Chapter 2

Theoretical Background

In this chapter, we first discuss the fundamental theories, working principles and techniques used in the existing digital image watermarking schemes. At the latter stage of this chapter an extensive review of the literature in watermarking of digital images for copyright protection, authentication and restoration is presented along with their merits and demerits. The different types of digital watermark are discussed in section 2.1. Section 2.2 discusses the important applications of watermarking, and performance evaluation metrics are explained in section 2.3. Section 2.4 categorizes the attacks on watermarking system while issues and challenges are described in section 2.5. In section 2.6, various watermarking methodologies are explained. Section 2.7 provides a detailed survey of image watermarking schemes and future directions are presented in section 2.8. Finally, conclusions of this paper are summed up in section 2.9.

2.1 Classifications of Watermarking

Digital watermarking can be classified on the basis of different criteria.

2.1.1 Attached Media

On the basis of attached media, digital watermarking can be divided into image watermarking, text watermarking, audio watermarking, video watermarking and graphic watermarking. Video and audio watermarking refers to embedding digital watermark in the video or audio stream to control video or audio applications. Image watermarking is used to embed watermark in a still image. Text watermarking refers to embedding watermark in DOC, PDF and other text file to prevent changes of text whereas in graphic watermarking the watermark is embedded into two-dimensional or three-dimensional computer-generated graphics to prevent changes [5].

2.1.2 Visibility or Perceptibility

On the basis of visibility (or perceptibility) watermarking is categorized into visible and invisible watermarking schemes. In case of visible (or perceptible) watermarking, the watermark is embedded in a host image in such a way that the watermark is noticeable to a human observer while in the case of invisible (or imperceptible) watermarking the embedded data is not detectable, but can possibly be extracted by some software or a computer program [6].

2.1.3 Resist Attack

On the basis of resist attack, watermarking is categorized into robust, fragile and semi–fragile watermarking. A robust watermark should be able to resist intentional or unintentional manipulations. Some important applications for robust watermarking are fingerprinting, data mining, copyright protection and ownership verification [6, 7, 8]. A fragile watermark is intended to be destroyed even after the minor unintentional or intentional manipulation in the watermarked image [8, 9, 10]. The third category, semi-fragile watermarking uses watermarks that have the ability to resist unintentional manipulations caused by common image processing operations like JPEG compression and is fragile against intentional, malicious manipulations [11, 12, 13, 14, 15]. The main applications field of fragile and semi-fragile watermarking are image and video content authentication [8].

2.1.4 Watermark Embedding Method

On the basis of embedding method, watermarking is categorized into spatial and transform domain techniques. Transform domain techniques perform the watermarking by changing the coefficients in the frequency domain of host image. In the spatial domain techniques, watermark is directly applied on pixel value of the host image. Spatial domain algorithms are easier to implement. The transform domain techniques are relatively more reliable and robust to various attacks [16, 17, 18]. However, many researches demonstrated that watermarking in the transform domain is not robust to geometrical attacks like cropping, rotation, scaling and translation. Discrete Cosine Transform (DCT) [19, 20, 21, 22] and Discrete Wavelet Transform (DWT) [23, 24, 25] are commonly used frequency domain techniques. A brief summary on the basis of watermark embedding method is listed in Table 2.1.

TABLE 2.1: Spatial Domain vs. Transform Domain watermarking Schemes.

Seq. No.	Characteristic	Spatial Domain	Transform Domain
1.	Robustness	Fragile in nature	More robust
2.	Capacity	High	Low
3.	Perceptual quality control	High	Low
4.	Computational complexity	Low	High
5.	Computational time	Less	More

2.1.5 Requirements for Watermark Extraction or Detection

On the basis of requirements for watermark detection or extraction, watermarking is categorized into blind, semi-blind and non-blind schemes. The non-blind watermarking schemes (also called private watermarking scheme) require the original host image and secret key(s) to identify the watermark. Semi-blind watermarking schemes demand the presence of both the secret key(s) and the watermark bits sequence. On the other hand, the blind (or public) watermarking schemes require only the secret key(s) for extraction [26, 27].

2.2 Application of watermarking

Digital watermarking schemes are application dependent. The following applications of watermarking are very common:

2.2.1 Copyright Protection

This is the most well-known application of watermarking. With the advancement of editing software, illegal operations such as duplication, manipulation have become easy and they are difficult to prevent. So, copyright protection becomes a very important issue. The basic idea is to embed information about the copyright proprietor into the data to prevent parties from claiming to be the rightful owners of the data [28]. The watermarks used for this purpose are very robust against various attacks intended to remove the watermark.

2.2.2 Content Authentication

Another important application of watermarking is content authentication and tamper detection. Embedding watermark in original content that can be later checked to verify it has been tampered or not. For this application, digital watermarks which are fragile in nature can be used.

2.2.3 Broadcast Monitoring

Broadcast monitoring is used to verify the programs broadcasted on radio or television. It particularly helps the advertising companies to realize wheather their advertisements appeared for the right duration or not. Watermarking methods can be used in consolidation with the active approach by adding the watermark (i.e. identification information) inside the host content, so that it is not lost during the transmission even if its format changes. Thus, the verification of the usage of the broadcasted material will be more reliable [3].

2.2.4 Transaction Tracking

Watermarks can be used to identify people who obtain content legally but redistribute it illegally. The owner can use a fingerprinting technique for tracing the source of the illegal copies. In this consequence, the owner can embed a unique watermark in the copies of the data that are supplied to each purchaser. At the time of verification, using fingerprinting techniques, the embedding identity can be compared with the purchaser's identity. So we can identify purchasers who have broken their license agreement by supplying the data to third parties. Using this method, the source of illegal copying or distribution can be tracked [29].

2.2.5 Copy Control

Watermarking can be used as a strong tool to prevent illegal copying. In this application, the watermark represents a copy-prohibit and watermark detectors in the recorder determine whether the data provided to the recorder may be stored or not.

2.3 Performance Evaluation Metrics

By definition it is clear that the watermarking techniques are based on imperceptible distortion into host signals. Perceptual quality is a measure of imperceptibility, obtained by determining both the amount of distortion introduced into a host signal by a watermarking algorithm, and how detectable the distortion is.

