Digital Watermarking and Attacks: A Review

Gangadhar Tiwari
NIT, Durgapur

Debashis Nandi
NIT, Durgapur

Madhusudhan Mishra
NERIST, Itanagar

Abstract

The present era of internet has made it easy to obtain the intellectual property, and reproduce it. This creates a high demand for content protection technique like digital watermarking, which has proved to be an important tool to protect the digital properties in recent years. Watermarking is a technique of hiding information which is used to safeguard proprietary information in digital media like photographs, digital music or video. Watermarking has seen a lot of research interest recently and so the attacks. It is used for content protection, copyright management, content authentication and tamper detection. This paper presents the survey of watermarking techniques and attacks with special reference to copyright protection.

I. Introduction

Due to advances in information technology, there has been increased use of multimedia and it has become an important way to deliver services to people. But this has generated problems for intellectual property rights (IPR) management. Protecting information has not received the attention that it deserves. It has only been recently that copyright laws have been established however, they are yet felt to be inadequate. Consequently, copyright abuse is rampant among multimedia users. Digital watermarking seems to be an important tool to protect IPR. The paper introduces Digital Watermarking its reviews and attacks with reference to copyright protection in its section - I. The section-II deals with embedding, extraction of watermark and its requirements. Section–III describes the various types of watermarking techniques and algorithms. Section-IV deals in various attacks. Section–V discusses Digital watermarking for copyright protection and conclusion has been drawn in section VI.

II. Digital Watermark

A digital watermark is a signal permanently embedded into digital data that can be detected or extracted by computing operations in order to make assertions about the data.

A. Watermark Embedding and Extraction

It involves Watermark embedding for copyright and Watermark detection to identify the owner. Embedding a watermark requires watermark carrier, watermark generator and a carrier modifier **[1].** Let I_{orig} denote the original multimedia signal before watermarking, let W denotes the watermark that the copyright owner wishes to embed, and let I_{water} denote the signal with the embedded watermark. The watermark **W** is encoded into I_{orig} using an embedding function **E**:

$$\boldsymbol{E} = \left(\boldsymbol{I_{orig}}, \boldsymbol{W}\right) = \boldsymbol{I_{water}} \qquad (\mathbf{1})$$

During the second stage of the watermarking system, the detecting function **D** uses knowledge of **W**, and possibly I_{orig} , to extract a sequence **W'** from the signal **R** using

$$\boldsymbol{D} = \left(\boldsymbol{R}, \boldsymbol{I_{orig}}\right) = \boldsymbol{W'} \qquad (\mathbf{2})$$

The extracted sequence **W'** is compared with the watermark **W** to determine whether **R** is watermarked. The comparison is based on normalized correlation coefficient ρ, and a threshold λ_0 used to make the binary decision (Z). To check the similarity between **W**, the embedded watermark and **W'**, the extracted one, the correlation measure between them can be found using

$$\rho(\mathrm{W}, \mathbf{W'}) = \mathrm{W} * \frac{\mathrm{W'}}{\sqrt{\mathrm{W} * \mathrm{W'}}} \qquad (\mathbf{3})$$

However, decision function is

$$Z(W,W') = 1\ if\ \rho > \lambda_0\ and = 0\ otherwise$$

'1' indicates watermark is detected, while '0' indicates absence of watermark.

B. Watermarking Requirements

The requirements are application-dependent and compete with each other. It includes Security, Invisibility, Robustness, False Negative/Positive Error Probability, and Capacity Issue.

III. Types of Watermarking

Christine et al defined types of Digital Watermarks as below

A. Based on Visibility

It includes visible watermark and invisible watermark.

B. Based on the availability of original signal during extraction

Private watermarking systems require the original file **A** in order to recover the watermark **W**.

Semi-private watermarking schemes do not use the original file for detection, but they also answer the yes/no question shown above. This could be described by the relation

$$A' * K * W \rightarrow \{\text{Yes,No}\}$$

Public watermarking requires neither the original file A, nor the embedded watermark W.

C. Watermarking Based on Transparency

Fragile watermarks- It does not survive lossy transformations to the original host signal and their purpose is to tamper detection of the original signal.

