Indiscernible Text Digital Watermarking

Lakshman Ji
M.Tech Student Biet Lucknow,
Uttar Pradesh, India

Ravi Asthana
Asst. Professor Biet Lucknow,
Uttar Pradesh, India

DOI: 10.23956/ijarcsse/SV7I5/0272

Abstract: Digital watermarking is a method in which the information is embedded directly and indiscernibly into digital data, e.g., image, video or audio signals, also called original data or host data, to form watermarked data. Loosely analogous to watermarks in paper documents, the embedded information is bound to the watermarked data wherever it goes. The embedded information should still be decodable from the watermarked data, even if the watermarked data is processed, copied, or redistributed. Potential applications of digital watermarking include copyright protection, distribution tracing, authentication and conditional-access control.

Digital watermarking of images can be done either using spatial techniques or transform techniques. Spatial techniques are simple and efficient but not robust whereas transform techniques are robust but complex. The algorithms developed are either for some specific applications or have very limited robustness. Hence, an attempt has been made to develop a generic watermarking scheme for grayscale digital images that can embed a meaningful text as watermark.

Keywords: RSA, DSA, OCR, DCT

I. INTRODUCTION

The flowering of sophisticated computer networks and that of Internet has prospected a new means of scientific, entertainment, business, and social opportunities in the form of electronic publishing and advertising, real-time information delivery, transaction processing, product ordering, web newspapers and magazines, digital repositories and libraries, network video and audio, personal communication, and lots more. The users are enjoying the convenience and advantages of advancement in technology. Digital media offer several benefits over analog media: the quality of digital audio, images and video signals are far more superior to that of their analog versions. Editing is a child’s play as one can access the exact distinct locations that one wishes to alter. Copying is simple with no loss of allegiance. A copy of a digital media is identical to the original.

Multimedia authentication techniques are usually designed either using digital signature or watermarking. A digital signature is a data string which associates a message with some originating entity to simulate the security properties of a handwritten signature on paper. Digital signature technique is a section of cryptography, and is based on algorithms like RSA, DSA etc. It is an encrypted version of the message digest extracted from the data and is usually stored in a separate file, which can be attached to the data to prove integrity and ingenuity. Normally, digital signature schemes (figure 1-1) provide two algorithms, one for signing which involves the user's secret or private key, and one for verifying signatures which involves the user's public key. Conceptually, the hash of data to be transmitted is generated along with originator information and is encrypted and stored in a separate file called “digital signature”. The receiver gets both the data and the signature file. The signature is decrypted at the receiver end and the hash is compared with the received data. If both match, the received data is treated as authentic.

This traditional manner of image authentication requires an auxiliary channel for the storage and transmission of a digital signature for each image. This increases the bandwidth requirements and imposes restrictions on implementation. Further, change in single bit of data results in non-authorization of the data. Thus, any kind of compression technique or the use of software that intentionally/ un-intentionally strips the header information during format conversion of the multimedia data – results in the failure of technique.
Digital watermarking is a method in which the information is embedded directly and imperceptibly into multimedia data, called original data or host data, to form watermarked data. Loosely analogous to watermarks in paper documents, the embedded information is bound to the watermarked data wherever it goes (figure 1). The embedded information should still be decodable from the watermarked data, even if the watermarked data is processed, copied, or redistributed. Potential applications of digital watermarking include distribution tracing, copyright protection, authentication and conditional-access control. Thus, the embedded information could be for instance a user-ID, a serial number for a certain copy of a document, or authentication information, etc.

Digital Watermarking offers an auspicious substitute to digital signatures in image authentication applications. The use of watermarks instead of digital signatures offers additional functionality by exploiting inherent properties of image content. The major advantage is the direct embedding of authentication information into the image data (figure 2). As a result, the authentication information persists even when the watermarked image goes under different attacks and format conversions.