2.3.1 Mean Square Error (MSE)

Mean Square Error (MSE) is defined as average squared difference between a distorted image and a reference image. It is calculated by the Eq.(2.1):

$$MSE = \frac{1}{M \times N} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \Big(f(m,m) - \hat{f}(m,n) \Big)^2 \tag{2.1}$$

Where, MSE is Mean Square Error which is for $M \times N$ two monochrome images f and $\hat{f}$ in which one is the original host image and another one is the noisy approximation of the first one.

2.3.2 Peak-Signal to-Noise Ratio (PSNR)

Peak-Signal to-Noise Ratio (PSNR) is the ratio between the maximum possible value of an image and the power of corrupting noise that affects the fidelity of its illustration. PSNR is usually expressed in terms of the logarithmic decibel scale.

$$PSNR = 10log_{10}\frac{(max)^2}{MSE} \tag{2.2}$$

Where, max is the maximum pixel value of the image and MSE is the mean square error, given in Eq.(2.1).

2.3.3 Normalized Correlation Coefficient(NCC)

The similarity of the two monochrome images f and $\hat{f}$ can be evaluated by normalized correlation coefficient defined by equation (2.3).

$$Corr = \frac{\sum_{m=0}^{M-1}\sum_{n=0}^{N-1}\left((f(m,n)-f_1)\left(\hat{f}(m,n)-f_2\right)\right)}{\sqrt{\left(\sum_{m=0}^{M-1}\sum_{n=0}^{N-1}\left(f(m,n)-f_1\right)^2\right)\left(\sum_{m=0}^{M-1}\sum_{n=0}^{N-1}\left(\hat{f}(m,n)-f_2\right)^2\right)}} \tag{2.3}$$

where f_1 and f_2 are the mean value of images f and $\hat{f}$. The correlation coefficient (Corr) has the value range from -1 to 1. $Corr = 1$, if two images are absolutely identical, $Corr = 0$ if they are completely uncorrelated, and $Corr = -1$ if they are completely anti-correlated.

2.3.4 Bit Error Rate (BER)

BER is a method of determining the similarity of two given signals. The BER is the ratio of the total number of bits error to the total number of bits embedded and is

calculated by:

$$r_{ber} = \frac{\left(\sum_{n=0}^{N-1} \tilde{W}_n \oplus W_n\right)}{N} \tag{2.4}$$

Where $\tilde{W}_n$ and W_n are the bits at the n^{th} position of the decoded watermark and the original watermark respectively. N is the total number of bits in watermark W.

2.4 Attacks on Watermarking Schemes

Attacks on watermarking schemes aim to remove or destroy any watermark signals in the cover data. In other words, we can say attacks on watermarking schemes aim to prevent the use of a watermark while preserving the visual quality of the media [30, 31, 32]. Therefore, it is important to understand how these attacks work in order to design a better and more robust watermarking technique. Table 2.2 illustrates the attacker's goal against digital watermarking for different applications. Attacks on watermarking schemes can be classified into four distinct categories namely removal attacks, geometrical attacks, protocol attacks and cryptographic attacks.

TABLE 2.2: Attacker's goal against watermark for different applications.

Factors	Fragile Watermarking	Robust Watermarking
Applications	Multimedia Authentication	Copy Control, Copyright Protection, Evidence of Ownership, Fingerprinting
Requirements	(1) Determine if the Work has been changed. (2) It is difficult for an unauthorized person to insert a valid watermark.	The watermark can still be detected even after severe processing.
Attacker's goal	(1) Make the watermark still valid after alteration of work. (2) Generate a valid work for new data.	Make the detector unable to detect the watermark while keeping the perceptual quality.

2.4.1 Removal Attacks

Removal attacks aim at the removal of a watermark from the watermarked image without attempting to break the security of the watermarking techniques [31]. This category includes quantization, de-noising, histogram equalization, lossy compression, collusion, remodulation, blur, sharpen and averaging attacks. In these types of attacks attackers does not attempt to know how the watermark has been embedded. Hence, there is no simple post processing can recover the watermark from the attacked watermarked image.

2.4.2 Geometrical Attacks

Geometrical attacks aim at not removing the embedded watermark itself (as like removal attacks), but to distort it through spatial or temporal alterations of the watermarked image. So, Geometrical attacks aim is increasing the difficulty to detect watermark [32]. This category includes rotation, scaling, translation, cropping and skewing attacks.

2.4.3 Protocol Attacks

Protocol attacks aim to interfere with its intended application rather than remove or distort the watermark [30]. For example by re-watermarking an image with a second owner, and so confusing the medias original ownership. This can create a situation of ambiguity with respect to the real ownership of the work. Another example of protocol attack is the copy attack. The copy attack estimates a watermark from watermarked image without having any specific knowledge about the watermarking technology and copies it to some other image called the target image.

2.4.4 Cryptographic Attacks

Cryptographic attacks are very similar to the attacks used in cryptography. These attacks aim to crack the security methods in watermarking schemes. Thus this

attack finds a way to remove the embedded watermark or to embed misleading watermarks. Example of this categories are brute-force attack, oracle attack etc. The aim of the brute-force attack is to find secret information or to crack the watermark security through an exhaustive search. The aim of oracle attack is to create a non-watermarked image when a watermark detector device is available.

2.5 Issues and Challenges

The important issues that occur in the study of digital watermarking techniques are:

- **Robustness:** Robustness is the watermarking technique's tolerance to common image processing methods i.e. how to embed and retrieve data such that it would survive malicious or accidental attempts at removal.
- **Transparency:** Embedded watermark should be invisible to the user.
- **Trustworthy detection:** Watermark detection result is able to supply a highly reliable decision as to the presence of assured watermark information.
- **Capacity:** Capacity means the maximum amount of data that can be embedded into the image to guarantee appropriate retrieval of the watermark during extraction.
- **Computational efficiency:** Computational efficiency is the efficiency of the implementation of the watermarking techniques. That is, the watermarking procedure must be implemented in a prompt manner for its utility in the real world.

The main challenge of the watermarking schemes is to achieve a better trade-off among robustness, capacity and imperceptibility. Robustness can be achieved by increasing the capacity or strength of the embedded watermark, but the imperceptibility would be decreased as well. In transform domain, to attain imperceptibility, the watermark should be embedded into the high frequency components of the cover signal. On the other hand, for robustness watermark can be embedded into the low frequency components only.