Robust Watermarks- It provides a mark that can only be removed when the original content is destroyed as well.

Semi-fragile watermarking: It differentiates between lossless and lossy transformations.

D. Based on Watermarking Media

Image Watermarking

Here watermark embedding techniques are designed to insert the watermark directly into the original image data. Requirements include imperceptibility, robustness to common signal processing operations, and capacity **[2]**.

Audio Watermarking

The requirements include inaudibility and robustness to signal alterations. Four techniques for audio watermarking are proposed viz. **Spread spectrum technique, Echo coding, Phase coding** and **Masking**. **Watermark embedding** consists of adding a perceptually weighted PN-sequence to the audio file while watermark detection consists of a correlation detector to determine whether the watermark exists.

Video Watermarking

The watermark must satisfy robustness and transparency along with copy generation management. A cost-effective solution for watermark detection is another critical requirement. Other issues include the placement of the detector. There exist two different approaches for detector placement viz. Watermark detection in drive and Watermark detection within the application/MPEG decoder. With respect to copy generation management the solutions includes using Secondary watermark and Tickets.

E. Watermarking Based on Application

It includes Copyright protection, Broadcast Monitoring, Image authentication, Data hiding and Covert communication. However the paper discusses only Copyright Protection.

F. Based on domain used for embedding process

I. Spatial domain techniques

It embeds the watermark by directly modifying pixel values of original image. These techniques are based on adding fixed amplitude pseudo noise (PN) sequences to an image. In this case, E and D are simply the addition and subtraction operators.

A. Least Significant Bits

It modifies the LSB of the host data. Two LSB techniques are proposed where the first replaces the LSB of the image with a PN sequence, while the

second adds a PN sequence to the LSB of the data. However, it is not secure against signal processing attacks.

B. Spread Spectrum

The main idea is to embed the watermark into a wide-band channel **[10]**. It offers the possibility of protecting the watermark privacy by using a secret key to control the pseudorandom sequence generator. Spread spectrum techniques allow the frequency bands to be matched before embedding the message. It is of two types-Direct sequences and Frequency Hopping.

In Direct sequences cover signal **A** is modulated by the watermark message '**m'** and a pseudorandom noise. As a consequence, the spectrum of the resulting message **m**' is spread over the available band. Then, the spread message **m**' is attenuated in order to obtain the watermark '**W'**. This watermark is then added to original file in order to obtain watermarked version **A'**. In order to recover watermark, the watermarked signal **A'** is modulated with the PN-sequence to remove it. The demodulated signal is then **W**.

In Frequency hopping, we select a pseudorandom subset of the data to be watermarked. The watermark **W** is then attenuated and merged with the selected data. As a result, the modulated watermark has a wide spectrum. For detection process, PN generator used to alter the cover frequency is used to recover the parts of the signal where the watermark is hidden. Then watermark is recovered by using detection method that corresponds to embedding mechanism.

II. Feature Domain Techniques

It takes into account region, boundary and object characteristics. It presents additional advantages in terms of detection and recovery from geometric attacks, compared to previous approaches. More research is yet to be done.

III. Transform Domain Techniques

It embeds the data by modulating the transform domain signal coefficients **[10]**. It has greater robustness, when the watermarked signals are tested after having been subjected to common signal distortions. To embed a watermark, a transformation is first applied to the host data, and then modifications are made to the transform coefficients.

A. Discrete Cosine Transform for DWM

DCT turn over the image edge to make the image transformed into the form of even function. Let $f(x, y)$ is the 8-bit image value at coordinates (x, y), and $F(u, v)$ is the new entry in the frequency matrix.

$$F(u,v) = \frac{1}{4C(u)C(v)[\sum_{x=0}^{7}\sum_{y=0}^{7}\frac{f(x,y)\,(\cos(2x+1)\,u\pi}{16}\left(\frac{\cos(2y+1)\,v\pi}{16}\right)}$$

The corresponding inverse transformation is defined as

$$f(x,y) = \frac{1}{4\left[\sum_{x=0}^{7}\sum_{y=0}^{7}C(u)C(v)\frac{F(u,v)\,(\cos(2x+1)\,u\pi}{16}\left(\frac{\cos(2y+1)\,v\pi}{16}\right)\right]}$$

The 2D-DCT can not only concentrate the main information of original image into the smallest low-frequency coefficient, but also it can cause the image blocking effect being the smallest. So it obtains the wide spreading application in the compression coding.

