A Short Summary of Digital Watermarking in Multimedia Information

Professor Irwin King, Professor F Y Duan
Department of Computer Science & Engineering
The Chinese University of Hong Kong

Introduction

Digital media facilitate efficient distribution, reproduction, and manipulation over networked information systems for image, audio clips, and videos. However, the fact that an unlimited number of perfect copies can be illegally produced is a serious threat to the rights of content owners. These efficiencies also increase the problems associated with copyright enforcement. A number of technologies are being developed to provide protection from illegal copying. They include: (1) encryption methods — the use of a public and private keys to encode the data so that the image can only be decoded with the required key, (2) site security methods — the use of firewalls to restrict access, (3) using publicly accessible low quality “thumbnail” images, and (4) digital watermarking, this includes the robust unobtrusive labeling of an image with information pertaining to copyright, and the use of image checksums or other techniques to detect the manipulation of image data.

To address the non-obtrusive copyright enforcement issue, digital watermarks (i.e., author signatures) are under investigation. Watermarking is the process of encoding hidden copyright information in an image by making small modifications to its pixel content. Unlike encryption which protects content during the transmission of the data from the sender to receiver, digital watermarking does not restrict access to the image information. Watermarking compliments encryption by embedding a signal directly into the data. Thus, the goal of a watermark is to always remain present in the data to provide solid proof of ownership. It should be noted that embedded signaling or watermarking can be used for a variety of other purposes other than copyright control. For example, it can be used for owner identification, to identify the content owner, fingerprinting, to identify the buyer of the content, for broadcast monitoring to determine royalty payments, and authentication, to determine whether the data has been altered in any manner from its original form. However, here we restrict our discussion here to issues that are related to copyright control.

Although there are two main divisions of watermarks, e.g., visible and invisible, this paper focuses on algorithms and techniques for invisible watermarks. In general, there are two basic requirements of invisible watermarks. The watermarks should be (1) perceptually invisible and (2) robust to common signal processing and intentional attacks. Early research on digital watermarking concentrates on the first objective without considering the second one. Recently much work has been devoted to designing robust watermarking schemes [17, 7, 23]. Perceptual models have also been incorporated to make the best tradeoff between perceptual invisibility and robustness to signal processing [23].

The goal of this paper is to give a brief summary of various digital watermarking techniques available for the purpose of authentication, forgery detection, and copyright...
enforcement. The paper is organized as follows. In the next section, we outline desirable properties of a watermark for copyright control, which can be quite different from watermarks for authentication purposes. Section 3 introduces a framework in which we will discuss the many different proposed watermark techniques that have been invented in recent years.

2 Properties of Watermarking Techniques

To be effective, the watermark should be: (1) perceptually invisible within the host media; (2) statistically invisible to thwart unauthorized removal; (3) readily extracted by the image owner; and (4) robust to accidental and intended signal distortions incurred by the host image, e.g., filtering, compression, re-sampling, retouching, cropping, etc. These characteristics are discussed in more detail.

2.1 Unobtrusive (Difficult to notice)

The watermark should be perceptually invisible to the viewer nor should the watermark degrade the quality of the content. In earlier work [7, 9, 6, Cox et al., had used the term “imperceptible”, and this is certainly the ideal. However, if a signal is truly imperceptible, then perceptually-based lossy compression algorithms should, in principle, remove such a signal. Current state-of-the-art compression algorithms probably still leave room for an imperceptible signal to be inserted. This may not be true of next generation compression algorithms. Thus, to survive the next generation of lossy compression algorithms, it will probably be necessary for a watermark to be noticeable to a trained observer.

2.2 Robustness

The watermark must be difficult (hopefully impossible) to remove. Of course, in theory, any watermark may be removed with sufficient knowledge of the process of insertion. However, if only partial knowledge is available, for example, the exact location of the watermark within an image is unknown, then attempts to remove or destroy a watermark by say, adding noise, should result in severe degradation in data fidelity before the watermark is lost.

In particular, the watermark should be robust to the following attacks and characteristics:

- Universality: The same digital watermark algorithm should apply to all three media types. This is potentially helpful in the watermarking of multimedia products. Also, this feature is conducive to implementation of audio, image, and video watermarking algorithms on common hardware.

- Tamper-resistance: The watermarking techniques should be robust to legitimate signal distortions as well as intentional attacks to remove or tamper with the digital watermark.

- Common Signal Processing: The watermark should still be retrievable even if common signal processing operations are applied to the data. These include, digital-to-analog and analog-to-digital conversion, resampling, requantization (including dithering and recompression), and common signal enhancements to image contrast and color, or audio bass and treble, for example.