II. DIGITAL WATERMARKING

Unlimited number of replicas of the original content can be made from unprotected digital content. This makes the content creators and content owners more concerned about the copyrights management of their digital contents. Apprehensions for the protection of copyrighted digital intellectual properties have been high, since 1980s. The cryptographic algorithms can resolve many of these issues. However, these solutions can only protect the digital contents if they never leave the digital domain and till the content is not decrypted. Once content is decrypted and copied, there is no protection offered. This situation calls for technological solutions to be used in addition to the cryptography technology. Digital Watermarking technologies, a descendent of steganography, have evolved as a major solution for the resolution. This group of technologies provides methods to “imprint” additional data or messages into multi-media contents such as still image, video and audio data. Generally, the imprinted data is invisible (or inaudible) to the ordinary users, and is difficult to be separated from the host media. The imprinted data can be extracted, but not removed, from the imprinted host media by the legitimate owner, as and when required.
Digital watermarking is divided into two categories: perceptible watermarks and imperceptible watermarks. Perceptible watermarks are visible watermarks and have been around for centuries in the form of stamps, seals, etc. Visible watermarks are used in much the same way as their bond paper counterparts. An ideal visible watermark is the one that can only be noticed on the careful examination of the image and should not be lost by simple image processing operations. The main advantage of using visible watermark (figure 1-4) is that they convey an immediate claim of ownership and discourages unauthorized use of precious stuff.

![Lena Image](image1.jpg) ![Watermark Image](image2.jpg)

**Figure 3:** Figure showing visible watermarking

Imperceptible watermark are invisible watermark and is a more complex concept to implement. The major advantage of invisible watermark is lesser threat of buccaneer trying to manipulate the stuff for removing watermark and gives a clear image with no identifiable noise. Invisible watermarks do not change the signal to a perceptually great extent, i.e., there are only minor variations in the output signal. An example of an invisible watermark (figure 5) is when some bits are added to an image modifying only its least significant bits.

![Watermark Image](image3.jpg)

**Figure 4:** Figure showing invisible watermarking using LSB substitution.
Visible and invisible watermarks both serve to deter theft but in different ways (Berghel and Gorman, 1996). Visible watermarks are like a “Do not trespass” sign whereas invisible watermarks are like the dye bank used to indelibly mark the hands and cloths of bank robbers.

**Basic Watermarking Principles**

All watermarking methods share the same building blocks: a watermark embedding system and a watermark recovery system and/or a comparison system. Figure 1-6 below shows the generic watermark embedding process. The input to the scheme is the watermark (or signature), the image (or cover data) and an optional key. The key is used to enforce security, which is preserving unauthorized parties from recovering and manipulating the watermark. The output of the watermarking scheme is the watermarked data.

![Image](image.png)

*Figure 5: Generic digital watermarking scheme*

The generic watermark recovery process is depicted in figure 1-7. Inputs to the scheme are the test image data, the original image data and the optional key. The output is the recovered watermark, if the test image data is watermarked and is an authentic one. The recovered watermark is subjected to the comparator to verify the authenticity of the watermark embedded.

![Image](image.png)

*Figure 6: Generic watermark recovery and comparison scheme*

**Application of Digital Watermarking**

The digital watermark should ideally be inseparable from the content. Therefore, digital watermarking seems to be a suitable method for associating additional information with the content. There are numerous watermarking applications:

- **Fingerprinting:** It is a technique by which the copy of the content is made unique to the person whom it is delivered. Thus, if the content is shared or distributed, source can be easily identified. Since a digital watermark is inseparable from the content and can be inserted in an imperceptible manner, it is an appropriate solution for fingerprinting.

- **Copyright protection:** The objective is to control access to and prevent illegal copying of copyrighted document. Digital watermarks enable a means to embed information about the rules of usage and copying of a given digital document.

- **Content ownership:** When the contents are made available through Internet and the content owner desires to indicate the ownership of the underlying materials, so that an observer might be encouraged to sponsor the owner of the content, visible watermarks can be deployed. In another scenario where the seller of the content suspects that his content can be edited and republished, invisible watermarks can be deployed.