In **Cox et al model,** the image is first subjected to a global DCT **[5].** Then, the 1,000 largest coefficients in the DCT domain are selected for watermarking. They used a Gaussian sequence of pseudo- random real numbers of length 1,000 as a watermark. This approach achieves good robustness against compression and other common signal processing attacks but not secure against the invariability attack proposed by Craver (1997) besides being computationally expensive.

The **Koch watermarking model** operates on 8x8 DCT coefficient blocks and manipulates a pair of coefficients to embed a single bit of watermark information. This approach modifies the difference between randomly selected mid-frequency components in random image blocks. However this is not a robust algorithm because two coefficients are watermarked from each block.

Using DCT for Watermarking has many drawbacks like, only spatial correlation of the pixels inside the single 2-D block is considered and the correlation from the pixels of the neighbouring blocks is neglected. Secondly it is impossible to de-correlate the blocks at their boundaries using DCT. There are undesirable blocking artefacts. Moreover scaling as add-on is an additional effort.

B. Digital Watermarking Using Wavelet Decomposition:

The basic idea of DWT in image processing is to multi-differentiated decompose image into sub-image of different spatial domain and independent frequency district. Then transform coefficient of sub-image **[9]**.

After the original image has been DWT transformed, it is decomposed into 4 frequency districts which is one low-frequency district(LL) and three high-frequency districts(LH,HL,HH). If the information of low-frequency district is DWT transformed, sub-level frequency district information will be obtained. Some features of DWT that makes it suitable for watermarking are multi-resolution analysis, good for signal having high frequency components for short durations and low frequency components for long duration, has higher compression ratios and avoids blocking artefact, allows good localization both in time and spatial frequency domain, provides for inherent scaling, higher flexibility to choose Wavelet function and good PSNR value.

In Wang algorithm et al the watermark is added to significant coefficients in significant sub-bands. First, the multi-threshold wavelet code is used to achieve the image compression purpose that adopts different initial thresholds in different sub-bands. This approach picks out coefficients whose magnitude is larger than the current sub-band threshold. The sub-band's threshold is divided by two after watermarking a sub-band.

A method for multi-index decision based watermarking is proposed by Zhihui and Liang. It is designed and implemented in the DCT domain as well as the wavelet domain utilizing HVS. Their experimental results show that the watermark based on the wavelet transform more closely approaches the maximum data hiding capacity in the local image.

Although Wavelets yield very good results in terms of compression ratio and the PSNR value, it has certain disadvantages like it requires longer compression time and cost compared to DCT and using larger DWT basis functions produces blurring and ringing noise near edges in images.

C. Fractal Transform

Fractal geometry is based on the idea of self-similar forms. To be self-similar, a shape must be able to be divided into parts that are smaller copies similar to the whole. Because of the smaller similar divisions of fractals, they appear similar at all magnifications. Fractals are defined by recursive formulas. Fractals are Self-Similar, have Non- Integer Dimension, are Invariant under scaling, cannot be described analytically and are produced by *iteration*. The key advantages of Fractals include Resolution and device independency, fast decompression and higher compression ratios. Fractal compression became a practical reality with the introduction **by Jacquin1** of the partitioned IFS **[3]**.

I. Partition schemes

The first decision to be made is type of image partition used for the range blocks. Since domain blocks must be transformed to cover range blocks, this decision, together with the choice of block transformation, restricts the possible sizes and shapes of the domain blocks. Some of the Partition schemes include Fixed Size Square Blocks, Quadtree, Horizontal-Vertical partition and Polygonal Blocks. On comparison quadtree was found to provide the best rate distortion results.

II. ENCODING

Fractal coding is achieved by representing a signal X by a quantized representation of a contractive transform T, which is chosen such that the fixed point X_T of T is close to X.