- Common Geometric Distortions: Watermarks in image and video data should also be immune from geometric image operations such as rotation, translation, cropping, and scaling.

- Subterfuge Attacks: Collusion and Forgery: In addition, the watermark should be robust to collusion by multiple individuals who each possess a watermarked copy of the
data. That is, the watermark should be robust to combining copies of the same data set to destroy the watermarks. Further, if a digital watermark is to be used as evidence in a court of law, it must not be possible for colluders to combine their images to generate a different valid watermark with the intention of framing a third-party.

3. A Framework for Watermarking

A watermarking framework consists of three parts: (1) the watermark, (2) the marking algorithm, and (3) the verification algorithm. Each owner has a unique watermark which the owner would like to embed into his/her proprietary work. The marking algorithm incorporates the watermark into the multimedia medium. The verification algorithm authenticates the watermarked information, determining both the owner and the integrity of the image.

Figure 3: The schematic of watermark testing algorithm.

There are currently numerous techniques for applying a digital watermark to an image. The techniques can be divided into two major categories based on the desired application for the watermark (1) to detect image tampering and (2) to embed copyright information. The classification shown in Fig. 4 highlights a number of interesting characteristics of the various watermarking techniques. The techniques used to detect image tampering tend to be fragile and introduce insignificant data loss. Robust watermark algorithms used to embed copyright data tend to introduce increased visible artifacts, the notable exception are the spread spectrum methods of digital watermarking which are particularly useful for copyright labeling being both robust and invisible. Further classification of digital watermarking can be achieved by categorising the image data by the robustness of the watermarking technique and the obtrusiveness, the amount of visible artifacts (or data loss) introduced. With the limited scope of this paper, we will focus only on digital watermarking algorithms that embed copyright information into the targeted media.

There are several major algorithm areas with many variations. One major algorithm is based on the modification of Least Significant Bit (LSB) of the pixel content [20, 27, 28, 26]. Watermarks also can modify the image’s spectral or transform coefficients directly. These algorithms often modulate Discrete Cosine Transform (DCT) coefficients according to a sequence known only to the owner [7, 13]. Watermarking techniques may be image dependent. These techniques increase the level of the watermark in the image while maintaining the imperceptibility of the mark [23, 3, 11]. For example, one of the wavelet methods incorporates features from most of the above techniques [18, 19]. Its implementation lends itself to watermarking data rate-scalable video [21]. Time stamps thwart a clever attack proposed by IBM [10] on all of these watermarking schemes.

Visible watermarks also exist; IBM has developed a proprietary visible watermark to protect images that are part of the digital Vatican library project [16]. The watermarking itself is only a small part of any controlled access and
distribution scheme; a method for secure distribution would combine encryption with digital watermarking [22]. Lastly there is the hybrid technique where many of these techniques may be used in combination with each other. The sections below describe these watermarking algorithms in detail.

3.1 Digital Watermarking Techniques

3.1.1 Least Significant Bit Modification

The most common and early watermarking approaches modify the least significant bits (LSB) of an image based on the assumption that the LSB data are insignificant. Two LSB techniques are described in [20]. The author first replaces the LSB of the image with a pseudo-noise (PN) sequence, while the second adds a PN sequence to the LSB of the data. And another early watermarking method obtains a checksum of the image data, then embeds the checksum into the LSB of randomly chosen pixels [26]. Others add a modified maximal length linear shift register sequence to the pixel data. They identify the watermark by using the spatial cross-correlation function of the modified sequence and part of the watermarked image [20, 27, 28]. The Digimarc Corporation describes a method that adds or subtracts small random quantities from each pixel. Addition or subtraction is determined by comparing a binary mask of L bits with the LSB of each pixel. If the LSB is equal to the corresponding mask bit, then the random quantity is added, otherwise it is subtracted. The watermark is subtracted by first computing the difference between the original and watermarked images and then by examining the sign of the difference, pixel by pixel, to determine if it corresponds to the original sequence of additions and subtractions. The Digimarc method does not make use of perceptual relevance and is probably equivalent to adding high frequency noise to the image. As such, it may not be robust to lowpass filtering.

Turner [25] proposed a method for inserting an identification string into a digital audio signal by substituting the “insignificant” bits of randomly selected audio samples with the bits of an identification code. Bits are deemed “insignificant” if their alteration is inaudible. Such a system is also appropriate for two dimensional data such as images, as discussed in [20]. Unfortunately, Turner’s method may easily be circumvented. For example, if it is known that the algorithm only affects the least significant two bits of a word, then it is possible to randomly flip all such bits, thereby destroying any existing identification code.