Watermarking schemes are also specified for fixed applications. So another practical challenge is the proper choice of watermarking techniques. Our objective is to draw attention to these issues and the trade-off involved in them.

2.6 Watermarking Methodology

The various watermarking techniques can be divided into two parts spatial domain and transform domain as shown in Figure 2.1.

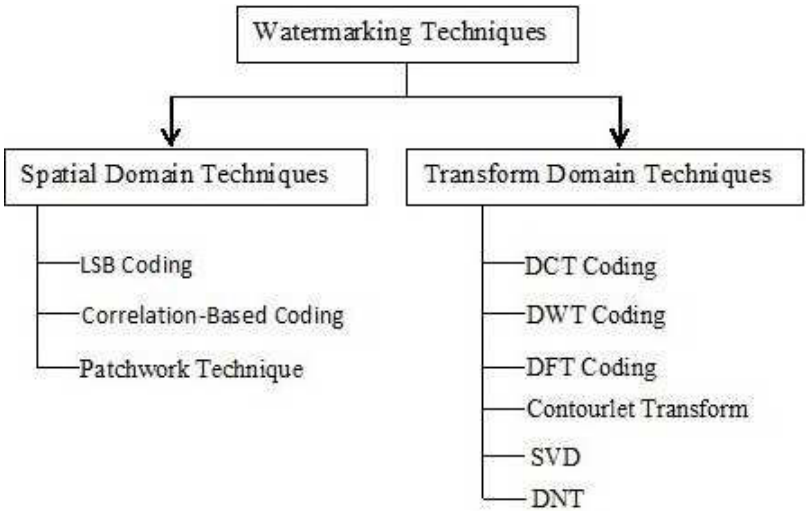

FIGURE 2.1: Different watermark techniques.

2.6.1 Spatial Domain Techniques

Spatial domain techniques are based on direct modification of pixels in an image. The drawback of spatial domain technologies is limited robustness. It is difficult for spatial-domain watermarks to survive under attacks such as lossy compression and low-pass filtering. Some spatial domain techniques are discussed below:

2.6.1.1 Least Significant Bits (LSBs) based technique

In this technique the LSBs of the host image are substituted with the watermark information. The watermark bits are embedded in a sequence that is decided by the one or some seed values known as watermark key(s). In order to retrieve it back this

watermark key(s) must be known. The watermark embedder first selects a subset of pixels on which the watermark has to be embedded. It then embeds the watermark in the LSBs of the pixels from this subset. LSB modification proves to be a simple and fairly powerful tool for watermarking and steganography but these techniques are vulnerable to any types of attacks and the watermark can be easily destroyed [33]. This approach is also very sensitive to noise and common signal processing.

2.6.1.2 Correlation based technique

In this method a pseudo random noise with a pattern W is embedded to an image I. At the receiver side the correlation between the random noise with the image is found out and if the value exceeds a certain threshold value the watermark is detected else it is not. This technique is also known as additive spread spectrum.

$$I_w(i,j) = I(i,j) + \alpha * W(i,j), i = 1,2,...,M; j = 1,2,...,N \tag{2.5}$$

where α represents the gain factor, I_w represents watermarked image, (i, j) represents spatial position and I represents host image.
Here, if we increase the gain factor then it increases the robustness of watermark but the perceptual quality of the watermarked image will decrease.

2.6.1.3 Patchwork Technique

Patchwork developed by Bender et al. [34], is a data hiding statistical approach based on a pseudo random, statistical approach. It works by imperceptibly embedding a particular statistic, with a Gaussian distribution, into the cover image. In this technique, two sets of pixels or patches, P_a and P_b of the cover image are pseudo randomly chosen.

$$d = a - b \tag{2.6}$$

Where a and b are brightness at patch P_a and P_b.

$$E[d] = 0 \tag{2.7}$$

Here Eq (2.7) indicate that average value of d after repeating this procedure a large number of times ($\sim$ 10000) is expected to be zero. Then the algorithm works by slightly brightening points in P_a, while darkening of the same factor (say k) those in P_b. So, the expected value is going to be shifted to some right (that is depending on factor k and repetition times) with respect to the average one of zero. This unique statistic indicates the presence or absence of a watermark. Patchwork is independent of the contents of the cover image. The important characteristics of patchwork are its resistance to cropping and to gamma corrections. Patchwork is destroyed by any affine transformation, like scaling, rotation or translation.

2.6.2 Transform Domain Techniques

Transform (i.e. Frequency) domain based techniques are very robust against attacks involving image compression and filtering because the watermark is actually spread throughout the image, not just operating on an individual pixel. Image transform can be applied either on whole image or to block by block manner, depending on application. Techniques for attaining transform domain watermarking would modify the selected coefficients in the transformed domain. Generally multiplicative spread spectrum is used, that is given in Eq(2.8).

$$I_w(i,j) = I(i,j) + \alpha * I(i,j) * W(i,j), i = 1,2,...,M; j = 1,2,...,N \tag{2.8}$$

Where I_w and I denotes the watermarked and original host image, W denotes the watermark and α is the watermark strength factor.

2.6.2.1 Discrete Cosine Transformation (DCT)

DCT is often used in signal processing and image, especially for lossy data compression, because it has a strong energy compaction property [35, 36]. The DCT has distinct property that most of the significant information of the image is concentrated in just a few low frequency coefficients of the DCT. It is referred as energy compaction property [37]. The energy compaction property of DCT is utilized in the JPEG compression technique to separate and remove insignificant high frequency

components in images. The two-dimensional DCT forward transform formula is presented as follow:

$$F(u,v) = c(u)c(v)\frac{2}{N}\sum_{x=0}^{N-1}\sum_{y=0}^{N-1} f(x,y)cos\left(\frac{2x+1}{2N}u\pi\right)cos\left(\frac{2y+1}{2N}v\pi\right) \tag{2.9}$$

where, x,y,u,v =0,1,2,. . . , N-1, and

$$c(u) = c(v) = \begin{cases} \frac{1}{\sqrt{2}} & u = 0, v = 0; \\ 1 & otherwise. \end{cases}$$

The inverse transform (IDCT) formula is described as follow.

$$f(x,y) = \frac{2}{N}\sum_{u=0}^{N-1}\sum_{v=0}^{N-1} c(u)c(v)F(u,v)cos\left(\frac{2x+1}{2N}u\pi\right)cos\left(\frac{2y+1}{2N}v\pi\right) \tag{2.10}$$

where $f(x, y)$ is the gray value of a pixel and $F(u, v)$ is the DCT coefficient. The top left corner coefficient of the frequency domain matrix represent the DC value, and the remaining coefficients represent the AC values of the image. In compared to spatial domain watermarking techniques, DCT based watermarking techniques are more robust. DCT is a linear transform, which maps an M-dimensional vector to a set of M coefficients.

