

Received 23 January 2024, accepted 20 February 2024, date of publication 26 February 2024, date of current version 15 March 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3370411

RESEARCH ARTICLE

Optimal Semi-Fragile Watermarking Based on Maximum Entropy Random Walk and Swin Transformer for Tamper Localization

P. ABERNA AND L. AGILANDEESWARI

School of Computer Science Engineering and Information Systems, VIT, Vellore, Tamil Nadu 632014, India

Corresponding author: L. Agilandeswari (agila.l@vit.ac.in)

ABSTRACT In the multimedia arena, image tampering is an uncontrollable process that necessitates content authentication and tamper detection in a variety of applications. One method that is recommended for meeting all of those needs in the multimedia arena is watermarking. Mobile cameras may now be used to effortlessly take high dynamic range (HDR) photographs, which increase the image's visual quality and realism. Watermark visibility and recognition algorithms built for standard images may be affected by this introduction of perceptual variations in the image relative to the source. In order to overcome those shortcomings, we introduced a novel, an optimal semi-blind watermarking technique that works for both colour and HDR compressed JPEG images. A unique quaternion dual-tree complex wavelet transform technique is used to extract the highly informative features from the original image. The optimal embedding region in the low frequency sub-band is determined using the maximal entropy random walk (MERW) algorithm. In order to detect tampering and to localize the tampered region a watermark is generated using the swin transformer model and watermark embedding is carried out in the selected optimal blocks. A dual scrambled image is encoded in the effective principal component coefficient values of the singular value decomposition (SVD) Transform in order to authenticate the watermarked image prior to watermark extraction. The semi-blind extraction process is intended to confirm the content's authenticity by comparing the recovered scrambled watermark with the regenerated original watermark. The process of extraction is merely the opposite of the process of embedding. When compared to previous research, the experimental results demonstrated good imperceptibility with an average PSNR of 65 dB and SSIM of 0.999 and strong robustness against attacks.


INDEX TERMS High dynamic image, maximum entropy random walk algorithm, quaternion dual-tree complex wavelet transform, Swin transformer.

I. INTRODUCTION

The vulnerability of multimedia content has grown in recent decades as its usage has grown. Due to the fast-developing digital technology, it urges to secure multimedia images against tampering applications. Watermarking is recommended as a highly concerned strategy for securing multimedia content [1], [2]. Digital watermarking is the process where the image visual quality is maintained by embedding the additional information in the original image. The additional

information used as a watermark is either an image or meta-data or a binary logo. Much research has been done in the past two decades concentrating on copyright protection [3] and content authentication applications and currently, it focuses on color images and high dynamic range (HDR) images for tamper detection applications. In the current trend due to the advancement of photographic technology and the availability of less-cost mobile cameras, high quality HDR images are produced with more color and detail pixel information.

Most modern mobile cameras have an HDR mode or an Auto HDR feature in the camera settings to capture an HDR image and compress the final HDR image into a

The associate editor coordinating the review of this manuscript and approving it for publication was Sawyer Duane Campbell .

JPEG format. In the early stage, multi-exposure low dynamic images (LDR) are captured and combined using software resulting in HDR images. But in the current digital world, the JPEG (Joint Photographic Experts Group) is a widely used file format for storing digital images, including HDR images. Some advanced camera apps or newer mobile devices may offer the option to save HDR images in other formats, such as HEIF (High-Efficiency Image Format) or DNG (Digital Negative). Thus, the rising trend of HDR images on real-time applications increases demand for securing multimedia data against intentional or unintentional attacks.

Few studies have been done explicitly to analyse HDR images against tone mapping attacks; in particular, very few studies have examined HDR images against image processing attacks employing watermarking approaches. At the very first, in HDR based watermarking technique Yu et al. [4] presented a spatial domain-based distortion-free data concealing approach. Using a secret key in the spatial domain homogeneity value of the superpixel, the secret message is directly placed on the 32-bit RGBE image format. Since every HDR image has a unique wide dynamic range and every pixel has an infinite number of decimal values, it is difficult to directly alter the pixels in an HDR image [5]. To get over this problem, bilateral filtering techniques based on HDR images are used, followed by a logarithmic wavelet domain to classify the scalable high-frequency image [6]. Their objective is only to detect watermarks using threshold values for which a unique method is used to compute them. The HDR image watermarking approaches utilised in this literature, which processed the images in three formats—RGBE [7], LogLuv [8], and OpenEXR format [9]—are mostly focused on tone mapping attacks. To defend against tone mapping attacks, Perez-Daniel et al. [10] presented a spatial domain watermarking technique. The super-pixels in the Y channel are separated according to their texture, colour, and semantic content. To insert the binary visible watermark in the adaptable area, a luma fluctuation tolerance threshold curve based on the transfer function is utilised. Luminance sensitivity in each luma code area found in the Y-channel is used to incorporate the watermark to accomplish invisible HDR watermarking. Several tone mapping operators and a few signal processing attacks were investigated in the experiment. Using Tucker decomposition, an auto-regressive robust watermarking model is presented in [11]. The auto-regressive model is used to calculate the local similarity among the initial feature map of Tucker Decomposition and to remove the non-adaptive zone to embed the watermark in a suitable region. A spatial domain image watermarking method for HDR images in the OpenEXR extension was presented by Lin et al. [9]. The HVS features take advantage of an adaptive low luminance region, where the watermark bits are encoded in a 10-bit mantissa with three channels. The outcomes demonstrated that there were no perceived variations in the visual quality between the generated watermarked HDR image and its tone-mapped low dynamic range (LDR)

