Multimedia Tools and Applications (2024) 83:31829–31891
https://doi.org/10.1007/s11042-023-16843-3

A review of image watermarking for identity protection and verification

Sunpreet Sharma[1] · Ju Jia Zou[1] · Gu Fang[1] · Pancham Shukla[2] · Weidong Cai[3]

Received: 28 September 2022 / Revised: 21 June 2023 / Accepted: 31 August 2023 /
Published online: 19 September 2023

Abstract
Identity protection is an indispensable feature of any information security system. An identity can exist in the form of digitally written signatures, biometric information, logos, etc. It serves the vital purpose of the owners' verification and provides them with a safety net against their imposters, so its protection is essential. Numerous security mechanisms are being developed to achieve this goal, and information embedding is prominent among all. It consists of cryptography, steganography, and watermarking; collectively, they are known as data hiding (DH) techniques. In addition to providing insight into various DH techniques, this review prominently covers the image watermarking works that have positively influenced its relevant research area. To that end, one of the main aspects of this study is its inclusive nature in reviewing watermarking techniques, via which it *aims* to provide a 360° view of the watermarking technology. The main contributions of this study are summarised below.

– The proposed study covers more than 100 major watermarking works that have positively influenced the field and continue to do so. This approach makes the discussion effective as it allows us to pivot on the vital watermarking works that have positively influenced the research area instead of just highlighting as many existing methods as possible. Moreover, it also empowers us to provide the readers with an insight into the current research trends, the pros and cons of the state-of-the-art methods, and recommendations for future works.
– In addition to reviewing the state-of-the-art watermarking works, this study solves the issue of reverse-engineering the main existing watermarking methods. For instance, most recent surveys have focused primarily on reviewing as many watermarking works as possible without probing into the actual working of the techniques. This approach can leave the readership without a vital understanding of implementing or reverse-engineering a watermarking method. This issue is especially prevalent among newcomers to the watermarking field; hence, this study presents the breakdown of the well-known watermarking techniques.
– A new systematisation of classifying existing watermarking methods is proposed. It classifies watermarking techniques into two phases. The first phase divides watermarking methods into three categories based on the domain employed during watermark

✉ Sunpreet Sharma
90931966@westernsydney.edu.au

Extended author information available on the last page of the article

embedding. The methods are further classified based on other watermarking attributes in the following phase.

Keywords Image watermarking review · Robust watermarking · Image verification · Copyright protection · Fragile watermarking · Identity protection

Nomenclature

ACC	Accuracy
BER	Bit error rate
$BPAP$	Block-level pixel adjustment process
C_b	Chrominance blue channel
C_r	Chrominance red channel
CE	Contrast enhancement
CR	Compression ratio
CRC	Cyclic redundancy check
$dB4$	Daubechies 4 wavelet
dB	Decibel
DC	Direct-current
DCT	Discrete cosine transform
DFT	Discrete Fourier transform
DH	Data hiding
DNA	Deoxyribonucleic acid
$DTCWT$	Dual-tree complex wavelet transform
DW	Direct watermarking
DWT	Discrete wavelet transform
ECC	Error correction coding
EV	Energy vector
FN	False-negative
FNR	False-negative rate
FOA	Fruit fly optimisation algorithm
FP	False-positive
FPP	False-positive problem
FPR	False-positive rate
FRT	Finite ridgelet transform
GA	Genetic algorithm
GC	Gamma correction
GN	Gaussian noise
HE	Histogram equalisation
HF	High-frequency
$HFCM$	High-frequency component modification
HH	High-high subband
HL	High-low subband
HVS	Human visual system
$IDCT$	Inverse of the DCT
$IDWT$	Inverse of the DWT
ISB	Intermediate significant bit
ISR	Insignificant region
$ISVD$	Inverse of the singular value decomposition

LBP	Local binary pattern
LF	Low-frequency
LH	Low-high subband
LL	Low-low subband
LPF	Low pass filter
LSB	Least significant bit
LWT	Lifting wavelet transform
$MD-5$	Message digest-5
MF	Mid-frequency
ML	Machine learning
MLP	Multi-layer perception
MRA	Multi-resolution analysis
MSB	Most significant bit
NCC	Normalised cross-correlation
$NSGA-II$	Non-dominated sorting genetic algorithm-II
NSW	New South Wales
PAV	Pseudo-random address vector
PHF	Perceptual hash function
$PSNR$	Peak-signal-to-noise ratio
QF	Quality factor
QIM	Quantisation index modulation
$RBAs$	Random bending attacks
RGB	Red-green-blue channels
$S\&P$	Salt and pepper noise
$SHA-256$	Secure hash algorithm 256
$SIRD$	Simple image region detector
SNs	Social networks
SR	Significant region
$SSIM$	Structural similarity index
SVD	Singular value decomposition
SVM	Support vector machines
$SVMW$	Singular value matrix watermarking
TN	True-negative
TP	True-positive
TPR	True-positive rate
VQ	Vector quantisation
XOR	Exclusive-or operation
Y	Luminance channel

1 Introduction

Since the beginning of the Internet, its usage has been on a hike. The Internet has influenced almost every aspect of human life, and their dependence on it is increasing daily. However, the Internet has never been as prominent since 2020. COVID-19 has altered how people interact in their professional and personal lives. This pandemic has severely curtailed the use of offices and places to socialise, forcing the world into a lockdown. This left people with the Internet as their primary mode of communication, ramping its usage to new heights.

The Internet is vital for keeping people in touch via social networks (SNs) and tools like Zoom™, and Microsoft Teams™. On the flip side, this is also the prime time for hackers to flex their muscles, and their actions' impact is being felt worldwide. For instance, data breaches exposed 36 billion records in the first half of 2020 [89]. A breach is even more detrimental when performed on SNs. For instance, the Twitter ™ breach in July 2020 aimed to ruin the image of politicians and business tycoons [15]. Moreover, such actions are the worst when carried out on sensitive data such as medical images, passports, licenses, and other legal documents [95]. Thus, thwarting them is vital. To this end, many data hiding (DH) techniques: cryptography, steganography, and watermarking, that fall under the umbrella of cybersecurity are blooming. A brief insight into these DH techniques is as follows.

Cryptography alias "secret writing" is derived from Greek words *kryptos*: secret and *graphein*: writing [45]. It is a way of transmitting a secret message by concealing it within a cover medium. Note a cover medium can exist in various forms: a video, an image, a speech signal, and others. However, as this review focuses primarily on images thus, the cover medium corresponds to an image in this discussion. Before transmission, a cryptography process consists of scrambling the secret message using a key, known as encryption, followed by its embedding in a cover image. After transmission, the secret message is extracted from the cover image and then unscrambled using the aforementioned key, decryption. Note that the secret key is generally transmitted separately from the encrypted image to minimise the chances of hacking. During the transmission, the primary purpose of encryption is to make the data unintelligible to unauthorised personnel.

Steganography is made from the Greek word *steganos*, which means "hidden". Some current research works classify cryptography and steganography as the same [22]. However, there are fundamental differences between the two. First and foremost, in the former's case, the secret message, also known as the *planetext*, is encrypted and converted into the *ciphertext* before it is concealed in a cover medium. In the latter's case, the secret message never changes its state and is embedded as it is but confidentially into a cover medium. Second, cryptography aims to hide the message content from a hacker but not the message's existence. Steganography even hides the very existence of the message within the communicating data. On the same note, the security of a cryptography process is assumed to be compromised when the encrypted message is hacked. In the case of steganography, it is considered compromised the moment the very existence of the hidden message is confirmed.

It is evident from the previous discussion that the primary concern of both steganography and cryptography processes is the security of the concealed message but not that of the cover medium. This is where watermarking comes into the limelight. In hindsight, steganography and cryptography are means of covert communication, whereas watermarking primarily focuses on media copyright protection and verification. Moreover, a watermark's embedding can be visible or invisible; however, the embedded message in steganography and cryptography schemes must be invisible or hidden. As this review focuses on image watermarking, thus it has the center stage in the rest of this discussion.

1.1 Our contributions

– In addition to reviewing the state-of-the-art watermarking works, this review solves the issue of reverse-engineering the main existing watermarking methods. For instance, most recent surveys have focused primarily on studying as many watermarking works as possible without probing into the actual working of the techniques. This approach can leave the readership without a vital understanding of implementing or reverse-engineering a

watermarking method. This issue is especially prevalent among newcomers to the watermarking field; hence, this study presents the breakdown of the well-known watermarking techniques. To the best of our knowledge, this study is the first review in the watermarking field that attempted to do so. It is assumed that the study can provide the necessary tools to the new entrants to kick-start their research and equally serve their experienced peers as their go-to study whenever they want to revisit essential watermarking concepts.
- In line with the above-mentioned contribution, this review probes into the watermarking works which have shaped the field and continue to do so. This approach makes the discussion effective as it allows us to pivot on the vital watermarking works that have positively influenced the lot instead of just highlighting as many existing methods as possible. Moreover, it also empowers us to provide the readers with an insight into the current research trends, the pros and cons of the state-of-the-art methods, and recommendations for future works.
- A new systematisation of classifying existing watermarking methods is proposed. It classifies watermarking techniques into two phases. The first phase divides watermarking methods into three categories based on the domain employed during watermark embedding. The methods are further classified based on other watermarking attributes in the following phase. More on this systematisation is within Section 5.1.

The rest of this discussion is as follows. Section 2 covers the general watermarking concepts, and Section 3 presents the commonly used performance metrics to evaluate watermarked images. Section 4 introduces watermarking attacks, and the subsequent Section 5 reviews the well-known watermarking works. The next is Section 6, wherein a summary of the methods discussed in this review is provided. Moreover, a questionnaire is also developed within this section that facilitates the evaluation of the existing processes and provides guidelines or recommendations for designing new ones. Finally, Section 7 concludes the discussion.

2 Watermarking

2.1 Definition and applications

The image watermarking process embeds subtle information known as the "watermark" to a host/original image. The embedded watermark can successively be extracted to validate the host image [96, 98]. A successful extraction proves the intactness of the host image or vice-versa. The upcoming Section 2.2 discusses the watermark's embedding and extraction procedures in detail.

The term "Digital Watermarking" dawned in the early 1990s, and since then, it has been an active research topic [113]. Its applications are continuously branching out to new advents in technology; for example, the process of watermarking a neural network is known as "passporting" [11, 111], securing the cloud storage systems [102, 103], electronic money transfers, e-governance [52]. Various state-of-the-art watermarking applications and their description are presented in Table 1.

Notwithstanding the successes of watermarking in the aforementioned applications, many prominent industries are still missing out on the benefits of this technology. For instance, according to Bertini et al. [9], only one of 13 main SNs uses watermarking technology. The same study also highlights that these platforms are the major sources of information leaks and identity theft. To this end, the Facebook™ security breach at the beginning of

Table 1 Watermarking applications

Application	Description	Example
Packaging and tracking	Watermarks offer brand authenticity and traceability to products throughout the global supply chain.	Digimarc™ [20]
Combating piracy	Forensic video watermarking helps combat piracy for premium videos and live sports.	NexGuard™ [67]
Neural networks' protection	Resolving copyright issues related to deep neural networks by watermark embedding.	White box embedding [111]
Medical devices' protection	Thwarting counterfeiting of medical devices and pharmaceutical products.	Ghost™ [26]
Plastic recycling	Automated sortation of plastic packaging by recycling facilities.	Holygrail 2.0 [33]
Maintaining electoral integrity	The ballot papers are watermarked during voting.	Ballot voting [115]
Medical record authentication	Authentication of digitally preserved patient's medical record	DICOM [82]
Currency protection	Watermarks ensure effective document and banknote protection.	G+D [25]
Traitor tracing	Watermarking traces the source(s) of leaks when proprietary data is illegally sold.	Renewable traitor tracing [87]
Identity documents	Watermarking is employed to maintain security standards for proof-of-identity (POI) documents.	Identity security [42]
The peer review process	Academic journals use watermarking to safeguard the manuscripts during the peer review process.	Conftool [16]

2020 impacted its 50 million users. These users had their email accounts compromised, pictures or images were stolen, and the same goes for the Twitter™ breach of July 2020 [92]. Subsequently, Services New South Wales (NSW), Australia's information systems were infiltrated, and numerous sensitive documents were stolen. Consequently, almost a quarter of a million Australians lost their personal information in the form of driver's licenses, handwritten signatures, and marriage and birth certificates [95]. Moreover, data breaches exposed billions of records in 2020, whereby 86% of violations were financially motivated, and 10% were motivated by espionage [19]. These are only a handful of snippets of the wide range of persisting cyber-attacks that have inspired this review, as thwarting them is pivotal.

To sum up, most of the above-mentioned incidents happened due to organisations' lacking copyright protection and authentication mechanisms. Therefore, the need for watermarking to address this shortfall is vindicated. Image copyright protection and authentication have been a critical focus of watermarking technology ever since its arrival [4, 65]. To that end, as this research study is focused on reviewing the image watermarking methods, its significance is therefore justified.

2.2 Watermarking process

The watermarking process primarily consists of two parts. The first part, as shown in Fig. 1, is that of the watermark's embedding, and the other is that of the watermark's extraction. The watermark embedding happens on the sender's side, whereas the extraction occurs on the receiver's side. Each of these parts is discussed below.

Firstly, the watermark is encrypted using an encryption algorithm. Note that this step is optional but is a common practice within image watermarking. The main reason for such encryption is that it uplifts the security of a watermarking scheme by making the watermark unintelligible to hackers. Consequently, even if a hacker can detect the presence of the watermark, it is simply impossible to make any sense of it as encryption scrambles it before embedding. To this end, the watermark can only be unscrambled by applying the inverse of the encryption algorithm that scrambled it in the first place. This is why an encryption algorithm is called the "key" in watermarking. In other words, the watermark can not be extracted without knowledge of the encryption algorithm employed during the embedding phase. Moreover, it is well established in the literature that the combination of watermarking and encryption is an indispensable tool that certainly limits, if not eradicates, the watermark's duplication or removal. Some of the widely used encryption algorithms are duly acknowledged in the later parts of this review.

Secondly, once the watermark is encrypted (if it is encrypted), it is embedded into the host image. The watermark embedding follows a set of rules generally called the embedding rules. Some researchers within the field are actively working on optimising the existing watermark embedding rules, and others are focused on developing new ones. Irrespective of who is doing what, an embedding rule is designed by considering several requirements which need to be addressed by a watermarking scheme. These requirements are discussed below in Section 2.3. Once the watermarked image is achieved, it is transmitted, and so is the encryption key(s). In most cases, the watermarked image and the encryption key(s) are transmitted separately to minimise hacking-related risks.

Thirdly, similar to the embedding phase, the watermark extraction follows a set of rules. These rules are known as the extraction rules. The watermarked image is decoded on the receiver's side, and the embedded watermark's bits are extracted. Subsequently, the extracted bits are unscrambled using the above-mentioned key(s), culminating in the extraction process.

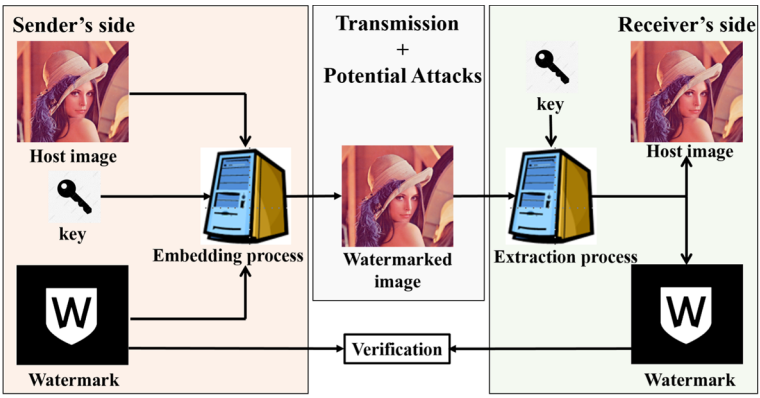

Fig. 1 An overview of the watermarking process. In this example, the embedded watermark is invisible, and the watermark extraction is blind. These attributes are discussed in detail in Section 5.1

Note, based on the extraction type, both the host and the watermarked images are sometimes required for the watermark extraction, and sometimes only the watermarked image is sufficient. This difference between the two is addressed in detail in Section 5.1.

2.3 Watermarking requirements

A successful image watermarking scheme needs to address three main requirements [98]. Firstly, in the case of invisible watermarking, adding a watermark to the host signal (original image) has to be imperceptible. This avoids any deformities perceived by the human visual system (HVS). Secondly, the watermark needs to be secure against unauthorised modifications. Thirdly, a watermarking scheme should have a healthy capacity, i.e., its ability to embed large watermark(s).

These three requirements are closely correlated; changing one can significantly affect the other. For instance, high capacity can improve security but degrades imperceptibility. In other words, the lower the capacity, the better the imperceptibility, and the weaker the security. Thus, reaching an equilibrium amongst these requirements is a significant challenge in the field, especially between imperceptibility and security, as they are conflicting in nature. The existing trade-offs between the watermarking requirements are illustrated in Fig. 2. Most current watermarking methods are developed by considering the trade-offs between these watermarking requirements.

3 Performance baseline and metrics

The efficacy of the performance of the watermarking methods was measured using several performance metrics. To this end, an insight into some of the widely cited performance metrics is given below [6, 39, 41, 43, 46, 47, 60] and [114].

Fig. 2 Illustration of trade-offs in watermarking. The 1^{st} column contains the host image (Lena, 512×512 in size) and watermark (jetplane). The 2^{nd} column shows Lena's image watermarked with jetplane (256×256 in size). The 3^{rd} and 4^{th} columns illustrate Lena's images watermarked with 128×128 and 64×64 sized watermarks, respectively. In the 1^{st} row, it can be observed that the watermark's imperceptibility is increasing from left to right because the watermark size is decreasing. In contrast, the 2^{nd} row shows that the extracted watermark's quality deteriorates as the size decreases. Best viewed when zoomed in

3.1 Imperceptibility measures

The embedded watermark's imperceptibility is measured via the peak-signal-to-noise ratio (*PSNR*). The *PSNR* values are calculated in decibels (dB) via (1)—the higher the *PSNR* value, the better the imperceptibility:

$$PSNR = 10\log_{10}\frac{(2^b-1)^2 wh}{\sum_{i=1}^{w}\sum_{j=1}^{h}[I(i,j)-I'(i,j)]^2}, \tag{1}$$

where b, w, and h represent the number of bits used to represent the pixel value, image width, and image height, respectively. Furthermore, $I(i,j)$ and $I^{'}(i,j)$ indicate pixel values of the host and the watermarked images, respectively.

Another parameter that measures the embedded watermark's imperceptibility is the structural similarity index (*SSIM*), calculated as per (2):

$$SSIM(I,I^{'}) = l(I,I^{'})c(I,I^{'})s(I,I^{'}), \tag{2}$$

here or at any other instance in this discussion, I and $I^{'}$ stand for the host and the watermarked images, respectively. Moreover, $l(I,I^{'})$, $c(I,I^{'})$, and $s(I,I^{'})$ are the functions comparing the luminance, contrast and the overall structure of the host image and the watermarked image, respectively. To this end, if there is no difference (in terms of luminance, contrast, and structural) between I and $I^{'}$, then the value attained by *SSIM* is '1' else, it is less than one. Note that the higher the *SSIM*, the better the imperceptibility. Further insight into *SSIM* can be gained from [2].

3.2 Security measures

The security of the embedded watermark is tested through normalised cross-correlation (*NCC*), given by (3), where W and $W^{'}$ stand for the original and the extracted watermarks of dimensions $P \times Q$, respectively:

$$NCC = \frac{\sum_{i=1}^{P}\sum_{j=1}^{Q}(W[i,j]\times W^{'}[i,j])}{\sqrt{\sum_{i=1}^{P}\sum_{j=1}^{Q}(W^2[i,j])}\times\sqrt{\sum_{i=1}^{P}\sum_{j=1}^{Q}(W'^2[i,j])}}. \tag{3}$$

Note, sometimes in the literature, the *NCC* is also addressed as "NC", and for the sake of consistency, the former is adopted throughout this discussion. The *NCC* values should range between [0 1], with '0' being the least in similarity and '1' being the highest. Further insight into the *NCC* and its theoretical basis can be gained from [66] and [125].

Another security parameter that measures the similarity between the embedded and the extracted watermarks is the bit error rate (*BER*). It is calculated as per (4);

$$BER = \left(\frac{\sum_{i=1}^{P}\sum_{j=1}^{Q}[(W[i,j]-W^{'}[i,j])^2]}{P\times Q}\right)\times 100. \tag{4}$$

The *BER*'s value lies between 0 and 1. The watermark extraction is perfect if the *BER* is '0'. In such a case, the extracted watermark bits are identical to the embedded ones. In contrast, the *BER* value of '1' indicates a total mismatch between the former and the latter. The symbols in (4) are similar to the ones in (3), i.e. W and $W^{'}$ stand for the original and extracted watermarks of dimensions P and Q, respectively.

3.3 Tamper detection and localisation measures

The false-positive rate (*FPR*), the false-negative rate (*FNR*), and the true-positive rate (*TPR*) are employed to measure tamper detection and tamper localisation attributes, facilitated only by a fragile watermark [78]. The *FPR, FNR,* and *TPR* are defined by (5), (6), and (7), respectively:

$$FPR = \frac{FP}{FP + TN}, \tag{5}$$

$$FNR = \frac{FN}{FN + TN}, \tag{6}$$

$$TPR = \frac{TP}{TP + FN}. \tag{7}$$

Here false-negative (*FN*) is the number of tampered pixels (which should be judged as tampered) that are judged as non-tampered. False-positive (*FP*) is the number of non-tampered pixels (which should be judged as non-tampered) that are judged as tampered. True-positive (*TP*) is the number of tampered pixels (which should be judged as tampered) that are judged as tampered. True-negative (*TN*) is the number of non-tampered pixels (which should be judged as non-tampered) that are judged as non-tampered.

Finally, another parameter that measures a watermarking scheme's effectiveness in tamper detection and tamper localisation is known as the accuracy (*ACC*) [78]. It is defined as per (8):

$$ACC = \frac{TP + TN}{FP + TN + TP + FN}. \tag{8}$$

The *ACC* should have values between [0 1]. The closer the *ACC*'s value to '1', the better the watermarking scheme's accuracy in detecting the tampering and locating or localising the regions it affects.

4 Watermarking attacks

Before delving into the intricacies of the watermarking attacks, we like to shed light on two critical terms. The first is the spatial domain, and the other is the transform domain. In image processing, the spatial and transform are two fundamental domains employed for analysing and manipulating digital images [31, 107]. Moreover, as the proposed study targets image watermarking, these terms are frequently used in the rest of this discussion.

The spatial domain refers to the original image representation, where each pixel value corresponds to a specific location in an image. In this domain, image processing operations are performed directly on the pixel values to obtain the desired outcome [31, 107]. In the spatial domain-based watermarking, the watermark is embedded directly into the host image's pixel values through various techniques covered in Section 5.2.

In contrast, the transform domain involves applying a mathematical transform to convert the image from the spatial domain to a different domain [31, 107]. For instance, the Fourier transform is a commonly used technique that converts an image from the spatial domain to the frequency domain, representing the image or pixel information as a set of coefficients [7]. In the transform domain-based watermark embedded, these coefficients are manipulated through

various techniques mentioned in Section 5.3. During the extraction phase, the watermark is extracted from the manipulated coefficients and transformed back into the spatial domain by taking an inverse of the applied transform technique.