2.6.2.2 Discrete Wavelet Transform(DWT)

The DWT transforms are based on wavelet. Wavelet is the small waves of varying frequency and limited duration [38, 39]. The wavelet transform based watermarking techniques decompose the image into four sub bands a low resolution approximation (LL) of the tile component and the three spatial directions components i.e. horizontal (HL), vertical (LH) and diagonal (HH) frequency characteristics. Wavelet techniques provide excellent space and frequency energy compaction. This energy tends to a cluster spatially in all sub-bands. At every level of decomposition the magnitude of the DWT coefficients is larger in the approximation sub band (LL), and smaller for other high resolution sub bands (HL, LH and HH). The high resolution sub bands help in locating the edge and texture patterns of any image.

Similar to the human visual system (HVS), DWT-based watermarking techniques enable good spatial localization and have multi-resolution characteristics. DWT provides higher compression ratio than DCT which is relevant to human perception. However, the disadvantage is that it is take longer compression time and more computing cost.

2.6.2.3 Discrete Fourier Transform(DFT)

The DFT and inverse DFT are the primary numerical transforms relating time and frequency in digital signal processing. The forward DFT of an image I of size $M \times N$ and the corresponding Inverse- DFT (IDFT) are defined as follows [40]:

$$F(u,v) = \frac{1}{MN} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x,y) exp\left(-j2\pi \left(ux/M + vy/N\right)\right) \tag{2.11}$$

The Fourier magnitude spectrum and phase angle are defined as follows:

$$f(x,y) = \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} F(u,v) exp\left(j2\pi \left(ux/M + vy/N\right)\right) \tag{2.12}$$

$$|F(u,v)| = \sqrt{R^2(u,v) + I^2(u,v)} \tag{2.13}$$

$$\phi(u,v) = tan^{-1}\left[\frac{I(u,v)}{R(u,v)}\right] \tag{2.14}$$

where $R(u,v)$ and $I(u,v)$ are the real and imaginary parts of $F(u,v)$, respectively. It is understood that the phase information is considerably more important than the amplitude information in preserving the visual intelligibility of the image. DFT is scaling, rotation and translation invariant. Therefore, it can be used to recover from geometric distortions, whereas the spatial domain, DCT and the DWT are not rotation, scaling and translation (RST) invariant and hence it is difficult to overcome from geometric distortions. DFT uses complex numbers, while DCT uses just real numbers. So in DFT cost of computing may be higher.

2.6.2.4 Contourlet Transform(CT)

The contourlet transform, proposed by Do et al. [41], provides an efficient representation for two-dimensional signals with smooth contours and in this case outperforms the wavelet transform which fails to recognize the smoothness of the contour. With the comparison of wavelet, contourlets offer a much richer set of directions and shapes [42]. Hence, contourlets are more effective than wavelets in capturing smooth counters and geometric structures in images. The contourlet transform also has multi scale and time-frequency localization features of the wavelet transform.

2.6.2.5 Singular Value Decomposition (SVD)

SVD is a numerical analysis tool used to diagonalize matrices. SVD has been developed for a variety of applications. The main properties of SVD in terms of image processing applications are: Singular values (SVs) of the image have very good stability to know when a small perturbation is made in the image of the singular value does not change significantly; Singular value is an algebraic intrinsic property [43, 44, 45]. The singular value decomposition processing in a matrix A can be decomposed into three matrices of the same size as the initial matrix; two orthogonal matrices U and V^T and a diagonal matrix S.

$$A = U * S * V^T \tag{2.15}$$

The columns of U and V^T are called left and right singular vectors of A respectively. They essentially determine the detailed geometry of the original image. The diagonal values of the matrix S are ranked in descending order i.e. $\sigma_1 \geq \sigma_2 \geq \sigma_3 ... \geq \sigma_N \geq 0$ [43].

2.6.2.6 Divisive Normalization Transformation(DNT)

Divisive Normalization Transform (DNT) is an invertible nonlinear transformation technique that provide a better perceptual sensitivity of the human visual system [45, 46, 47, 48]. Divisive Normalization Transform has been used in watermarking because its coefficients are statistically independent and homogeneous in comparison

to the discrete wavelet transforms. It also provides superior perceptual quality of watermarked image. Divisive Normalization Transform is constructed upon linear image decomposition, followed by a divisive normalization stage in which each linear transform coefficient is divided by a normalization factor. The linear transformations may be wavelet-type transforms or DCT.

2.7 Comprehensive review of some significant watermarking schemes

Digital watermarking has a received a great interest in the research and many watermarking methods have been proposed in the literature. In this section we have presented a comprehensive review of significant digital watermarking techniques for the application of image authentication, restoration (or recovery), and copyright protection.

2.7.1 Watermarking techniques for image authentication

To protect the authenticity of digital images, several approaches have been proposed. These approaches include fragile watermarking, semi-fragile watermarking, conventional cryptography digital signatures that are based on the image content. Methods are classified according to the service they provide, that is strict or selective authentication, localization, tamper detection, reconstruction proficiencies and robustness against different desired image processing operations. A good authentication system should have the following necessary features [49]:

- It should be capable to determine whether an image has been altered or not;
- It should be accomplished to locate any modification made on the image;
- The embedded authentication data should be invisible under normal viewing conditions.

