same watermark? In other words, did it survive the distortion?

If the watermark is encoded in perception holes, lossy compression or other distortions can render a negative response to one of these questions and therefore, as argued by Cox et al [5], the most robust solution for a watermark is to encode it in a perceptually significant area. If this is perpetrated, any significant change to the watermark that would affect either its detection or integrity can affect the integrity of the document, effectively reducing the document's quality and in most cases its usability. The watermark's robustness becomes then directly related to the document's perception quality.

2.10 Fragility

Sometimes a watermark need not be robust but frail. If a watermark is used to guarantee authenticity of a file, it might be inserted in such a way as to only be detected if there were no modifications in a document whatsoever, but the occasional digital copy. A digital copy, of course, does not trigger a fragility defense, since it is in effect duplicating the original data bit by bit and does not modify the watermark or the document as a whole at all.
Such a watermarking scheme is described in detail by Nikolaidis [6] in their ring-shaped watermarking scheme for images. When an image is scaled or rotated, the watermark ring follows the scaling and rotation without changes due to its circular nature. Nevertheless, if the image is cropped the ring will become only an arc, will not contain all the watermark's information, and will not guarantee the document's authenticity.

2.11 Tamper-Resistance

Analogous to robustness, tamper-resistance measures the watermark's capacity of surviving intentional destructive attacks, or tampering. The major difficulty in measuring tamper-resistance comes from not knowing in advance what kind of tampering will the attacker use. As a result, tamper-resistance is only useful as another indication of robustness of a method.

Some tampering attacks on images are expected ahead of time, such as scaling, cropping, rotation, JPEG compression and decompression, brightness and contrast modifications and more. On movies and sound files, other typical attacks are common, some of them derived from the 2D equivalent, such as MPEG compression or cropping, others typical of dynamic media, such as resampling or frame-rate reduction through the destruction of intermediate frames. Though complete tamper resistance is impossible, a measurement method can be determined by limiting the tampering attacks to the ones expected ahead of time.

However, tamper-resistance scales are limited to the attacks known at the time this scale is created and as such it has a major flaw. Attackers invent new attack methods frequently and it is improbable that any person or company would willingly spend the resources needed to track all possible types of attacks and to track only a few attacks would be an incomplete and potentially useless job.

2.12 Key Restrictions

To decode a watermark, frequently a key of some sort is needed. This key can be either unrestricted, meaning it will be available for a variety of detector, or restricted, meaning it will be kept secret by a small number of detectors, as detailed in [1].

When there is a need for the key to be unrestricted, it is usually protected by some method to make it difficult for a possible attacker to tamper with the watermark. One such method is the one described in [7]. This method uses a combination of restricted and unrestricted keys, which are referred to as private and public keys respectively. The method makes only some parts of the pseudo noise public, while keeping the more vital information private.

2.13 False Positive Rate

False Positive Rate (FPR) is the chance of detecting a watermark where one does not exist. It is a dangerous pitfall of watermark design and one of the main reasons for criticism against watermark verification as a legal tool.

Most watermarking methods check to see if a watermark has been inserted in the medium before attempting to decode it. That is done through a faster, limited version of the watermark-extracting algorithm or by simply studying the transmission medium's characteristics. Consider for example an encoding method that concentrates the watermark's energy on a particular frequency, typical of spread-spectrum methods. A decoding method would first check that frequency against the average frequency distribution of the whole image. If there were a significant difference, the algorithm would accuse the presence of a watermark. However, the algorithm would also detect any image that happened to have a frequency distribution resembling the one expected by the decoding mechanism. The method would then "decode" a non-existent watermark, which would in most cases amount to rubbish but in awkward cases could actually resemble some other watermark. The FPR can be measured by checking the image's frequency distribution against other images that could server the same purpose.

Also, any linearly inserted watermark has its pixels' colors defined by a sum that allows for similar results coming from different sources. Mathematically translating this assertion makes its concept easier to grasp. Let us imagine a linearly inserted watermark. If the original image's pixel color is represented by **O**, the watermark is **W** and the watermarked picture is **T**, then:

$$\mathbf{T = O + W} \qquad (2)$$

Concurrently, a picture similar to the original has a pixel color **O'** at the same position as in the original picture and is applied a watermark **W** identical to the original watermark, thus reaching a pixel color **T'**.

$$\mathbf{T' = O' + W} \qquad (3)$$

It becomes than clear that a malicious attacker could then create a watermark that have a **W'** value such that when combined to **O'** would reach **T**.

$$\mathbf{O' + W' = T = O + W} \qquad (4)$$

Thus, a malicious attacker could create a counterfeit original picture, which combined with a tailor-made watermark would achieve the same final picture. This way, when passed through the same detection process, the final picture will accuse a watermark that was not originally there in the first place, disaccrediting the method.

2.14 Possibility of Watermark Modification without Destruction

Sometimes an algorithm might be created to allow for watermark modification by sources with permission to modify it. For example, the DVD specification proposed initially asked for a variety of qualities for each watermark, which would signify commands such as "data cannot be copied". The DVD standard implements four different watermarks, CF (Copy-Free), CN (Copy-Never), CO (Copy Once). The CO watermark was designed to be modified after the copy.