images and that a high embedding capacity was attained. It is evident from the description above that distortion-free systems (i.e., virtually undetectable watermarking systems), higher payload systems, and watermark detecting systems are offered. However, the above solution is still underperforming, because it does not take the robustness requirements into account. As a solution, Bakhsh and Moghaddam [8] proposed an artificial bee colony (ABC) optimisation technique to create a robust HDR image watermarking system. The watermark is embedded in the LH and HL sub-bands of the discrete wavelet (DWT) transform, and the best embedding block is chosen using an optimization technique. Recently, a saliency detection method was used to investigate a trade-off watermarking system [12]. The saliency object is generated using a unique ResNet architecture on an HDR image. The original and watermarked images' foreground and background are divided into distinct sections based on the saliency mask. The quantized index modulation approach is used to include the permuted random segment watermark blocks into the lifting wavelet transform (LWT) host image bit plane. Better results were obtained while using different metrics to evaluate the experiment result in order to examine the trade-off performance.

The challenges in HDR images are (i) Tone Mapping techniques that compress the wide dynamic range into a standard dynamic range by altering the image's characteristics, including color and contrast. (ii) When tone mapped and saved as JPEG the bit depth is reduced from 10 or 12-bit to 8-bit per channel, potentially impact on watermark imperceptibility and robustness. (iii) Due to more visually appealing and realistic representation of the HDR image produces higher perceptual differences than to the low dynamic image (LDR) image, which affects the watermark visibility of the algorithms designed for standard images [6]. (iv) Research on HDR images is not exposed much in the case of robustness, content authentication, tamper detection, and localization applications. This motivated us to attempt a tamper detection based HDR image watermarking system by embedding the watermark features.

Further the existing color image watermarking techniques where the better resulted techniques and the challenges still exist are discussed in detail. Generally, the watermarking is categorized based on its types namely, Robust watermarking [13], [14], Fragile watermarking, and Semi-fragile watermarking [15], [16]. Notably, a powerful watermarking system can endure attacks related to signal processing, but it cannot withstand against original image manipulations or pinpoint the tampered region, instead, it verifies whether the image has been tampered with or not and it is termed as "Robust watermarking". Eventhough the watermark fails to be robust against unintentional attacks, it can effectively detect tampering, which is termed as "Fragile watermarking". The Semi-fragile watermarking addresses the flaws of both watermarking systems. It stays robust by tolerating image processing attacks and fragile for intentional attacks.

As a result, the semi-fragile system has been suggested for content authentication, and tamper detection system.

Whereas, domain of embedding is classified as, spatial and transform domain. Compared with spatial domain, transform domain has highly suggested because spatial domain fails to provide robustness and imperceptibility. In the literature, various transform domain techniques are proposed, namely Discrete Fourier Transform (DFT), Discrete Cosine Transforms (DCT), Discrete Wavelet Transform [17], Integer Wavelet Transform (IWT) [18], Stationary Wavelet Transform (SWT) [19] and Dual-Tree Complex Wavelet Transforms [20], Discrete Contourlet Transforms [5], Curvelet Transform [21], [22]. Mathematical tool transform are namely Singular Value Decomposition (SVD) [23], QR Decomposition [19], and hybrid of these transform such as, DWT-SVD [24], DCT-DWT-SVD [25], Hilbert-IWT [26], DWT-BAM [27]. These transform domain techniques are able to process grayscale and colour images with qualities like imperceptibility, resilience, and capacity. However, they are unable to handle colour information in an image effectively. Thus, it's critical to preserve the colour information prior to watermark insertion. To preserve the color information, a color processing model has been suggested by many researchers, where the luminance Y component is considered when embedding the watermark. However, the disadvantage of ignoring the correlation between the color channels makes the system vulnerable towards color attacks. To avoid this problem, quaternion form of transform domain techniques is suggested. They are Quaternion Fourier Transform (QFT) [28], Quaternion Discrete Cosine Transform [29], Quaternion Wavelet Transform (QWT) [30], Quaternion Curvelet transforms [31], Quaternion Hadamard Transform [32], Quaternion Singular Value Decomposition [23] and so on. Hence, to treat the color images in a holistic manner the Quaternion Discrete Fourier Transform (QDFT) is introduced using quaternions algebraic form by Li et al. [28] for copyright protection system. An additional tensor decomposition is used after a third order tensor is created from the three imaginary frequencies of QFT. Odd-even quantization is the technique used to incorporate the watermark. To improve the extraction process accuracy, the geometric distortion is rectified by multiple output least squares support vector regression (MLS-SVR) network model and pseudo-Zernike moment features. The experimental results shown better imperceptibility and robustness for image-processing and geometric attacks. Multi-level and multi-region watermarking systems are presented in [33]. The experimental results show better results, but selecting an optimal SURF keypoint is a challenge as it affects the texture region during the extraction process.