2.7.1.1 Fragile watermarking schemes

Lim et al. proposed a pixel-wise fragile watermarking scheme for image authentication in [50]. In this technique seven most significant bits (MSBs) of a pixel are given as an input to the hash function. A single bit value either 0 or 1 for each pixel is calculated using a secret key and hash value. This calculated value was embedded in to the LSB of corresponding pixel. If any modification is found in image, the receiver end calculated hash value and extracted hash value from LSB will mismatch and hence tampered pixel can be identified. Similarly, a pixel-wise fragile watermarking scheme for image authentication was proposed by Shivani et al. [51]. Here three bits called as Associative, Relational and Authentication bits were generated by three different methods for each pixel. After that, Singh et al. [10] proposed a fragile watermarking for image authentication. In this approach the three bits watermark were generated from the five MSBs of each pixel using three different algorithms and embedded into the three LSBs of the corresponding pixel in the host image. In [52], Zhang et al. proposed a fragile watermarking scheme that was capable of accurately localizing the pixels. In this scheme, a folded version of the authentication data derived from five MSBs of each pixel and some additional test data used for estimating the tampering strength are embedded into the three LSBs of each pixel. In [53] a discrete wavelet-based fragile watermarking scheme for secure image authentication was given. In this proposed algorithm, the embedded watermark was generated by the DWT, and embedded into the LSB of the host image after the enriching security watermark was scrambled by chaotic systems. The fragile watermarking schemes proposed in [10, 50, 51, 52, 53] are very sensitive to any alteration, so they do not resist any types of attack. In [53] accuracy of tampered localization is also reduces due to using large non-overlapping blocks.

2.7.1.2 Semi-Fragile watermarking schemes

Hu and Han [54] proposed a DWT based semi-fragile watermarking technique for image authentication. In this scheme, the image features were extracted from the low frequency domain(LL3 component) using a de-noising algorithm and MD5 hash function to generate two watermarks. Here, one copy of watermark was used for

classifying the intentional content alteration and the other one for indicating the altered location. Further Wu et al.[55] proposed an Integer Wavelet Transform (IWT) with parameters based semi-fragile watermarking technique for image authentication. This scheme can tolerate lossy JPEG compression to a quality as low as 40%, and locates the tampered area accurately. Another DWT based semi-fragile watermarking scheme was proposed by Jin et al. [56] where adaptive quantization is used for better image authentication and tamper detection. In this scheme, watermark was generated from the low frequency sub bands (LL3 component) of three-level wavelet decomposition. This scheme is capable to resist common image processing (such as JPEG compression, noise-adding, filtering), but is sensitive to malicious changes of the image. Hu et al. [57] proposed a tamper detection scheme for block truncation coding (BTC). In this scheme, one bit authentication code of each image block was generated from the quantization levels. Several copies of the authentication data were embedded into the bit maps of compressed image blocks based on the block permutations. These block permutations were generated by using the random sequences induced by the selected random number seeds. Focus is needed on optimum number of the copy of the authentication code that should be set, otherwise image qualities of the embedded images decrease when the copy of the authentication code increases.

2.7.1.3 Reversible watermarking schemes

Lo and Hu proposed a block-based reversible image authentication scheme for digital image in [58]. The authentication codes were embedded into the residual values block by block by using the prediction-based histogram shifting process. Similarly, reversible watermarking schemes are proposed in [59, 60, 61]. Due to data reversibility properties these scheme can be employed to protect the image integrity of the general-purposed images as well as the special-purposed images like medical images, remote sensing images and military images.

2.7.2 Watermarking techniques for image authentication and restoration

After successful tampered image detection and localization, tampered image restoration is extremely essential especially when it is utilized in evidence of court and medical imaging. The main concept of the restoration schemes is that the compacted form of the principal content should be hidden in another region of the image itself. This sub-section provides a review of significant watermarking schemes for image authentication and restoration.

2.7.2.1 DCT based watermarking schemes

He et al. [62] proposed a DCT based self-embedding fragile watermark scheme of 8×8 block size. This method only works under the circumstances that regions storing the original information of tampered areas must be unchanged. So, this scheme suffers from reconstruction dependency(i.e.tampering coincidence) problem. The schemes proposed in [63, 64, 65, 66, 67] are proficient to resolve the reconstruction dependency problem. Qian et al. [64] proposed a self-embedding watermarking technique based on DCT to reduce the embedding data for self-recovery. The main operation is to encode different types of blocks with varied sizes and use the inpainting technology to recover the blocks with few details. The main defect in [64] is that a complete image cannot be reconstructed with the extracted bits if the tampering rate is greater than 35% of the total size of the image. Randall et al. [68] proposed a DCT based semi-fragile watermarking scheme and embeds a small size of payload data for restoration in the frequency domain. Then an iterative method was implemented for restoring the altered regions in the image. The data embedded was very restricted to keep the better image quality but making the scheme insecure. Lin and Lin [69] proposed a semi-fragile watermarking scheme with self-restoration ability based on sharing and lattice embedding techniques. In this scheme, the recovery data were shared among many shadows, then lattice-embedding was utilized to embed each shadow in the DCT domain of an 8×8 block. This watermarking scheme was shown to be moderately robust to image processing operations including noise and brightness adjustment, JPEG compression and resist certain security attacks such as a collage attack, a vector quantization attack and a cut-and-paste attack. The

drawback of this scheme is that the percentage of the tampered region is limited to around 16.66%. Wang et al. [70] proposed a DCT based semi-fragile self-restoration watermarking scheme. In this scheme, the mean value of each 4×4 block is used as the restoration watermark. After a tampered 8×8 block is identified, its four lowest DCT coefficients are restored from linear combinations of mean values of its four 4×4 sub-blocks, where the weights of linear combinations are obtained via linear regression. This technique is shown to be moderately robust to JPEG compression and restoration capable of the tampered region can go up to 50%. A watermarking scheme with flexible self-recovery quality was proposed by Zhang et al. in [65]. In this scheme, the embedded watermark data for content recovery are computed from the DCT coefficients of the cover image. When a part of a watermarked image is altered, the watermark bits in the area without any modification can be extracted. The recovered image quality by this scheme depends on the tempered area. If the smaller the tampered area, the more the amount of available watermark data will be, leading to a better quality of recovered content. This scheme is capable to restore up to tampering rate 60%. Korus and Dziech in [63] proposed a self-recovery watermarking scheme based on an erasure communication channel. This scheme is capable to recover even when 50% of the image area becomes tampered. Zhang et al. [71] proposed a self-embedding fragile watermarking scheme which was based on DCT and fractal compression coding. In this scheme, an interleaved and overlapped block structure was used to improve the accuracy of localization. This scheme is capable to restore tampered image up to tampering rate 80% but drawback of this scheme is, it is fragile in nature and not resist any common signal processing operations.