In image watermarking, an attack is defined in the form of manipulations, if performed on a watermarked image, have the potential to harm the embedded watermark. In other words, it may impair the watermark detection on the receiver's side after transmitting the watermarked image [38]. Image watermarking attacks exist in a wide range; however, they can be classified as either geometrical attacks or non-geometrical attacks [21].

The geometrical attacks are the ones that occur within the spatial domain, i.e., via direct manipulation of the pixels. Because of their simplicity, geometrical attacks are the most commonly used ones. Some readily used geometrical attacks are shown in Fig. 3. These attacks are relatively easier to apply and can be applied using readily available software, such as Microsoft Paint™, Adobe Photoshop™, etc. In fact, some of the examples in Fig. 3 are attained using Microsoft Paint™. Moreover, these attacks are easily perceived by the HVS.

In contrast, the non-geometrical attacks are relatively sophisticated and can be executed in both the spatial and the transform domains. To this end, their implementation requires some knowledge from hackers. Consequently, these attacks are generally more severe than the geometrical attacks and inflict more damage on the watermark. Moreover, in some instances, they are so discrete that it is pretty much impossible to tell by the naked eye whether the watermarked image is attacked or not. Some commonly used non-geometrical attacks are shown below in Figs. 4 and 5.

In addition to the aforementioned watermarking attacks, other well-known manipulations are covered here. Vector quantisation (VQ), copy-move, and protocol attacks have been in the limelight over the last few years. Due to space constraints, this discussion does not elaborate on the intricacies of these attacks, and only a brief overview is provided here. However, Haghighi et al.'s study offers an excellent insight into these attacks [30].

- In the VQ attack, a section of a watermarked image(s), achieved using a particular watermarking method, is inserted into another watermarked or target image acquired by the same method. Illustrations within the red boundaries in Fig. 6 depict images exposed to the VQ attack.
- In the copy-move attack, a part(s) from a watermarked image is copied and subsequently placed within the same watermarked image. Illustrations within the orange boundaries in Fig. 6 show a few examples of the images attacked via copy-move.

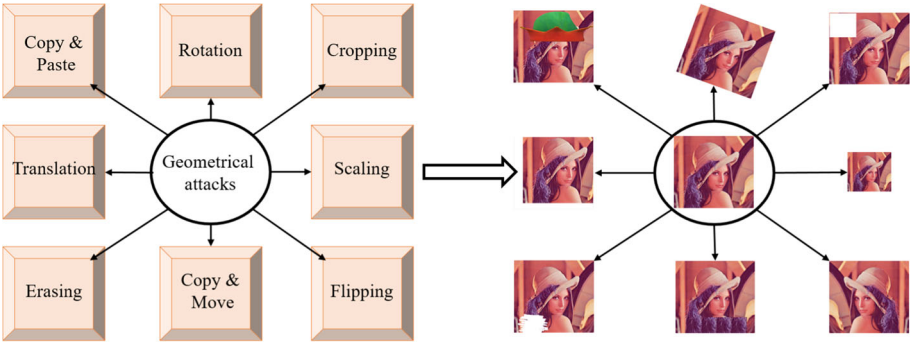

Fig. 3 A few examples of the well-known geometrical attacks. Best viewed when zoomed in

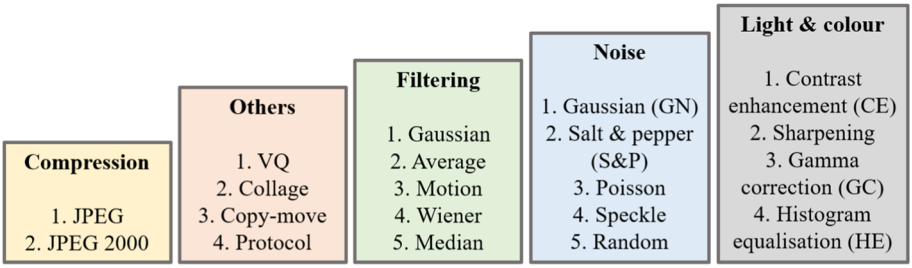

Fig. 4 Commonly used non-geometrical attacks

– The protocol attack, also known as the watermark copy or ambiguity attack, is one of the significant watermarking manipulations. In this attack, external information is inserted into a target image so that the least significant bits (LSBs) of the target image remain unaltered. Consequently, the attack often leads to ambiguity during the watermark extraction process, and the attack may remain unnoticed. Despite the attack's effectiveness, many state-of-the-art methods have not been tested against this attack. The effects of the protocol attack are evident from illustrations within the green boundaries of Fig. 6.

5 Review of the existing methods

The year-wise distribution of the methods discussed in this paper is illustrated in Fig. 7.

5.1 Classifications of the existing methods

The watermarking methods discussed in this review are classified in two phases: phase-1 and phase-2. In phase-1, methods are classified based on the domain employed for watermarking

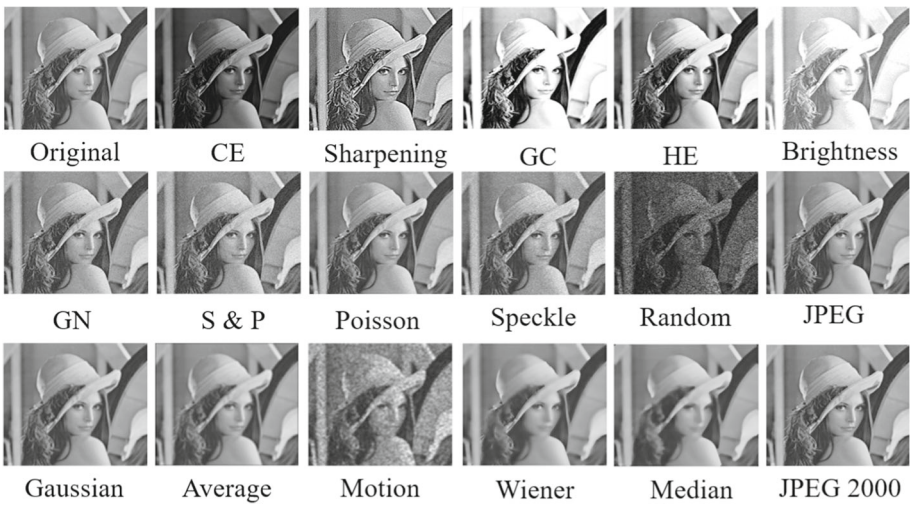

Fig. 5 Examples of commonly used non-geometrical attacks. Best viewed when zoomed in

Fig. 6 Illustrations (top to bottom) of the VQ, copy-move, and protocol attacks. This figure is inspired by Sharma et al.'s study [100]. Best viewed when zoomed in

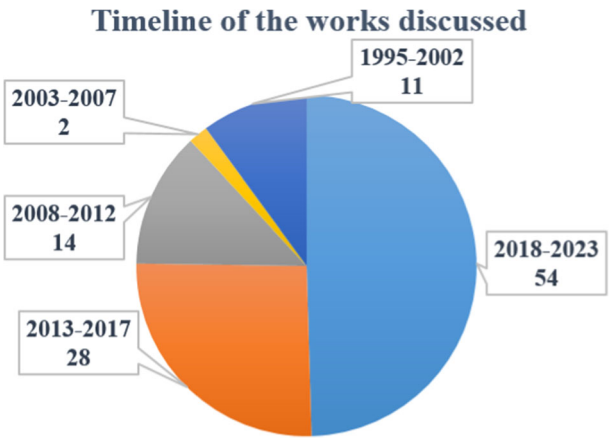

Fig. 7 The year-wise distribution of the discussed methods

embedding. In phase-2, they are further classified based on several attributes. These phases are briefly illustrated in Fig. 8.

In the first phase (phase-1), the methods classified based on the embedding domain belong to one of the following three categories. Firstly, the techniques for embedding the watermark in the spatial domain. Secondly, the ones in which embedding occurs in the transform or frequency domain. Finally, the others that employ both the spatial and the transform domains during embedding are the hybrid domain-based methods. Some well-known existing watermarking methods in each category or domain are discussed later in this review.

Once a method is classified in the first phase, it is further classified in the second phase (phase-2) based on the following attributes.

The first attribute is based on the watermark's security. Based on this attribute, a method is further divided into two sub-categories. The methods wherein the embedded watermark can withstand watermarking attacks are known as robust watermarking methods. In other words, the embedded watermark in such methods is robust and can be extracted after the watermarked image is exposed to any attack. The techniques wherein the embedded watermark has zero tolerance towards watermarking attacks are fragile. In other words, the embedded watermark in these methods is fragile and can not be extracted after the watermarked image is exposed to an attack.

The second attribute is based on the watermark's extraction process. Based on this attribute, image watermarking methods are further divided into two sub-categories. Ones that require both the original and the watermarked images during the watermark's extraction are called non-blind. The others in which only the watermarked image suffices for the watermark's extraction are called blind.

The third attribute is based on the watermark's visibility. Image watermarking methods are further divided into two sub-categories based on this attribute. Ones in which the embedded watermark is visible to the HVS, and in others, it is invisible.

5.1.1 Visible and invisible attributes

Generally, the watermark embedding process can be expressed as (9):

$$I_{Watermarked} = I_{Host}(1 + \beta W_{Total}), \tag{9}$$

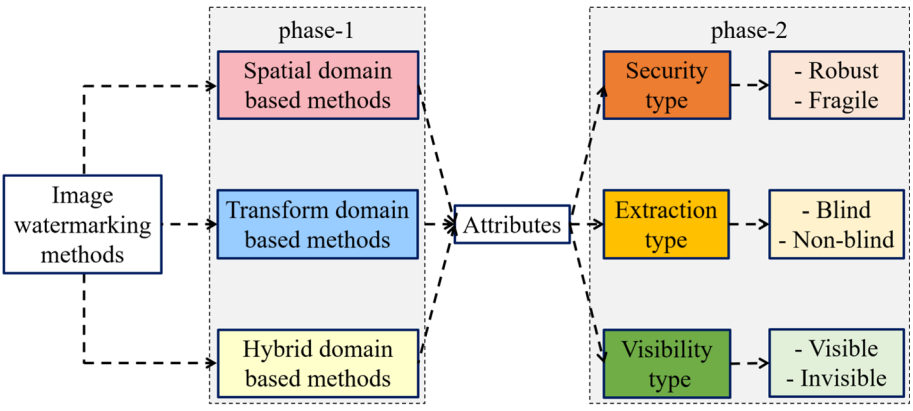

Fig. 8 Existing image watermarking methods' classifications

where $I_{Watermarked}$, I_{Host}, W_{Total}, and β stand for the final watermarked image, the original or host image, the total watermark embedded, and the watermark's embedding strength or scaling parameter, respectively. Note that in (9), the range of β is (0 1], specifying the watermark's visibility. To this end, an obvious watermark is represented by '1' [10]. An illustration of the watermark's visibility in response to different embedding strength factors is given in Fig. 9.

It can be observed in the first row of Fig. 9 that the watermark appears as it is, i.e., unscrambled. This indicates that the watermark was not scrambled before the embedding. In contrast, the second row shows how a scrambled watermark appears in a host image and responds to different embedding strength factors. Note that a greyscale host image is used here in the second row because the discussed changes are visually more prominent (in terms of illumination) in a greyscale image than in its colour counterpart.

Note that (9) is only a general representation of the watermark embedding process and does not explain various intricacies. For instance, the overflowing issue happens when the embedding process (in the case of a greyscale image) causes some pixels to have values greater than 255. Such complications are prominent in the spatial domain-based techniques, rectified by improvising the embedding process. Specifically, the embedding rules, a unique aspect of the overall embedding process, are tailored to limit, if not nullify, the embedding-related issues. Further insight into various embedding processes and rules is provided later in this discussion.

5.1.2 Blind and non-blind attributes

In the case of the non-blind watermark extraction, (9) can be rearranged, and the watermark can be extracted as per (10):

$$W_{Total} = \frac{I_{Watermarked} - I_{Host}}{\beta I_{Host}}. \tag{10}$$

Fig. 9 The watermark's response to different embedding strength factors. β's value from left to right is 0.1, 0.5, and 0.9. The embedded watermark is not scrambled in the first row but in the second row. Note that this illustration is achieved from Sharma et al.'s method in [99]. Best viewed when zoomed in

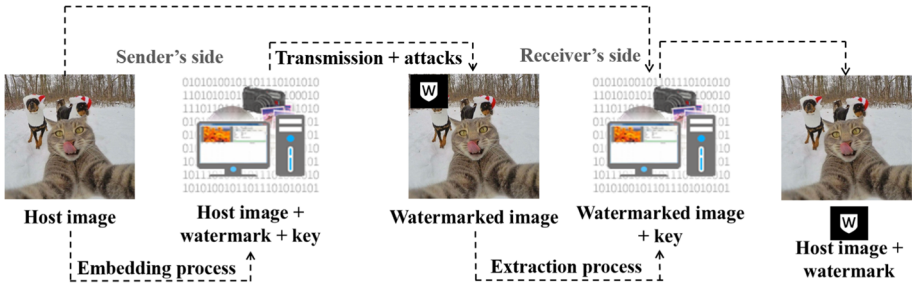

Fig. 10 A working illustration of a non-blind watermark extraction process

It is essential to realise that (10) only outputs the watermark(s) in a scrambled state. The final step in watermark extraction is unscrambling the former by an inverse execution of the aforementioned secret key. A pictorial representation of a non-blind extraction is presented in Fig. 10.

In the case of blind watermarking, the extrication process is not as straightforward as it is in non-blind watermarking. The blind watermark extraction generally follows extraction rules to extract the embedded watermark bits. In general terms, the extraction rules go hand in hand with the embedding rules (discussed in detail in Sections 5.2, 5.3, and 5.4) as the latter varies from method to method; therefore, the former also changes. To this end, it is difficult to express a blind extraction process in a generalised manner. However, the specific steps in any blind extraction process are illustrated in Fig. 11. To this end, insight into the execution of various blind extraction processes is provided as this discussion progresses.

5.1.3 Robust and fragile attributes

An application dictates whether a watermarking scheme needs to be fragile or robust. In other words, robust watermarking achieves copyright protection, and media authentication or verification is performed through fragile watermarking. To this end, the watermark embedding into the host image is tailored to meet the application's requirements. For instance, the resultant strategy is robust when the watermark is embedded into the host image's features that are not easily manipulated or affected by an attack; otherwise, it is fragile.

Fragile watermarking is subdivided into two categories based on the integrity criteria [13]. The first is called semi-fragile watermarking, which provides soft authentication, i.e., has relaxed integrity criteria. The watermark embedded using semi-fragile watermarking

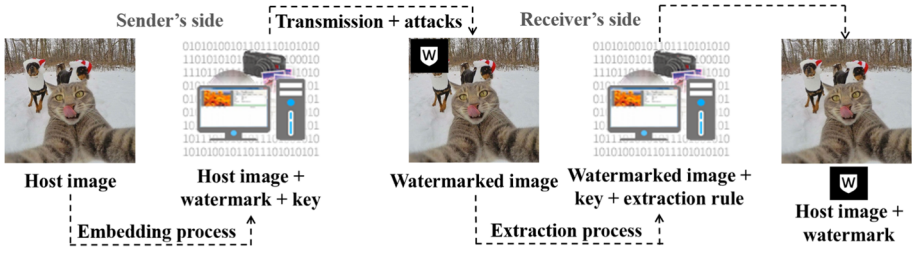

Fig. 11 A working illustration of a blind watermark extraction process

Fig. 12 Robustness illustration of a robust watermark when exposed to different attacks. Solid blue, yellow, and purple boundaries contain the watermarked images under rotation attack at 45°, Gaussian noise (GN) at 0.001, and JPEG compression with a quality factor (QF) of 40, respectively. All dashed borders represent the extracted watermarks from the attacked watermarked images. This illustration is achieved using Sharma et al.'s methods in [99]. Best viewed when zoomed in

techniques is tailored to entertain certain modifications or attacks, such as JPEG or JPEG 2000 compression and luminosity changes. Methods in the other category are considered to be ultimately fragile or hard fragile. These methods follow hard integrity criteria against all modifications-more on these categories is provided as the discussion progresses.

Illustrations of how robust and fragile watermarks respond to watermarking attacks are provided in Figs. 12 and 13, respectively.

The watermark's (*DICTA 2020*) survival or robustness against various attacks is evident in Fig. 12. This ability of a watermark to withstand or survive attacks helps prove an image's copyright information. Moreover, as the extracted watermark (in the case of robust watermarking) is intelligible and resembles the original or embedded watermark, the achieved *NCC* values are high or close to '1' on a scale with a range of [0 1]. To this end, readers may refer to [99] for an insight into the original watermark (*DICTA 2020*) and the *NCC* performance of the watermarks illustrated in Fig. 12.

In contrast, when an image embedded through fragile watermarking is attacked, the embedded watermark becomes unintelligible. This phenomenon is highlighted in Fig. 13, where the *WSU* watermark is employed for fragile watermarking, and its successful extraction (in dashed green borders) is evident from an unattacked image (in solid green boundaries). However, extracted watermarks (in dashed red borders) are unintelligible from images that are attacked (in solid red boundaries), confirming the existence of an attack on the watermarked image and invalidating its authenticity. Moreover, *NCC* values attained by unintelligible watermarks are insignificant and close to '0', the lower end of a scale ranging from [0 1]. One may refer to [97] for further insight into Fig. 13.

A preview of the methods discussed within this review and how they are classified under phase-1 and phase-2 is presented in Table 2.

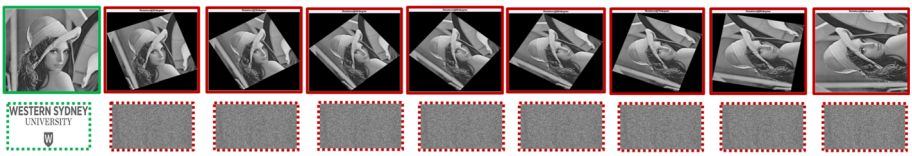

Fig. 13 Fragility illustration of a fragile watermark when exposed to the rotation attack. The solid red boundaries contain watermarked images after modification or attack, and the solid green boundaries contain the watermarked image with no modification. Subsequently, the extracted fragile watermarks from these images are contained within their corresponding coloured dashed boundaries. This illustration is achieved using Sharma et al.'s methods in [97]. Best viewed when zoomed in

Table 2 A preview of the methods discussed within this review and how they are classified under phase-1 and phase-2

Study ↓	Spatial	Transform	Hybrid	Robust	Fragile	SF	Visible	Invisible	Blind	NB
Abraham and Paul [1]	✓	×	×	✓	×	×	×	✓	×	✓
Benrhouma et al. [8]	×	✓	×	×	✓	×	×	✓	✓	×
Celik et al. [12]	✓	×	×	×	✓	×	×	✓	✓	×
Chang et al. [13]	✓	×	×	×	✓	×	×	✓	✓	×
Chang et al. [14]	✓	×	×	×	✓	×	×	✓	✓	×
Dadkhah et al. [17]	×	✓	×	×	✓	×	×	✓	✓	×
Fridrich and Goljan [24]	×	✓	×	×	✓	×	×	✓	✓	×
Gul and Ozturk [27]	✓	×	×	×	✓	×	×	✓	✓	×
Haghighi et al. [28]	×	✓	×	×	✓	×	×	✓	✓	×
Haghighi et al. [29]	×	×	✓	✓	×	✓	×	✓	✓	×
Haghighi et al. [30]	×	×	✓	×	✓	×	×	✓	✓	×
Hsu and Tu [36]	✓	×	×	×	✓	×	×	✓	✓	×
Hurrah et al. [39]	×	×	✓	✓	✓	×	×	✓	✓	×
Hurrah et al. [40]	✓	×	×	×	✓	×	×	✓	✓	×
Islam et al. [43]	×	✓	×	✓	✓	×	×	✓	✓	×
Kamili et al. [46]	×	✓	×	✓	×	×	×	✓	✓	×
Kang et al. [47]	×	✓	×	✓	×	×	×	✓	✓	×
Li et al. [53]	×	✓	×	×	✓	×	×	✓	✓	×
Li et al. [54]	✓	×	×	×	✓	×	×	✓	✓	×
Lin et al. [57]	×	✓	×	✓	×	×	×	✓	✓	×
Loan et al. [60]	×	✓	×	✓	×	×	×	✓	✓	×
Lu and Liao [61]	×	×	✓	✓	✓	×	×	✓	✓	×
Nguyen et al. [68]	×	✓	×	×	✓	×	×	✓	✓	×
Ni et al. [69]	✓	×	×	✓	×	×	×	✓	✓	×
NR and Shreelekshmi [70]	✓	×	×	×	✓	×	×	✓	✓	×
Pal et al. [71]	✓	×	×	✓	×	×	×	✓	✓	×

Table 2 continued

Study ↓	Spatial	Transform	Hybrid	Robust	Fragile	SF	Visible	Invisible	Blind	NB
Pal et al. [72]	✓	×	×	✓	×	×	×	✓	✓	×
Pal et al. [73]	✓	×	×	✓	×	×	×	✓	✓	×
Parah et al. [75]	✓	×	×	✓	×	×	×	✓	✓	×
Parah et al. [76]	×	✓	×	✓	×	×	×	✓	✓	×
Prasad and Pal [77]	✓	×	×	×	✓	×	×	✓	✓	×
Prasad and Pal [78]	✓	×	×	×	✓	×	×	✓	✓	×
Preda and Vizireanu [79]	×	✓	×	×	×	✓	×	✓	✓	×
Preda [80]	×	✓	×	×	×	✓	×	✓	✓	×
Preda et al. [81]	×	✓	×	×	×	✓	×	✓	✓	×
Qi and Xin [83]	×	✓	×	×	×	✓	×	✓	✓	×
Qi and Xin [84]	×	✓	×	×	×	✓	×	✓	✓	×
Raj and Shreelekshmi [85]	✓	×	×	×	✓	×	×	✓	✓	×
Rawat and Raman [86]	✓	×	×	×	✓	×	×	✓	✓	×
Rhayma et al. [88]	×	✓	×	×	×	✓	×	✓	✓	×
Sharma et al. [97]	×	×	✓	✓	×	✓	×	✓	✓	✓
Sharma et al. [99]	×	✓	×	×	×	✓	×	✓	✓	✓
Sharma et al. [101]	×	✓	×	×	×	✓	×	✓	✓	✓
Sharma et al. [100]	×	×	✓	✓	✓	×	×	✓	✓	×
Singh and Singh [105]	×	✓	×	×	✓	×	×	✓	✓	×
Singh and Singh [106]	×	✓	×	×	✓	×	×	✓	✓	×
Thanki and Borra [108]	×	✓	×	×	×	✓	×	✓	×	✓
Ullah et al. [112]	×	✓	×	×	×	✓	×	✓	✓	×
Verma et al. [114]	×	✓	×	✓	×	×	×	✓	✓	×
Wang et al. [117]	×	✓	×	×	✓	×	×	✓	✓	×
Wenyin and Shih [118]	✓	×	×	×	×	✓	×	✓	✓	×
Wong [119]	✓	×	×	×	✓	×	×	✓	✓	×

Table 2 continued

Study ↓	Spatial	Transform	Hybrid	Robust	Fragile	SF	Visible	Invisible	Blind	NB
Wong [120]	✓	×	×	×	✓	×	×	✓	✓	×
Wong and Memon [121]	✓	×	×	×	✓	×	×	✓	✓	×
Xiang et al. [122]	✓	×	×	✓	×	×	×	✓	✓	×
Xiao and Wang [123]	✓	×	×	×	×	✓	×	✓	✓	×
Yeung and Mintzer [124]	✓	×	×	✓	×	×	×	✓	✓	×
You et al. [126]	×	✓	×	✓	×	×	×	✓	✓	×
Zhang et al. [127]	✓	×	×	×	✓	×	×	✓	✓	×
Zong et al. [129]	✓	×	×	✓	×	×	×	✓	✓	×
Zong et al. [130]	✓	×	×	✓	×	×	×	✓	✓	×

Here, semi-fragile is denoted by *SF* and non-blind by *NB*

5.2 Spatial domain-based methods

Several spatial domain methods exist in the field of image watermarking. They are easy to implement and faster than methods executed in other domains. The majority of these methods belong to one of the following categories.