Chen et al. [23] provided a strong dual-color watermarking that protects the copyright by utilising quaternion singular value decomposition (QSVD). This scheme resulted in highly correlated color channels through quaternion form which not only has strong anti-attack performance but is also robust

and imperceptible to some common attacks. In addition to the above, an optimal system needs to be adapted to balance and enhance the performance of watermark characteristics using popular optimization techniques in [17], [25], and [31].

In watermarking techniques, watermark embedding, extraction, and watermark generation are the major key aspects. Most of the researchers used a different image, logo, partial image feature as a watermark. For instance, different images like RGB, Grayscale, or binary logo, and partial image features like transform domain feature, block-based average feature values are generated as watermarks. Many studies have been conducted in the tamper detection field in recent years, revealing that, when compared to standard watermarking systems, deep learning-based watermarking systems are more efficient in the watermark embedding and extraction process due to their learnable traits [34]. As a result, watermark embedding and extraction are carried out using the CNN model to design a robust and blind watermarking system. Thirteen CNN layers have preprocessing, embedding, and extraction networks are designed. The binary watermark is properly embedded in the host images using an adaptive scaling factor. The resultant showed high invisibility and robustness against pixel-value change attacks and geometric attacks. Yu et al. [35] recently proposed a convolutional-neural-network (CNN)-based multiple residuals learning model for a robust median filter forensic approach using compressed JPEG and lower-size images. Reference [36] presented a deep learning-based image forgery detection technique focused on splicing and copy-move detection using CNN-SRM hybrid model. The resultant feature is then input into an SVM classifier using a feature fusion method. According to the aforementioned study, content privacy and tamper detection challenges received more attention in the CNN model than tamper location and recovery.

To solve the above issue, Lee et al. [21] proposed an algorithm that generates the watermark from CNN features due to its traits. The VGG16 network is suggested to generate a mask decoder using the high-frequency image traits. The low-frequency sub-band images and the root mean squared high-frequency feature are provided as input to the VGG16 model from the stationary wavelet transform. Using two batch-normalized inception-based mask deconvolutions followed by bilinear upsampling and two simple bilinear upsamplings, the mask decoder model detects tamper localization.

For more than a decade, convolutional neural networks (CNN), have demonstrated persistent success in the field of computer vision. In recent times, the Transformer model has demonstrated its dominance in numerous applications, such as text summarization and machine translation [22], [37]. The transformer model relies on self-attention rather than convolution, which is a straightforward parallel processing technique that outperformed the CNN models performance. Currently, Transformer models are intentionally designed for

NLP applications to address long term dependency issue. The essential self-attention mechanism that picks up on global traits, is what drives the evolution of the transformer model. Attention is the key advantage in transformer model, which is extended to deep learning models like convolution. By marrying both the models advantage together, it is termed as convolution-attention model. Reference [38] presented a tamper detection and recovery system using convolution attention model. Later from the transformer models the image classification model is proposed by only using the encoder block, which is a an image-specific model called Vision and Swin Transformer exists. Vision Transformers have recently demonstrated ground-breaking performance in a number of different tasks, including image classification [39], object detection [40], and semantic segmentation [41]. A transformer-based watermarking model was attempted by Palani and Loganathan [19] for tamper detection applications. Using the transformer model, multi-image watermark features are generated. The DWT-QR decomposition is used for embedding, and stationary wavelet transform (SWT) techniques are then applied. For attacks like image processing, better outcomes were obtained. While the authentication key is embedded in the LH sub-band of the singular value matrix, the most important six-bit vision feature maps are embedded in the Schur decomposition matrix. Because of its effective deep global features performance versus the tamper detection system, this system has demonstrated improved outcomes in terms of robustness and better-tampered image recoverability.

Later a compromised image recovery system, where a vision transformer-based multi-watermark feature map is used to detect localize tampering and to recover the region [18]. Embedding is done using schur decomposition and singular value decomposition transforms. As the feature learning process is limited to local patches, a hierarchical extended vision transformer, also referred as Swin transformer, produces deeper global characteristics than vision through shifted window traits. Due to its strong potential, we incorporated the Swin transformer model in the proposed work to extract the watermark feature map.

A. LIMITATIONS AND CONTRIBUTIONS

From the above literature, the flaws (or) limitations of the existing system are listed as,

- 1) Most of the existing quaternion-based color image watermarking systems suggested for robust watermarking system [23], [32].
- 2) Apart from robustness most of the system have not concentrated on content authentication, Tamper detection applications [29], [33].
- 3) Robust global characteristics of the original image have not been taken into consideration as a watermark by any of the existing methods.
- 4) The majority of HDR image research work focused on various Tone mapping attacks, with very few systems evaluated for unintentional attacks and none of the systems evaluated for intentional attacks [9].