III. DECODING

Reconstruction of the encoded image is achieved by computing the fixed point of the image transform from its encoded coefficients. Some methods for decoding include Standard Decoding, Successive Correction Decoding, and Hierarchical Decoding.

However, using fractal transform for watermarking has certain limitations like slow compression rate due to iteration, being an unproven technology and Patented and expensive for a small user. It is less effective compared to simpler zero-tree recently introduced in wavelet scalar quantization. Moreover, all images do not have self-similar patterns; hence it may not be suitable for general use.

D. Inference on Transform Domain Techniques

Each of the transform domain techniques has its own limitations which make them unsuitable for watermarking One solution to this can be to use the hybrid techniques like Combining DCT and DWT, Combining SVD and DWT and using wavelet domain fractal encoding.

However the first two ideas had not produced significant result and it was concluded that Using fractal domain wavelet encoding provides for better signal quality at fast rate along with high compression ratio at low cost. Hence the review further discusses wavelet domain fractal encoding **[11]**.

Iano et al proposed another method where Fast fractal encoding using Fisher's domain classification is applied to the low-pass sub-band of wavelet transformed image and a modified set partitioning in hierarchical trees coding, on the remaining coefficients. The proposed scheme promoted an average of 94% reduction in encoding-decoding time comparing to the pure accelerated fractal coding results. XIAO et al proposed another model where encoding leads to deal the image by 3 level wavelet decomposition, and the high frequency section of the decomposed image adopts scathe-less predicting encoding, while the low frequency section adopts fractal compression encoding. In the process of decoding, the high frequency section carries out scathe-less predicting decoding, while the low frequency section is decoded to IFS code which is used for sub-image reconstruction.

IV. Watermarking Attacks

It is categorized under following groups

Removal attacks

It aims at complete removal of watermark information from data without cracking the security. This includes denoising, quantization, remodulation and collusion attacks **[4].**

Geometric attacks

It aims to distort the watermark detector synchronization with the embedded information. The detector could recover embedded watermark information only when perfect synchronization is regained. Robustness to geometric distortions relies on the use of either a transform invariant domain or specially designed periodic watermarks whose Auto-covariance function allows estimation of the geometric distortions**.**

Cryptographic attacks

It aims at cracking the security methods in watermarking schemes or to embed misleading watermarks. It includes brute-force search and Oracle attack.

Protocol attacks

It aims at creating ownership deadlock. It includes invertible watermarks attack where the attacker subtracts his watermark from the watermarked data and claims to be owner. Another protocol attack is the copy attack where the goal is to estimate a watermark from watermarked data and copy it to target data.

V. Digital Watermarking and Copyright Protection

Unauthorized detection and decoding of watermark should be prevented for copyright protection. However it impractical to make such a watermarking system because the design of watermarking algorithm is a compromise between conflicting requirements of robustness, security, imperceptibility and capacity etc. One solution to this can be combining Chaotic Cryptography with the Watermarking System **[5]**. Thus message is encrypted before embedding and decrypted after it is detected. Such a system requires two keys. The watermark Key K_w for controlling the watermarking layer and the encryption key K_c controlling the encryption layer. Thus at the embedder the watermarked work is given by:

$$c_w = E_{k_w}(c_0, m_c) = E_{k_w}\left(c_0, E_{k_c}(m)\right) \qquad (4)$$

While at the detector end, the reverse process is employed

$$m = D_{k_c}(D_{k_w}(c_w) \qquad (5).$$

The encryption ensures message security besides preventing the Adversary from determining the watermark. Security can be further improved using Timestamp from Trusted Third Parties or developing Non-Invertible Watermarks and combining them respectively with Chaotic Encryption.

Timestamp Based Watermarking

Hung et al 2005 proposed a method that combines image feature extraction and timestamp technique **[6]**. This scheme can resist both geometric distortion, signal processing and protocol attacks. It has three characteristics- Not modifying original image, Register at fair third party, and using timestamp to prove who the real owner is. However the scheme is unsuitable as due to instability of feature points, high cost of using Bi-orthogonal Wavelet and a trade-off between false alarm probability and miss probability.

