Survey on Digital Image Watermarking: Techniques and Attacks

Amitoj Kaur[1], Jagroop Kaur[2]
[1]Student, M.Tech CE, UCoE, Punjabi University Patiala
[2]Assistant Professor, CE, UCoE, Punjabi University, Patiala

Abstract

Increase of the use network i.e. the internet led to increase of digitization. Now a day's digital media is preferred over the analog media. But this increase led to threat that exact copies of the data can be made by the hackers without any loss of information. So to provide copyright protection to the owner digital watermarking is used. The copyright content is hided in the data such that data is still in readable form and is not tampered. In this paper various techniques of digital watermarking are discussed. Also the various types of attacks on digital image watermarking are listed.

Keywords: Digital watermarking, DCT, DWT, copyright protection.

1. Introduction

During the past decade, with the development of information digitalization and internet, digital media increasingly predominate over traditional analog media [1]. The expansion in multimedia will continue at even more steep rate with the availability of advance multimedia services such as e-commerce, telemarketing, etc [3]. The growth in usage of multimedia is result of their notable benefits in efficient storage, ease of manipulation and transmission [2]. But data authentication is one of the major issues in the exchange of digital data over the internet. As one of the concomitant side-effects, it is also becoming easier for some individual or group to copy and transmit digital products without the permission of the owner [1]. Hackers and the pirates are able to make perfect copies of the digital media without loss of the information.

As a matter of fact, the future development of networked multimedia systems is conditioned by the development of efficient methods to protect data owners against unauthorized copying and redistribution of the material put on the network. Whereas encryption systems do not completely solve the problem, because once encryption is removed there is no more control on the dissemination of data [3]. The solution to this problem is digital watermarking.

Covering many subjects such as signal processing, communication theory and Encryption, the research in digital watermark is to provide copyright protection to digital products, and to prevent and track illegal copying and transmission of them [1]. A digital watermark is a code carrying information about the copyright owner, the creator of the work, the authorized consumer and whatever is needed to handle the property rights associated to any given piece of information. The watermark is intended to be permanently embedded into the digital data so that authorized users can easily read it. At the same time, the watermark should not modify the content of the work but slightly (it should be unperceivable or almost unperceivable by human senses), and it should be virtually impossible for unauthorized users to remove it. By means of watermarking the work is still accessible, but permanently marked [3].

Digital watermarking can hide copyright information (watermark) into the "essence" of the multimedia object. And the hidden information should be imperceptible and robust against malicious attacks. Digital watermarking is possible because of the imperfections in the Human Vision System (HVS). Digital watermark utilizes the limitation of HVS to make it invisible, thus avoiding degrading original digital products, as well being hard to get identified or destroyed [2]. It is realized by embedding data that is invisible to the human visual system into a host image. Thus digital image watermarking is the process by which watermark data is hidden within a host image imposing imperceptible changes to the image. The root of watermarking as an information hiding technique can be traced from ancient Greece as Steganography [4].

1.2. Characteristics of digital image watermarking

- *Visibility*: an embedded watermark can be either visible or not visible according to the requirement [5].
- *Unobtrusive:* It should be statistically and perceptually invisible so that data quality is not degraded and attackers are prevented from finding and deleting it[3].
- *Robustness*: piracy attack or image processing should not affect the embedded watermark. Robustness might also incorporate a great degree of fragility to attacks, i.e. multimedia cover object is totally destroyed if it detects any tapering.
- *Readability*: A watermark should convey as much information as possible. A watermark should be statistically undetectable. Moreover, retrieval of the digital watermark can be used to identify the ownership and copyright unambiguously.
- *Integrity*: No loss of original multimedia carrier.
- *Accessibility*: both types of watermarking must permit for accessibility. Visible type allows information handling for any interested entity to call attention to the copy/reproduction rights, while the invisible type necessitates extra authorization information in order to access the watermark.
- *Security*: Security: watermarking accounts for the protection of ownership against forgery and unlawful threats. Invisible watermark should be secret and must be undetectable by an unauthorized user in general [5].

2. Process of watermarking

The watermarking process is divided in 3 parts:

1. Watermark embedding
2. Transmission
3. Watermark extraction

In the embedding process, the watermark may be encoded into the cover data using a specific key. This key is used to encrypt the watermark as an additional protection level. The output of the embedding process, the watermarked image, is then transmitted to the recipient. During this transmission process, the watermarked image may be subjected to attacks either deliberately or due to transmission error or noise. Therefore, there is no guarantee that the watermarked image received by the recipient is exactly the same data as that sent by the transmitter. This data nonetheless need to be decoded to extract the watermarked image. In the model shown in Figure 2, the original cover data is needed in the extraction process. This process is therefore called a blind technique. In a non-blind technique, the original cover data is unknown to the recipient hence the decoding process will have just rely on the watermark [6].