- 5) Lack of research on HDR images in the areas of tamper detection, location and content authentication [8].

The above research results are widely explored in hand-crafted conventional watermark features generated through transform domain techniques. These features make it difficult to express the characteristics of the entire host image. The key factor for improving the robustness is by, generating a robust global watermark feature instead of embedding an image or a logo that depicts the characteristics of the host image. Taking this factor into consideration, convolution watermarking models were designed to embed and extract the watermark, and few systems employed convolution features or convolution masks as watermarks due to their feature learning traits. Thus, the deep learning-based watermarking system attained better results in terms of content privacy and tampered image prediction. The flaw with the convolution model is that the existing system learns local features using kernels that are connected locally. In contrast, the Transformer model learns features globally using self-attention for all the tokens. The difference between CNN and Transformer is the feature interacting mechanism [42]. The computational complexity increases when a large number of tokens are processed. To solve this issue, several recent works, such as Swin Transformer which works locally inside the patch window. We intend to generate robust deep global watermark features that express the characteristics of the original image. The Swin transformer model is effective in learning deep global features by shifting and windowing processes. The Swin transformer traits match our intention, which motivated us to incorporate the Swin-based optimal watermarking model in our proposed system.

In addition, to address the difficulties of HDR images, it's important to consider watermarking techniques specifically designed to handle the limitations and characteristics of JPEG and HDR-compressed JPEG images. Adaptive watermarking algorithms, robust feature extraction methods, and careful consideration of compression parameters and quality settings can help improve the performance and resilience of watermarks in these challenging scenarios. The contributions towards this work are listed below:

- 1) Design an adaptive and robust feature extraction watermarking techniques to handle both the color image and compressed HDR images against tamper detection and localization application.
- 2) Attempted Swin transformer-based watermark generation model for the first time on color and HDR images, which extracts invariant global feature maps as watermarks to achieve robustness against intentional and unintentional attacks and achieved better results for tamper detection applications.
- 3) The optimal region is determined using a novel maximum entropy random walk algorithm.
- 4) Quaternion dual complex tree wavelet transform technique is derived to maintain the color information intact and also to embed the watermark, where the correlation among color channels is maintained.

- 5) Payload is improved by embedding the watermark features repeatedly to enhance the tamper detection accuracy.
- 6) The effectiveness and efficiency of the Swin features is evaluated in terms of robustness and imperceptibility metrics for various attacks on both RGB and HDR images.

B. ARTICLE ORGANIZATION

The rest of the article is organized as: Section II discussed the materials and methods, an overview of proposed model is discussed in Section III. Section IV explicates the Experimental Results and comparative Analysis in Section V. Finally concludes with future scope in Section VI.

II. MATERIALS AND METHODS

A. SWIN TRANSFORMER

Swin transformer [41], [43] is a hierarchical mechanism that processes images using the concept of shifted windowing. Fixed image patches are inappropriate attributes for image processing transformer models since it increases the number of tokens for high-resolution images, resulting in quadratic computational complexity. Swin transformers are built to avoid such issues where a hierarchical dense feature maps are generated by combining image patches represented by window shifting traits, where the model’s performance is limited by self-attention computation within the local window as well as cross windowing connections. In addition, swin transformer has the advantages of hierarchy, locality, translation invariance and an inductive bias that suits for task targeting [44]. Therefore, it is suitable for large datasets that generalize the process by learning features. The architecture selected in this paper, is the Swin Transformer Base, and its structure is shown in Fig. (2).

According to swin transformer block as in Fig. (2), first the original image is divided into patches of size 4×4 through the patch partition module where each patch ‘ X_i ’ treated as “tokens”. The features of each channel are linearly embedded by flattening it with dimension $\frac{H}{4} \times \frac{W}{4} \times C$. After linear embedding four stages of Swin modules with self-attention are applied on each patch to construct different size feature maps. So, the number of original image patches and the linear embedding layer together referred as “Stage 1”. Apart from Stage 1, the last three stacked Swin modules are downsampled by patch merging layer which reduce the number of tokens into halves. This first hierarchical patch merging layer combine the features into 2×2 patches and apply linear layer which downsample the number of tokens by $2 \times C$. The Swin transformer block is applied on resultant patch merged tokens with resolution of $\frac{H}{8} \times \frac{W}{8}$ represented as “Stage 2”. The process was repeated twice in stage 3 and stage 4 with patch dimension of $\frac{H}{16} \times \frac{W}{16}$ and $\frac{H}{32} \times \frac{W}{32}$. The Swin transformer blocks comprises of four components namely, multilayer perceptron (MLP), window multi-head self-attention (W-MSA), shifted