5.2.1 Fragile and other attributes-based methods in the spatial domain

The LSB-based watermarking is one of the most, if not the most widely used, watermarking techniques in the spatial domain. The LSB watermarking methods are divided into two categories. The first is the LSB substitution, and the other is LSB matching [110]. In the former's case, the LSBs of the host image are straightaway substituted with the LSBs of the watermark. In the latter's case, the LSBs of the host and watermark images are matched in the first instance, and the only LSBs that differ from each other are substituted. Consequently, the former has higher watermarking capacity but is prone to unnecessary noise, whereas the latter is the reverse. The degradation in the watermarked image's quality also depends on how many bits are utilised during the embedding process. For instance, if only the LSB (the far right bit in Fig. 14) is utilised in an eight-bit greyscale pixel, the difference between the watermarked images produced using the substitution and the matching-based techniques is insignificant. In this case, the *PSNR* of the images achieved using either of these techniques is around 51 dB [3]. This *PSNR* value drops to 44 dB if the LSB and an intermediate significant bit (ISB) are employed for embedding [17, 23]. Even in this scenario, the difference between the watermarked images produced using each method is negligible. However, a further *PSNR* drop (from 44 dB to around 37 dB and 41dB in the cases of substitution and matching-based techniques, respectively) happens when three of the rightmost bits are involved [105, 109].

Most existing approaches in this category follow similar steps as illustrated in Fig. 15. However, the embedding rule mainly differentiates them from each other. It is often the novel feature that separates one method from the other. Moreover, the LSB-based techniques have an excellent watermarking capacity and are primarily used in fragile watermarking. In the case of tampering, fragile watermarking methods can detect tampering and locate the regions affected by it. In other words, the former characteristic is known as tamper detection, and the latter as tamper localisation. More on these characteristics is provided as the discussion progresses.

Moreover, the watermarked image's degree of degradation is further based on the type of watermark, i.e., whether it is a foreign object or self-generated. It is well-known that a watermark's imperceptibility is considered better when a self-generated watermark is employed during embedding [100]. Merely because, in the foreign watermark's case, the foreign noise is added to the host image. By the way, the term *foreign* refers to the watermark that does not

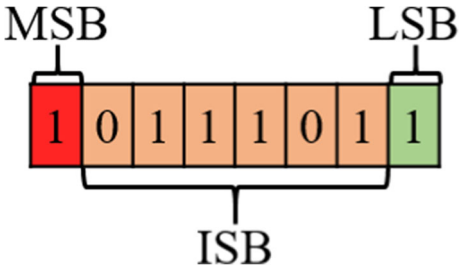

Fig. 14 Different bits within an eight-bit pixel. MSB, LSB, and ISB are the most significant, least significant, and intermediate significant bit(s)

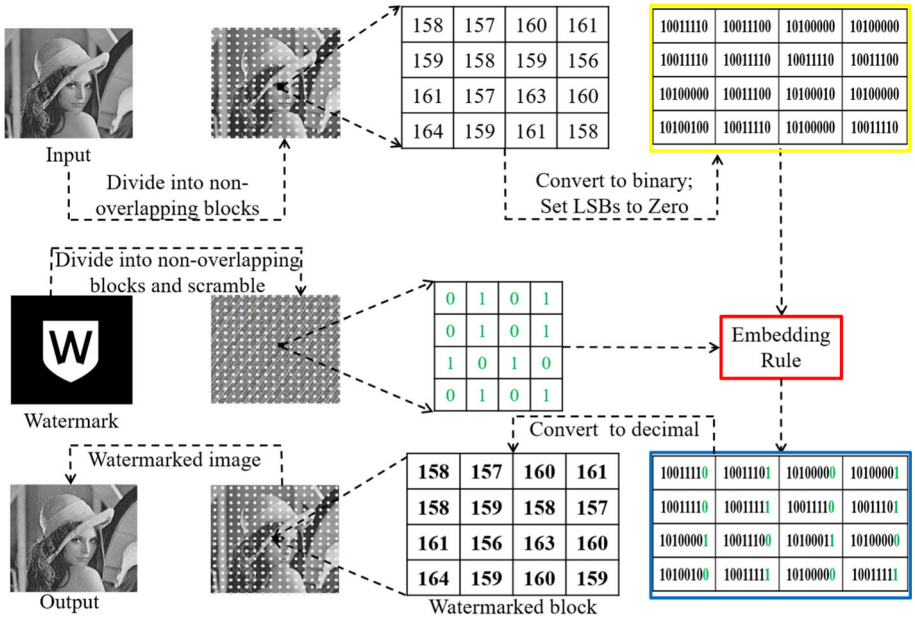

Fig. 15 A working illustration of the LSB substitution-based method. In this illustration, the watermarking is invisible

originate from the host image. In other words, it does not carry any information associated with the host image. A TV channel's logo displayed during the news and the university's emblem on the transcripts are two of the many use cases of foreign watermarks. In contrast, a self-generated watermark is generated from the host image and generally carries some information associated with the host image. That said, a foreign watermark is usually employed for copyright protection or ownership claims, facilitated by robust watermarking schemes. In contrast, the self-generated watermark is used for verification or authentication, enabled by fragile watermarking. In addition, the self-generated watermarks exhibit better tamper detection and localisation performance when compared to their foreign counterparts. These are some reasons why preference is given to self-generated watermarks for fragile watermarking. Needless to say that self-generated watermark-based schemes are also referred to as self-embedding watermarking schemes in the literature. Hence, from this point onward, most discussion on fragile watermarking revolves around self-generated watermark-based methods. Readers may refer to earlier surveys [31] and [107] as well as methods [39] and [46] to gain further insight into foreign watermark-based fragile watermarking. Moreover, Fig. 15 is tailored to show the usage of a foreign watermark in an LSB-based method.

Walton's work in [116] is one of the first well-known works in fragile watermarking. Their technique uses the checksum approach, wherein the sum of the first seven bits (starting from the leftmost bit or the MSB) in an eight-bit pixel is employed to detect whether or not the pixel is tampered with. Although their method is laborious and struggles with limited tamper detection, being a pioneer inspired many later works, such as [119, 120] and [124]. These later works curbed the limitations of Walton's method, but they suffered from the VQ and collage attacks. Wong and Memon address the shortfall in [121], wherein authors have divided the host image in a block-wise manner and then used a hashing algorithm to establish the inter-block dependency. The method ignited the use of hashing in fragile

watermarking or encryption in general. Interblock reliance is necessary to deal with VQ and collage attacks; however, the technique requires a binary equivalent of the host image to complete the verification process, incurring additional communication costs. Celik et al.'s method in [12] eradicated the downsides associated with the earlier methods [119] and [121]. Their process used a hierarchical approach to tackle VQ and collage attacks but performed poorly in tamper detection. The such poor performance resulted from using large pixel blocks while conducting the detection procedure. Note that the details on hashing or other encryption algorithms mentioned in this review are not provided. Because the watermarking technology uses these techniques as a tool, an in-depth discussion on encryption techniques is redundant in this review. However, readers can refer to [48] and [49] to gain insight into various encryption techniques.

Regarding hashing and hierarchical approaches, Hsu and Tu have used message digest-5 (MD5) hashing to generate the authentication bits, which are subsequently embedded into the host image [36]. These bits are then used for tamper detection and localisation in two hierarchical phases, wherein the detection results of the first phase are improved in the second phase. To this end, if the tampering rate is 7.64%, the method's *FPR* and *FNR* performances are 0.22% and 1%, respectively. Unfortunately, these values degrade significantly when the tampering rate is $> 40\%$. Subsequently, Li et al.'s method in [54] extends the work in [36]. The extended process is implemented block-wise, wherein a 64-bit authentication code exclusive to each block is computed using the MD5 hashing algorithm and finally embedded via the LSB substitution. The improvised technique can outperform the method in [36] regarding *FNR* and *FPR* performance. For instance, even for 80% tampering, the extended process can achieve the *FNR* and *FPR* values of 3.1% and 16%, respectively. Another hashing technique readily used in watermarking is secure hash algorithm-256 (SHA-256) [27, 85]. Recently, a combination of MD5 and SHA-256 hashing techniques has been used by Neena and Shreelekshmi in [85]. The combination has not only improved the scheme's overall fragility against most watermarking attacks, but the *FPR* and *FNR* performances have also surpassed that of the above-discussed works. In contrast, in their study, Gul et al. [27] proved that Neena et al.'s method struggled with accurately detecting the tampered regions and shared that the combination of two hashing techniques leads to a hike in the processing time. Inspired by these reasons, Gul et al. employed only the SHA-256 hashing in their method, via which they could maintain the watermark's fragility against the majority of attacks in a streamlined fashion. However, the tamper detection accuracy suffered as the employed size of the block-wise division was 32×32.

By the way, it's not only the hash-based approaches hired for securing the watermark before embedding but also the chaos-based approaches. Some known spatial domain-based fragile watermarking works employing chaos-based encryption are [13, 77, 78] and [86]. In their non-blind approach, Raman and Rawat used Arnold cat map and logistic mapping [86]. They implemented the Arnold cat map on the host image, the resultant scrambled image is divided into 8×8 blocks, and the LSBs of the pixels within these blocks are embedded with the watermark. However, before embedding, an encrypted version of a binary watermark is prepared using an XOR operation between the watermark image and a chaotic sequence obtained using a logistic map. Undoubtedly, this approach makes removing the watermark by a hacker highly unlikely, but the major drawback is that the employed binary logo watermark is foreign.

On the other hand, Chang et al.'s method uses a self-generated watermark in their fragile watermarking scheme [13]. The approach also utilises a novel two-pass logistic map along with Hamming code. Their method exhibits excellent tamper detection and localisation abilities, shown via the *FNR* and *FPR* performances of 0.07 % and 0.43 %, respectively. Above

all, their approach proved that the VQ attack could be nullified even without inter-block dependency. The main shortfall of the method is that it is operable only on greyscale images. Motivated by [13] and [86], Prasad et al. in 2020 presented their work on fragile watermarking in [78]. Their approach generated the authentication code by combining MSBs and Hamming code. The generated code is further encrypted by using a logistic map. Subsequently, the encrypted code is embedded into the LSBs using a novel block-level pixel adjustment process (BPAP). Prasad et al.'s approach achieves high tamper detection and localisation while maintaining the required visual quality of watermarked images. The reported *FNR, FPR,* and *ACC* are 0.08%, 1.45%, and 99.89%, respectively. Another study by Prasad et al. in late 2020 presents an active forgery detection scheme using fragile watermarking, which works at the pixel level [77]. The watermark preparation and embedding procedures in this method are very similar to their predecessor work in [78]; however, the main difference is that the predecessor method is implemented at the block level, whereas the other is at the pixel level. To this end, this scheme's tamper detection precision is higher than the previous method [78], whereas it lacks tamper localisation ability. To this end, the *FNR, FPR,* and *ACC* values reported by [77] are 0.45%, 0.01%, and 99.71%, respectively.

5.2.2 Semi-fragile and other attributes-based methods in the spatial domain

Several semi-fragile watermarking methods exist in the spatial domain but not as many as in the other domains. Some of the most influential semi-fragile watermarking works within the spatial domain are discussed here.

Schlauweg et al.'s semi-fragile watermarking utilises a self-generated watermark [91]. Firstly, the host image is processed using lattice quantization to generate the watermark data, which is then encrypted using the MD5 hash algorithm. The encrypted watermark is subsequently embedded using the novel dither modulation-based approach and error correction coding (ECC). The method performs well when exposed to desirable manipulations such as JPEG compression but fails to provide soft authentication to other non-malicious attacks, such as rotation.

Xiao and Wang proposed a scheme tailored to accommodate the sharpening attack [123]. The method has a direct use case as image sharpening is a commonly used image modification. To this end, Laplacian sharpening, or sharpening in general, is used for edge enhancement in images without altering the actual (image) content. Xiao and Wang argued that their method could withstand Laplacian sharpening to any degree and distinguish it from other attacks. Their proposed algorithm is low in time complexity and high in watermark imperceptibility because only the LSB value of pixels is altered. Moreover, the watermark is embedded by modifying the parity of the pixel value and its Laplacian sharpening result, making it tolerant to the Laplacian sharpening but fragile to other attacks. Conversely, the method's main flaw is that it requires an external or foreign watermark and cannot achieve tamper detection and localisation.

Local binary pattern (LBP) based watermarking is another widely employed technique [118]. The LBP can be perceived as a particular case of the LSB substitution; however, the main difference is in their applicability. For instance, the LBP is mainly used in semi-fragile watermarking, whereas the LSB serves hard-fragile watermarking. As mentioned earlier, the watermark can withstand specifically authorised modifications in semi-fragile watermarking. To this end, as the LBP-based watermarking is immune to luminosity changes, it is an excellent candidate for scenarios wherein the watermark must withstand watermarking attacks, such as CE, brightness, HE, and gamma correction.

An illustration of how to calculate the LBP from a pixel block is shown in Fig. 16. Here, N_P and C_P stand for the neighbouring pixel(s) and the center pixel, respectively. Note that there are several ways via which the C_P (represented using red ink in Fig. 16) can be calculated, and peers in the field are actively working on finding novel ways to improvise its selection. However, for explanation simplicity, $C_P = 158$ is selected for illustration in Figs. 16 and 17.

Once the LBP is obtained, it can be embedded using the steps outlined in Fig. 17. There are a lot of commonalities between these steps and those related to the LSB substitution-based methods (shown above in Fig. 15). However, in Fig. 17, the authors have deliberately demonstrated an example of the LBP-based watermarking wherein the employed watermark is self-generated. In other words, the watermark itself is generated from the host image. This approach has many benefits, such as improving tamper detection and localisation capabilities.

Wenyin and Shih presented the LBP-based semi-fragile watermarking in [118]. It was a breakthrough work that emphasised using the LBP for semi-fragile watermarking. Their proposed work starts with single-level watermarking, wherein a logo-based watermark is embedded using the LBP. In the beginning, the study shows the working using the LBP that is 3×3 in block size. Subsequently, the working is also demonstrated using LBPs of other dimensions, for instance, 5×5 or bigger. The results achieved by their approach revealed that the proposed scheme is robust against readily used image manipulations, such as additive noise, luminance change, and contrast adjustment. At the same time, the method is fragile against other attacks, such as filtering, translation, and cropping. To this end, the technique exhibits tamper detection and localisation abilities against unentertained attacks. The scheme's success positively influenced many later semi-fragile watermarking works, such as [14, 127] and [128]; however, a significant flaw is common in these methods. Specifically, these methods employ LBPs that are odd in pixel numbers or dimensions, for instance, 3×3, 5×5, and more-resulting in an issue when dealing with a host image whose dimensions are in powers of two. Above all, these methods are operable only on greyscale images. These shortfalls are addressed by Pal et al. in a series of their works [71, 72] and [73]. As mentioned earlier, a higher ratio of the semi-fragile watermarking methods exists in other domains; hence, the rest of the semi-fragile works are discussed later in this review.

5.2.3 Robust and other attributes-based methods in the spatial domain

The transform domain is generally preferred over the spatial domain when the focus is robust watermarking. That said, some spatial domain-based robust watermarking works still have left their mark. A few of those are summarised below.

In contrast to the spatial domain-based fragile watermarking methods, which primarily tend to employ the LSBs during embedding, the robust watermarking techniques prefer using the ISBs. If robustness is the main requirement, embedding into the MSBs may seem perfect, but such is not the case. The MSB-based embedding significantly degrades the image quality,

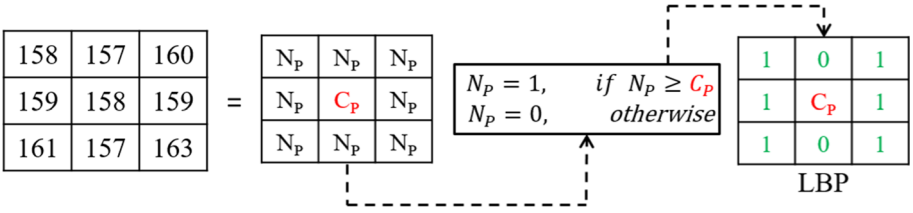

Fig. 16 An illustration showing the generation of the LBP. Best viewed when zoomed in

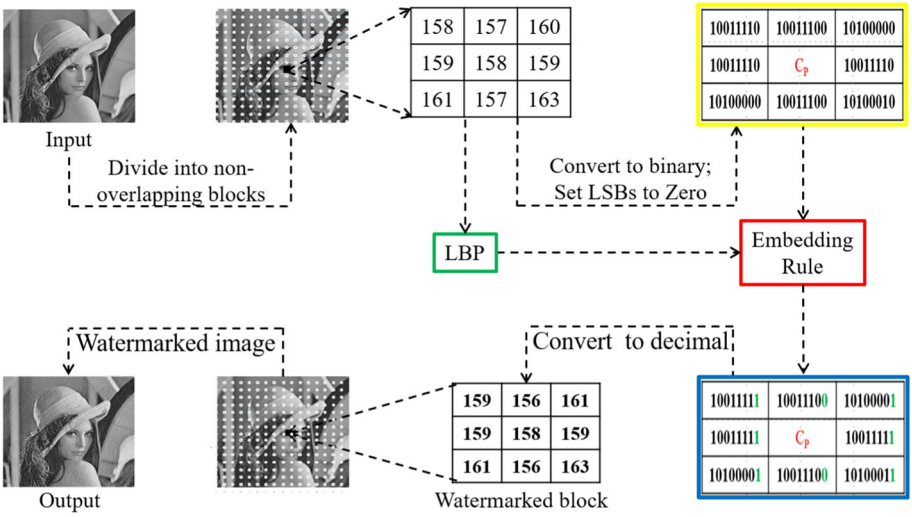

Fig. 17 A working illustration of the LBP-based watermarking. In this illustration, the watermark is self-generated

compromising the vital balance between the watermarking requirements of imperceptibility and robustness. Hence the use of ISBs in achieving spatial domain-based robust watermarking is justified.

Parah et al. in 2017 proposed an ISB-based robust watermarking scheme for grayscale images [75]. Their strategy is blind and demonstrates how a watermarking scheme's robustness changes when a watermark is embedded using the ISBs instead of the LSBs. Moreover, the method employs a foreign binary logo watermark encrypted using a pseudo-random address vector (PAV). Details on PAV are present in Parah et al.'s other work in [74]. In [75], the method's robustness is tested through some commonly used watermarking attacks, such as histogram equalisation (HE), median filtering, low pass filtering (LPF), JPEG compression, GN, salt and pepper (S&P) noise, and rotation but not against other readily used attacks such as cropping and scaling. Abraham and Paul presented their work in 2019 to address this shortfall [1]. Their non-blind approach achieves watermarking in colour images, wherein only the blue channel is employed during watermark embedding. That is because the HVS is less sensitive to changes to the blue channel than to the red and green channels. The method in [1] utilises a block-based approach, wherein each block is exposed to a sub-region selection process using a simple image region detector (SIRD) before the watermark embedding. SIRD facilitates the selection of the most appropriate region or sub-region for watermark embedding within an 8×8 pixel block. Selected pixels in a sub-region are subsequently modified to achieve watermark embedding. Moreover, two embedding masks M1 and M2, are used during the embedding process. M1 modulates or adjusts the blue channel with respect to the watermark bit, and M2 is the compensating mask that changes red and green color channels in response to the blue channel's modulation. In a nutshell, M1 and M2 masks aim to maintain the balance between imperceptibility and robustness.

The experimental results of the method in [1] show its robustness against several geometric and non-geometric attacks. Many complex watermarking attacks are also addressed, including cropping, resizing, and flipping. However, the robustness evaluation does not cover the scheme's effectiveness against simultaneously occurring multiple attacks or a combination of

attacks. Hasan et al., in 2021, presented one of the most recent works on the ISB-based robust watermarking [32]. Considering this method caters only to greyscale images, the authors emphasised using the host image's black pixels for watermark embedding. This technique balances watermarking requirements by employing the third ISB plane of the black pixels (the third ISB from the right in Fig. 14) and Pascal's triangle during the embedding process. To this end, Pascal's triangle selects the most suitable black pixels for embedding by achieving a minimum trade-off between imperceptibility and robustness. The study's experimental analysis has proved that embedding using black pixels instead of white results in better *PSNR* and *NCC* performances. Moreover, the scheme's $\mathcal{O}(n^2)$ time complexity is low enough to be adopted for real-time applications.

Histogram shifting is another widely accepted watermarking scheme. It was devised by Ni et al. [69] in 2006, and since then, it has been vastly employed. The main advantage of the technique is that it produces watermarked images with superb imperceptibility. To put into perspective, the average *PSNR* value of the watermarked images achieved by Ni et al.'s method is at least 48 dB, which was higher or on par with any other existing method(s) at that time. On the flip side, such a great imperceptibility came at the price of low capacity. A general representation of the histogram shifting is shown in Fig. 18. Here, the highest point (with respect to the *y-axis*) within the histogram is termed as the *peak point*. In other words, the peak point depicts the most frequently occurring greyscale value within the host image. In contrast, it is also well established that there is always an absence of a grey level (sometimes more than one) in a natural image. The figure defines such an absent grey level as the *zero point*. In the literature, grey levels within a histogram are also referred to as *bins* [69].

A step-wise breakdown of the histogram shifting technique's methodology is given in Fig. 19. In this scheme, the histogram of the host image is plotted at first. Subsequently, the peak and the zero points are located. After that, the greyscale values between the peak and zero points are shifted to create a gap next to the peak point. In other words, this shifting can be perceived as the zero point's shifting from its initial greyscale position to the one next to the peak point's. In Fig. 18, the grey levels are shifted to the right as the zero point in the original histogram is located on the peak point's right. Finally, all the pixels corresponding to the peak point's grey level are located, and watermark bits are embedded using an embedding rule. Several embedding rules have been devised since Ni et al.'s method in 2006; however, a straightforward version is expressed below.

Suppose pixels associated with the peak point have a greyscale value of 150. If the watermark bit to be embedded is 0, then no change is made. However, if the watermark bit to be embedded is 1, then a pixel with a value of 150 is incremented by 1, so it can be placed at the grey level of 151 in the (watermarked) histogram. These steps are repeated for other pixels

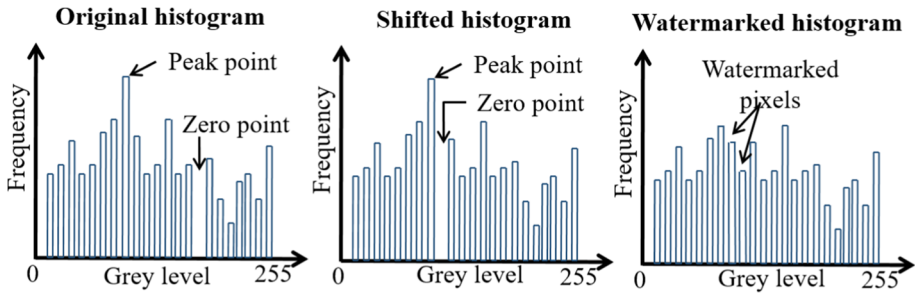

Fig. 18 Illustration of the histogram shifting technique

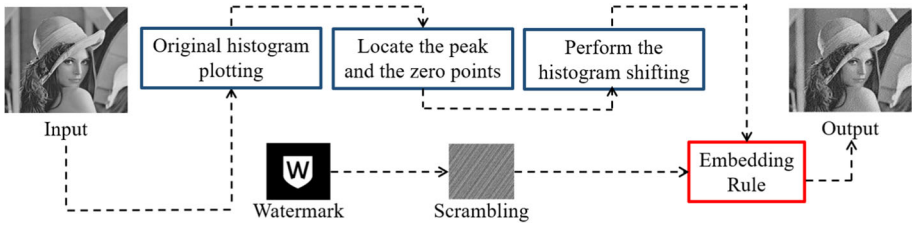

Fig. 19 A step-wise breakdown of the histogram shifting technique

(with 150 as their grey level) until all the watermark bits are embedded and the watermarked image is generated.