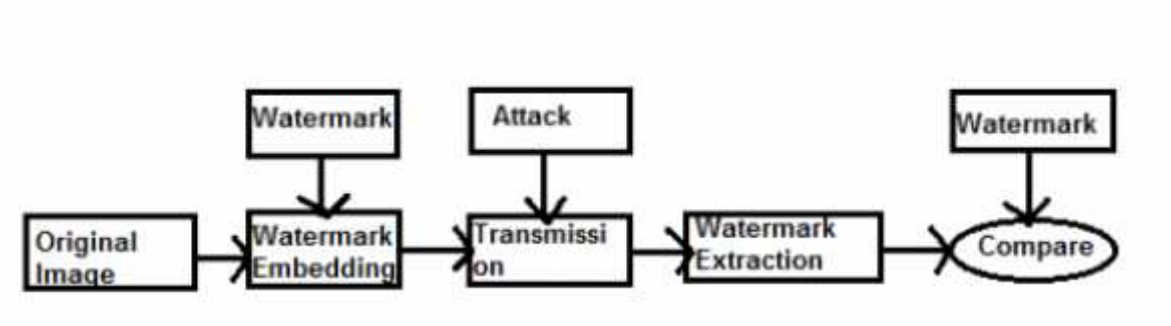

Fig.1. Process of Digital Watermarking

3. Classification of Digital Watermarking

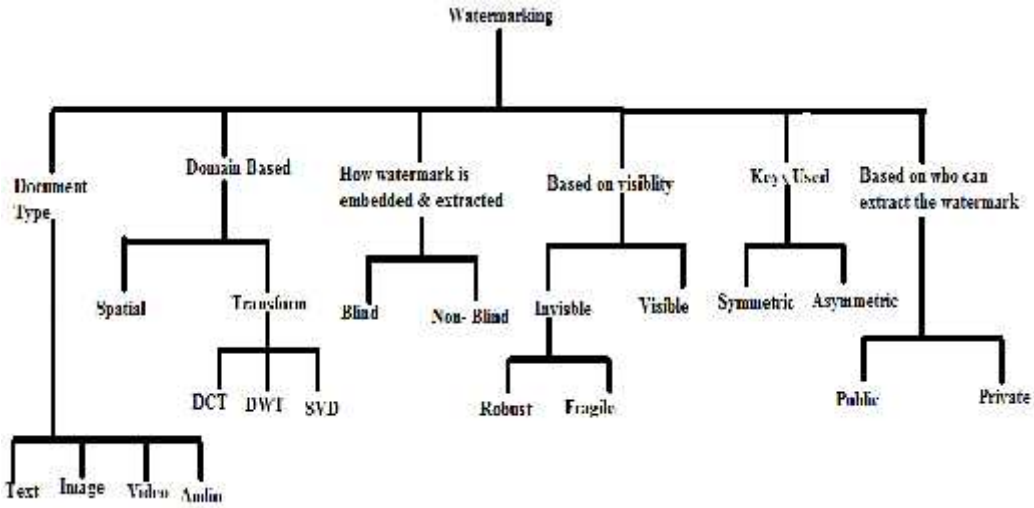

Fig.2 Classification of Digital Watermarking

- **According to the domain for watermark embedding**
  Spatial-domain watermarking technologies change the intensity of original image or gray levels of its pixels. This kind of watermarking is simple and with low computing complexity, because no frequency transform is needed. However, there must be tradeoffs between invisibility and robustness, and it is hard to resist common image processing and noise. A common technique used in spatial domain watermarking is LSB. Frequency-domain watermarking embeds the watermark into the transformed image. It is complicated but has the merits which the former approach lacks [1]. Techniques which use frequency domain watermarking are Discrete Cosine Transform and Discrete Wavelet Transform.

- **According to how watermark is detected and extracted**
  Blind-extracting watermarking means watermark detection and extraction do not depend on the availability of original image. The drawback is when the watermarked image is seriously destroyed; watermark detection will become very difficult. Non-blind-extracting watermark can only be detected by those who have a copy of original image. It guarantees better robustness but may lead to multiple claims of ownerships [1].