window-based multi-head self-attention (SW-MSA), layer normalization (LN) and Drop-Path. The W-MSA and SW-MSA are the key components of Swin used together for each Swin module. The LN will normalize the feature distribution data and Drop-path has regularization effect. Unlike multi-head attention (MSA) layer, shifted-MSA is employed along with two MLP layer where GELU activation function is induced between the MSA and MLP layer. The SA computation process is provided in Eq. (1), (2) and (3). The Key (K), Value (V), Query (Q) computed using Equation 1,

$$Q_i = W_q P_i, K_i = W_k P_i, V_i = W_v P_i \quad (1)$$

where W_q, W_k, W_v are the trainable parameters, which will be the same for all the input patch vector $X_{i:n}$ sequence. N number of heads referred as self-attention mechanism is combined together as multi-head attention (MHA) shown in Eq (4).

Scaling Dot product (Q, K, V)

$$= \frac{K_{i:n}^T Q_i}{\sqrt{d_K}} \quad (2)$$

$$Self - Attention(SA_i) = softmax \left(\frac{K_{i:n}^T Q_i}{\sqrt{d_K}} + B \right) V_i, \quad (3)$$

where $i=1,2, \dots n$

$$MultiHead Attention(MHA) = Concat(SA_1 + SA_2 +, \dots, SA_n) \quad (4)$$

where B represents relative position code, d represents query dimension. The W-MHA process is given in equation (5) and (6),

$$\hat{P}^l = Droppath(W - MHA(LN(P^{l-1}))) + P^{l-1} \quad (5)$$

The attention weight score is given as input MLP classifier to obtain the learned feature maps as in Eq.(6),

$$P^l = Droppath \left(MLP \left(LN \left(\hat{P}^l \right) \right) \right) + \hat{P}^l, l = 1, 2 \dots L \quad (6)$$

Within W-MSA, the global interaction is limited as self-attention is calculated within the local window. The S-MHA is used to divide the L layer by moving half of the window to the L+1 layer and then dividing it again in order to prevent this. The local window attention weight will interact globally as a result of doing this. The computation is given in the equation below,

$$\widehat{P}^{l+1} = Droppath \left(MLP \left(LN \left(P^l \right) \right) \right) + P^l \quad (7)$$

$$P^{l+1} = Droppath \left(MLP \left(LN \left(\widehat{P}^{l+1} \right) \right) \right) + \widehat{P}^{l+1} \quad (8)$$

From the trained Swin Transformer model, robust features are obtained to enhance the robustness of the proposed watermarking scheme.

B. MAXIMUM ENTROPY RANDOM WALK (MERW) ALGORITHM

Graph emerged from the spectral graph theory concept, representing the relation between objects or samples. Graph spectral theory is linked with signal processing from where it is stimulated to graph-based transform. Due to their ability to mathematically model the interconnections between pixels, they have recently gained adoption in computer vision applications. Applications like Citation analysis, social networks, and link-structure analysis have highly benefited from using graph-based algorithms. Edge identification and segmentation are a few more image processing challenges that have been solved using graph-based techniques [45].

A graph algorithm determines a vertex’s importance, power, or energy inside a graph by considering global information rather than local vertex-specific information. Mihalcea [46] has shown adapting graph-based ranking technique such as Kleinberg’s HITS algorithm (Kleinberg, 1999) or Google’s PageRank (Brin and Page, 1998) for the weighted and unweighted graph.

Random walks (RW) and maximal entropy random walk (MERW) are the suitable algorithms for the undirected image which has been successfully adopted for image segmentation [47], link prediction [48], and object detection [49], object localization [50], visual saliency region [51], tampering detection [52]. Since MERW assigns the uniform probability distribution of all pathways in a given graph by globally maximising entropy, it performs better than Random Walk; Whereas, RW selects uniform probability distribution for every vertex among its outgoing edges which the maximizes the entropy rate locally. MERW is a popular random walk algorithm that works based on the transition probabilities which are chosen according to the maximum entropy principle. The distribution of transition probability on graph structure indicates the region’s importance in terms of maximum entropy. In the proposed model, we utilised the MERW algorithm to determine the adaptive region due to its localization property on a graph. Graph is represented as $G = \{V, E\}$, where V & E represent vertices and edges, $V = \{v_0, v_1, \dots, v_n\}$, where $V \in R^{i \times j}$ and $E = \{i, j\}$, where $\{i, j\} \in V$. The graph method is extended to images by representing each image pixel as vertex V and interconnection between pixels represented as edge E . The adjacency weight matrix A_{ij} defined from the connectivity of two vertex through edges E (i.e., incoming and outgoing links between two vertices) which is computed as follows:

$$A_{ij} = \begin{cases} (a_{ij}, & (i, j) \in E \\ 0, & otherwise \end{cases} \quad (9)$$