Ni et al.'s method motivated many later works, such as [34, 37, 44, 56] and [90], but these studies focused on reversible image watermarking. The reversibility attribute is exclusive to blind watermarking methods that allow the host image's reconstruction once the watermark is extracted from it. The study of these methods is out of this review's scope; however, readers may refer to Sreenivas et al.'s survey to understand the intricacies of reversible watermarking techniques [107]. Nonetheless, Ni et al.'s methods also inspired other histogram-based watermarking schemes, focusing mainly on robust watermarking. Most procedures are devised to tackle severe attacks such as cropping and random bending attacks (RBAs). Common examples of RBAs are global bending, high-frequency bending, and jittering. These attacks are responsible for causing de-synchronisation between the embedding and extraction processes, making the watermark extraction hard or sometimes impossible [122]. A few histogram-based methods developed to deal with such harsh attacks are discussed below.

Xiang et al. mentioned that geometric attacks, including RBAs, only shift pixels' position [122]. Consequently, they do not affect the histogram's shape as it is independent of the pixels' position but dependent on their grey levels. To this end, even after a geometric attack, the histogram's shape is barely modified; thereby, robustness is guaranteed. Moreover, this analogy is verified by Zong et al. in [129], wherein they compared histograms of unattacked images with those of geometrically attacked. Through this comparison, the authors illustrated that histograms hardly varied from each other. That said, in Xiang et al. method [122], the host image (I) is first exposed to the Gaussian low-pass filter because it allows for combating the high-frequency-based attacks. Subsequently, the yielded low-frequency image's (I_{Low}) mean value (A) is calculated, and the histogram is constructed. After that, the population of the pixels corresponding to the grey level of A is quantified, which also defines the length of the watermark or the number of bits that can be embedded. Subsequently, the watermark embedding is achieved based on the embedding rules, which tend to manipulate a pair of neighbouring bins or greyscales within the histogram. Readers are urged to refer to Xiang et al. 's study [122] for further insight into the workings of the relevant embedding and extraction procedures.

Xiang et al.'s method (mentioned above) gained a lot of attraction and has also been extensively used in the field. However, Zong et al. highlighted some of its flaws in [129]. The major weakness is its inability to use the histogram's shape to its fullest during the embedding process. This inability results in low watermarking capacity and, even worse, uncertain fluctuations within the embedding capacity. For instance, Zong et al., in their aforementioned study, proved that the watermark's length in Xiang et al.'s method is dependent on the population of the pixels with grey level corresponding to the mean value (A). Therefore, the lower the population, the lower the embedding capacity, and the lower the robustness. To this end,

Zong et al. tackled this issue by not letting only the mean value of I_{Low} dictate the embedding capacity but by employing as much of the histogram's shape as possible.

Similar to Xiang et al.'s approach, Zong et al. in their novel histogram-based watermarking method [130], employ a Gaussian low-pass filter to preprocess the host image. Moreover, the watermark bits are embedded only into the low-frequency components of the filtered image to withstand various non-geometrical attacks. In addition, the geometrical manipulations, including the RBAs, are tackled using a technique called the histogram-shape-related index, which selects the most suitable pixel groups for watermark embedding. Consequently, a safe band is introduced between the selected and non-selected pixel groups, further suppressing the effects of geometric attacks. Moreover, during watermark embedding, a novel high-frequency component modification (HFCM) scheme is implemented to compensate for the side effects of Gaussian filtering. Even though the embedding rules in Zong et al.'s methods [129] and [130] are not much different from Xiang et al.'s work but the distinct features discussed in this paragraph are exclusive to Zong et al.'s approaches. Thanks to these unique features, Zong et al.'s methods have the excellent embedding capacity and exhibit robustness superiority over Xiang et al.'s approach.

Needless to say that the above-mentioned histogram-based watermarking methods are the backbone of the other existing histogram-based watermarking techniques. Readers are encouraged to explore relatively recent studies in [35, 55], and [62].

5.3 Transform domain-based methods

In the context of robust watermarking, the transform domain-based watermarking techniques are considered a better candidate than the spatial domain-based techniques. Several reasons justify this superiority; however, their immunity to geometric attacks is the main one. That is because the geometric attacks result in a direct altercation with the pixels, thereby damaging the watermark embedded in the spatial domain. However, in the transform domain-based methods, the watermark is embedded using the frequency coefficients, which are unlikely to be damaged via direct manipulation of the pixels [97, 100, 101]. Consequently, the transform domain-based methods are more resilient to attacks and suitable for robust watermarking. The most prominent and widely employed transform domain-based watermarking techniques are discussed below.

Firstly, discrete cosine transform (DCT) is a readily used technique in the transform domain. The general sequence of the steps involved in the DCT-based watermarking methods is given in Fig. 20. Here, the host image is first divided into 8×8 non-overlapping blocks. Subsequently, the DCT is carried on each block to yield the respective DCT coefficients. Based on frequencies, the DCT coefficients are categorised as low-frequency (LF), mid-frequency (MF), and high-frequency (HF). Moreover, the first low-frequency coefficient is the direct-current (DC) coefficient. To this end, these coefficients are depicted using different colour codes in Fig. 20.

The extracted DCT coefficients are exposed to a selection procedure that selects the suitable coefficients for the watermark embedding. In most existing DCT-based watermarking works, the MF coefficients are preferred for embedding the watermark. That is because the MF coefficients, unlike their counterparts (LF and HF coefficients), allow alterations while maintaining an appropriate balance between imperceptibility and robustness. A complete account of how the host image's behaviour changes when a watermark is embedded into different DCT coefficients can be found in [76]. The selected coefficients are then manipulated per an embedding rule to achieve the watermark embedding. Of course, embedding rules

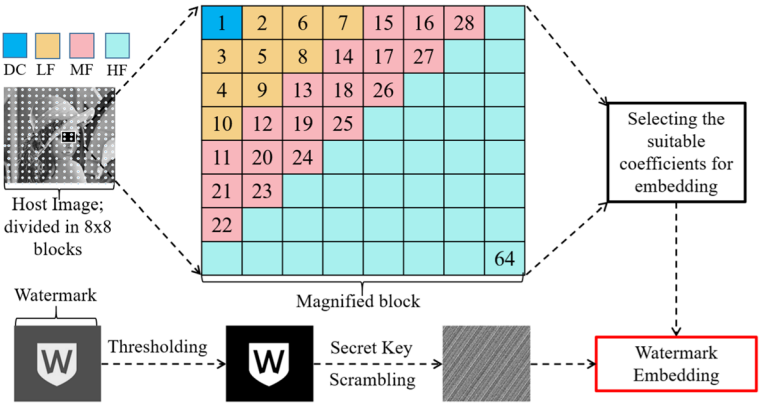

Fig. 20 General steps involved in a DCT-based watermarking method. Digits within the magnified block are the numbers allocated to the DCT coefficients, where DC, the lowest frequency component, is labeled as 1, and 64 is dedicated to the highest frequency component

vary from method to method, making them unique. Finally, the inverse of the DCT (IDCT) is performed, and the watermarked image is achieved.

Secondly, singular value decomposition (SVD) is another popular transform domain-based technique many in the field use. SVD is a numerical tool that decomposes a matrix into two orthogonal matrices and a diagonal matrix. If $\mathbf{I}$ is the matrix representation of I_{Host} and $\mathbf{I}$ is a real-valued matrix of $m \times n$ dimensions, i.e., $\mathbf{I} = \mathbb{R}^{m \times n}$, then its SVD is formulated as per (11):

$$SVD(\mathbf{I}) = \mathbf{USV}^T. \tag{11}$$

Here, $\mathbf{U} \in \mathbb{R}^{m \times m}$ and $\mathbf{V} \in \mathbb{R}^{n \times n}$ are two unitary or orthogonal matrices, referred to as the left and right singular matrices, respectively. These two matrices represent the geometrical features of I_{Host}. Moreover, T denotes the transpose operation, and $\mathbf{S} \in \mathbb{R}^{m \times n}$ is the diagonal matrix that contains the positive (non-negative) singular values of $\mathbf{I}$ in descending order. To this end, $\mathbf{S}$ controls the luminosity attribute of I_{Host}. The main advantages of employing SVD for image watermarking are below.

The first benefit is that the singular values in $\mathbf{S}$ are highly stable, and a (slight) change made to them generally goes unnoticed by the HVS. Hence, these values serve as an excellent candidate for achieving imperceptible watermarking. Another benefit is that whenever a data matrix is distorted, its element values are changed, but the singular values have little to no changes. These singular values withstand geometrical and non-geometrical attacks, making them suitable for robust watermark embedding.

The present SVD-based watermarking methods are divided into two categories. The first category is singular value matrix watermarking (SVMW), and the other is direct watermarking (DW). In the former's case, the singular values of the watermark (S_w) and the host image (S_H) are extracted and combined to create S_{new}. The S_{new} is subsequently combined with (U_H) and (V_H^T) to achieve the watermarked image ($I_{Watermarked}$). In the latter's case, only the S_H values are used and directly combined with the watermark (W) to create S_{new}. The SVD is subsequently performed on S_{new} to achieve S_{HNew}, which is then combined with (U_H) and (V_H^T) to achieve ($I_{Watermarked}$). This difference between the two is further highlighted using the figures below. Here, Fig. 21 shows the steps involved in the SVMW technique, whereas Fig. 22 is for the DW-based SVD approach.

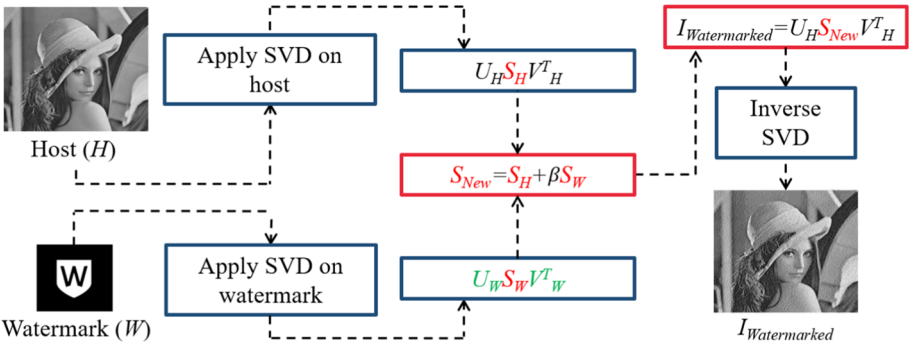

Fig. 21 Steps involved in the SVMW-based SVD approach. The key (s) or the side information required during the extraction phase is in green ink. Best viewed when zoomed-in

The benefits of using the SVD for image watermarking are evident from the discussion above; however, they come at a price. The primary issue amongst the SVD-based watermarking methods is the false-positive problem (FPP). This problem leads to an ambiguous situation where a hacker can obtain a counterfeit watermark and unlawfully obtain the rights to an image. For instance, the SVD-based techniques, such as SVMW and DW, tend to use the left and right singular matrices (shown using the green ink in Figs. 21 and 22) as the key(s) or side information during the extraction phase. Hackers understand that the diagonal singular values can be extracted from the left and right singular matrices. To this end, hackers use this significant limitation to gain access to the original watermark and then replace it with their own. By doing so, the adversaries can claim ownership of an image or a media in more general terms.

Thirdly, discrete wavelet transform (DWT) is another well-versed transform domain-based technique. Almost the whole image processing space has benefited from the arrival of DWT, and its advantages in achieving image watermarking are immense. A general step-by-step breakdown of DWT-based watermark embedding is shown in Fig. 23.

Here, the first step is to expose the host image to the DWT operation. Precisely, the DWT of an image yields four frequency subbands, termed and represented in Fig. 23 as low-low (LL), low-high (LH), high-low (HL), and high-high (HH). Note that the wavelet's ability to decompose an image is called multi-resolution analysis (MRA), via which the DWT

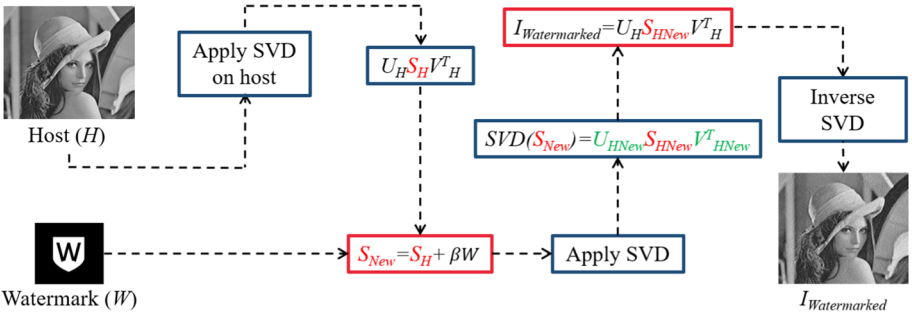

Fig. 22 Steps within the DW-based SVD approach. The key (s) or the side information required during the extraction phase is in green ink. Best viewed when zoomed-in

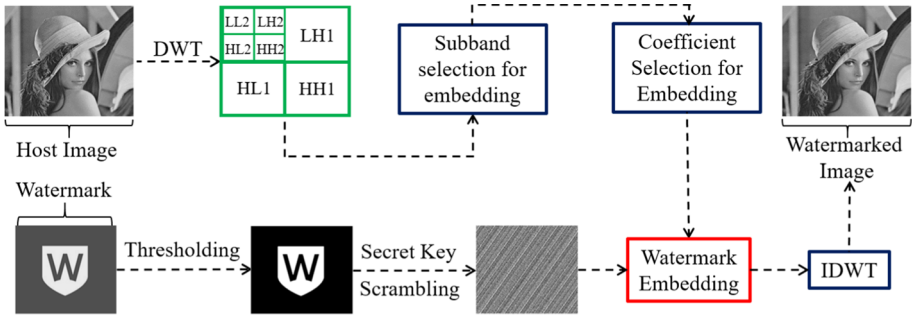

Fig. 23 General steps involved in a DWT-based watermarking method. In this illustration, 2-level DWT decomposition is performed, and the digits within green boundaries depict different subbands at different decomposition levels. Best viewed when zoomed-in

coefficients at different decomposition levels are extracted (see [18] and [94] to gain an insight into the MRA). It is well-known that the HVS is more receptive to low-frequency modulations; as the LL subband comprises the low-frequency DWT coefficients, it is generally considered unfit for watermark embedding. Similarly, the HH subband contains high-frequency coefficients, which can easily be oppressed by the usual watermarking attacks, such as compression and high-pass filtering, rendering it unsuitable for embedding. So choosing an appropriate subband(s) in the DWT-based watermarking is an essential step in the embedding procedure. Subsequently, the DWT coefficients within the selected subband(s) are determined and successively manipulated to achieve the actual watermark embedding. Needless to say, the embedding itself is reached via defined embedding rules (red box in Fig. 23). The final step is to perform the DWT inverse (IDWT) to achieve the final watermarked image.

5.3.1 Fragile and other attributes-based methods in the transform domain

Fridrich and Goljan have left their mark on DCT-based fragile watermarking through their work in [24]. A step-by-step breakdown is necessary because this pioneering work is heavily cited. In the first step, pixels in the host image are stripped of their LSBs, i.e., set to zero. Subsequently, these stripped-off pixels are divided into 8×8 blocks, and each block is exposed to the DCT operation. The extracted DCT coefficients are then subject to a quantization procedure with the help of the quantization table (see [24] for an insight into the actual quantization table). The first 11 of the total quantized DCT coefficients are converted into 64 bits using binary encoding. To this end, the 64 bits are self-generated watermark bits. Fridrich and Goljan, in their study, proved that binary encoding of the first 11 coefficients guarantees a binary sequence of 64 bits.

Moreover, they have also shown that only these 11 coefficients are sufficient to represent information of an 8×8 block, even though they are compressed to 50% via JPEG compression. Subsequently, these 64 bits are inserted into the LSBs of the pixels belonging to an 8×8 block. Note that in this technique, 64 bits of one block are embedded into the LSBs of another block by using the concept of block mapping. That is vital, less so for the tamper detection and localisation, but more for the reconstruction or recovery of the tampered regions. For instance, if a block is tampered with, its recovery information (64 bits) can be found in the other block's LSBs, which can be extracted and employed to reconstruct the tampered region(s). Even though the quality of the restored blocks is lower than 50%, it is sufficient to inform the user about the original content.

Furthermore, the tamper detection and localisation capabilities of the method in [24] are satisfactory, thanks to 8×8 blocks employed as authenticators. In the same study, Fridrich and Goljan proposed improving the quality of the restored regions by utilising two LSBs (the LSB and an ISB) to hide 128 bits, but this comes at the price of a poor-quality watermarked image. Notwithstanding the successes of the method, it is vulnerable to standard attacks, such as setting all LSB bits to zeros. This issue is addressed by Li et al. in [53].

The initial steps in Li et al. method [53] are similar to the ones in [24], but the 64-bit watermark used in the former's case is achieved from 14 of the total quantised DCT coefficients instead of 11. Once selected, 64 bits are embedded using a novel block-mapping-based approach, termed by the authors as the dual-redundant-ring structure [53]. Unlike in [24], the novel technique allows 8×8 blocks of the host image to form a cycle, wherein watermark bits of the 1^{st} block are hidden into the LSBs of its adjacent block. Subsequently, a copy of the watermark bits is also embedded into the ISBs of another block whose position is dictated by the 1^{st} block. The approach results in a ring-like formation, wherein block dependency (dependence of a block's information on the other) is achieved. Moreover, the existence of multiple copies of a watermark bit elevates the survival chances of the watermark, which leads to an improvement in tamper detection and localisation performances. Needless to say, that block dependency helps in combating the VQ and collage attacks. However, the method is only applicable to greyscale images, and the fixed block size of 8×8 is also responsible for increasing the *FPR*, which is undesirable.

Singh et al. in late 2015, proposed a self-generated watermark-based fragile watermarking scheme [106]. Their strategy is blind and implemented block-wise, wherein the size of each block is 2×2. Each of the four pixels in a 2×2 block is stripped off its three far-right bits (LSB and two ISBs), whereas the remaining five bits (the MSB and four ISBs) are utilised to achieve a self-generated watermark. The generated watermark is a blend of authentication and multitasking bits. Authentication bits verify the host image and provide tamper detection and localisation characteristics. In contrast, multitasking bits can perform the function of the authentication bits but also carry the information required to restore or recover tampered regions. To this end, the primary purpose of multitasking bits is restoration, whereas their authentication ability is generally used as a backup in case an attack damages actual authentication bits. Note that five bits (MSB and four ISBs) of each pixel in a 2×2 block are used to generate multitasking and authentication bits in the following manner.

Firstly, ten of the total multitasking bits are generated using a combination of the DCT and quantization techniques. By the way, the same combination is used in the above-mentioned methods [24] and [53]. Subsequently, the first of two authentication bits is produced by combining five bits (the MSB and four ISBs) and a cyclic redundancy check (CRC) bit with the help of a key. The second authentication bit is achieved by combining five bits (the MSB and four ISBs) and the Hamming code using another key. Readers are encouraged to refer to [106] for the intricacies of the CRC and Hamming code and how they are gelled with five bits (the MSB and four ISBs) to produce two authentication bits. Once 12 watermark bits (ten multitasking and two authentication bits) are generated from a 2×2 block, they are embedded into pixels of another 2×2 block, selected with the help of a block-mapping procedure. Similar steps are executed on the remaining 2×2 blocks to achieve the final watermarked image. When operating on color images, the system is performed on one of the red, green, and blue (RGB) channels and replicated to the other two channels. Subsequently, the processed channels are concatenated to achieve the final (colour) watermarked image.

Singh et al.'s method has excellent fragility and sensitivity to even small changes, which is desirable [106]. Furthermore, the tamper detection and localisation performances are high due to small-sized blocks. However, the method has a significant flaw in using multiple keys;

specifically, six such keys are used throughout the process. Since these keys must be shared with the receiver through the transmission channel, it poses a severe threat of being hacked. Not to mention the overhead imposed on the overall processing time by multiple keys is also significant.

Singh et al. presented another DCT-based fragile watermarking technique in late 2016 [105]. This scheme can be perceived as an extension of their previous work in [106]. To this end, when it comes to embedding and extraction procedures, the new method follows the footsteps of its predecessor. However, the main difference is that the practice employs three secret keys instead of six. That makes the new process more balanced from its application viewpoint; however, the three keys are still far too many.

Dadkhah et al. in [17] proposed an SVD-based fragile watermarking scheme capable of tamper detection and localisation. The scheme begins by dividing the host image into 4×4 blocks, and each block is further subdivided into four blocks, each of which is 2×2 in size. Subsequently, the SVD is performed, and a 3-bit authentication code from each sub-block is generated. After concatenation, a 12-bit authentication code is achieved from four sub-blocks. Furthermore, the average value of these sub-blocks is calculated, and the first five far-left bits (MSB and four ISBs) are extracted and concatenated, resulting in a 20-bit block recovery code. In the next step, previously achieved 12-bit authentication data is placed within the LSBs of the pixels belonging to the aforementioned sub-blocks. However, the 20-bit recovery data is placed in ISBs of pixels in a different 4×4 block. To this end, the different 4×4 block is found using the mapping operation.

The main strength of Dadkhah et al.'s method is its ability to use pixel blocks (in the form of blocks and sub-blocks) of different sizes. This difference implements tamper detection and localisation hierarchically. For instance, the first set of tamper detection results is achieved using a bigger block (4×4 in size), which can then be fine-tuned using smaller blocks. Moreover, as there are more blocks to play with, it also increases the overall watermarking capacity, which helps avoid collage attacks and aids in the recovery of tampered regions. In contrast, the advantages of this study are challenged in [8].

Benrhouma et al. in [8] highlighted several flaws associated with Dadkhah et al.'s method. The most significant drawback is the false alarm problem, i.e., recognising tampering even when there is none. This issue arises because the singular values are very sensitive to changes. In Dadkhah et al.'s method, the singular values are calculated before the pixels are stripped of their LSBs. In this case, if the watermark bits are embedded into the LSBs, it is almost assumed that the produced singular values would be different from the first calculated ones-consequently giving rise to the false alarm issue. Thereby, Benrhouma et al. proposed that this issue can be fixed by stripping off the LSBs before the singular values are calculated. Moreover, they also clarified that the 20-bit recovery information should be embedded into a 4×4 mapped block using a combination of the LSBs and ISBs. This combination contradicts Dadkhah et al.'s method, as they only use ISBs of the mapped block for embedding the recovery bits. Readers may refer to these studies for further details on the highlighted differences. However, a significant shortfall shared by methods [8] and [17] is that they do not operate on colour images, which are often used nowadays.