- **According to the ability of watermark to resist attack**

Fragile watermarks are ready to be destroyed by random image processing methods. The change in watermark is easy to be detected, thus can provide information for image completeness. Robust watermarks are robust under most image processing methods and can be extracted from heavily attacked watermarked image. Thus it is preferred in copyright protection [1].

- **According to media in which watermark is embedded:**
  Digital watermarking can also be classified on the basis of type of media in which it is embedded. That is whether it is embedded in text, image, video etc.

- **Public or private watermarking:**
  In public watermarking the users of the content are authorized to check the watermark or to detect it whereas in private watermarking the users of the content are not authorized to check or detect the watermark.

- **According to the keys used:**
  On the basis of the keys used digital watermarking can be categorized as symmetric and asymmetric key watermarking. In asymmetric watermarking different keys are used for embedding and extracting the watermark. In symmetric key watermarking same key is used for embedding and extracting the watermark.

- **According to the visibility of the watermark**
  On the basis of visibility watermark can be divided into visible watermarking and invisible watermarking. In visible watermarking watermark is visible to the user whereas in case of invisible watermarking watermark is not visible to the user.

Fig 3 a. Original Image

Fig3 b. Watermark to be embedded

Fig.3 c. Invisible watermarking

Fig.3d Visible Watermarking

Figure 3(a, b, c, d) shows the difference between visible and invisible watermarking. Figure 3a shows the original image in which watermark is to be embedded. Fig 3b shows the watermark to be embedded into the original image. Figure 3c, 3d shows the original image "Lena.jpg" after embedding the watermark into it. In figure 3c invisible watermarking is done and in figure 3d visible watermarking is done.

4. Techniques used in Digital Image Watermarking

Watermarking techniques can be grouped into the following two categories according to the watermark insertion mode. One is to embed the watermark by directly modifying the pixel values of the original image also referred as spatial domain method. The second is to is to transform the image, then add the watermark into the transform domain, especially the orthogonal transforms, such as the DCT domain, the Wavelet domain, Fourier domain, etc. This method is also referred as transform domain method [7].

4.1. Spatial Domain methods

Spatial-domain technologies refer to those embedding watermarks by directly changing pixel values of host images. Some common spatial-domain algorithms include Least Significant Bit (LSB) Modification, Patchwork, Texture Block Coding, etc. The most serious drawback of spatial-domain technologies is limited robustness [1]. In spatial domain method the watermark is applied to pixel domains. During the watermark embedding process no transforms are applied to the host image or the cover image. Cover image is combined with simple operations in the pixel domain. Detection of the watermark can be done by correlating the pixels values with expected pixel values. The spatial method is less robust to geometric distortion and less resistant to noise and compression. It is faster as the transformation is not required [7]. It is difficult for spatial-domain watermarks to survive under attacks such as lossy compression and low-pass filtering. Also the information can be embedded in spatial domain is very limited [1].

The most straight-forward method of watermark embedding is LSB. Given the extraordinarily high channel capacity of using the entire cover for transmission in this method, a smaller object may be embedded multiple times. Even if most of these are lost due to attacks, a single surviving watermark would be considered a success [1].

LSB substitution however despite its simplicity brings a host of drawbacks. Although it may survive transformations such as cropping, any addition of noise or lossy compression is likely to defeat the watermark. An even better attack would be to simply set the LSB bits of each pixel to one, which fully defeating the watermark with negligible impact on the cover object. Furthermore, once the algorithm is discovered, the embedded watermark could be easily modified by an intermediate party [1].

To use a pseudo-random number generator to determine the pixels to be used for embedding based on a given "seed" or key would be an improvement on basic LSB substitution. As the watermark could no longer be easily viewed by intermediate parties therefore security of the watermark would be improved as the watermark could no longer be easily viewed by intermediate parties. The algorithm however would still be vulnerable to replacing the LSB's with a constant. Even in locations that were not used for watermarking bits, the impact of the substitution on the cover image would be negligible. LSB modification proves to be a simple and fairly powerful tool for stenography, however lacks the basic robustness that watermarking applications require [1].

4.2. Transform domain Method

Transform domain method is divided into two methods: frequency domain method and wavelet domain method. In this first the cover image or the host image is transformed and then the watermark is embedded into its coefficients. In frequency domain the watermark is embedded in DCT, DFT, and FFT domains etc whereas in wavelet transform DWT is used for embedding the watermark. The transform domain method takes advantage of properties of alternate domains to address the limitations of spatial domain methods or to support additional features. To embed a watermark, a frequency transformation is applied to the host data. Then, modifications are made to the transform coefficients.