Hu et al, 2009 proposed another method that uses the important features of the original image to construct watermark, without amendment to the images of these features and protects the copyright of image better based on time-stamp and digital signature **[7]**. It includes DHWM which is a cult of watermark and key exchange algorithm, and TTP. When there is dispute in copyright, TTP can nail down copyright ownership through technical means and Key exchange. However it has some disadvantages like increase in network delay and reduced performance.

Non-Invertible Watermarking

Craver et al proposed first non-invertible watermark scheme by applying cryptographic primitives. The main idea is to compute the watermark key in a one-way manner from the original work so that an attack would have to invert one-way function to compute a suitable fake original, fake watermark and fake watermarking-key. Adelsbach et al analysed the impact of watermarking scheme's false-positives rate on security of Craver's scheme and found that it is insecure unless false-positives rate is low.

Tang et al 2009 proposed another method that combines a non-invertible watermark scheme with an HVS model adaptive watermark **[8]**. It resists attacks by making a restriction on the adaptive coefficients thereby making the adaptive and noninvertible watermark scheme appropriate.

VI. Conclusion

Digital watermarking can be achieved by using both transform domain and spatial domain techniques. For optimum signal quality and faster transmission in very low bandwidth channels, the demand for high compression ratios and greater speed in the coding/decoding techniques has increased. Wavelet based Fractal coding can be one of the solution to it as it provides for better perceptual quality at higher compression ratio. This paper also discusses two important watermarking techniques of Time stamping and Non-Invertible Watermarks as measures for Copyright Protection. Creating robust watermark which can be used on a commercial scale, and that stands in a court of law is still a challenging research problem as not one algorithm is secure against all types of attacks.

References

[1] Chun-Shien Lu, Multimedia Security: Steganography and Digital Watermarking Techniques for Protection of Intellectual Property, Idea Group Publishing, 3 Henrietta Street, Covent Garden, London

[2] Christine I. Podilchuk and Edward J. Delp, "Digital Watermarking: Algorithms and Applications", IEEE Signal Processing Magazine, July 2001

[3] Brendt Wohlberg and Gerhard de Jager, "A Review of the Fractal Image Coding Literature", IEEE Transactions on Image Processing, Dec. 1999.Vol. 8

[4] Nikolaidis A., Tsekeridou S., Tefas A., Solachidis, V, "A survey on watermarking application scenarios and related attacks" International Conference on Image Processing Proceedings, IEEE, 2001, Vol. 3,pp. 991 - 994

[5] Ingemar *Cox*, Matthew *Miller*, Jeffrey *Bloom,*: *Digital Watermarking*, 2nd Edition, Academic Press, San Diego, USA, 2002

[6] Hung-Mn Sun, Chen-Jung Hong, Chiung-Hsun Chen, "A New Approach to Feature-based Copyright Protection of Images", Information Technology: Research and Education, 3rd International Conference, IEEE, September 2005, pp. 233 - 237

[7] Hu Chaoju, Wang Xuning, "Zero Watermark Protocol Based on Time-stamp and Digital Signature", Information Technology and Applications, IEEE, May 2009, Vol. 3, pp. 193 - 196

[8] Xiaojun Tang, Qingxiu Du, Shuwu Zhang, "Non-Invertible DWT Domain Image Watermarking Using Human Visual Model", Information, Communications and Signal Processing, 2009, pp.1 - 5

[9] Brannock, E., Weeks, M., Harrison, R., "Watermarking with wavelets: Simplicity leads to robustness:" Georgia State Univ., Atlanta, South-eastcon, IEEE, April 2008, pp. 587 - 592

[10] Wei Zhihui, Xiao Liang, "An evaluation method for watermarking techniques" International Conference on Multimedia and Expo, IEEE, 2000, Vol.1, pp. 373 – 376

[11] Li Yang, Du Sidan, "Research on Wavelet Domain Fractal Coding in Digital Watermarking" IEEE International Conference on Multimedia and Expo, 2005, pp. 61 - 64