In 2022, Neena and Shreelekshmi proposed a fragile watermarking scheme for tamper localisation in images [70]. The method is blind and uses logistic mapping and SVD. The host image in the proposed approach is divided into non-overlapping 2×2 pixel blocks, and eight watermark bits are generated from each block. To this end, in a block, pixels' six bits (MSB and five ISBs) are extracted and permuted using the logistic map, followed by an SVD operation. This combination generates eight watermark bits, further exposed to an encryption operation. Subsequently, these watermark bits are placed in the LSBs and ISBs (two rightmost

bits) of pixels in a block, and the other six bits are left unchanged, thereby achieving the watermarked block. The same steps are executed for the remaining 2×2 blocks, and the final fragile watermarked image with tamper detection and localisation abilities is gained. The method's experimental simulations show its sensitivity towards several severe attacks, such as copy-paste, content removal, text addition, noise addition, vector quantization, collage, content only, and constant feature. Moreover, compared to other state-of-the-art methods, the approach has shown improved precision and accuracy in tamper detection and localisation. The main drawback of the scheme is its inability to restore the affected areas. Furthermore, Neena and Shreelekshmi have claimed that the method is operable on greyscale and colour (RGB space) images. Still, the working illustrations on colour images are absent. Similarly, the execution time analysis or time complexity is not presented.

In 2016, Nguyen et al. proposed a DWT-based blind watermarking scheme [68]. In addition to providing authentication, the method can detect tampering and localise the affected area. In this scheme, the host image is divided into 8×8 blocks, and each block is subject to a two-level DWT decomposition. Subsequently, the extracted DWT coefficients from the second-level subbands (*LL2, HL2*, and *LH2* in Fig. 23) are further divided into 2×2 blocks, which are then employed for watermark embedding. To this end, the watermark is generated by a secret key, ensuring that its length is equal to the number of the 2×2 blocks employed during embedding. Moreover, in a 2×2 bock, two coefficients are embedded with authentication bits and the third coefficient with a recovery bit. To clarify, the generated watermark comprises authentication and recovery bits; readers may refer to Nguyen et al.'s study to understand how these watermark bits are formulated. That said, three watermark bits are embedded in an 8×8 block. Subsequently, the rest of the 8×8 blocks are processed using similar steps, and finally, the inverse of the DWT yields the watermarked image. The method has good tamper detection and localisation performances, and the quality of recovered or restored images was also on par with the state-of-the-art techniques at the time. However, the study fails to explain why the image is decomposed to the *second-level* using the DWT. In other words, why is the first, third, or any other decomposition level not chosen?

In 2018, Wang et al. proposed a fragile watermarking scheme based on LBP and DWT [117]. Firstly, the host image's pixels are stripped off their LSBs, resulting in a stripped-off host image, which is then exposed to single-level DWT. Secondly, the DWT coefficients are divided into 3×3 blocks, utilised to produce the LBP. The LBP's binary bits are encrypted using a logistic map, and then a pseudo-random sequence is achieved. This sequence is arranged to form a chaotic image that is resized to a quarter of the host image. Note that this resized chaotic image is a self-generated watermark in itself. After that, the initially stripped-off image is divided into 2×2 non-overlapping blocks because the self-generated watermark is one-fourth of the host image. In the next step, the maximum valued pixel in a 2×2 block is selected, and its LSB is replaced with a watermark bit from a self-generated watermark. Similarly, the remaining blocks are watermarked, culminating in the watermarking process. The experimental analysis of the method has proven its ability to withstand several watermarking attacks, such as content removal, collage, and content-only. The *PSNR* evaluation of watermarked images produced by the method has shown its superiority over the aforementioned methods [8] and [86]. In contrast, the method's limitation lies in using a conventional LBP operator, which cripples the technique in processing the image edges. To this end, if the tampering is in one such region, it goes undetected.

Thanki et al. proposed a non-blind and fragile watermarking scheme for colour images, wherein they employed a combination of finite ridgelet transform (FRT) and DWT [108]. Here, the authors have claimed that this combination can achieve better watermarking capacity and imperceptibility than techniques solely based on DWT. Another distinct feature of

the method is that the employed watermark is also a coloured logo. Initially, the host image is divided into RGB channels, each exposed to FRT to produce the FRT coefficients. Subsequently, DWT is performed on the FRT coefficients of each channel, and the resultant LL subband is used for watermark embedding. To this end, the coloured watermark is split into RGB channels, and each is scrambled using the Arnold encryption. After that, scrambled watermark bits of a colour channel are embedded into the LL subband of the corresponding colour channel associated with the host image. Once the embedding process is over, the final watermarked image is achieved by executing inverses of DWT and FRT. In the experimental section, the fragility of the watermark is tested against several geometrical and non-geometrical attacks. The *NCC* values of the extracted watermark are closer to zero, highlighting the fragility aspect of the watermark. Readers may refer to Thanki et al. study for an insight into actual embedding and extraction rules.

Notwithstanding the successes of Thanki et al.'s method, it does not provide tamper detection and localisation and is quite laborious in terms of processing time. Moreover, we believe using a coloured watermark is excessive and unnecessary, especially for fragile watermarking. A binary watermark is preferred in such a scheme due to its narrower bit-depth. To be precise, a binary watermark (only 2-bit in depth) can easily be embedded throughout the host image without affecting the image's quality. Moreover, such a spread is necessary for effective tamper detection and localisation.

5.3.2 Semi-fragile and other attributes-based methods in the transform domain

In 2013, Preda proposed a DWT-based semi-fragile watermarking method [80]. The method is blind and equipped to provide tamper detection and localisation. In the beginning, the host image is decomposed using DWT. The second-level DWT coefficients are extracted and concatenated to form a one-dimensional (1D) vector (C), which is then permuted or scrabbled using the secret key (K) to generate $C^{'}$. The coefficients in $C^{'}$ are divided into groups of length d. The total number of such groups dictates the watermark size, which is a binary random sequence generated using the above-mentioned secret key; K. Subsequently, watermark embedding is initiated group-wise, wherein the maximum valued coefficient in a group is embedded with a watermark bit so that the group's mean value remains unchanged. The same steps are repeated for the rest of the coefficients groups, and the watermark embedding is accomplished. Finally, the IDWT is taken, and the watermarked image is achieved. The experimental results of the method show its resistance to VQ, (mild to moderate) JPEG compression, and other non-malicious attacks. Preda has also demonstrated through simulations how *PSNR* and *BER* values change in response to changing decomposition levels of DWT. To this end, the performance of watermarked images is tested for DWT decomposition levels from one to three. Moreover, in terms of tamper detection and localisation, only the subjective results are shown in the study. Unfortunately, the study does not include the objective results using parameters such as *FPR, TPR*, and *ACC*. Furthermore, the method is only operable on greyscale images. Despite these shortfalls, the technique has more upsides than downs; therefore, it has been widely cited since its arrival.

In late 2015, Preda et al. published two studies back to back on semi-fragile watermarking, but this time they were based on DCT [79, 81]. Surprisingly, both studies' embedding and extraction procedures are almost identical, but the experimental results in [81] are more comprehensive than [79]. The host image is divided into 8×8 non-overlapping blocks in these methods. Each block is utilised to gain a self-generated watermark, achieved by XORing two components; a pseudo-random binary component developed with the help of a secret key and a block-dependent feature. The block-dependent feature protects the scheme against

cut-and-paste attacks like VQ and collage. Once the self-generated watermark is obtained, it is embedded into the host image using the following steps. Firstly an 8×8 block is exposed to DCT operation, and the DCT coefficients are extracted. Secondly, the DCT coefficients are quantised using a quantization matrix representing the JPEG compression's QF. Note that for illustration purposes, QF of 50 is selected in Preda et al.'s methods [79] and [81]. Subsequently, the quantised DCT coefficients (within the low to mid-frequency range except for the DC coefficient) are selected using a secret key during embedding. Once selected, the watermark is embedded by manipulating the selected DCT coefficients with the help of a modified quantisation index modulation (QMI) approach, given in [79]. The rest of the 8×8 blocks are processed using similar steps, and after that, the inverse of DCT (IDCT) is performed, and the final watermarked image is achieved.

Preda et al.'s methods [79] and [81] are resilient toward copy-paste and JPEG compression attacks. The average *PSNR* value of the produced watermarked images is > 40 dB. Moreover, tamper detection and localisation results are desirable. For instance, in most cases or attacks, the *FPR* values are nil, and the *FNR* values are close to zero. Despite all these successes, we wonder why the method is not tested using JPEG 2000 compression, considering it is a widely adapted compression strategy. Furthermore, in their studies, Preda et al. mentioned that the proposed strategy could be extended to colour images by using the luminosity component. However, it is unclear that the luminosity component of which colour space, i.e., YC_bC_r, *HSV, LUV, YUV*, etc. To this end, working illustrations on colour images are also missing; hence there is ambiguity on this front.

Qi and Xin proposed a DWT-based semi-fragile watermarking scheme that achieves image authentication and provides tamper localisation [83]. Here, the authors chose DWT over other transforms because of its contribution to the JPEG-2000 image coding standard. The scheme embeds a self-generated watermark into the low-frequency wavelet coefficients. Initially, the host image is divided into 4×4 non-overlapping blocks, which are utilised in the watermark generation using the Mersenne twister algorithm [63]. Subsequently, each 4×4 non-overlapping block is exposed to the DWT operation, and extracted low-frequency coefficients in the LL subband are selected for watermark embedding. This selection is made because most watermarking attacks easily affect the high-frequency coefficients. During embedding, the selected coefficient (X) is quantised, i.e., X divided by q obtains X_q, where q is the quantisation or threshold value and the X_q is the quantised value. Subsequently, the parity of X_q is calculated by dividing X_q by 2. If divisible by 2, the parity value is 0; otherwise, it is 1. To this end, as the employed watermark is also binary, therefore, if the embedded bit matches the parity value of 0, then X_q is manipulated to $X_q \times q$; otherwise, it is changed to $(X_q \times q) + q$. The method uses this strategy to embed a single watermark bit in a 4×4 block, and the total number of blocks needing manipulation is proportional to the watermark's length.

Qi and Xin's scheme uses a binary error map to achieve authentication and localisation [83]. To this end, the study does not explicitly mentions whether the used system is blind or non-blind, but as the binary error map generation requires the extracted and the original watermarks; therefore, in our opinion, the scheme is non-blind. The binary error map is generated block-wise by taking an absolute difference between the original and extracted watermark bits. The difference can be 1 or 0, where the former defines a tampered block, and the latter represents a non-tampered block. To this end, a tampered region is hierarchically subdivided into two classes. For instance, a 3×3 region is considered strongly tampered with; if four or more pixels are altered, else the tampering is mild. The method performs well against several non-malicious attacks, such as compression (JPEG and JPEG 2000), Gaussian LPF, median filtering, blurring, and S&P. However, the major flaw of this method

is in selecting the q value. Specifically, the selection is made empirically, and the selection procedure itself is non-adaptive, thereby giving rise to issues that may occur because of the manual thresholding [97, 101]. Moreover, the scheme is only operable on greyscale images and does not address non-malicious geometrical attacks. Lastly, the testing against various copy-paste attacks is also not documented.

In their subsequent work, Qi and Xin focused again on semi-fragile watermarking, but this time they used a combination of SVD and DWT [84]. Overall, this work's embedding and extraction strategies are similar to those in [83]; still, the following differences exist. Firstly, the self-generated watermark in [84] is achieved with the help of an XOR operation between the singular values (SVs) and the bits achieved via the Mersenne twister algorithm. Such a logic-based operation allows a self-generated watermark to have features dependent and independent of the host image's content, equipping the scheme with robustness and authentication attributes. Secondly, unlike their previous work, the quantisation strategy employed in this work is adaptive. Finally, a similar concept of binary error mapping from [83] is utilised in [84]; however, it's been further improvised in the latter method. For instance, the method in [83] uses only two authentication measures (*M1* and *M2*), whereas the scheme in [84] employs five such measures (*M1*-*M5*). The authors have claimed that more of these measures uplift the scheme's authentication ability and elevate its tamper detection and localisation performances. These improvements are justified through the scheme's experimental analysis.

In contrast to their strengths, two main issues exist within Qi and Xin's schemes [83] and [84]. The first is that these schemes use several secret keys, more than five, to be specific. To this end, the authors have not explained how these keys are managed and transmitted. The second is that the impact of employing these many keys remains undiscussed, specifically in terms of the overhead imposed on the overall processing time, which is a vital aspect in determining the real-time applicability of a scheme.

Ullah et al. proposed a DWT-DCT-based semi-fragile watermarking method in [112]. Their approach is blind, and one of the pioneering works wherein the properties of DWT and DCT are utilised to target non-malicious JPEG compression. The following steps are involved in achieving watermark embedding. Firstly, the approximate subband (*LL1*) reached via the host image's single-level DWT is utilised to achieve a self-generated watermark. To this end, the extracted DWT coefficients of the *LL1* subband are exposed to the DCT operation, and as a consequence, the DCT coefficients are born. Subsequently, these DCT coefficients are quantised, but unlike other aforementioned DCT-based methods, which tend to use quantisation tables, in Ullah et al.'s method, the quantisation is achieved via Huffman coding. As Huffman coding is a compression strategy, the quantised image here is a compressed image represented in a binary pattern. The binary bits from the compressed image are matched against the original but quantised DWT coefficients from the *LL1* subband with the help of an XOR operation, thereby yielding the self-generated watermark bits. Once all such watermark bits are generated, they are permuted using a secret key to obtain the ultimate watermark. Secondly, the DWT coefficients from the detail subbands (*LH1, HL1*, and *HH1*) are selected, concatenated, and divided into groups. After that, with the help of embedding rules outlined in Ullah et al.'s method, the eligible coefficient in each group is embedded with a watermark bit. The size of the coefficient group is inversely proportional to that of the watermark. i.e., the larger the watermark, the smaller the group size. Finally, each coefficient group is processed, and the watermark embedding is completed.

The experimental results of Ullah et al.'s method demonstrate its ability to distinguish between malicious and non-malicious tampering. Moreover, it is illustrated that the scheme's semi-fragility attribute can withstand JPEG compression and cut-and-paste attacks. To this

end, the method shows promising results in authenticating the transmitted image and localising the affected areas. In contrast, the scheme's operability is demonstrated only on greyscale images, and its extension to colour images is unclear. In the case of cut-and-paste attacks, the cut-off tampering percentage beyond which the method's semi-fragility attribute can decide whether the tampering is malicious or non-malicious is not defined.

It is evident from the discussion above that image compression is one of the most, if not the most, targeted manipulation when designing a semi-fragile watermarking scheme. Most of the aforementioned schemes focus only on JPEG compression rather than its successor JPEG 2000. To this end, Rhayma et al. in 2021 presented semi-fragile watermarking that caters specifically to JPEG 2000 compression [88].

The scheme proposed in [88] is blind and uses a combination of DWT and QIM during watermark embedding. The scheme begins by decomposing the host image into five levels using DWT, whereby the *LL5* subband is extracted. Subsequently, the coefficients in *LL5* are exposed to a perceptual hash function (PHF) to gain a self-generated watermark. The beauty of PHF-based watermarking is that it primarily responds to geometric manipulations, such as cropping, collage, VQ, and others which directly alter the image's content during the authentication phase. In contrast, the acceptable changes, such as non-malicious compression and rotation, remain unnoticed. Hence, PHF is very much desired in semi-fragile watermarking. Finally, the generated watermark is embedded back into the *LL5* subband by carefully manipulating the coefficients with the help of embedding rules that follow the principles of QIM. Once all the watermark bits are embedded, the watermarked image is obtained through the IDWT operation.

At the receiver's end, the watermarked image is decomposed into five levels using the DWT, and after that, the watermark bits are extracted using the extraction rules given in [88]. Subsequently, the PHF is executed on the *LL5* subband's coefficients, and a hash is generated. The generated hash is matched against the extracted watermark bits, and a successful match authenticates the transmitted image. The method is one of its kind that has exclusively targeted JPEG 2000 compression and has shown promising watermarking results. Moreover, the embedded watermark is also tested against Gaussian noise and rotation attacks, which has proven to withstand the rotation attack as long as it is non-malicious. In contrast, it's not immune to the GN attack. Despite all the positives of the method, the authors have not explained why the host image is decomposed into five levels when performing DWT. It is well known that as the decomposition level increases, the capacity of the watermarking scheme decreases. To this end, the smaller the watermark, its ability to provide authentication is limited. These are some significant doubts associated with this work and require further investigation.

5.3.3 Robust and other attributes-based methods in the transform domain

In the past, Lin et al. developed a DWT coefficient difference-based robust watermarking scheme in [57]. At the time, the approach was the first of its kind, and since then, it has been widely adopted. There are a few critical steps in Lin et al.'s method; the breakdown is below.

Firstly, as the scheme aims to protect copyright, the employed watermark is foreign. The watermark, a binary logo, is encrypted using a secret key before embedding. Secondly, a greyscale host image is decomposed into three levels using DWT, and the *LH3* subband is selected for watermark embedding. The coefficients within the *LH3* subband are divided into non-overlapping groups of seven, i.e., vectors of length seven. Subsequently, the highest and the second-highest valued coefficients are located in each group, and the difference between them is calculated. Note that the number of groups selected is proportional to the length of the

watermark, i.e., is equal to the number of watermark bits. Once the difference is calculated, the highest and the second-highest valued coefficients within a group are quantised per the embedding rules outlined in [57]. Finally, the other groups are similarly processed, and the final watermarked image is achieved via the IDWT operation.

The watermark extraction in Lin et al.'s method is blind, and the extraction rules are formulated using entities such as the significant difference, thresholding, and the scaling factor. The technique is imperceptible and produces watermarked images with PNSR $>$ 40 dB. Moreover, the method is robust against geometrical and non-geometrical attacks. It can especially resist JPEG compression even with QF as low as 20. Still, the technique has some security-related concerns. In fact, Meerwald et al. dedicated a study wherein they pinpointed several flaws of Lin et al.'s method [64]. The shortfalls ignited the interest of many others in the field, resulting in several later works such as [43, 99, 114] and [126].

You et al. tailored their approach by considering that human eyes are insensitive to coefficients belonging to high-frequency DWT subbands [126]. They highlighted that adding watermark bits to these coefficients can hike the imperceptibility attribute of a watermarking scheme. To this end, they pivoted their approach around the idea that the ability to capture such coefficients is vital in designing wavelet-based watermarking. However, at the time, they were unhappy with the limited ability of the existing wavelet-based methods to capture coefficients only in three directions. Hence it motivated their study, wherein they constructed new wavelet filter banks that can capture coefficients more efficiently than tensor wavelets, such as Haar, Daubechies 4 (dB4), and Lin et al.'s bi-orthogonal 5.5. They devised filter banks that are built on the concepts of non-tensor product-based wavelet filter banks, explained thoroughly in [51]. That said, the embedding and extraction processes in You et al.'s method are closely related to the ones in Lin et al.'s.

In You et al.'s study, the performance of the proposed scheme is compared to Lin et al.'s method and others. The technique is superior to its counterparts in terms of imperceptibility and robustness. Moreover, its capacity is comparable to Lin et al.'s method. The experimental analysis also shows that the blindly extracted binary watermark resists geometrical and non-geometrical manipulations. In contrast to these successes, neither Lin et al. nor You et al.'s methods are operable on colour images. Moreover, in our opinion, if You et al.'s aim was to capture coefficients in more than three directions, why didn't they use other variations of the wavelet transforms that existed back then? For instance, the dual-tree complex wavelet transform (DTCWT) can capture coefficients in six directions, and the same is true for the shearlet transform. To this end, even though these transforms can capture coefficients in different directions. Still, this facility comes at the price of a significant overhead in processing time. Moreover, the processing time analysis or time complexity is absent in You et al.'s study, which leaves the readers in a dilemma of whether the proposed scheme is acceptable for real-time applications.

Another unclear aspect of methods in [57] and [126] is the selection of the non-overlapping coefficient blocks used for watermark embedding. For instance, if the watermark bits to be embedded are smaller than the number of non-overlapping coefficients blocks, which of the total blocks are employed for embedding, and how are they selected? Verma et al.'s method is another significant difference-based approach that has attempted to answer this question [114]. The initial steps (up to the selection of the *LH3* subband) are identical to the ones in Lin et al.'s method [57]. After that, the coefficients within the *LH3* subband are divided into 2×2 non-overlapping blocks. Subsequently, the difference between the two smallest coefficients in each 2×2 block is calculated, and these differences are then sorted. By the way, sorting can either be in ascending or descending order. Once sorted, the difference values are employed to select a threshold to differentiate the significant regions (SR) from the insignificant (ISR) ones.

Next, the watermark embedding procedure utilises non-overlapping blocks corresponding to the difference values within the SR. Once the embedding blocks are selected, the highest and the second-highest valued coefficients within each block are quantised using the same embedding rules outlined in [57]. Note that the number of watermark bits decides the number of coefficient blocks that are ultimately employed (for the watermark embedding) from within the SR. Once all the watermark bits are embedded, the final watermarked image is achieved by executing IDWT.

In their study, Verma et al. have shown that their proposed approach has superior imperceptibility and robustness to its counterparts. Moreover, they have included the processing time analysis in their research, demonstrating the potential usage of their method for real-time applications. There is no doubt that Verma et al.'s approach addresses some of the issues associated with [57] and [126]. However, their method still has the following ambiguities. First and foremost, the selection of thresholding value that differentiates the SR from the ISR is manual, i.e., empirically selected. Such a choice is undesirable as it requires a manual adjustment every time a new host image is introduced. Secondly, the method uses three separate secret keys to secure the overall scheme, whereas other techniques in [57] and [126] have used a single key throughout. To this end, the intricacies of dealing with multiple keys are not discussed within Verma et al.'s method. Finally, the technique is operable only on greyscale images, and a discussion on its potential extension to colour images is also absent.

Islam et al. further improvised Verma et al.'s approach in [43]. The main contribution of their work that sets them apart from their predecessors [57] and [114] is that they presented a practical methodology for selecting wavelet subbands. To this end, they demonstrated a change in the behaviour of the watermark when embedded in different wavelet subbands at different decomposition levels. This aspect of Islam et al.'s study has served many later wavelet-based watermarking methods and continues to do so. It gives the readers a clear indication of which subband(s) to employ to fulfil the watermarking requirement(s). Even though the watermark embedding and extraction rules in [43] are almost similar to the ones in [57] and [114]. Still, in [43], the employed blind extraction procedure has incorporated an extra step of utilising the support vector machine (SVM), a step missing in the counterpart methods.

Despite its aforementioned successes, Islam et al.'s study has the following downsides. Firstly, it fails to explain how the non-overlapping coefficient blocks used for watermark embedding are selected. Secondly, it suffers from the shortfall of manual thresholding and employing multiple keys. These same issues are also prevalent in approaches [57, 114] and [126]. Finally, using the SVM generally imposes an overhead in terms of the overall processing time, which remain undiscussed discussed in Islam et al.'s study. Hence, how can one ensure its suitability for real-time applications?

Sharma et al. in late 2020, proposed a novel signature-based watermarking scheme for identity protection [99]. The approach is motivated by the issues in methods [43, 57] and [114]. The technique uses significant difference-based watermarking, wherein embedding rules are inspired by the ones within [43, 57] and [114]. However, it differs from its inspirators in the following ways. Firstly, it presents a novel median-based embedding block selection procedure. This procedure is adaptive and selects the most suitable non-overlapping coefficient blocks (from the total blocks) for watermark embedding, thereby eradicating the fundamental problem in [43, 57], and [114]. Secondly, the method follows a non-blind watermark extraction, and as a result, the method is fast. No doubt there is a debate within the watermarking community regarding the blind and non-blind extraction procedures, i.e., which one is better?