In a recent paper, Macq and Quisquater [14] briefly discuss the issue of watermarking digital images as part of a general survey on cryptography and digital television. The authors provide a description of a procedure to insert a watermark into the LSB of pixels located in the vicinity of image contours. Since it relies on modifications of the least significant bits, the watermark is easily destroyed. Further, their method is restricted to images, in that it seeks to insert the watermark into image regions that lie on the edge of contours.

3.1.2 Information Tagging

Caronni [5] suggests adding tags—small geometric patterns—to digitized images at brightness levels that are imperceptible. While the idea of hiding a spatial watermark in an image is fundamentally sound, this scheme is susceptible to attack by filtering and redigitization. The fainter such watermarks are the more susceptible they are to such attacks and geometric shapes provide only a limited alphabet with which to encode information. Moreover, the scheme is not applicable to audio data and may not be robust to common geometric distortions, especially cropping.

Brassil et al. [4] propose three methods appropriate for document images in which text is common. Digital watermarks are coded by: (1) vertical shifting text lines, (2) horizontally shifting words, and (3) altering text features such as the vertical endlines of individual characters.
Unfortunately, all three proposals are easily defeated, as discussed by the authors. Moreover, these techniques are restricted exclusively to images containing text.

3.1.3 Quantization Noise Embedding

Tanaka et al. [24, 15] describe several watermarking schemes that rely on embedding watermarks that resemble quantization noise. Their ideas hinge on the notion that quantization noise is typically imperceptible to viewers. Their first scheme injects a watermark into an image by using a predetermined data stream to guide level selection in a predictive quantizer. The data stream is chosen so that the resulting image looks like quantization noise. A variation on this scheme is also presented, where a watermark in the form of a dithering matrix is used to dither an image in a certain way. There are several drawbacks to these schemes. The most important is that they are susceptible to signal processing, especially requantization, and geometric attacks such as cropping. Furthermore, they degrade an image in the same way that predictive coding and dithering can.

In [24], the authors also propose a scheme for watermarking facsimile data. This scheme shortens or lengthens certain runs of data in the run length code used to generate the coded fax image. This proposal is susceptible to digital-to-analog and analog-to-digital attacks. In particular, randomizing the LSB of each pixel's intensity will completely alter the resulting run length encoding. Tanaka et al. also propose a watermarking method for "colorscaled picture and video sequences". This method applies the same signal transform as JPEG (DCT of 8 x 8 subblocks of an image) and embeds a watermark in the coefficient quantization module. While being compatible with existing transform coders, this scheme is quite susceptible to requantization and filtering and is equivalent to coding the watermark in the least significant bits of the transform coefficients.

3.1.4 Statistical Techniques

Bender et al. [2] describe two watermarking schemes. The first is a statistical method called "Patchwork" that somewhat resembles the statistical component of Cox's proposal. Patchwork randomly chooses n pairs of image points, (ai, bi), and increases the brightness at ai by one unit while correspondingly decreasing the brightness of bi. The expected value of the sum of the differences of the n pairs of points is then claimed to be 2n, provided certain statistical properties of the image are true. In particular, it is assumed that all brightness levels are equally likely, that is, intensities are uniformly distributed. However, in practice, this is very uncommon. Moreover, the scheme may (1) not be robust to randomly jittering the intensity levels by a single unit, and (2) be extremely sensitive to geometric affine transformations.

In [24], the authors also propose a scheme for watermarking facsimile data. This scheme shortens or lengthens certain runs of data in the run length code used to generate the coded fax image. This proposal is susceptible to digital-to-analog and analog-to-digital attacks. In particular, randomizing the LSB of each pixel's intensity will completely alter the resulting run length encoding. Tanaka et al. also propose a watermarking method for "colorscaled picture and video sequences". This method applies the same signal transform as JPEG (DCT of 8 x 8 subblocks of an image) and embeds a watermark in the coefficient quantization module. While being compatible with existing transform coders, this scheme is quite susceptible to requantization and filtering and is equivalent to coding the watermark in the least significant bits of the transform coefficients.

3.1.5 Frequency Spectrum-Based Methods

Koch et al. [12] propose two general methods for watermarking images. The first method breaks up an image into 8 x 8 blocks and computes the Discrete Cosine Transform (DCT) of each of these blocks. A pseudorandom subset of the blocks is chosen, then, in each such block, a triple of frequencies is selected from one of 18 predetermined triples and modified so that their relative strengths encode a 1 or 0 value. The 18 possible triples are composed by selection of three out of eight predetermined frequencies within the 8 x 8 DCT block. The choice of the 8 frequencies to be altered within the DCT
block is based on a belief that the “middle frequencies... have moderate variance”, i.e., they have similar magnitude. This property is needed in order to allow the relative strength of the frequency triples to be altered without requiring a modification that would be perceptually noticeable.