- **Authentications and Integrity:** In this case, images have been scanned and the content owner desires the ability to detect any alteration of the images, digital watermarks can verify that content is genuine and from an authorized source, as well as verify that the content has not been altered or falsified. The presence of unhampered digital watermark can ensure that the content has not been altered. In this case, the watermark must be designed in a way that any alteration of the content lead to the destruction of the watermark or creates a mismatch between the content and the watermark that can be easily detected.
Covert Communication: The practice of embedding message into another message in a manner that prevents the observer from suspecting that anything unusual is happening. It is one of the earliest applications of watermarking. The application has been formulated by Simmons (1984) as the “prisoner’s problem”.

General Digital Watermarking Methods

This section outlines general digital watermarking methods for text, images, audio and video data. These medium differ in ways that present unique problems for watermarking. Still the principle of watermarking remains the same.

- **Image**: Watermarking of images typically modify pixel intensities or transform coefficients. An image may be subjected to a great deal of manipulation such as filtering, cropping, geometric transformations, and lossy compression etc. Thus imperceptibility, robustness are usually the most important properties of image watermarks. One of the plausible hardships in image watermarking is the availability of finite bandwidth, thus, as the image size decreases, the permissible watermark length decreases.

- **Audio**: As in images, the watermarking desirable characteristics for audio data are imperceptibility, watermark bit rate, robustness, security, and computational cost. Audio watermarking uses the time and frequency masking properties of the human ear to conceal the watermark, and make it inaudible. One of the techniques for audio watermarking is by hiding short echoes, of size a few milliseconds, that can’t be perceived by human auditory system. One of the plausible hardships is the need to keep the watermark even after a certain number of re-encoding and digital-analog-digital transfer processes.

- **Video**: Digital video is a sequence of still images, and many image watermarking techniques, those can be implemented in real-time, can be extended to video. In contrast to single images, the large video bandwidth means that long messages can be embedded in video. Speedy embedding and extraction to watermark is one of the key issues because of the huge amount of data that must be processed. Beyond those for still image compression, video watermarking poses some unique requirements like frame shuffling, inter-frame collusion, etc.

- **Text Document**: Watermarking raw texts are very difficult because of the difficulty to define the appropriate place in which to embed hidden information. Brassil et. al., (1999 July) have investigated text watermarking and proposed a variety of methods for embedding hidden messages in PostScript documents. They exploited layout information of the text like word spacing, line spacing, text formatting etc. One of the major challenges is that optical character recognition (OCR) can remove any layout information but OCR is expensive, imperfect and often requires manual supervision.

Attacks on Digital Watermarks

In most watermarking applications, the watermarked data is likely to undergo under various attacks or processing like lossy compression, signal enhancement or digital-analog and analog-digital conversion etc. before it reaches the intended receiver. Such processing intentionally or unintentionally impairs the embedded watermark. In watermarking terminology, an attack is any processing that may impair detection of the watermark or communication of the information conveyed by the watermark. The processed, watermarked data is then called attacked data. The most important aspect of any watermarking scheme is its robustness against attacks. Watermarking attacks can be categorized widely under four classes of attacks: removal attacks, geometric attacks, cryptographic attacks, and protocol attacks.

- **Removal attacks**: These attacks targets removal of the watermark information from the watermarked data without deciphering the security of the watermarking algorithm. Hence, no processing, even prohibitively complex, can recover the watermark information from the attacked data. This category includes de-noising, remodulation, quantization, and collusion attacks. Though these methods do not succeed in complete watermark removal, but they may nevertheless damage the watermark information significantly. Sophisticated removal attacks try to optimize operations, generally using statistical models, to damage the embedded watermark as much as possible while maintaining high quality of the attacked document. Collusion attacks are applicable when many copies of a given data set, each watermarked with a different watermark, can be obtained by an attacker. In such a case, a successful attack can be achieved by averaging all copies or taking only small parts from each different copy. Recent results show that even a small number of different copies can lead to successful watermark removal.