In our opinion, the question is irrelevant as each is unique, and the application generally dictates whether the extraction needs to be blind or non-blind. For instance, vital documents

such as passports are physically verified at airports. Therefore, the host signal's (passport's) presence is essential for the watermark's extraction and verification. Similarly, the original currency note must be held against the light source so the hidden watermark can emerge. These are two examples of many watermarking applications wherein non-blind extraction is the only option. In contrast, there are many watermarking applications wherein the extraction has to be blind because the original signal is absent in such cases. For instance, in art, retailers use many authorised or legitimate (electronic) replicas of famous artworks. Art organisations often employ watermarking to keep the artwork's legitimacy intact; however, as it is infeasible to use the original artwork or signal for watermark extraction, the only option is to use blind extraction. To sum up, blind and non-blind extractions are unique in their ways, serve different purposes, and both have pros and cons. Thereby, it's not fair to compare the two.

In 2016, Parah et al. proposed a DCT-based robust watermarking strategy in [76]. Since its arrival, the approach has been widely accepted and cited, deserving a breakdown. Firstly, the method divides the host image into 8×8 non-overlapping blocks. Each block is then exposed to the DCT operation, and the DCT coefficients are yielded. Secondly, in a block, one of the nine *LF* coefficients closest to the *DC* coefficient (see Fig. 20) is determined. Similarly, a coefficient in the remaining 8×8 blocks is selected. Subsequently, the difference between the coefficients selected from the first and second (adjacent) blocks is calculated. Thirdly, based on the watermark bit (0 or 1), the calculated difference is manipulated by modifying one of the coefficients employed to calculate the difference in the first place. The coefficient modification within a block is achieved via the help of the scaling factor, the block's *DC* coefficient, and the median value of nine *LF* coefficients. The readers are encouraged to refer to Parah et al.'s study for an insight into the use of these parameters to modify selected DCT coefficient(s) [76]. The procedure is repeated for the rest of the 8×8 blocks or until all the watermark bits are embedded. Finally, the watermarked image is achieved via the IDCT operation.

Parah et al.'s method achieves imperceptibly watermarked images, whose robustness is tested against several geometrical and non-geometrical attacks. The blindly extracted logo watermark from attacked images shows resilience and immunity to several hybrid attacks. Another advantage of Parah et al.'s strategy is the ability to operate on grayscale and colour images. These qualities contributed to the method's wide acceptance via the later works such as Loan et al. and Hurrah et al.'s studies in [60] and [39]. However, Parah et al.'s technique lacks in the capacity aspect as only one watermark bit is embedded within an 8×8 block. To this end, if the host image is 512×512 in size, the maximum watermark that can be embedded is 64×64 in size. Another limitation of Parah et al.'s method is the ambiguity in selecting the vital parameters. For instance, one can observe in the study that the selection of the essential entities, such as the scaling factor and the threshold value(s), is unclear and non-adaptive. In other words, non-adaptive choices are manual adjustments that generally lead to several issues discussed within [97] and [101]. Last but not least, as Parah et al.'s study has not provided the processing time analysis; hence its potential to be used for a real-time application remains inconclusive. Similar issues persist within Loan et al.'s study in [60], but some are addressed within and Hurrah et al.'s work in [39].

In 2019, Hurrah et al. proposed a dual watermarking framework for privacy protection and content authentication of multimedia [39]. The study presents two watermarking schemes, Scheme 1 and Scheme 2; the former targets robust watermarking, whereas the latter targets hybrid or multipurpose watermarking. As the current section focuses on robust watermarking in the transform domain, Scheme 1 is expanded here. The main attribute that differentiates Hurrah et al.'s Scheme 1 from its inspirators (studies [60] and [76]) is the usage of the DWT-DCT-based combination. The effectiveness of such a combination is illustrated via the

robustness performance of the watermark in Hurrah et al.'s study, whereby it can withstand various watermarking attacks. Especially in the case of JPEG and JPEG 2000 compression attacks, the watermark extracted via Hurrah et al.'s method achieves higher *NCC* and *BER* values than those by [60] and [76]. The credit for this superiority goes to the DWT-DCT-based combination, as these tools are the founding blocks of image compression strategies. In addition to these advantages, Hurrah et al. have also covered the processing time analysis and presented the overall time taken by Scheme 1. To this end, the scheme is fast enough to be employed in real-time watermarking greyscale and colour images.

Notwithstanding the benefits offered by Hurrah et al.'s method [39], it suffers from the same ambiguity-based issues that exist within methods [60] and [76]. Specifically, the embedding and extraction rules employed in Hurrah et al.'s Scheme 1 are almost identical to the ones used by [60] and [76]. Therefore, selecting several vital parameters, such as the scaling factor and the threshold value(s), is unclear and non-adaptive. Hurrah et al.'s Scheme 1 also lacks watermarking capacity. In fact, its capacity is half of the methods in [60] and [76]. To this end, one may assume that the watermark's robustness must suffer with low capacity, but it is not valid for Hurrah et al.'s Scheme 1. For instance, despite having low watermarking capacity, Hurrah et al.'s Scheme 1 has better or higher watermarking robustness than counterpart methods [60] and [76]. The secret to this performance superiority is in employing the DWT-DCT-based combination.

Hurrah et al.'s study [39], along with many others [47, 97, 101, 104], prove that the disadvantages of methods based only on one technique (let it be DWT, DCT, SVD, etc.) are limited by combining them with other techniques. For instance, tools such as DCT, SVD, and machine learning (ML) produce watermarked images with excellent imperceptibility and robustness when combined with other transform domain-based processes. However, these combinations may also suffer from some flaws. For instance, machine or deep learning-based techniques require intense computation power, data, and training, making such procedures laborious. Hence, integrating multiple schemes into one is a cumbersome task. Based on their current performance and ability to achieve robust watermarking, Begum et al.'s review study has sorted the primary transform domain-based techniques as $DCT > SVD > DWT > DFT$ [7]. Such sorting, wherein DFT stands for discrete Fourier transform, can be used as a guide when an application aims for watermarking robustness. It must be acknowledged that the mentioned sorting guide is not a law or rule but is established empirically. However, in the author's opinion, it is helpful as it points the reader (especially a newcomer in the field) in a well-defined direction and allows them to employ or choose the best among the existing techniques.

In 2018, Kang et al. proposed a robust watermarking scheme wherein a combination of DWT, DCT, and SVD is employed [47]. Firstly, the host image is subject to the DWT operation, yielding an approximate and three detail subbands. Subsequently, the approximate (*LL*) subband is divided into non-overlapping blocks, each of which is 8×8 in size. Secondly, each non-overlapping bock is exposed to the DCT operation, and eight of the total *MF* coefficients are selected. The selected *MF* coefficients are then split into two groups, each comprising four coefficients. Thirdly, the SVD is performed on both coefficient groups, yielding a series of singular values that are then sorted (ascendingly or descendingly). Subsequently, the highest singular value in each coefficient group is selected and manipulated as per the watermark bit. For instance, if the watermark bit to embedded is one ($W_{em} = 1$), then the highest singular value in the first coefficient group is scaled up by multiplying with the watermark strength factor. The highest singular value in the other group is scaled down by dividing with the same watermark strength factor. In contrast, if the watermark bit to embedded is zero ($W_{em} = 0$), then the highest singular value in the first coefficient group is scaled down by dividing

it by the watermark strength factor. The highest singular value in the other group is scaled up by multiplying with the same watermark strength factor. The procedure is repeated for the remaining 8×8 non-overlapping blocks until all watermark bits are embedded. Finally, the watermarked image is achieved by executing inverse transformation operations in the following sequence of IDWT, IDCT, and ISVD.

Kang et al.'s approach has several benefits [47], but the major is in its ability to select the watermark strength parameter adaptively. The novel optimisation strategy that aims to achieve the balance between *PSNR, SSIM, NCC* and *BER* values is behind the adaptive selection procedure. To this end, as the watermark strength parameter contributes directly to watermark embedding, therefore the produced watermarked images are robust and imperceptible. Similarly, the blind extraction rules utilise the watermark strength parameter to achieve high watermark reconstruction. Another advantage of the adaptive selection process is limiting the side information requirement. For instance, in non-adaptive procedures, the manually chosen watermark strength parameter must be communicated or shared with the receiver to extract the watermark. To this end, apart from the watermarked signal, other information shared with the receiver is considered side information in watermarking. Generally, the limited the side information, the limited or brief the transmission, and the fewer the chances for a hacker to intercept. Hence, Kang et al.'s adaptive approach is streamlined and lighter in transmission than its counterparts [39, 43, 76, 114] and many others that are mentioned above.

In contrast to its discussed benefits, Kang et al.'s approach suffers from three main flaws [47]. Firstly, it's not operable on colour images. Secondly, the process is ill-equipped to deal with several geometrical attacks, such as rotation and translation. Lastly, the method's capacity is at the lower end when compared to its counterpart methods. Specifically, Kang et al.'s technique requires an 8×8 block to embed a single watermark bit. In other words, in a 512×512 host image, a watermark with a maximum size of 32×32 can be embedded. Despite the method's shortfalls, its ability to adaptively choose the watermark embedding strength parameter outweighs its inabilities. Thereby, Kang et al.'s study motivates several subsequent studies [50, 58], wherein selecting the watermark embedding strength parameter is adaptive. For instance, Liu et al. use the *fruit fly optimisation* algorithm (FOA) for selecting the watermark embedding strength parameter [58], and the same is chosen in Koley's study via the *adaptive alpha-beta blending* technique [50]. At this stage, one may ask for the best adaptive algorithm for choosing the watermark embedding strength parameter. In short, there is no such method that is the best. In other words, no way is better than the other, as each is unique and targets different issues. In our opinion, the suitability of the adaptive selection procedure is generally based on the techniques or tools employed during watermarking or the overall watermarking process. For instance, WSMN is an ML-based watermarking method that uses a non-dominated sorting genetic algorithm (NSGA-II) evolutionary algorithm to select the watermarking embedding strength parameter [29]. Similarly, other ML-based watermarking methods prefer to employ evolutionary algorithms because they tend to complement each other. To this end, getting into the intricacies of the existing adaptive selection procedures is beyond the scope of this discussion. However, if a method can adaptively choose the watermark embedding strength parameter, the readers are encouraged to perceive this ability as a highly favorable quality.

It is evident from the previous discussion that spatial domain-based methods generally serve the requirement of fragile watermarking. The transform domain-based methods are preferred for robust watermarking. In this instance, a question may arise: What happens when a watermarking scheme is expected to fulfill both requirements?

The answer to this question lies within hybrid watermarking methods or hybrid domain-based watermarking methods, as discussed in the section below.

5.4 Hybrid domain-based methods

In hybrid watermarking methods, more than one watermark (one robust and the other fragile) is embedded in the host image, generally via a combination of transform and spatial domains. Although such techniques are great as they can solve multiple issues simultaneously, they are laborious and contribute significantly to the implementation timing, both of which are undesirable for real-time applications.

Recently, hybrid domain-based methods have gained attention, and some prominent ones are discussed here. These methods are also known as multipurpose watermarking methods, as they can simultaneously address multiple issues, such as copyright protection and authentication. A general framework of a hybrid domain-based watermarking method is given in Fig. 24. A robust watermark is embedded in the first instance in hybrid methods, followed by a fragile watermark. However, the authentication and copyright checks (in the extraction phase) are independent and can be executed in any order. An insight into some of the selected hybrid watermarking methods is presented below.

Researchers Lu and Liao presented the first significant idea on multipurpose watermarking in 2001 [61]. The method employs DWT and embeds two distinct watermarks, one robust and the other fragile, via which multiple authentication and copyright protection goals are achieved. Because the study is a pioneer in the field, it gained much attention and is widely recognised. Moreover, the method produces imperceptible and secure watermarked images and has the advantage of operating on greyscale and colour images. In contrast, the scheme underperforms in tamper detection and localisation accuracy. Furthermore, the time-complexity analysis is absent in the study.

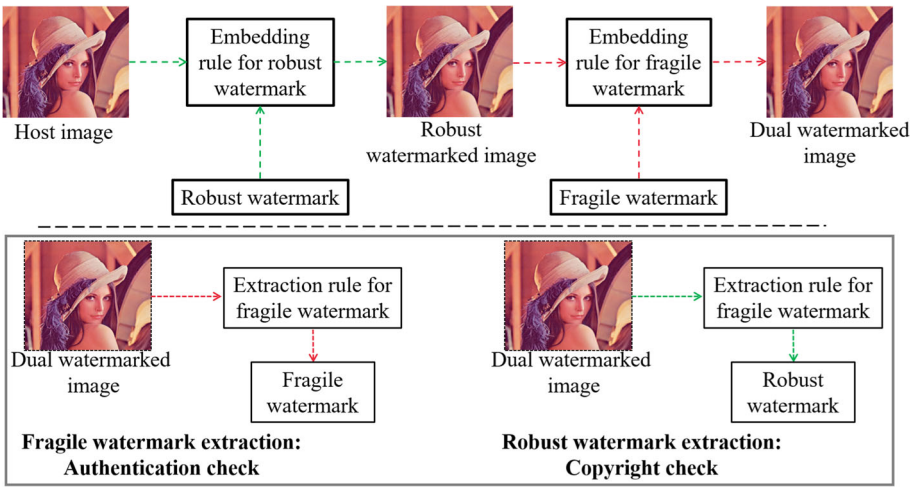

Fig. 24 A generalised blueprint of a hybrid domain-based watermarking method. The green arrows represent steps that deal with the robust watermark(s), whereas the red arrows are associated with those with the fragile watermark(s). The extraction phase is within solid grey borders. Note that the authentication and copyright checks are independent of each other and can be performed in any order. Best viewed when zoomed in

The tamper detection and localisation accuracy-related issues of [61] are tackled by Liu et al. in their study [59], wherein a multipurpose watermarking scheme for colour images is presented. Their methods exposed the Y channel of the YC_bC_r colour model to the DWT operation, whereby extracted low-frequency wavelet coefficients are manipulated to achieve robust watermark embedding. Subsequently, a robust watermarked image, acquired via the IDWT, is split into *RGB* channels. Using a novel LSB-based embedding method, each channel is embedded with a fragile watermark. The scheme produces imperceptible watermarked images which can provide copyright protection and authentication. Moreover, the tamper detection and localisation performances were on par (if not better) with other multipurpose watermarking methods that existed at the time of the scheme's arrival. Despite its several benefits, the approach lacks in the following aspects. Firstly and similar to [61], the study has not presented or covered the processing time analysis. Secondly, the system only works on colour images, not greyscale ones, raising doubts about its application versatility. Finally, as the method only employs the DWT, it suffers from well-known issues, such as aliasing [97, 101]. These issues are detrimental to the image reconstruction process, thereby affecting the watermark's imperceptibility in the watermarked image.

In 2019, Hurrah et al. presented a dual watermarking framework for privacy protection and multimedia content authentication in [39]. It is worth re-establishing that the study proposes two watermarking schemes: scheme 1 and scheme 2. The former is a robust watermarking scheme, and the latter is multipurpose; thereby, scheme 2 is more relevant to the discussion in this section and expanded upon here. Before embedding, Arnold transform-based encryption is used to scramble the logo watermarks, which are further secured using a novel encryption approach proposed within the study [39]. When embedding in a colour image, scheme 2 embeds one of the *RGB* channels with the robust watermark, and embedding itself is achieved via the combination of DWT and DCT. Subsequently, one of the remaining two channels is embedded with a fragile watermark in the spatial domain. Scheme 2 has tamper detection and localisation abilities and can maintain a vital balance between the watermarking requirements of robustness and imperceptibility. However, considering that the method uses only one of the three colour channels for fragile watermarking, it fails to explain how it safeguards the other two channels. In other words, if there is tampering, how does it verify whether the other two channels have been tampered with or not?

Hurrah et al.'s scheme 2 in [39] motivated Kamili et al.'s study [46], wherein a novel multipurpose watermarking scheme called DWFCAT is introduced. Firstly, in DWFCAT, chaotic and deoxyribonucleic acid (DNA) encryption techniques are employed to encrypt logo watermarks (one robust and the other fragile). Secondly, the DCT is performed on the Y channel, and the extracted DCT coefficients are manipulated to achieve robust watermark embedding. Subsequently, the C_b channel is divided into the non-overlapping (8×8) blocks in the spatial domain. A bit from the fragile or second watermark logo is embedded by replacing the LSB of a randomly selected pixel within a block. Note that a replica or copy of the already embedded (fragile) watermark bit is placed in another (randomly chosen) pixel's LSB in the same block. Here, the original bit is responsible for authentication, and the replica bit provides tamper detection and localisation. Finally, these steps are repeated, and a multipurpose watermarked image is achieved. Kamili et al.'s approach uses a single colour space; therefore, it is faster than Liu et al. and Hurrah et al.'s methods [39] and [59]. It also outshines the performance of Liu et al.'s method in the context of *PSNR*. In contrast, the scheme's tamper detection and localisation performances are not presented using well-known parameters, such as *FPR*, *FNR*, *TPR*, and *ACC*.

In 2020, Hurrah et al. came up with another multipurpose watermarking approach applicable to medical images [40]. The scheme is motivated by Kamili et al.'s method and is tailored

to operate within the spatial domain. Even though no robust watermarking is involved, the scheme is still multipurpose as it targets image authentication, tamper detection, and localisation, and restores tampered regions. Re-establishing that reversibility in watermarking allows the recovery and reconstruction of the areas affected by attacks. Notwithstanding that Hurrah et al.'s scheme fulfils multiple purposes, it lacks two main aspects. The first is that the method can not provide copyright protection; hence an image can be stolen. The second is its use of the foreign logo in fragile watermarking. As mentioned above, a foreign logo is undesirable for fragile watermarking because anything foreign corrupts the host image and ultimately degrades watermarking imperceptibility. To this end, most fragile watermarking schemes in the literature prefer using self-embedding watermarks, i.e., those generated from the host image [107]. Moreover, how does one apply a foreign logo-based watermarking if, in case, no such logo is available?

Most of the above-mentioned reversible watermarking methods suffer from a significant shortfall of having no backup for the recovery or digest information. To this end, the approach can no longer be reversible when the recovery information gets destroyed. This major limitation is tackled in Haghighi et al.'s study [30], wherein a self-embedding image watermarking scheme called TRLG is proposed for tamper detection, localisation, and recovery. The method produces four digest images via a combination of techniques, such as the lifting wavelet transform (LWT) and halftoning. Subsequently, the quality of the digest images is enhanced with the help of genetic algorithm (GA) optimisation. Multiple digest images provide numerous (four) chances to recover a tampered block. To this end, the Chebyshev system is used, which not only shuffles the watermark but also maps or correlates the shuffled information with (embedding) blocks. Moreover, other devised techniques, such as mirror-aside and partner-block, further improve the recovery of tampered regions. The method provides high image authentication and surpasses several state-of-the-art methods (such as Hurrah et al.'s scheme 2 in [39]) in tamper detection and localisation performances. Similarly, the technique produces imperceptible watermarked images, and an average *PSNR* value of 46 dB is obtained from test images. The only but significant limitation of Haghighi et al.'s method is its inability to provide copyright protection.

In late 2020, Haghighi et al. responded to the limitations of TRLG by proposing WSMN [29]. To this end, WSMN differs from TRLG because it uses robust and fragile watermarks to achieve copyright protection and authentication, whereas TRLG is reversible and only provides authentication. That said, WSMN and TRLG are multipurpose watermarking schemes, each capable of achieving tamper detection and localisation. WSMN is based on shearlet transform and employs smart algorithms such as multi-layer perception (MLP) and non-dominated sorting genetic algorithm (NSGA-II). In WSMN, quantisation and correlation techniques are used to watermark approximate and detail coefficients with robust and authentication watermarks, respectively. Moreover, suitable blocks for embedding are differentiated from non-suitable ones with the help of K-Means clustering, and the watermark embedding strength parameter is optimally selected via NSGA-II. Furthermore, WSMN's potential to withstand geometrical and non-geometrical attacks is elevated by MLP's learning ability, which also contributes to its high tamper detection and localisation performances. In other words, it makes the robust watermark(s) immune to several hybrid attacks and uplifts WSMN's ability to authenticate. In contrast, it impairs the processing time and lengthens the overall process. Last but not least, WSMN's inability to recover the tampered regions puts it on the back foot compared to techniques such as TRLG or other reversible methods.

In 2022, Sharma et al. developed a first-of-its-kind multipurpose watermarking scheme, wherein a single watermark achieves multiple goals of copyright protection and authentication [97]. The method's application versatility reflects in its ability to operate across various colour

spaces, such as greyscale, RGB, and YC_bC_r. Firstly in the approach, a binary logo watermark is encrypted with the help of the Fisher and Yates algorithm. Subsequently, the encrypted watermark's embedding into the host image is achieved within the transform domain using the DWT-DCT-based combination. Secondly, the spatial domain is also employed, wherein a novel concept of checkpointing is devised. A watermarked image is exposed to defined watermarking attacks during checkpointing. After each attack, the energy of the attacked watermarked image is calculated and stored in an array. Such an array in the study is termed as the energy vector (EV). Before the watermarked image is transmitted, the EV is shared with the receiver as the side information and a secret key. Note that here the secret key is the number of iterations used by the Fisher and Yates algorithm to shuffle the logo watermark. Once the receiver receives the watermarked image, its energy is calculated and matched against the energy values within the EV. The received watermarked image is deemed authentic if a match exists, triggering the extraction process. Otherwise, it is inauthentic, and the extraction is terminated. Note that embedding and extraction processes are implemented in the transform domain, whereas checkpointing is executed within the spatial domain. The method is fast, produces imperceptible watermarked images, and can prove their copyright information and verify their integrity. However, the process is irreversible and cannot achieve tamper detection and localisation. Moreover, the method is also non-blind and may result in security issues caused by such methods [5, 99] and [101].

Sharma et al.'s most recent study [100] addressed issues of their previous work in [97]. In [100], a novel multipurpose image watermarking scheme capable of protecting and authenticating images with tamper detection and localisation abilities is presented. The method is operable on greyscale and colour images; however, the study has explained the working functionality using colour images. Firstly, a colour host image is converted into a YC_bC_r space, wherein the Y channel is embedded with a robust watermark. To this end, the robust logo watermark is encrypted using the Fisher and Yates algorithm, and the encrypted bits are embedded via the DWT-DCT-based combination. Once the robust watermarked image is achieved, it is split into *RGB* channels, and each channel is exposed to a halftoning operation. Subsequently, acquired halftoned equivalents are used as fragile watermarks using which each *RGB* channel is watermarked. Secondly, in [100], two 16-bit seeds are used during the fragile watermarking process. The first seed is the mean seed, extracted from the greyscale equivalent of a colour channel, and the other is the fragile watermark seed, extracted from the halftone equivalent. These two seeds are exposed to an XOR operation, and a 16-bit "XOR-seed" is attained. This step is vital because it creates an interdependency between the watermark and the host image. Specifically, such interdependency complicates the removal process if a hacker tries to remove the watermark. Even if a hacker somehow eliminates the watermark, the action seriously harms the host image, giving the receiver a straightaway impression of tampering and nullifying hacking attempts.