Superficially, this scheme is similar to our own proposal and, in fact, also draws analogies with spread spectrum communication. However, the structure of their watermark is different from ours. The set of frequencies is not chosen based on any perceptual significance or relative energy considerations. Further, because the variance between the eight frequency coefficients is small, one would expect that their technique may be sensitive to noise or distortions.

This is supported by the experimental results which report that the “embedded labels are robust against JPEG compression for a quality factor as low as about 50%”. An earlier proposal by Koch and Zhao [13] used not triples of frequencies but pairs of frequencies, and was again designed specifically for robustness to JPEG compression. Nevertheless, they state that “a lower quality factor will increase the likelihood that the changes necessary to superimpose the embedded code on the signal will be noticeably visible”.

In a second method, designed for black and white images, no frequency transform is employed. Instead, the selected blocks are modified so that the relative frequency of white and black pixels encodes the final value. Both watermarking procedures are particularly vulnerable to multiple document attacks. To protect against this, Koch and Zhao propose a distributed 8 x 8 created by randomly sampling 64 pixels from the image. However, the resulting DCT has no relationship to that of the true image and consequently may be likely to cause noticeable artifacts in the image and be sensitive to noise.

3.1.6 Checksum Technique

This watermark is formed from the checksum value of the seven most significant bits of all pixels [26]. A checksum is the modulo-2 addition of a sequence of fixed-length binary words. It is a special type of hash function. In this technique, one word is the concatenation of eight 7-bit segments, which come from eight different pixels. Each pixel is involved in the checksum only once. The final checksum is fifty-six bits. The technique then randomly chooses the locations of the pixels that are to contain one bit of the checksum. The pixel locations of the checksum, together with the checksum itself, form the watermark. The last bit of each chosen pixel is changed (if necessary) to equal the corresponding checksum bit. This value must be kept secret.

To verify this watermark the checksum of a test image is obtained, and compared to the ideal version in watermark. Any discrepancy invalidates the image. The advantages of this technique are: (1) the embedding watermark only changes (on average) half of the pixels covered by watermark, (2) an image may hold many watermarks as long as they do not overlap, and (3) this method is very fast. On the other hand, the disadvantages of this technique are: (1) this watermarking method is fragile. Any change to either the image data itself or the embedded checksum can cause the verification procedure to fail, (2) the checksum method does not detect pixels swaps or similar attacks. A forger could replace a section with one of equal size and checksum, and (3) an attacker could remove the entire watermark by replacing the LSB plane.

3.1.7 Hybrid and Other Techniques

The hybrid approach combines several techniques together to synthesize a new variation of the watermarking algorithm. For example, Walton [26] uses a checksum on the image data which is embedded in the least significant bits of certain pixels. Others add a maximal length linear shift register sequence to the pixel data and identify the watermark by computing the spatial cross-correlation function of the sequence and the watermarked image [20]. Watermarks can be image dependent, using independent
visual channels [11], or be generated by modulating JPEG coefficients [3]. These watermarks are designed to be invisible, or to blend in with natural camera or scanner noise. Visible watermarks also exist; IBM has developed a proprietary visible watermark to protect images that are part of the digital Vatican library project [27].

In addition to direct work on watermarking images, there are several works of interest in related areas. Adelson [1] describes a technique for embedding digital information in an analog signal for the purpose of inserting digital data into an analog TV signal. The analog signal is quantized into one of two disjoint ranges, (0, 2, 4,...,1, 3, 5,..., for example) which are selected based on the binary digit to be transmitted. Thus Adelson’s method is equivalent to watermark schemes that encode information into the least significant bits of the data or its transform coefficients. Adelson recognizes that the method is susceptible to noise and therefore proposes an alternative scheme wherein a 2 x 1 Hadamard transform of the digitized analog signal is taken. The differential coefficient of the Hadamard transform is offset by 0 or 1 unit prior to computing the inverse transform. This corresponds to encoding the watermark into the least significant bit of the differential coefficient of the Hadamard transform. It is not clear that this approach would demonstrate enhanced resilience to noise. Furthermore, like all such least significant bit schemes, an attacker can eliminate the watermark by randomization.

4 Conclusion

The proliferation of network multimedia systems dictates the need for copyright protection of digital property. To conclude, any successful watermarking algorithm would have to exploit properties of the human visual system and combine these with effective modulation and channel coding. Future work will concentrate on producing watermarks that are robust to filtering, lossy image compression, noise corruption and changes in contrast. In addition these algorithms must anticipate possible attacks on the integrity and security of the watermark and to devise suitable countermeasures. This paper serves as a brief summary on several more recent and popular digital watermarking techniques for multimedia information systems.

References