4.2.1. DCT

The first efficient watermarking scheme was introduced by Koch et al. In their method, the image is first divided into square blocks of size 8x8 for DCT computation [1] as shown in the figure.

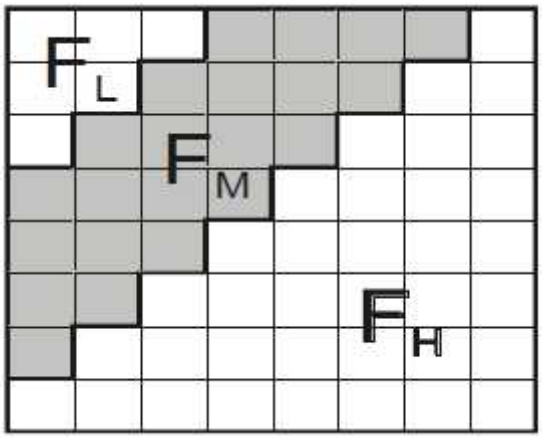

Fig.4. DCT coefficients

In general, the DCT coefficients are divided into three bands (sets), namely low frequencies, middle frequencies and high frequencies. Fig. 4 visualizes these bands. Low frequencies (F_L) are correlated with the illumination conditions and high frequencies (F_H) represent noise and small variations (details). Middle frequencies (F_M) coefficients contain useful information

and construct the basic structure of the image. Middle frequencies F_M is chosen to embed the watermark as the embedding of watermark in a middle frequency band does not scatter the watermark information to most visual important parts of the image i.e. the low frequencies. It does not overexpose them to removal through compression and noise attacks where high frequency components are targeted [8]-[9]. Other reason for inserting watermarks in high frequency band is that they tend to have less influence on the quality of original image, while watermarks in low band will achieve a better robustness (since a large portion of high frequency components may be quantized to zero under JPEG compression). And the mid-bind scheme is right a trade-off between the imperceptibility and robustness [1].

The steps involved in any technique which is based on DCT [10] are as follows:

Step 1 Divide the entire image into 8x8 sized non-overlapping blocks.

Step 2 Take the DCT of each block of size 8x8.

Step 3 Apply a block selection criterion based on the knowledge of Human Visual System (HVS).

Step 4 Use some coefficient selection criteria for embedding.

Step 5 Embed the watermark by modifying the selected coefficients.

Step 6 Take the inverse DCT of each block.

Almost all the algorithms for digital watermarking based on DCT are classified on the basis of step 3 and 4 i.e. the main differentiation between these algorithms is on the basis of block selection criteria or coefficient selection criteria.

4.2.2. Wavelet Transform

The new JPEG2000 standard has adopted a new technique, the wavelet transform. Though this standard has not been widely used yet, any new watermarking algorithm that intends to survive in the future should get along with it. Here come the watermarking schemes based on wavelet transform [1].

In DWT the image is high and low-pass filtered along the rows and the results of each filter are down- sampled by two. Those two sub-signals correspond to the high and low frequency components along the rows and are each of size N by N/2. Each of those sub-signals is then again high and low-pass filtered, but this time along the column data. The results are again down-sampled by two [11].

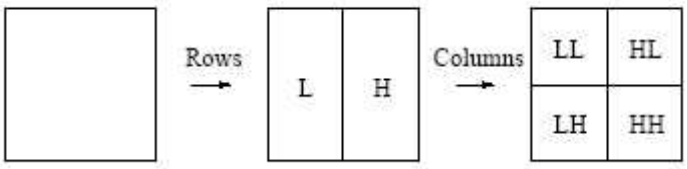

Fig.5. One DWT decomposition step.

The LL sub-band is the result of low-pass filtering both the rows and columns and contains a rough description of the image. Therefore the LL sub-band is also called the approximation sub-band. The HH sub-band was high-pass filtered in both directions and contains the high-frequency components along the diagonals. The HL and LH images are the result of low-pass filtering in one direction and high-pass filtering in the other direction. LH contains mostly the vertical detail information, which corresponds to horizontal edges. HL represents the

horizontal detail information from the vertical edges. All three sub-bands HL, LH and HH are called the detail sub-bands, because they add the high-frequency detail to the approximation image [11].

The decomposition of image is done up to three decomposition steps as shown in the figure 6.