The random walk refers to a random walker $\{R_n\}$, where $n \in 0, 1 \dots$. The random walker moves in random transition from current node $i \in V$ to neighbour node $j \in V$ based on transition probability ‘ P ’ where $P_{ij} : P(R_{n+1} = j | R_n = i)$. The MERW method hopping is a Markov process that hops

longer m-steps from one node i to the other node j . As a result, the walker distribution towards maximal entropy achieves the MERW global structure. The suggested model discovers the maximal entropy region based on the transition probability distribution in a graph. The average entropy rate of the stochastic transition process can be calculated by taking the probability of visiting each vertex and averaging it using Shannon entropy equation:

$$E = - \sum_{i=1}^n \rho_i^* \sum_{j=1}^n P_{ij} \log P_{ij} \quad (10)$$

The transition probability matrix $\{P_{ij}\}$ is computed using equation (11)-(16). From the real symmetric matrix L , non-negative eigenvalues and eigenvector are generated as:

$$L = \varphi \lambda \varphi^T = \sum_i^n \varphi_i \lambda \varphi_i^T \quad (11)$$

where $\lambda = \{\lambda_1, \lambda_2 \dots \lambda_n\}$ are eigenvalues and $\varphi_i = \{\varphi_1, \varphi_2, \dots, \varphi_n\}$ are the corresponding eigenvector. The graph laplacian matrix L is obtained by $L = D - A_{ij}$, where D is a diagonal matrix, where i^{th} diagonal element d_i is sum of all edge weights of vertices V . Then transition probability matrix are given by,

$$P_{ij} = \frac{A_{ij} \varphi_i}{\lambda \varphi_j} \quad (12)$$

where λ represents eigenvalue of the weight matrix, and φ_i, φ_j denote the i^{th} and the j^{th} node corresponding eigenvector. According to the Frobenius-Perron theorem, both the φ and λ are non-negative sign, so that one can choose $\varphi_i > 0$. The stationary probability distribution $\rho_i(t)$, find the particle at node i at time t generally through stationary state ρ_i^* using the equation:

$$\rho_i^* = \sum_j \rho_j^* P_{ji} \quad (13)$$

Due to the non-negative weight matrix, all the eigen vectors are normalised to $\sum \varphi_i = 1$, so the transition matrix are also normalized. Therefore, the stationary distribution of finding the walker at node i by the following equation:

$$\rho_i^* = \varphi_i^2, \quad (14)$$

where $\rho_i^* : i \in V$ and $\sum_i \rho_i^* = 1$. Further the quantity of interest in generating the probability P of trajectory $\delta_{v_0 v_n}^t$ of length t passing through nodes (v_0, v_1, \dots, v_n) .

$$P(\delta_{v_0 v_n}^t) = (P_{v_0 v_1}, P_{v_1 v_2}, \dots, P_{v_{t-1} v_n}) \quad (15)$$

The transition probability distribution of trajectory ‘ $P(\delta_{v_0 v_n}^t)$ ’

$$P(\delta_{v_0 v_n}^t) = \frac{1}{\lambda^t} \frac{\varphi_i}{\varphi_j} \quad (16)$$

Start the random walk process from a specific region and iteratively perform the random walk using the transition probabilities. At each step, the random walk can transition

to neighbouring regions based on the probabilities defined in the adjacency matrix. After several iterations, the random walk process will converge, and you can compute the MERW scores for each region. The MERW scores represent the adaptive high-energy values of the regions, indicating their suitability for watermark embedding. Based on the MERW scores, identify the regions with the highest scores, as these regions are considered adaptive high-energy regions which are suitable for embedding watermarks.

C. QUATERNION FORM

For color image watermarking systems, most of the system utilised Quaternion form of representation mathematical notion suggested by Huang et al. in 1843 [53]. Specifically in watermarking, the frequency domain technique is the first preprocessing step in acquiring more information to further process the data, to embed the watermark either in a single or in a RGB channel. When the watermark is embedded either in single color channel or an entire color image, it losses the correlation among the channels which affects the performance of robustness. To avoid this, the quaternion form of image representation is taken as a preprocessing step before embedding the watermark. Quaternion form of an image is suggested as a solution where it converts it to a vector field. Generally, the quaternion form is expressed as a sum of real and vector form,

$$q = \mu + qi + qj + qk \tag{17}$$

where μ represents real part (q_{real}), and $qi + qj + qk$ as imaginary part, which satisfies $i^2 = j^2 = k^2 = ijk = -1$. The pure quaternion form expressed as $q = qi + qj + qk$, when $q_{real} = 0$, if q has unit norm $q=1$, then q is unit quaternion form.

$$q = \sqrt{q \cdot q^*} = \sqrt{q^2 r} + \sqrt{q^2 i} + \sqrt{q^2 j} + \sqrt{q^2 k} \tag{18}$$