Furthermore, a 32-bit embedding seed is achieved by concatenating the XOR and mean seeds. The first 16 bits of the embedding seed are placed into the LSB plane, and the remaining 16 are placed into the ISB plane of greyscale pixels. In this way, each pixel in a 4×4 block belonging to a colour channel is embedded with a mean seed bit and an XOR seed bit. Here, the former provides tamper detection and localisation ability, whereas the latter offers authentication or verification. This series of steps is repeated for the remaining 4×4 blocks of a colour channel until the channel is watermarked. Finally, the procedure is replicated in the other two colour channels. Once watermarked, all three (RGB) colour channels are combined to form a dual watermarked colour image, wherein the first watermark is robust and the other fragile.

The watermarked images produced using Sharma et al.'s method [100] achieve high *PSNR* and *SSIM* values. For instance, when a large watermark (256×256 in size) is embedded in the host image that is 512×512, the smallest *PSNR* and *SSIM* values attained are > 41 dB and > 0.9, respectively. The proposed multipurpose watermarking scheme's robustness attribute is commendable as the robust watermark can resist most geometric and non-geometric attacks. To this end, the watermark's robustness is also top-notch when tested against several hybrid or complex manipulations, such as VQ, copy-move, and protocol. Moreover, an average *NCC* value of > 0.95 is achieved by a 32×32 watermark when tested against 70 odd watermarking attacks. Similarly, the fragile watermarking aspect of the proposed multipurpose scheme is high in fragility and desirably sensitive to minute changes. The scheme's precision and accuracy in tamper detection and localisation are superb and superior to counterpart methods covered in the study. In other words, the average *ACC* value achieved in the study is 0.9394 or 93.94%, which highlights the method's ability to recognise a wide range of image manipulations that often happen in an industrial environment. Notwithstanding the several advantages of the scheme, the main drawback is that it does not have the tamper restoration ability. That said, the study has already highlighted this drawback, wherein the authors have called the limitation the focus of their future work.

6 Summary and recommendations

Several questions have been raised in the course of the discussion so far. In this section, those concerns are summarised and arranged in a questionnaire. Subsequently, responses to the questions are provided, based on which recommendations are made. The readers can use the questionnaire and recommendations as guidelines when evaluating an existing watermarking scheme and developing new ones.

A summary of the significant questions raised in this review and our responses in the form of recommendations are outlined below.

Q1. Does the watermarking scheme use a foreign or self-generated watermark?

Recommendation: It is worth re-establishing that here in the discussion, a foreign watermark is referred to as one that is not generated from the host image. In contrast, the self-generated watermark is generated from the host image. Mainly a foreign watermark is used for achieving copyright protection through robust watermarking. For instance, degree testamur generally has a university emblem or logo to prove the copyright. Hence, employing a foreign logo as a watermark is desirable in robust watermarking, even though adding anything foreign to the host image is regarded as noise that ultimately degrades the *PSNR*. However, in the case of fragile watermarking, as there is no such requirement, using a foreign watermark is redundant. To this end, self- generated watermarks achieve high imperceptibility and exhibit better tamper detection and localisation ability than their foreign counterparts. To sum up, using foreign watermark(s) for robust watermarking is recommended, whereas the self-generated watermark(s) in fragile and semi-fragile watermarking.

Q2. Does a watermarking scheme justify the experimental analysis and comparison with other methods?

Recommendation: If a watermarking study compares the proposed method with existing methods, it is essential to justify the comparison. To this end, several aspects must be considered to validate the experimental results of a study. For instance, comparing

two methods is fair if conducted on the same test images with identical dimensions and colour space. The same is true for a watermark, i.e., the same-sized watermark (if binary, then the same number of black and white bits) must be employed during embedding by all methods involved in a comparison. Moreover, it is also vital to use a common performance baseline and metrics during the comparison. Last but not least, each method must be executed using a common machine while conducting the time complexity or processing time analysis.

Q3. Is the selection of the watermark embedding strength parameter or other parameters adaptive?

Recommendation: The watermark embedding strength parameter is one of the critical parameters in watermarking process. In other words, it dictates the balance between security and imperceptibility attributes; thereby, its correct selection is vital. To this end, the watermark embedding strength parameter is selected empirically or manually in many existing watermarking schemes. Consequently, several issues (mentioned above in Section 5.3.3) may occur because of manual selection approaches. Hence, it is recommended to employ a watermarking strategy wherein the selection of the parameters is adaptive, and some examples are [28–30] and [100]. By the way, selecting the watermark embedding strength parameter is primarily of high importance within the transform domain-based techniques, wherein it decides the visibility of the watermark. However, in spatial domain-based watermarking, the bit plane determines the watermark's visibility. For instance, embedding into the MSB plane of a greyscale image achieves the highest watermark visibility, whereas embedding into the LSB plane results in the most negligible watermark visibility.

Q4. Does the watermark embedding procedure establish a dependency between the watermark and the host image?

Recommendation: Combining the watermark's information with the host image's establishes a dependency between the two. Such dependence strengthens the security aspect of a watermarking scheme [100]. In other words, removing or destroying the watermark from the combined information is tedious for an attacker. Assuming that a hacker somehow manages to remove the watermark, in which case, the image's imperceptibility would be seriously compromised, giving the receiver a straightforward impression of the tampering. To sum up, an embedding process is recommended to ensure a dependency between the watermark and the host image. Watermarking schemes in [30, 97] and [100] are a few candidates who can do so.

Q5. Does the watermarking scheme or study cover the computational complexity or processing time analysis?

Recommendation: The question may seem obvious, but it is evident in the above discussion that almost half of the schemes mentioned in this chapter have ignored the processing time analysis or time taken by a method from start to finish. It is also common in several other studies that are not included in this review. Such analysis is necessary as it dictates whether a technique can be employed in real-time. Therefore, one must acknowledge the processing time analysis in their research; this way, the application feasibility of the research work can be vindicated.

Q6. Does the watermarking scheme attain the balance between tamper detection and localisation attributes?

Recommendation: Tamper detection and localisation performances are pixel-block size dependent. To this end, the pixel-block size is inversely proportional to tamper detection performance and directly proportional in the case of tamper localisation. In other words, the smaller the block size, the more accurate the tamper detection results and the more inaccurate the tampered region localisation. However, the larger the block size, the better the tampered region localisation and the more inaccurate the tamper detection results. That said, as tamper detection and localisation are exhibited by a fragile watermarking scheme, maintaining a balance between the two is a must (see Fig. 25). One way of achieving such equilibrium is through a method that follows a hierarchical approach. For instance, the procedure may start with large pixel blocks and gradually reduce to smaller ones. This approach fine-tunes the balance between tamper detection and localisation attributes and improves their performances. Readers may refer to Sharma et al.'s study in [100] for further insight into the recommendation.

Q7. Does the robust watermark have immunity against geometrical and non-geometrical attacks?

Recommendation: In the case of robust watermarking, the robust watermark must withstand geometrical or non-geometrical attacks. Using transform domain-based watermarking is recommended to achieve immunity against both categories of attacks. In transform domain-based watermarking, embedding is achieved by manipulating the frequency coefficients. Such embedding makes it highly unlikely for geometrical or non-geometrical attacks to destroy every coefficient representing the watermark information.

Q8. Does the watermarking scheme maintain the balance between the watermarking requirements of imperceptibility, security, and capacity?

Recommendation: As mentioned above in Section 2.3, a watermarking scheme must fulfil the requirements of imperceptibility, security, and capacity and attain a balance between the three. To this end, Fig. 26 can be used to interpret the trade-offs

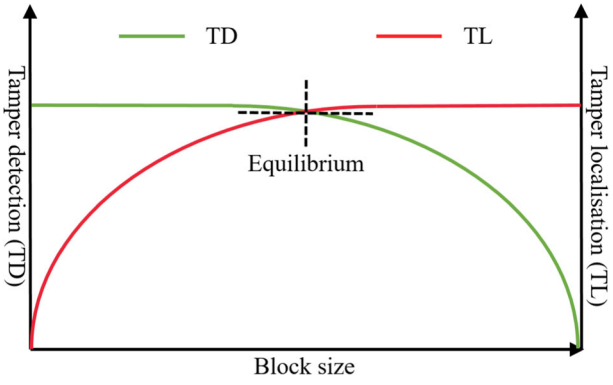

Fig. 25 Existing trade-offs in tamper detection and localisation. Here, if the equilibrium shifts to the left, the scheme has high precision in tamper detection but low in tampered region localisation, and the other way around if the shift is to the right. Note that the curves in this figure are not obtained via the experimental simulations. In contrast, the author has hand-drawn these curves for pictorial representation of tamper detection and localisation trade-offs

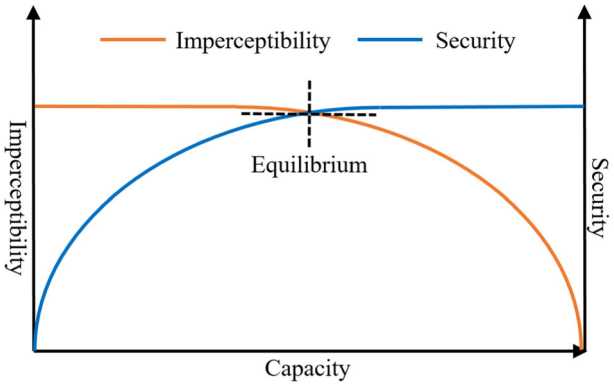

Fig. 26 Existing correlations and trade-offs in watermarking. Here, if the equilibrium shifts to the right, then the scheme is high in security but low in imperceptibility, and it is the other way around if the change is to the left. Note that the curves in this figure are not obtained via the experimental simulations. In contrast, the authors have hand-drawn these curves to demonstrate watermarking security and imperceptibility trade-offs. Note that graphics here in the figure are inspired by Sharma et al.'s study [99]

between these requirements. The exact figure can also be used to evaluate an existing watermarking technique and develop a new one.

Q9. Does the scheme have enough watermarking capacity to achieve tamper detection and localisation?

Recommendation: Effective tamper detection and localisation can only be achieved if a fragile watermarking scheme has sufficient watermark information or capacity to represent both attributes. In other words, there need to be enough watermark bits, of which one portion must be significant enough to perform the authentication or tamper detection operation. The other portion must be ample to accomplish the localisation of tampered regions. One recommendation that provides further insight into this aspect is Sharma et al.'s method [100], wherein two (an ISB and the LSB) of the total eight bits in a greyscale pixel are used for achieving tamper detection and localisation.

Q10. Does the watermarking scheme use multiple encryption keys? If yes, does it tackle the issues that may arise from them?

Recommendation: Firstly, to the best of our knowledge, using a key or encryption key is optional in watermarking. However, it's a must in other data-hiding techniques, such as cryptography. Notwithstanding that encrypting the watermark before embedding increases the robustness, it is nowhere mentioned in the literature that a certain number of encryption keys can guarantee a watermark's immunity against all possible manipulations. That being said, even if a watermarking scheme has employed multiple keys, there is always a possibility for the watermark to be compromised.
Secondly, we agree that if robustness is the only objective, it is desirable to use multiple keys; however, multiple keys call for strict key management protocols, giving rise to several questions about storage and transmission. For instance, how are the multiple keys stored and transmitted? What are the overheads imposed during the transmission process? How is the receiver ensured which key to use in which sequence? So the recommendation is to use as few keys as possible, but if multiple keys are used in a study, it must also address the above-mentioned questions.

Q11. Does the method work on both greyscale and colour images?

Recommendation: The question may seem obvious, but it is evident from the above discussion that several methods do not operate on colour images. That being said, most literature covering multipurpose watermarking techniques is focused on greyscale images. Statistically, in the last 25 years, 739 multipurpose watermarking works for images have been published, of which 643 are for greyscale images, and only 96 are for colour ones [93]. As illustrated in Fig. 27, this imbalance needs to be rectified as the greyscale images are rarely used today. To this end, a watermarking scheme need not be limited to only one colour space but should operate across different ones. It is an important quality that highlights a watermarking scheme's application and operations versatility.

Q12. Has the watermarking scheme been tested against hybrid or complex watermarking attacks?

Recommendation: Besides standalone geometric and non-geometric attacks, a watermarking scheme should also be secure against a combination of attacks. An attack that arises out of such a combination is also known as a hybrid attack, which is generally more lethal than the usual attacks. Despite this, several studies [39, 46, 60, 76] discussed in this review have ignored testing their methods against hybrid or complex attacks, such as VQ, copy- paste, collage, protocol, etc. It is recommended to consider these attacks while conducting the security evaluation of a watermarking scheme. Specifically when a strategy aims to achieve watermarking robustness.

Q13. Is the watermarking scheme tested against printing or scanning operations?

Recommendation: Sharma et al.'s methods are the only approaches discussed in this review that has considered the effects a printer can impose on the watermark if a watermarked image is exposed to a printing operation [96, 101]. To this end, this is a common trend amongst existing image watermarking methods that either

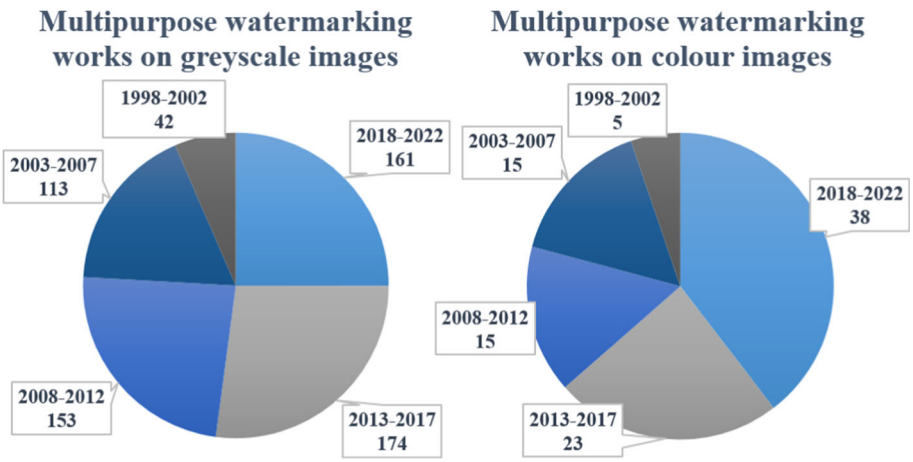

Fig. 27 The breakdown of the multipurpose watermarking works published in the last 25 years. The left pie chart shows the works that catered to the greyscale images, and the right is for the colour images. The data for generating these charts is extracted from Scopus®, available at [93]. Note that the graphics in the figure are inspired by Sharma et al.'s study [100]

focus on electronic media (e-media) or print media, thereby separating the two. Notwithstanding that some intricacies are exclusive to printing algorithms, in our opinion, printing needs to be viewed as another watermarking attack, and the same goes for scanning. That said, recording the change in the watermark's behaviour is recommended once exposed to printing and scanning attacks.

Q14. Does the watermark extraction process requires limited side information?

Recommendation: This question relates to using multiple keys and whether the watermark extraction is blind or non-blind. Side information is vital for successful watermark extraction and is transmitted separately to the watermarked image. A hacker gaining access to the side information can seriously threaten the security of a watermarking scheme and even jeopardise the whole mechanism. Hence, a watermarking technique that relies on less side information is preferable because the smaller the side information, the faster the transmission and the fewer transmission attempts. That said, using multiple keys inflicts an overhead regarding the side information, and the non-blind watermarking is at the back foot in a similar context.

A summary of the methods discussed in this review is presented in Table 3.

7 Conclusion

A thorough literature review of image watermarking for identity protection and authentication is presented in this paper. It covers several notable watermarking works which have left their mark on image watermarking research. The new systematisation is proposed and employed to classify various watermarking techniques. Moreover, existing studies are reviewed so that not only are their advantages and disadvantages presented to the readers, but they can also reverse-engineer those techniques. Furthermore, a questionnaire of vital questions must be acknowledged while evaluating an existing watermarking scheme, and developing a new one is compiled. Lastly, the recommendations within the questionnaire are outlined, providing readers with the potential solutions to the raised questions. Overall, the following conclusions are drawn from the proposed study.

- The proposed review covers over 100 prominent watermark ing works that have positively influenced and shaped the research area. Moreover, the methods discussed in the study are subject to a new systematisation, based on which a novel way of classifying watermarking techniques is proposed. A preview of the examined studies and how they are classified under the new systematisation is presented in Table 2.
- The methods are reviewed in a way that highlights their pros and cons and allows readers to reverse-engineer those methods. To that end, the study has the potential to provide the necessary tools to the new entrants in the area to kick-start their research and equally serve their experienced peers as their go-to study whenever they want to revisit essential watermarking concepts.
- Finally, a questionnaire is compiled wherein the vital questions that need to be acknowledged while evaluating an existing watermarking technique and developing a new one are presented. Moreover, the recommendations are also provided within the questionnaire, via which the questions raised can be tackled and addressed. A summary of the significant questions raised and their potential solutions in the form of our recommendations is outlined in Section 6.

Table 3 A summary of the main methods covered in this review is presented here

Study ↓	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14
Abraham and Paul [1]	FW	✓	–	✓	×	–	✓	✓	–	– |–	×	×	×	✓
Benrhouma et al. [8]	SW	✓	×	✓	×	✓	–	✓	✓	✓ |×	×	✓	×	×
Celik et al. [12]	SW	✓	–	✓	✓	✓	–	✓	✓	× |–	✓	✓	×	✓
Chang et al. [13]	SW	✓	–	✓	×	✓	–	✓	✓	× |–	×	×	×	✓
Chang et al. [14]	SW	✓	–	✓	×	✓	–	✓	✓	× |–	×	×	×	✓
Dadkhah et al. [17]	SW	✓	×	✓	×	✓	–	✓	✓	✓ |×	✓	✓	×	×
Fridrich and Goljan [24]	SW	✓	×	✓	×	✓	–	✓	✓	× |–	×	×	×	✓
Gul and Ozturk [27]	SW	✓	–	✓	×	✓	–	✓	✓	× |–	×	×	×	✓
Haghighi et al. [28]	SW	✓	✓	✓	×	✓	–	✓	✓	✓ |×	✓	✓	×	×
Haghighi et al. [29]	FW+SW	✓	✓	✓	×	✓	✓	✓	✓	✓ |×	✓	✓	×	×
Haghighi et al. [30]	SW	✓	✓	✓	×	✓	–	✓	✓	✓ |×	✓	✓	×	×
Hsu and Tu [36]	SW	✓	–	✓	×	✓	–	✓	✓	× |–	✓	✓	×	✓
Hurrah et al. [39]	FW	✓	×	×	✓	✓	✓	✓	✓	✓ |×	✓	✓	×	×
Hurrah et al. [40]	FW	✓	–	✓	✓	✓	–	✓	✓	✓ |×	✓	×	×	×
Islam et al. [43]	FW	✓	×	×	×	–	✓	✓	–	✓ |×	×	×	×	×
Kamili et al. [46]	FW	✓	×	×	✓	✓	✓	✓	✓	✓ |×	×	×	×	×
Kang et al. [47]	FW	✓	✓	×	×	–	✓	✓	–	× |–	×	×	×	✓
Li et al. [53]	SW	✓	–	✓	×	✓	–	✓	✓	✓ |×	×	✓	×	×
Li et al. [54]	SW	✓	–	✓	×	✓	–	✓	✓	✓ |×	×	✓	×	✓
Lin et al. [57]	FW	✓	×	×	×	–	✓	✓	–	× |–	×	×	×	✓
Loan et al. [60]	FW	✓	×	✓	×	–	✓	✓	–	✓ |×	✓	✓	×	✓
Lu and Liao [61]	SW	✓	✓	✓	×	✓	×	✓	✓	× |–	✓	✓	×	✓
Nguyen et al. [68]	SW	✓	×	✓	×	✓	–	✓	✓	× |–	×	✓	×	✓
Ni et al. [69]	SW	✓	–	✓	✓	–	–	–	–	– |–	×	–	–	✓
NR and Shreelekshmi [70]	SW	✓	×	✓	×	✓	–	✓	✓	✓ |×	×	✓	×	×
Pal et al. [71]	FW	✓	–	✓	✓	✓	✓	✓	✓	× |–	×	×	×	✓

Table 3 continued

Study ↓	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14
Pal et al. [72]	FW	✓	–	✓	✓	✓	✓	✓	✓	× | –	×	×	×	✓
Pal et al. [73]	FW	✓	–	✓	✓	✓	✓	✓	✓	× | –	×	×	×	✓
Parah et al. [75]	FW	✓	–	✓	✓	–	✓	✓	–	× | –	×	×	×	✓
Parah et al. [76]	FW	✓	×	✓	×	–	✓	✓	–	× | –	✓	✓	×	✓
Prasad and Pal [77]	SW	✓	–	✓	×	✓	–	✓	✓	× | –	×	×	×	✓
Prasad and Pal [78]	SW	✓	–	✓	×	✓	–	✓	✓	× | –	×	✓	×	✓
Preda and Vizireanu [79]	SW	✓	×	✓	×	✓	✓	✓	✓	× | –	×	–	–	✓
Preda [80]	SW	✓	×	✓	×	✓	✓	✓	✓	× | –	×	✓	–	✓
Preda et al. [81]	SW	✓	×	✓	×	–	–	✓	–	× | –	×	–	–	✓
Qi and Xin [83]	SW	✓	×	✓	×	✓	✓	✓	✓	✓ | ×	×	×	×	×
Qi and Xin [84]	SW	✓	×	✓	×	✓	✓	✓	✓	× | –	×	×	×	✓
Raj and Shreelekshmi [85]	SW	✓	–	✓	×	✓	–	✓	✓	× | –	✓	✓	×	✓
Rawat and Raman [86]	FW	✓	–	✓	×	✓	–	✓	✓	× | –	×	×	×	✓
Rhayma et al. [88]	SW	✓	×	✓	×	–	–	✓	–	× | –	×	–	–	✓
Sharma et al. [97]	FW	✓	×	✓	✓	–	✓	✓	–	× | –	✓	×	×	✓
Sharma et al. [99]	FW	✓	×	✓	✓	–	✓	✓	–	× | –	×	×	×	✓
Sharma et al. [101]	FW	✓	×	✓	✓	–	✓	✓	–	× | –	×	×	✓	✓
Sharma et al. [100]	FW+SW	✓	✓	✓	✓	✓	✓	✓	✓	× | –	✓	✓	×	✓
Singh and Singh [105]	SW	✓	×	✓	✓	✓	–	✓	✓	✓ | ×	✓	×	×	×
Singh and Singh [106]	SW	✓	×	✓	✓	✓	–	✓	✓	✓ | ×	✓	×	×	×
Thanki and Borra [108]	FW	✓	×	×	✓	–	✓	✓	–	× | –	×	×	×	✓
Ullah et al. [112]	SW	✓	×	✓	×	✓	–	✓	✓	✓ | ×	×	✓	×	×
Verma et al. [114]	FW	✓	×	×	✓	–	✓	✓	–	✓ | ×	×	×	×	×
Wang et al. [117]	SW	✓	×	✓	×	✓	–	✓	✓	✓ | ×	×	✓	×	×
Wenyin and Shih [118]	FW	✓	–	✓	×	✓	–	✓	✓	– | –	×	×	×	✓
Wong [119]	FW	✓	–	✓	×	✓	×	✓	✓	✓ | ×	×	×	×	✓

Table 3 continued

Study ↓	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14
Wong [120]	FW	✓	–	✓	×	✓	×	✓	✓	✓ | ×	×	×	×	✓
Wong and Memon [121]	FW	✓	–	✓	×	✓	×	✓	✓	✓ | ×	×	×	×	✓
Xiang et al. [122]	SW	✓	–	✓	×	–	✓	✓	–	× | –	×	✓	×	✓
Xiao and Wang [123]	FW	✓	–	✓	×	–	–	✓	–	– | –	×	×	×	✓
Yeung and Mintzer [124]	SW	✓	–	✓	×	✓	–	✓	✓	× | –	✓	×	×	✓
You et al. [126]	FW	✓	×	×	×	–	✓	✓	–	✓ | ×	×	×	×	×
Zhang et al. [127]	SW	✓	–	✓	×	✓	–	✓	✓	× | –	×	✓	×	✓
Zong et al. [129]	SW	✓	–	✓	×	–	✓	✓	–	– | –	×	✓	×	✓
Zong et al. [130]	SW	✓	–	✓	×	–	✓	✓	–	× | –	×	✓	×	✓

Note that some of the questions are irrelevant to robust watermarking, some to fragile watermarking, and so on. The notion of not applicable (−) is used in this case. Moreover, a foreign watermark is denoted by FW and a self-generated watermark by SW. On another note, Q10 is the only two-in-one question, and the desired response to each part of Q10 is either × | − or × | ✓

Acknowledgements The Western Sydney University Postgraduate Research Award supports this work.