When it is inserted in the factory, the DVD receives a CO watermark, meaning it can be copied once, but only once. On the process of copying this DVD, a DVD recorder would then create a DVD with a CN watermark meaning it could not be copied further and also change the watermark on the original DVD from CO to CN, signifying that this DVD had been copied and therefore could not be copied in the future, as explained in [5] and [8].

2.15 Insertion Method (manual/automatic for visible watermarks)

Visible watermarks are usually utilized for marketing purposes, and must be inserted in areas of a data file where it will not interfere with the content of the original watermark. This process was traditionally done manually, with an operator inserting the watermark on pre-designed areas of the original media file with 3 or 4 different possible locations that would be chosen according to the distribution of information in the file at hand.

Recently other processes arose [4] that solve that problem by analyzing the data prior to insertion, using predetermined assumptions regarding the file's edge distribution. The result is encouraging and might apply for more than simple marketing purposes.

2.16 Multiple Embedding

There is no rule defining the amount of watermarks to insert within a media file. In effect, many algorithms utilize more than one watermark to pursue a higher tamper-resistance. Every time a watermark is inserted, the modified file loses resemblance with the original file however, and its data is modified. Therefore, there must be a limit to the amount of watermarks that one could insert within a media file and still maintain the same aspect of the original data.

This property determines how many watermarks can be inserted in any one file. This number is then unique for each media file, relating to each possible watermarking method.

2.17 Data Payload

Data payload is the amount of information that the owner of the original data is willing to lose in order to insert a watermark. In effect, it is determined by the watermark's size and helps to establish the maximum number of possible values returned by the detector.

2.18 Computational Cost

If the watermark needs to be detected in real-time, for instance on a video or audio file on the DVD, the computational cost of decoding this watermark becomes a factor. The insertion cost is also important since the more complex the method, the higher the computational cost and the higher the monetary cost of the equipment needed.

In real commercial applications, there is an asymmetry in the computational cost for insertion and detection, due to the particularities of each situation. In the DVD process, for example, there is a need to decode the DVD watermarks in real time and through a simple chip into any DVD player. However, insertion of this watermark needs to be as costly as possible since often times insertion methods are also capable of removing watermarks and making this process expensive is another way to discourage an attacker of using the watermark's insertion process for removing it.

2.19 Standardization

Watermarks can also follow standards, which might be necessary for commercial utilization. The digital DVD-video introduced a big leap in quality over the analog VHS but also introduced a problem: how to make sure the digital media would not be copied without permission. Digital Watermarks were seen as a solution for that problem in themselves, which they are not. No watermarking scheme can guarantee impossibility of copyright infringement by itself, it must be combined with other security methods

(scrambling, encoding, statistics, database theory, etc.) in order to be successful and still it cannot be completely and absolutely full proof.

Therefore, it makes sense to use a standard for watermarking detection, especially in widespread copyrighting schemes. Since detection doesn't necessarily mean recovering of the watermark, such a standard would still allow for some freedom of choice regarding insertion methods. For example, there could be a variety of ways to encode 1000 bits, as long as they always have a frequency distribution peaking on 75 KHz.

References

1. Cox, I.; Miller, M.; Linnartz, J.P. and Kalker, T., "A Review of Watermarking Principles and Practices", in *Digital Signal Processing for Multimedia Systems* (Parhi, K. and Nishitani, T., eds.), Ch. 17, 1999.

2. "VIVA: Visual Identity Verification Auditor", Philips Research, Philips DVS, APTN, IMEC, MTG-Broadcast, AIM, website, http://www.intec.rug.ac.be/Research/Groups/hfhsdesign/viva/

3. Rivest, R. R., "Chaffing and Winnowing: Confidentiality without Encryption", MIT Lab for Computer Science, March 18, 1998 (rev. July 1, 1998), http://theory.lcs.mit.edu/~rivest/chaffing.txt

4. Kankanhalli, M. S. and Ramakrishnan, K.R., "*Adaptive Visible Watermarking of Images*", IEEE International Conference on Multimedia Computing and Systems, Florence, Italy, June 1999, pp. 568-572.

5. Cox, IJ and Miller, M.L., "A Review of Watermarking and the Importance of Perceptual Modeling", Proceedings of Electronic Imaging 97, February 19 1997.

6. Nikolaidis, N. and Pitas, I., "Digital Image Watermarking: an Overview", IEEE International Conference on Multimedia Computing and Systems, Florence, Italy, June 1999, pp. 1-6.

7. Hartung, F. and Girod, B., "Fast Public-Key Watermarking of Compressed Video", Proceedings IEEE International Conference on Image Processing (ICIP 97), Santa Barbara, October 1997.

8. Kalker, T., "System Issues in Digital Image and Video Watermarking for Copy Protection", IEEE International Conference on Multimadia Computing and Systems, Florence, Italy, 1999.