For color image the quaternion form is expressed as,

$$q = \mu + q_R(x, y)i + q_G(x, y)j + q_B(x, y)k \tag{19}$$

where $q_R(x, y)$, $q_G(x, y)$, $q_B(x, y)$ represent the red, green, and blue channels pixels respectively. The inverse quaternion form is just the conjugate of general quaternion form which can be represented as,

$$q = \mu - qi - qj - qk \tag{20}$$

Color image inverse quaternion form is represented as,

$$q = \mu - q_R(x, y)i - q_G(x, y)j - q_B(x, y)k \tag{21}$$

D. DUAL TREE COMPLEX WAVELET TRANSFORM (DTCWT)

Among those transform domain techniques, discrete wavelet transform has been suggested the most in the existing works due to its efficiency. However, it fails to achieve the shift-invariance and directionality properties resulting in performance degradation in some cases. To address this issue, a Dual-Tree Complex Wavelet Transform is developed by

Kingsbury [54] having many advantages like shift-invariance, strong directional selectivity, limited redundancy, and perfect reconstruction property with less computing complexity. The DTCWT complex values are obtained from DWT and complex wavelet transform. The inverse DTCWT is performed where the real and imaginary parts are conjugated to form real signal. By averaging the two real signals the final reconstruction is attained. First, the translation and dilation sequence is performed on image $I(x, y)$ using complex scaling function $\varphi(x, y)$ resulting six complex wavelet function $\Psi_k, l^\theta(x, y)$. For an image $I(x, y)$ the DTCWT is expressed as,

$$I(x, y) = l_{\in Z} R_{k0,l} \Psi_{k,l}^\Theta(x, y) + \sum_{\Theta} \sum_{k \geq k_0} \sum_{l \in Z^2} W_{k,l}^\Theta \psi_{k,l}^\theta(x, y) \tag{22}$$

where Θ represents six complex wavelet directionalities, $\Theta \in (\pm 15^\circ, \pm 45^\circ, \pm 75^\circ)$, Z represents natural number, k and l are shift and dilation, $W_{k,l}$ represents complex wavelet coefficients with complex $\vartheta_{k0,l}(x) = \vartheta_{k0,l}^{real}(x) + \sqrt{-1} \vartheta_{k0,l}^{img}(x)$, and $\vartheta_{k,l}(x) = \vartheta_{k,l}^{real}(x) + \sqrt{-1} \vartheta_{k,l}^{img}(x)$. So, the DTCWT applied on the image $I(x, y)$ produce two complex low frequency sub-band and six direction high frequency sub-bands expressed as shown in Eq. (23) and (24)

$$DTCWT(I(u, v)) = Low\ frequency(L_1) + High\ frequency(H_1) \tag{23}$$

$$DTCWT(I(u, v)) = x(\forall, L_1, u, v) + x(\gamma, H_1^\theta, u, v) \tag{24}$$

The above Eq. (24) is elaborated as,

$$x(\forall, L_1, u, v) = R(x(\forall, L_1, u, v)) + kImg(x(\forall, L_1, u, v)) \tag{25}$$

$$x(\gamma, H_1^\theta, u, v) = R(x(\gamma, H_1^\theta, u, v)) + kImg(x(\gamma, H_1^\theta, u, v)) \tag{26}$$

where R and Img stands for real and imaginary parts, \forall represents number of decomposition levels, here it is 1, L_1 represents Low frequency sub-bands, u, v represents coefficients sub-band locations where the values range between 0 to $\frac{N}{2^*}-1$. Thus, the result of DTCWT has shown good shift-invariance, selection strong directionality, with less redundancy and also improved the efficiency of imperceptibility than any other wavelet transforms.

E. QUATERNION DUAL-TREE COMPLEX WAVELET TRANSFORM

The advantage of both quaternion and dual-tree complex wavelet transform are combined as expressed in the given equation from (27) – (32). The color image ‘i’ RGB channel is denoted as q_R, q_G, q_B . The QDTCWT is obtained by

substituting Eq. (24) in Eq. (19) which is expressed as,

$$QDTCWT(q) = I(m, n) = \mu + DTCWT(q_R(m, n))i + q_G(m, n)j + q_B(m, n)k \quad (27)$$

$$QDTCWT(q) = \mu + DTCWT(q_R(m, n))i + DTCWT(q_G(m, n))j + DTCWT(q_B(m, n))k \quad (28)$$

Now substitute μ in Eq. (28),

$$QDTCWT = [re(DTCWT(q_R(m, n))) + Img(DTCWT(q_R(m, n)))]i + [re(DTCWT(q_G(m, n))) + Img(DTCWT(q_G(m, n)))]j + [re(DTCWT(q_B(m, n))) + Img(DTCWT(q_B(m, n)))]k \quad (29)$$

where $\mu = \alpha*i + \beta*j + \gamma*k$ denotes a unit quaternion subject to the constraint that $\mu^2 = -1$ where i, j, k are real numbers. The above Eq. (29) is same as in [31] and [55], so we have used the four-vector space quaternion dual-tree complex wavelet transform which is expressed as,

$$QDTCWT(q(m, n)) = Q_0(m, n) + Q_1(m, n)i + Q_2(m, n)j + Q_3(m, n)k \quad (30)$$