Funding Open Access funding enabled and organized by CAUL and its Member Institutions No funds, grants, or other support was received.

Data Availability The manuscript has no associated data.

Declarations

Conflicts of interest/Competing interests The authors have no conflicts of interest/competing interests to declare that are relevant to the content of this article.

References

1. Abraham J, Paul V (2019) An imperceptible spatial domain color image watermarking scheme. Journal of King Saud University-Computer and Information Sciences 31(1):125–133
2. Alzahrani A, Memon NA (2021) Blind and robust watermarking scheme in hybrid domain for copyright protection of medical images. IEEE Access 9:113714–113734
3. Aminuddin A, Ernawan F (2022) Ausr1: Authentication and self-recovery using a new image inpainting technique with lsb shifting in fragile image watermarking. Journal of King Saud University-Computer and Information Sciences
4. Anand A, Singh AK (2020) Watermarking techniques for medical data authentication: a survey. Multimedia Tools and Applications pp 1–33
5. Ansari IA, Pant M (2017) Multipurpose image watermarking in the domain of dwt based on svd and abc. Pattern Recogn Lett 94:228–236
6. Barr M, Serdean C (2019) Wavelet transform modulus maxima-based robust logo watermarking. IET Image Process 14(4):697–708
7. Begum M, Uddin MS (2020) Digital image watermarking techniques: a review. Information 11(2):110
8. Benrhouma O, Hermassi H, Belghith S (2017) Security analysis and improvement of an active watermarking system for image tampering detection using a self-recovery scheme. Multimedia Tools and Applications 76(20):21133–21156
9. Bertini F, Sharma R, Montesi D (2020) Are social networks watermarking us or are we (unawarely) watermarking ourself? arXiv preprint arXiv:2006.03903
10. Bhowmik D, Abhayaratne C (2019) Embedding distortion analysis in wavelet-domain watermarking. ACM Trans Multimed Comput Commun Appl (TOMM) 15(4):1–24
11. Boenisch F (2020) A survey on model watermarking neural networks. arXiv preprint arXiv:2009.12153
12. Celik MU, Sharma G, Saber E, Tekalp AM (2002) Hierarchical watermarking for secure image authentication with localization. IEEE Trans Image Process 11(6):585–595
13. Chang CC, Chen KN, Lee CF, Liu LJ (2011) A secure fragile watermarking scheme based on chaos-and-hamming code. J Syst Softw 84(9):1462–1470
14. Chang JD, Chen BH, Tsai CS (2013) Lbp-based fragile watermarking scheme for image tamper detection and recovery. In: 2013 international symposium on next-generation electronics, pp 173–176. IEEE
15. Twitter attack 2020 (2020). https://www.cnbc.com/2020/07/16/twitter-hackers-made-121000-in-bitcoin-analysis-shows.html
16. Conftool (2023). https://conftool.net/ctforum/index.php/topic,264.0.html

17. Dadkhah S, Abd Manaf A, Hori Y, Hassanien AE, Sadeghi S (2014) An effective svd-based image tampering detection and self-recovery using active watermarking. Signal Process Image Commun 29(10):1197–1210
18. Daubechies I, Han B, Ron A, Shen Z (2003) Framelets: Mra-based constructions of wavelet frames. Appl Comput Harmon Anal 14(1):1–46
19. Data breach investigation (2023). https://www.verizon.com/business/resources/reports/2020-data-breach-investigations-report.pdf
20. Digimarc (2023). https://www.digimarc.com/products/digimarc-watermarks
21. Evsutin O, Dzhanashia K (2022) Watermarking schemes for digital images: Robustness overview. Signal Process Image Commun 100:116523
22. Evsutin OO, Melman AS, Meshcheryakov RV (2020) Digital steganography and watermarking for digital images: a review of current research directions. IEEE Access
23. Fan M, Wang H (2018) An enhanced fragile watermarking scheme to digital image protection and self-recovery. Signal Process Image Commun 66:19–29
24. Fridrich J, Goljan M (1999) Images with self-correcting capabilities. In: Proceedings 1999 International conference on image processing (Cat. 99CH36348), vol 3. IEEE, pp 792–796
25. G+d (2023). https://www.gi-de.com/en/payment/cash/banknote-security-technology/watermarks
26. Ghost brand packing watermark (2023). https://paxxus.com/ghost/
27. Gul E, Ozturk S (2020) A novel triple recovery information embedding approach for selfembedded digital image watermarking. Multimedia Tools and Applications 79(41):31239–31264
28. Haghighi BB, Taherinia AH, Harati A (2018) Trlh: Fragile and blind dual watermarking for image tamper detection and self-recovery based on lifting wavelet transform and halftoning technique. J Vis Commun Image Represent 50:49–64
29. Haghighi BB, Taherinia AH, Harati A, Rouhani M (2021) Wsmn: An optimized multipurpose blind watermarking in shearlet domain using mlp and nsga-ii. Appl Soft Comput 101:107029
30. Haghighi BB, Taherinia AH, Mohajerzadeh AH (2019) Trlg: Fragile blind quad watermarking for image tamper detection and recovery by providing compact digests with optimized quality using lwt and ga. Inf Sci 486:204–230
31. Haouzia A, Noumeir R (2008) Methods for image authentication: a survey. Multimedia tools and applications 39(1):1–46
32. Hasan MK, Kamil S, Shafiq M, Yuvaraj S, Kumar ES, Vincent R, Nafi NS (2021) An improved watermarking algorithm for robustness and imperceptibility of data protection in the perception layer of internet of things. Pattern Recogn Lett 152:283–294
33. Holygrail2.0 (2023). https://www.digitalwatermarks.eu/
34. Hong W (2012) Adaptive reversible data hiding method based on error energy control and histogram shifting. Opt Commun 285(2):101–108
35. Hou X, Min L, Yang H (2018) A reversible watermarking scheme for vector maps based on multilevel histogram modification. Symmetry 10(9):397
36. Hsu CS, Tu SF (2010) Probability-based tampering detection scheme for digital images. Opt Commun 283(9):1737–1743
37. Hu X, Zhang W, Li X, Yu N (2015) Minimum rate prediction and optimized histograms modification for reversible data hiding. IEEE Transactions on Information Forensics and Security 10(3):653–664
38. Huang CH, Wu JL (2004) Attacking visible watermarking schemes. IEEE transactions on multimedia 6(1):16–30
39. Hurrah NN, Parah SA, Loan NA, Sheikh JA, Elhoseny M, Muhammad K (2019) Dual watermarking framework for privacy protection and content authentication of multimedia. Futur Gener Comput Syst 94:654–673
40. Hurrah NN, Parah SA, Sheikh JA (2020) Embedding in medical images: an efficient scheme for authentication and tamper localization. Multimedia Tools and Applications 79(29):21441–21470
41. Hurrah NN, Parah SA, Sheikh JA, Al-Turjman F, Muhammad K (2019) Secure data transmission framework for confidentiality in iots. Ad Hoc Netw 95:101989
42. Identity security (2023). https://www.homeaffairs.gov.au/criminal-justice/files/security-standards-proof-identity-documents.pdf
43. Islam M, Roy A, Laskar RH (2020) Svm-based robust image watermarking technique in lwt domain using different sub-bands. Neural Comput & Applic 32(5):1379–1403
44. Jaiswal SP, Au OC, Jakhetiya V, Guo Y, Tiwari AK, Yue K (2013) Efficient adaptive prediction based reversible image watermarking. In: 2013 IEEE International conference on image processing. IEEE, pp 4540–4544
45. Kadhim IJ, Premaratne P, Vial PJ, Halloran B (2019) Comprehensive survey of image steganography: Techniques, evaluations, and trends in future research. Neurocomputing 335:299–326

46. Kamili A, Hurrah NN, Parah SA, Bhat G, Muhammad K (2020) Dwfcat: Dual watermarking framework for industrial image authentication and tamper localization. IEEE Transactions on Industrial Informatics
47. Kang Xb, Zhao F, Lin Gf, Chen Yj (2018) A novel hybrid of dct and svd in dwt domain for robust and invisible blind image watermarking with optimal embedding strength. Multimedia Tools and Applications 77(11):13197–13224
48. Kocarev L (2001) Chaos-based cryptography: a brief overview. IEEE Circuits and Systems Magazine 1(3):6–21
49. Kocarev L, Lian S (2011) Chaos-based cryptography: theory, algorithms and applications, vol 354. Springer Science & Business Media
50. Koley S (2019) A feature adaptive image watermarking framework based on phase congruency and symmetric key cryptography. Journal of King Saud University-Computer and Information Sciences
51. Kovacevic J, Vetterli M (1992) Nonseparable multidimensional perfect reconstruction filter banks and wavelet bases for r/sup n. IEEE Trans Inf Theory 38(2):533–555
52. Kumar C, Singh AK, Kumar P (2018) A recent survey on image watermarking techniques and its application in e-governance. Multimedia Tools and Applications 77(3):3597–3622
53. Li C, Wang Y, Ma B, Zhang Z (2011) A novel self-recovery fragile watermarking scheme based on dual-redundant-ring structure. Comput Electr Eng 37(6):927–940
54. Li C, Wang Y, Ma B, Zhang Z (2013) Multi-block dependency based fragile watermarking scheme for fingerprint images protection. Multimedia tools and applications 64(3):757–776
55. Liang X, Xiang S, Yang L, Li J (2021) Robust and reversible image watermarking in homomorphic encrypted domain. Signal Process Image Commun 99:116462
56. Lin SL, Huang CF, Liou MH, Chen CY (2013) Improving histogram-based reversible information hiding by an optimal weight-based prediction scheme. J Inf Hiding Multim Signal Process 4(1):19–33
57. Lin WH, Horng SJ, Kao TW, Fan P, Lee CL, Pan Y (2008) An efficient watermarking method based on significant difference of wavelet coefficient quantization. IEEE Trans Multimed 10(5):746–757
58. Liu J, Huang J, Luo Y, Cao L, Yang S, Wei D, Zhou R (2019) An optimized image watermarking method based on hd and svd in dwt domain. IEEE Access 7:80849–80860
59. Liu XL, Lin CC, Yuan SM (2016) Blind dual watermarking for color images' authentication and copyright protection. IEEE transactions on circuits and systems for video technology 28(5):1047–1055
60. Loan NA, Hurrah NN, Parah SA, Lee JW, Sheikh JA, Bhat GM (2018) Secure and robust digital image watermarking using coefficient differencing and chaotic encryption. IEEE Access 6:19876–19897
61. Lu CS, Liao HY (2001) Multipurpose watermarking for image authentication and protection. IEEE Trans Image Process 10(10):1579–1592
62. Manikandan V, Masilamani V (2018) Histogram shifting-based blind watermarking scheme for copyright protection in 5g. Comput Electr Eng 72:614–630
63. Matsumoto M, Nishimura T (1998) Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator. ACM Transactions on Modeling and Computer Simulation (TOMACS) 8(1):3–30
64. Meerwald P, Koidl C, Uhl A (2009) Attack on "watermarking method based on significant difference of wavelet coefficient quantization". IEEE Transactions on Multimedia 11(5):1037–1041
65. Menendez-Ortiz A, Feregrino-Uribe C, Hasimoto-Beltran R, Garcia-Hernandez JJ (2019) A survey on reversible watermarking for multimedia content: A robustness overview. IEEE Access 7:132662–132681
66. Ncc calculations on matlab (2023). https://www.mathworks.com/help/images/ref/normxcorr2.html
67. Nexguard (2023). https://dtv.nagra.com/nexguard-video-watermarking-for-pay-tv
68. Nguyen TS, Chang CC, Yang XQ (2016) A reversible image authentication scheme based on fragile watermarking in discrete wavelet transform domain. AEU Int J Electron Commun 70(8):1055–1061
69. Ni Z, Shi YQ, Ansari N, Su W (2006) Reversible data hiding. IEEE Transactions on circuits and systems for video technology 16(3):354–362
70. NR NR, Shreelekshmi R (2022) Fragile watermarking scheme for tamper localization in images using logistic map and singular value decomposition. J Vis Commun Image Represent 85:103500
71. Pal P, Jana B, Bhaumik J (2019) Robust watermarking scheme for tamper detection and authentication exploiting ca. IET Image Process 13(12):2116–2129
72. Pal P, Jana B, Bhaumik J (2019) Watermarking scheme using local binary pattern for image authentication and tamper detection through dual image. Secur Priv 2(2):e59
73. Pal P, Jana B, Bhaumik J (2021) A secure reversible color image watermarking scheme based on lbp, lagrange interpolation polynomial and weighted matrix. Multimedia Tools and Applications 80(14):21651–21678
74. Parah S, Sheikh J, Hafiz A, Bhat G (2015) A secure and robust information hiding technique for covert communication. Int J Electron 102(8):1253–1266

75. Parah SA, Sheikh JA, Assad UI, Bhat GM (2017) Realisation and robustness evaluation of a blind spatial domain watermarking technique. Int J Electron 104(4):659–672
76. Parah SA, Sheikh JA, Loan NA, Bhat GM (2016) Robust and blind watermarking technique in dct domain using inter-block coefficient differencing. Digital Signal Processing 53:11–24
77. Prasad S, Pal AK (2020) Hamming code and logistic-map based pixel-level active forgery detection scheme using fragile watermarking. Multimedia Tools and Applications 79(29):20897–20928
78. Prasad S, Pal AK (2020) A tamper detection suitable fragile watermarking scheme based on novel payload embedding strategy. Multimedia Tools and Applications 79(3):1673–1705
79. Preda R, Vizireanu D (2015) Watermarking-based image authentication robust to jpeg compression. Electron Lett 51(23):1873–1875
80. Preda RO (2013) Semi-fragile watermarking for image authentication with sensitive tamper localization in the wavelet domain. Measurement 46(1):367–373
81. Preda RO, Vizireanu DN, Halunga S (2016) Active image forgery detection scheme based on semi-fragile watermarking. Revue Roumaine des Sciences Techniques-Serie Electrotechnique et Energetique, Romania 61(1):58–62
82. Qasim AF, Aspin R, Meziane F, Hogg P (2019) Roi-based reversible watermarking scheme for ensuring the integrity and authenticity of dicom mr images. Multimedia Tools and Applications 78(12):16433–16463
83. Qi X, Xin X (2011) A quantization-based semi-fragile watermarking scheme for image content authentication. J Vis Commun Image Represent 22(2):187–200
84. Qi X, Xin X (2015) A singular-value-based semi-fragile watermarking scheme for image content authentication with tamper localization. J Vis Commun Image Represent 30:312–327
85. Raj NN, Shreelekshmi R (2018) Blockwise fragile watermarking schemes for tamper localization in digital images. In: 2018 International CET conference on control, communication, and computing (IC4). IEEE, pp 441–446
86. Rawat S, Raman B (2011) A chaotic system based fragile watermarking scheme for image tamper detection. AEU Int J Electron Commun 65(10):840–847
87. Traitor tracing (2023). https://patents.google.com/patent/US20070067244
88. Rhayma H, Makhloufi A, Hamam H, Hamida AB (2021) Semi-fragile watermarking scheme based on perceptual hash function (phf) for image tampering detection. Multimedia Tools and Applications 80(17):26813–26832
89. Risk based security report (2023). https://pages.riskbasedsecurity.com/hubfs/Reports/2020/2020%20Q3%20Data%20Breach%20QuickView%20Report.pdf
90. Roy A, Chakraborty RS (2019) Toward optimal prediction error expansion-based reversible image watermarking. IEEE Transactions on Circuits and Systems for Video Technology 30(8):2377–2390
91. Schlauweg M, Pröfrock D, Palfner T, Müller E (2005) Quantization-based semi-fragile public-key watermarking for secure image authentication. In: Mathematics of data/image coding, compression, and encryption VIII, with applications, vol 5915. SPIE, pp 41–51
92. Van der Schyff K, Flowerday S, Furnell S (2020) Duplicitous social media and data surveillance: an evaluation of privacy risk. Computers & Security p 101822
93. Scopus (2023). https://www.scopus.com/search/form.uri?
94. Selesnick IW, Baraniuk RG, Kingsbury NC (2005) The dual-tree complex wavelet transform. IEEE Signal Proc Mag 22(6):123–151
95. Service nsw cyber incident (2023). https://www.service.nsw.gov.au/cyber-incident
96. Sharma S, Zou J, Fang G (2020) Significant difference-based watermarking in multitoned images. Electron Lett 56(18):923–926
97. Sharma S, Zou J, Fang G (2022) A single watermark based scheme for both protection and authentication of identities. IET Image Process 16(12):3113–3132
98. Sharma S, Zou JJ, Fang G (2019) Recent developments in halftone based image watermarking. In: 2019 International conference on electrical engineering research & practice (ICEERP). IEEE, pp 1–6
99. Sharma S, Zou JJ, Fang G (2020) A novel signature watermarking scheme for identity protection. In: Proceedings of the 20th international conference on digital image computing: techniques and applications (DICTA), 30 November-03 December, 2020, Melbourne, Australia, pp 1–5
100. Sharma S, Zou JJ, Fang G (2022) A novel multipurpose watermarking scheme capable of protecting and authenticating images with tamper detection and localisation abilities. IEEE Access 10:85677–85700
101. Sharma S, Zou JJ, Fang G (2023) A dual watermarking scheme for identity protection. Multimedia Tools and Applications 82(2):2207–2236
102. Sharma Y, Javadi B, Si W, Sun D (2017) Reliable and energy efficient resource provisioning and allocation in cloud computing. In: Proceedings of the 10th international conference on utility and cloud computing, pp 57–66

103. Sharma Y, Taheri J, Si W, Sun D, Javadi B (2020) Dynamic resource provisioning for sustainable cloud computing systems in the presence of correlated failures. IEEE Transactions on Sustainable Computing
104. Singh AK, Kumar B, Singh SK, Ghrera S, Mohan A (2018) Multiple watermarking technique for securing online social network contents using back propagation neural network. Futur Gener Comput Syst 86:926–939
105. Singh D, Singh SK (2016) Effective self-embedding watermarking scheme for image tampered detection and localization with recovery capability. J Vis Commun Image Represent 38:775–789
106. Singh D, Singh SK (2017) Dct based efficient fragile watermarking scheme for image authentication and restoration. Multimedia Tools and Applications 76(1):953–977
107. Sreenivas K, Kamkshi Prasad V (2018) Fragile watermarking schemes for image authentication: a survey. International Journal of Machine Learning and Cybernetics 9(7):1193–1218
108. Thanki R, Borra S (2018) A color image steganography in hybrid frt-dwt domain. Journal of information security and applications 40:92–102
109. Tong X, Liu Y, Zhang M, Chen Y (2013) A novel chaos-based fragile watermarking for image tampering detection and self-recovery. Signal Process Image Commun 28(3):301–308
110. Tran DN, Zepernick HJ, Chu TMC (2022) Lsb data hiding in digital media: a survey. EAI Endorsed Transactions on Industrial Networks and Intelligent Systems pp 1–50
111. Uchida Y, Nagai Y, Sakazawa S, Satoh S (2017) Embedding watermarks into deep neural networks. In: Proceedings of the 2017 ACM on international conference on multimedia retrieval, pp 269–277
112. Ullah R, Khan A, Malik AS (2013) Dual-purpose semi-fragile watermark: Authentication and recovery of digital images. Comput Electr Eng 39(7):2019–2030
113. Van Schyndel RG, Tirkel AZ, Osborne CF (1994) A digital watermark. In: Proceedings of 1st international conference on image processing, vol 2. IEEE, pp 86–90
114. Verma VS, Jha RK, Ojha A (2015) Significant region based robust watermarking scheme in lifting wavelet transform domain. Expert Syst Appl 42(21):8184–8197
115. Ballet voting (2023). https://www.aec.gov.au/elections/candidates/files/ballot-paper-formality-guidelines.pdf
116. Walton S (1995) Image authentication for a slippery new age. Dr. Dobb's Journal 20(4):18–26
117. Wang C, Zhang H, Zhou X (2018) Lbp and dwt based fragile watermarking for image authentication. Journal of Information Processing Systems 14(3):666–679
118. Wenyin Z, Shih FY (2011) Semi-fragile spatial watermarking based on local binary pattern operators. Opt Commun 284(16–17):3904–3912
119. Wong PW (1998) A public key watermark for image verification and authentication. In: Proceedings 1998 international conference on image processing. ICIP98 (Cat. No. 98CB36269), vol 1. IEEE, pp 455–459
120. Wong PW (1998) A watermark for image integrity and ownership verification. In: PICS, pp 374–379
121. Wong PW, Memon N (2001) Secret and public key image watermarking schemes for image authentication and ownership verification. IEEE Trans Image Process 10(10):1593–1601
122. Xiang S, Kim HJ, Huang J (2008) Invariant image watermarking based on statistical features in the low-frequency domain. IEEE Transactions on Circuits and Systems for Video Technology 18(6):777–790
123. Xiao J, Wang Y (2008) A semi-fragile watermarking tolerant of laplacian sharpening. In: 2008 International conference on computer science and software engineering, vol 3. IEEE, pp 579–582
124. Yeung MM, Mintzer F (1997) An invisible watermarking technique for image verification. In: Proceedings of international conference on image processing, vol 2. IEEE, pp 680–683
125. Yoo JC, Han TH (2009) Fast normalized cross-correlation. Circuits, Systems and Signal Processing 28(6):819–843
126. You X, Du L, Ym Cheung, Chen Q (2010) A blind watermarking scheme using new nontensor product wavelet filter banks. IEEE Trans Image Process 19(12):3271–3284
127. Zhang H, Wang C, Zhou X (2017) Fragile watermarking based on lbp for blind tamper detection in images. J Inf Process Systems 13(2):385–399
128. Zhang W, Shih FY (2017) 8 watermarking based on local binary pattern operators. In: Multimedia security: watermarking, steganography, and forensics. CRC Press, pp 141–163
129. Zong T, Xiang Y, Natgunanathan I (2014) Histogram shape-based robust image watermarking method. In: 2014 IEEE international conference on communications (ICC). IEEE, pp 878–883
130. Zong T, Xiang Y, Natgunanathan I, Guo S, Zhou W, Beliakov G (2014) Robust histogram shape-based method for image watermarking. IEEE Transactions on Circuits and Systems for Video Technology 25(5):717–729

Authors and Affiliations

Sunpreet Sharma[1] · Ju Jia Zou[1] · Gu Fang[1] · Pancham Shukla[2] · Weidong Cai[3]

1 School of Engineering, Design and Built Environment, Western Sydney University, Locked Bag 1797, Penrith 2751, NSW, Australia

2 Department of Computing, Faculty of Engineering, Imperial College London, London, UK

3 School of Computer Science, The University of Sydney, Sydney, NSW, Australia