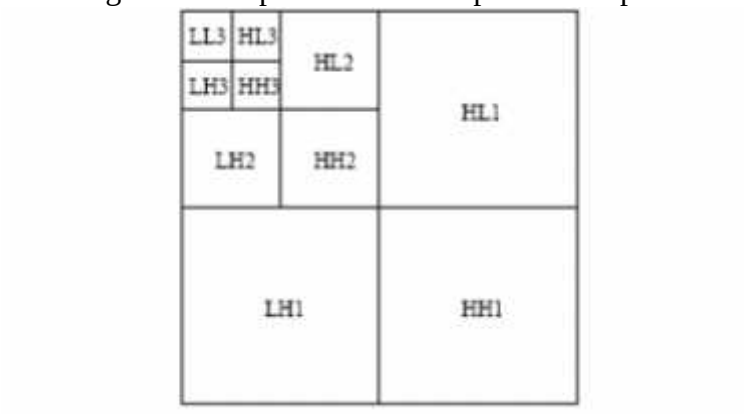

Fig 6. Three step decomposition

One of the many advantages over the wavelet transform is that that it is believed to more accurately model aspects of the HVS as compared to the FFT or DCT. This allows us to use higher energy watermarks in regions that the HVS is known to be less sensitive to, such as the high resolution detail bands {LH, HL, HH) [1].

Other advantages of wavelet transform are [11]:

- Image and video compression standards such as JPEG-2000 and MPEG4 are based on wavelets. Therefore high data compression is possible.
- Wavelet-based watermarking techniques have multi-resolution hierarchical characteristics. Hence an image can be shown at different levels of resolution and can be sequentially processed from low resolution to high resolution. The high frequency sub-bands of the wavelet transform include the edges and textures of the image and the human eye is not generally very sensitive to changes in such bands. This allows the watermark to be added to such sub-bands without being perceived by the human eye.
- High robustness to common signal processing.

5. Attacks on Watermarking

A robust watermark should survive a wide variety of attacks both incidental (means modifications applied with a purpose other than to destroy the watermark) and malicious (attacks designed specifically to remove or weaken the watermark) [12, 13]. Some of the best known attacks are [11]:

Simple attacks: (other possible names include “waveform attacks” and “noise attacks”) are conceptually simple attacks that attempt to impair the embedded watermark by manipulations of the whole watermarked data (host data plus watermark) without an attempt to identify and isolate the watermark. Examples include filtering, compression (JPEG, MPEG), and addition of noise, addition of an offset, cropping, Digital to analog and analog to digital conversion.

Detection-disabling attacks: (other possible names include “synchronization attacks”) are attacks that attempt to break the correlation and to make the recovery of the watermark impossible or infeasible for a watermark detector, mostly by geometric distortion like

zooming, shift in (for video) direction, rotation, cropping, pixel permutations, sub-sampling, removal or insertion of pixels or pixel clusters, or any other geometric transformation of the data.

Ambiguity attacks: (other possible names include "deadlock attacks," "inversion attacks," "fake watermark attacks," and "fake-original attacks") are attacks that attempt to confuse by producing fake original data or fake watermarked data. An example is an inversion attack that attempts to discredit the authority of the watermark by embedding one or several additional watermarks such that it is unclear which the first authoritative watermark was.

Removal attacks: are attacks that attempt to analyze the watermarked data, estimate the watermark or the host data, separate the watermarked data into host data and watermark, and discard only the watermark. Examples are collusion attacks, denoising, certain filter operations, or compression attacks using synthetic modelling of the image (e.g., using texture models or 3-D models). Also included in this group are attacks that are tailored to a specific watermarking scheme.

6. Performance Evaluation of Watermarking Systems:

The common metric used in signal processing industry for the performance evaluation of the watermarking systems is Signal to Noise Ratio (SNR). Suppose the original image is I(m,n), the output image is D(m,n), then generally SNR is defined as[1]:

$$SNR = 10\log_{10}[\sum_{m}\sum_{n} I(i,j)^2 / \sum_{m}\sum_{n} (I(i,j) - D(i,j))^2]$$

When SNR approaches infinity, the original image and output image are totally the same.