2.7.2.2 DWT based watermarking schemes

Chamlawi et al. [72] proposed a DWT based watermarking scheme for image authentication and restoration of the tampered image. In this scheme binary watermark image was embedded for authentication and image digests as a compressed version of the cover image itself for restoration. The scheme can authenticate and recover images at the cost of low security and imperceptibility. This scheme treats the lossy JPEG compression as a malicious alteration to the quality factor above 70. In [28] Preda proposed a DWT based self-recovery watermarking scheme. The LL wavelet

sub-band of the second wavelet decomposition of the original image is used as a recovery watermark and is embedded in the LH, HL and HH sub-bands of the first wavelet decomposition. This scheme is able to detect a good estimate of the original image, even if the watermarked image has severely been tampered. This scheme is need to be used some error correction codes to protect the recovery watermark better against quantization and compression attacks. Phadikar et al. [73] proposed a QIM dither modulation in the IWT domain based semi-fragile watermarking scheme for image authentication and recovery. In this scheme, a predefined binary pattern was used for authentication and halftone version of the lowest sub-band generated by second level of IWT decomposition was used for recovery. Digital halftone processing is a technique to convert gray scale images into two-tone binary images. This algorithm shows watermark robustness against several content preserving modifications using an adequate threshold value; however this threshold value causes high false negative error rates. A low quality of the recovered image is another disadvantage of this scheme because the extracted recovery watermark must be scaled four times.

2.7.2.3 DWT-DCT based watermarking schemes

In [74] Chamlawi et al. proposed a DWT-DCT based watermarking scheme, where the recovery watermark is generated using the DWT and the DCT from the original image. The main drawback of this self- recovery scheme is that it is not able to resist large intentional content modifications. In [75], the coefficients of the LL sub-band of the second integer wavelet decomposition are used as a recovery watermark compressed by DCT. These coefficients are embedded in the low frequency Wavelet sub-bands using a LSB approach, where the last 3 bits of these coefficients are used as a watermark. Because of the low resilience of the LSB approach, this technique is also not very resilient against large content modifications. Rosales-Roldan et al. [76] proposed two watermarking algorithms for image content authentication with localization and restoration capability of the tampered regions. The first algorithm watermark was embedded in Integer Wavelet Transformation (IWT) domain while in second algorithm DCT domain. In both algorithms, watermark was generated using error diffusion half-toning technique. Here, performance of both algorithms is almost same but differ only in the embedding domain. Both algorithms provide

the same tamper detection and recovery capability if the threshold is properly selected. The proposed scheme, up to 25% of the tampered region can be recovered without tampering/missing coincidence. When the tampering rate more than 25%, the quality of the restoration image decreases due to loss of restoration watermark caused by tampering or missing coincidence and false alarm errors.

2.7.2.4 Other techniques based watermarking schemes

Zhang et al. [66] proposed two self-embedding watermarking techniques called a reference sharing mechanism, in which the watermark was derived from the original principal content in different regions and shared by these regions for content restoration. In the first scheme, the watermark was derived from the original data in five MSBs planes. The second scheme was based on the hierarchical self-embedding scheme in which the host principal content is decomposed into three levels. The first method can restore the five most significant bits (MSBs) accurately, if the tampering rate is below 24% while second method is capable to restore up to tampering rate 66%. In [77] a fragile watermarking technique is proposed by Zhang and Wang, which was using a lossless difference expansion (DE) technique, for information embedding and restored the original image without any error. The main disadvantage of this approach is limited robustness against malicious attacks. The maximum supported tampering rate is only 3.2%.

Zhang et al. proposed a fragile watermarking technique for successful restoration of extensive tampering demonstrated in [78]. At the cost of low restoration quality, the technique allows for tampering rate up to 59%. Here the watermark data are made up of two parts, reference-bits for restoration and hash-bits matching the content of local blocks. The reference bits are obtained by quantization of selected lowfrequency DCT coefficients. In order to spread the reference information over the entire image, individual blocks are pseudo-randomly grouped into pairs, and the reference information is projected onto the watermark payload using a random binary matrix while the hash-bits are embedded into the local blocks. In [79] Chen et al. proposed a chaos based self-embedding fragile watermarking scheme of block size 2×2. This scheme provides tamper detection and recovery with the security against the maliciously modified by collage attack, constant-average attack. In [80] Chuan et

al. proposed adaptive bit allocation mechanism based a fragile watermarking scheme for image authentication and restoration. The authentication-bits are generated by a DCT-based image hashing algorithm to localize the tampered locations. The restoration of the image is achieved by restoration bits which are generated by encoding the non-subsampled contourlet transform (NSCT) [81, 82] coefficients of the cover image. The embedding of these bits is based on an adaptive bit allocation mechanism which embeds the bits based on the degree of smoothness of the host image. These bits are added only in the first LSB plane of the cover images making the watermark to be very fragile and easy to break. In [83], Hung et al., proposed a half-toning technique based fragile watermarking scheme by generating the signature bits from DCT coefficients of the host image and taking the Inverse-DCT of the image after embedding signature bits to create a watermarked image. For protection, the binary random image is embedded in the LSB of the watermarked image. The restoration is performed by substituting the damaged areas by inverse half-toning of the image from the same position. The scheme can authenticate and restore the tampered locations, however the quality of watermarked image and restored image are reduced.

Liu proposed a self-embedding block-wise watermarking scheme based on the moment preserving technique for color images in [84]. This scheme was used morphological closing operations and two-stage dual parity-check method. This scheme is able to provide high detection rate and recover the tempered region with high quality. Ho et al. proposed a Pinned Sine Transform (PST) based semi-fragile authentication watermarking scheme with self-restoration capability in [85]. The bits as restoration watermark were generated from the image coefficients after PST, and embedded into the LSBs plane of the original image. In PST, an image field is decomposed into two sub-fields, i.e., the boundary eld and a residual eld [86], known as the pinned eld which vanished at the boundaries. The tamper detection accuracy rate of this scheme was shown to be higher than 98% even with light non-malicious image processing operations. However, the LSB based self-restoration watermark is fragile and could be easily distorted.

Zhang et al. [67] proposed a self-recovery watermarking scheme. In this scheme, a block-wise mechanism was used for tampered area detection and a pixel-wise mechanism was used for original content recovery. The recovery data generated

by exclusive-OR operation on the original MSBs and authentication data. The recovery data are embedded into the three LSB planes. This scheme is capable to exact recovery up to tampering rate 54%. Qian et.al [87] proposed an inpainting assisted self-recovery watermarking scheme. The main operation is to encode different types of blocks with varied number of bits and an inpainting method are used to the recovery of blocks with few details. The strength of this scheme is the high quality reconstructed image can be achieved with a low embedding rate. But the drawback of this scheme is the existence of the risk that all the recovery information will be invalid once a block type is broken.