- **Geometric attacks**: These attacks aim to distort the watermark detector synchronization with the embedded information rather than actually removing the embedded watermark. The watermark detector could only recover the embedded watermark information when perfect synchronization is re-established, which is practically unfeasible.

- **Cryptographic Attacks**: These attacks intend at identifying the security methods used in the watermarking scheme. One of the common techniques for it is brute-force search. Once security method identified, the way to remove the embedded watermark information or to embed misleading watermarks is explored. Practically, application of these attacks is restricted due to their high computational complexity.

- **Protocol Attacks**: These attacks aim at attacking the entire concept of the watermarking application. One type of protocol attack is based on the concept of invertible watermarks (Craver et. al., 1997 October) whereby the attacker subtracts his own watermark from the watermarked data and claims to be the owner of it. This creates ambiguity regarding the true ownership of the data. Hence, for copyright protection applications, watermarks...
need to be non-invertible. The requirement of non-invertibility of the watermarking technology implies that it should not be possible to extract a watermark from a non-watermarked document. Another protocol attack is the copy attack where the goal is not to destroy the watermark or impair its detection, but to estimate a watermark from watermarked data and copy it to some other data, called target data (Kutter et al., 2000). The estimated watermark is adapted to the local features of the target data to satisfy its imperceptibility. The copy attack is applicable when a valid watermark in the target data can be produced with neither algorithmic knowledge of the watermarking technology nor the knowledge of the watermarking key.

Besides these generalized attacks, the major attack that a computer novice performs intentionally or unintentionally is to enhance image quality, reduce size, JPEG compression, brightness and contrast, cropping and noise.

- **JPEG compression**: JPEG (Joint Photographic Expert Group) is a standardized image compression mechanism. This technique provides a compression method that is capable of compressing color or grayscale images.

  The main characteristic of the JPEG compression is the ability to compress an image file by discarding information that is not perceived by human visual system. Slight changes in color are not perceived well by the human eye, whilst slight changes in intensity are noticeable. Thus, high quality compression is achieved for colored images (20:1 to 25:1) thru JPEG compression in comparison to grayscale images (5:1). There are two schemes of compression in JPEG:
  - **Lossy**: The compressed image when decompressed has not the same quality as the original one.
  - **Lossless**: This scheme does not lose any of the image data (the same image quality) in other words the compressed image when decompressed; the image looks the same with the original one.

![Figure 7: DCT Block](image)

In digital watermarking, robustness issue is mainly focused on the lossy compression. The lossy operation in the whole process of JPEG compression is during the quantization process carried on the DCT transformed data. As a result of this quantization process, many higher frequency components are rounded to zero, and the rest become small positive or negative numbers.

The DCT blocks (8x8), have 64 pixels containing 1 DC value and 63 AC values (Figure 8). The DC pixel (0x0) is very important to JPEG compression. Changing the DC coefficient affect the luminosity of the image and modification of this coefficient affects all the others coefficients into matrix. In the figure 1-8, the black pixel represents the DC coefficient and the grey ones represent the AC coefficients where there are not any significant changes after the JPEG compression.

Digital watermarking is claimed to be the ultimate solution for copyright protection over Internet. However, some problems related to robustness and security of watermarking algorithms to intentional or unintentional attacks still remain unsolved. These problems must be solved before digital watermarking can be accepted practically for copyright ownership protection in digital media.

### III. CONCLUSION

The objective of this work was to present a novel watermarking technique for colored images with high robustness without compromising the quality of the watermarked image. The state-of-the-art technique developed is cogent to incorporate both colored watermarks and text watermarks and resulted in high imperceptibility, excellent quality of watermarked image, greater sustenance to image compression like JPEG compression. The sustenance of the technique to common image manipulations like cropping, median filtering, Gaussian filtering, color replacement etc., that are done from various sophisticated image manipulating tools, is promising.

### REFERENCES