Another similar one is Peak SNR (PSNR). For images with 255 gray levels, the PSNR is defined as [1]:

$$PSNR = 10\log_{10}\left[\sum_{m}\sum_{n} (255)^2 / \sum_{m}\sum_{n} (I(i,j) - D(i,j))^2\right]$$

The similarity of extracted watermark W1 and original watermark W is computed by the following formula [1]:

$$SM = \sum_{m}\sum_{n} W(i,j) * W1(i,j) / \sqrt{\sum_{m}\sum_{n} W(i,j)^2 * \sum_{m}\sum_{n} W1(i,j)^2}$$

If the result is larger than some determined threshold, we consider W1 = W [1].

7. Conclusion

With the growth in digitization of data threat to the owner's copyright has also increased. Any one over the internet can copy the data without much loss of the information. To authenticate the owner and to provide copyright protection to the owner digital watermarking was introduced. The process of watermarking is divided into three parts i.e. watermark embedding, transmission of the watermarked data and watermark extraction. In the process of digital watermarking a code called watermark is embedded to the data, then the data is transmitted and then watermark extraction process is done to check whether any kind of tampering of data was done during the transmission. Depending upon the need watermarking

is classified into various types like visible and invisible watermarking, fragile and robust watermarking etc.

Various techniques that can be applied to embed the watermark into the image are broadly classified into two categories i.e. pixel or spatial domain and transform domain. The most commonly used technique of pixel domain method is LSB. Transform domain is further divided into frequency domain and wavelet domain. In frequency domain we have DCT, FFT and DFT. DCT is most commonly used now a days. DWT is a more robust technique than frequency domain techniques. It has many advantages over the DCT techniques. It is compatible with JPEG2000 standard as in JPEG2000 standard DWT is used.

There are various parameters to check the performance of the watermarking system. These parameters are SNR, PSNR. Hence digital watermarking is a process which provides owner copyright protection and also to protect data owners against unauthorized copying and redistribution of the material put on the network.

8. References

[1] Lin Liu, "A Survey on Digital Watermarking Techniques"
[2] J. Zhang, Anthony T. S. Ho, "An Efficient Digital Image-in-Image Watermarking Algorithm Using the Integer Discrete Cosine Transform (IntDCT)", ICICS-PCM 2003,15-18 Dec 2003 Singapore
[3] Mauro Barni, Franco Bartolini, Vito Cappellini, Alessandro Piva, "A DCT-domain system for robust image watermarking", 1998 Elsevier Science B.V., Signal Processing 66 (1998) 357-372
[4] Chunlin Song, Sud Sudirman, Madjid Merabti, "A robust region-adaptive dual image watermarking technique", J. Vis. Commun. Image R. 23 (2012) 549–568
[5] Robust Digital Image Watermarking Technique Based on Histogram Analysis, Hamza A. Ali, Sama'a A. K. Khamis, *World of Computer Science and Information Technology Journal (WCSIT), Vol. 2, No. 5, 163-168, 2012*
[6] Chunlin Song, Sud Sudirman, Madjid Merabti, "Recent Advances in the Classification of Watermarking Techniques in Digital Images", PGNet, 2009
[7] Y.J.Song T.N.Tan, "Comparison of Four Different Digital Watermarking Techniques+", IEEE 2000, Proceedings of ICSP2000

[8] Liwei Chen, Mingfu Li, "An Effective Blind Watermark Algorithm Based on the DCT", IEEE, Proceedings of the 7th World Congress Intelligent Control and Automation, June 2008, Chongquing, China.

[9] A. Hanaa , M. hadhoud, and A. Shaalan, "A Blind Spread Spectrum Wavelet Based Image Watermarking Algorithm" International Conference on Computer Engineering & Systems, pp. 251-256, 2009.

[10] Mehmet Utku Celik, Gaurav Sharma, Ahmet Murat Tekalp. " Lossless Generalized – LSB Data Embedding", IEEE Transactions on Image Processing, vol. 14, No.2, February 2005.

[11] Harpuneet Kaur, Mr. R. S. Salaria, "Robust Image Watermarking Technique to Increase Security and Capacity of Watermark Data", A thesis submitted in partial fulfilment of the requirements for the award of degree of Master of Engineering in Software Engineering, May 2006

[12] A. Nikolaidis, S. Tsekeridou, A. Tefas, V Solachidis, "A SURVEY ON WATERMARKING APPLICATION SCENARIOS AND RELATED ATTACKS", IEEE international Conference on Image Processing, Vol. 3, pp. 991 – 993, Oct. 2001.

[13] Frank Hartung, Martin Kutter, "Multimedia Watermarking Techniques", Proceedings of The IEEE, Vol. 87, No. 7, pp. 1085 – 1103, July 1999.