2.7.3 Watermarking techniques for copyright protection

Due to rapid growth in the availability of multimedia content in digital form, a major problem faced by owners is protection of their content or material. They feel worried and concerned about copyright protection and misuse of other intellectual property rights (IPR) of their digital content. A copyright protection scheme must meet the requirements of robustness, imperceptibility, security, blindness and unambiguity [88]. To meet these requirements, many watermarking techniques have been proposed in the past years. A review of important watermarking techniques for copyright protection can be classified into following:

2.7.3.1 DCT based watermarking schemes

Secure spread spectrum based a technique proposed by Cox et al. [16] in which modified 100 largest DCT coefficients excluding DC term, one for embedding the watermark to achieve the requirements of robustness and imperceptibility. This algorithm can extract a reliable copy of the watermark from imagery that are degraded with several common geometric and signal processing procedures. The main defect is that the scheme requires the original image to extract the watermark. Moreover, security is also another serious problem in [89]. In [90], Lin and Chen proposed a DCT based image watermarking technique for copyright protection. This scheme resists some image-processing operations and general JPEG compression to some degree. The main defect of this scheme is that the embedded watermarks may be

destroyed seriously in case of high JPEG compression. Lin et al. [35], proposed a watermarking technique that adjusts the DCT low-frequency coefficients by the concept of mathematical remainder. So this scheme preserves acceptable visual quality of watermarked image. The strength of this scheme is more suitable for images that will be highly JPEG-compressed. Suhail and Obaidat proposed a robust watermarking scheme based on DCT and image segmentation in [60]. In this scheme image is segmented by using Voronoi diagram and features extraction points. Then, a pseudorandom sequence was embedded in the DCT domain of each image segment. This scheme is robust against common image processing operations, JPEG compression for a compression ratio of up to 45 and some label of geometric manipulation. Lee and Li proposed a DCT based self-recognized and crop-resistant watermarking scheme in [91]. This scheme is resist several attacks, including blurring, brightness, contrast, color reduction, large-scale cropping (up to 75%), distortion, Gaussian noising, JPEG compression, rotation, scaling, and sharpening.

2.7.3.2 DWT based watermarking schemes

Wang et al. [92] proposed a wavelet-based watermarking scheme which embeds the scrambled watermark into the middle frequency of wavelet domain. It basically meets the requirements of the security and blindness. The main defect is that when their scheme suffers from some serious attacks, the extracted watermark is ambiguous. Joo et al. [93] proposed a more robust wavelet-based technique than the scheme of Cox et al. [16]. Their scheme embeds the watermark into low subband of wavelet domain by selecting visually insensitive location and repeatedly embedding the watermark to meet the requirement of the robustness. The main defect is that it requires the original image to extract the watermark. In addition, the result of repeatedly embedding the watermark is time-consuming.

A DWT based watermarking scheme proposed by Jianhong et al. in [94] which utilized human visual system (HVS), JPEG standard compression algorithm and DWT decomposition. This scheme is good against JPEG Compression but is not good against some geometric attacks like cropping, scaling, and rotation. DWT and discrete fractional random transform (DFRNT) based robust watermarking scheme proposed by Kim et al. in [95] using two-dimensional (2D) bar code. This scheme was

utilized the benefits of the frequency decomposition ability of DWT and the inherent randomness of DFRNT. The scheme is resist general image processing attacks such as image compression and noise adding. In [96] Hamghalam et al. proposed a geometric modeling based robust watermarking scheme in the wavelet domain. In this scheme, eight samples of wavelet approximation coefficients on each block were utilized to construct two line segments in the 2D space. Here Geometrical tools are used to solve the trade off between the transparency and robustness of the watermarked image. High robustness against gain, noise and compression attacks is the strength of this technique. This technique also shows the superiority against of the common attacks, such as median filtering, scaling, Gaussian filtering and rotation attacks.

2.7.3.3 DWT-SVD based watermarking schemes

Singular Value Decomposition (SVD) based first watermarking scheme was proposed by Liu et al. [97] in 2002. The main feature of SVD-based image watermarking is the stability of singular values, which contain most of the image energy. The DWT-SVD method that combines the SVD technique with the DWT technique is found to be more robust than the DWT-based method [98]. Lai et al. [99] proposed a hybrid DWT-SVD watermarking procedure. In this scheme watermark image was divided into two halves and embedded into the two singular value matrices of intermediate frequency sub-bands (HL1 and LH1 sub-bands) of the host image. After embedding the watermark, the two halves are combined to get the watermarked image. At the time of the extraction reverse procedure is adopted to get the extracted watermark image. This watermarking scheme is good enough in the robustness and imperceptibility against a variety of attacks but is suffering from watermark ambiguity/false positive problem. In false positive problem it was possible to extract a watermark which was not actually the embedded one [100].

A DWT-SVD based watermarking scheme proposed by Pandey et al. in [44] which avoid the pitfall, false-positive problem encountered by Lai et al. [99] while maintaining the imperceptibility and robustness. For avoiding the possibility of false watermark extraction this scheme used a secret key, which is necessary at the time of extraction. This scheme is also robust against the print and scan attack while lacking in case of forging and morphing attack. Pandey et al. proposed divisive

normalization transformation (DNT), DWT and SVD based another hybrid scheme in [45]. This scheme is able to solve the problem of statistically and perceptually redundant wavelet coefficients, used during watermarking with the help of DNT while maintaining the robustness and imperceptibility. In this hybrid scheme, the host image was transformed into discrete wavelet transform domain and then corresponding divisive normalization transform coefficients has been computed. Furthermore, SVD is applied to DNT coefficients of intermediate frequency sub-bands of host image. Finally a watermark has been embedded into singular values of DNT coefficients of intermediate frequency sub bands of host image. DNT is an invertible nonlinear image transformation technique in which all linear transform coefficients are divided by a weighted sum of coefficient amplitudes in a generalized neighborhood. This scheme is also robust in various intentional and non-intentional attacks.

2.7.3.4 Other techniques based watermarking schemes

Based on amplitude modulation, a precise and robust watermark scheme for color image was proposed by Lari et al. in [101]. A watermark is embedded in color image by modifying the pixel values in blue channel. The blue channel is preferred because of the human visual systems reduced sensitivity, and it also guarantees virtual imperceptibility. Because amplitude modulation is a spatial domain watermarking method, it may not be robust enough, i.e. incapable of exact watermark retrieval. In order to enhance the bit retrieval, authors apply a Gaussian mask to equalize the luminance intensity. In order to increase the robustness, authors used curvelet transform to detect singularities such as lines and curve and to prevent the system from using these locations in an image for embedding the watermark bits.

In [102], Wei Lu and Hongtao Lu proposed a novel robust digital image watermarking scheme with the aid of subsampling and nonnegative matrix factorization. Here, Sub-sampling is employed to create a sub-image sequence. The nonnegative matrix factorization (NMF) is applied to decompose the sequence on basis of the column similarity of the sub-image sequence. A Gaussian pseudorandom watermark sequence is embedded in the factorized decomposition coefficients. Due to

the high resemblance of sub-images and meaningful factorization for NMF, the proposed scheme is capable of achieving more robustness, predominantly towards common permutation attacks. The experimental assessment demonstrates the enhanced performance of the proposed scheme. An innovative digital watermarking method that works on basis of vector quantization (VQ) was projected by ChinChen Chang and Hsien-Wen Tseng [103]. On contrary to the traditional VQ-based watermarking schemes, the mean of sub-blocks is employed to train the VQ codebook. Further, the Anti-Gray Coding (AGC) technique is utilized to improve the robustness of the proposed watermarking system. Here, the secret keys are employed to conceal the related information between the original image and the watermark followed by the registration of a set of secret keys to a trusted third party for future validation. So, the original image remains unaltered even after the watermark being melted into the set of secret keys. The experimental evaluation illustrates that watermark is capable of surviving a range of possible attacks. Moreover, the size of the secret keys can be reduced as well.

A very high capacity algorithm proposed by Alattar in [104], based on the difference expansion (DE) of vectors of an arbitrary size has been developed for embedding a reversible watermark with low image distortion. Test results of the spatial triplet-based and spatial quad-based algorithms indicate that the amount of data one can embed into an image depends highly on the nature of the image. To maximize the amount of data that can be hidden into an image, embedding procedure can be applied recursively an across the color components. It does not support if the image is highly distorted. Schyndel et al. [105], generated a watermark using a m-sequence generator. The watermark was embedded to the least significant bit of the original image to produce the watermarked image. The watermark was extracted from suspected bits at the proper locations. Detection was performed by a cross-correlation of the original and extracted watermark. Schyndel showed that the resulting image contained an invisible watermark with simple extraction produces. The watermark was not robust to additive noise.

A QR factorization and DWT based a watermarking scheme proposed by Naderahmadian and Hosseini-Khayatin [106]. In this scheme watermark image was embedded by directly modifying the first row elements in the obtained R matrix after

performing QR factorization in the DWT domain. This scheme is efficient in computational complexity and good robustness against some image processing operations in comparison with SVD and DCT methods. A QR factorization and DCT using quantization index modulation technique based a robust watermarking scheme was proposed in [107]. This scheme is good in imperceptibility and also displays strong robustness against common signal processing operations, JPEG200 compression and geometric attacks.

Naderahmadian and Hosseini-Khayat proposed QR decomposition based a blind robust watermarking scheme in [108]. This scheme has better capacity and robustness in comparison with SVD and DCT methods. This scheme is being robust against different image processing attacks like rotation, median and average filtering and salt and pepper noise. Guo and Liu in [109] proposed a robust watermarking scheme for embedding a multitone watermark using low computational complexity. The proposed scheme is capable to guard against reasonable cropping or print-and-scan attacks. The drawback of this scheme is that the contrast of the decrypted watermark is somewhat unclear because of the interference with the host images texture. Su et al. [110] proposed a Scale-Invariant Feature Transform (SIFT) based robust watermarking scheme by using both the interest point extraction and the pilot signal embedding. The pilot signal detection in the proposed scheme was based on the local search to recover the regions for the subsequent watermark detection. This scheme resists against geometrical transformations.

2.8 Future Research Directions

Digital image watermarking is an emerging research area for authentication, recovery (or restoration) and copyright protection of electronic media documents. Most of the research is going on in the field of digital image watermarking. The cause might be that there are so many images available on Internet which needs to be protected due to advancement of multimedia editing techniques. The area of watermarking is well studied and researched, but existing watermarking schemes have own strengths and limitations.

Subsection 2.7.1 reviews significant watermarking schemes for image authentication. These schemes mostly are fragile in nature and unable to resist any types of attacks

and almost all block wise schemes have low tampered localization accuracy. In future, it is essential to improve the accuracy of tampered localization and to develop the image authentication schemes that survive unintentional attacks.
Subsection 2.7.2 reviews significant watermarking schemes for image authentication and restoration. Basis of that following limitations in the most of existing schemes are: (a) Low tampered localization accuracy, (b) In extensive tampered area, the restoration quality decreases, (c) Restoration is not possible at high tampering rate (d) Problem of reconstruction dependency and (e) Inability to resist any types of attack. Basis on these limitations, future requirements are to design and develop the new watermarking schemes for image authentication and restoration, which could resist common image operations and are capable to restore the tampered area even though at high tampering rate.
Subsection 2.7.3 reviews significant watermarking schemes for image copyright protection. Basis of that following limitations in the most of existing schemes are: (a) False positive problems (b) Ambiguous (c) Unable to resist geometric attacks and (d) Non Blind. The possible remedy might be a new watermarking scheme for copyright protection which would capable to resist geometrical attacks with reducing all limitations.

2.9 Conclusion

In this chapter an extensive discussion of various aspects of digital watermarking schemes like classification, methodology, applications, attacks, issues and challenges along with their needs and enhancements are given. This chapter classified and discussed digital watermarking in all the known aspects like robustness, watermark type, domain, and detection process. Further this chapter presents the comparative performance analysis of various watermarking methodologies along with the detailed discussion of significant existing watermarking schemes and their applications which are extremely diverse including authentication, restoration and copyright protection. The researchers have made great efforts in watermarking field as a security tool in the past decade; though, there is still a lot to be done.