where

$$Q_0(m, n) = -\alpha Img(DTCWT(q_R)) - \beta Img(DTCWT(q_G)) + \gamma Img(DTCWT(q_B)),$$

$$Q_1(m, n) = re(DTCWT(q_R)) + \gamma Img(DTCWT(q_G)) - \beta Img(DTCWT(q_B)),$$

$$Q_2(m, n) = re(DTCWT(q_G)) + \alpha Img(DTCWT(q_B)) - \gamma Img(DTCWT(q_R)),$$

$$Q_3(m, n) = re(DTCWT(q_B)) + \beta Img(DTCWT(q_R)) - \alpha Img(DTCWT(q_G))$$

Here, $re(x)$ denotes the real part of the usual complex number x , and $Img(x)$ denotes the imaginary component. The conventional DTCWT matrix of the red, green, and blue channels is represented by the symbols $DTCWT(q_R)$, $DTCWT(q_G)$, $DTCWT(q_B)$, respectively. For the sample image the quaternion dual-tree complex wavelet transform of red channel is visualized in Fig. (1). The inverse quaternion dual-tree complex tree wavelet transform (IQDTCWT) can be expressed as,

$$IQDTCWT(q(m, n)) = re(IDTCWT(Q_0(m, n))) + \mu Img(IDTCWT(Q_0(m, n))) + [re(IDTCWT(Q_1(m, n))) + \mu Img(IDTCWT(Q_1(m, n)))]i + [re(IDTCWT(Q_2(m, n))) + \mu Img(IDTCWT(Q_2(m, n)))]j + [re(IDTCWT(Q_3(m, n))) + \mu Img(DTCWT(Q_3(m, n)))]k \quad (31)$$

$$IQDTCWT(q(m, n)) = f(Q_0)(m, n) + f(Q_1)(m, n)i + f(Q_2)(m, n)j + f(Q_3)(m, n)k \quad (32)$$

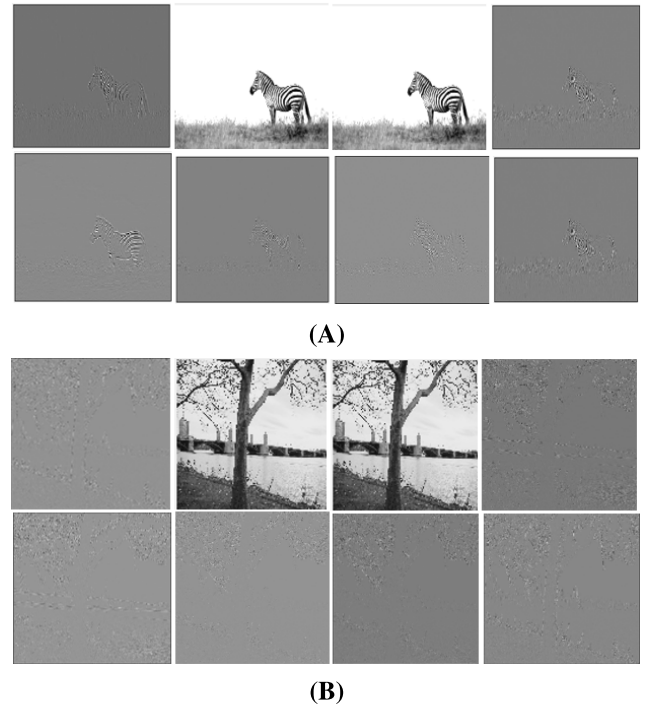


FIGURE 1. Red Channel of Quaternion Dual-Tree Complex Wavelet Transform (A) CASIA dataset images - Image 7, (B) LVZ-TMO HDR images - Image 7.

where,

$$f_{Q_0}(m, n) = re(IDTCWT(Q_0)(m, n)) - \alpha Img(IDTCWT(Q_1)(m, n)) - \beta Img(IDTCWT(Q_2)(m, n)) - \gamma Img(IDTCWT(Q_3)(m, n)),$$

$$f_{Q_1}(m, n) = re(IDTCWT(Q_1)(m, n)) + \alpha Img(IDTCWT(Q_0)(m, n)) + \gamma Img(IDTCWT(Q_2)(m, n)) - \beta Img(IDTCWT(Q_3)(m, n)),$$

$$f_{Q_2}(m, n) = re(IDTCWT(Q_2)(m, n)) + \beta Img(IDTCWT(Q_0)(m, n)) + \alpha Img(IDTCWT(Q_3)(m, n)) - \gamma Img(IDTCWT(Q_1)(m, n)),$$

$$f_{Q_3}(m, n) = re(IDTCWT(Q_3)(m, n)) + \gamma Img(IDTCWT(Q_0)(m, n)) + \beta Img(IDTCWT(Q_1)(m, n)) - \alpha Img(IDTCWT(Q_2)(m, n))$$

III. PROPOSED MODEL

In this work, a novel optimal watermarking scheme using a Swin transformer and maximum entropy random walk algorithm is suggested for HDR image-based tamper detection applications. Normally, the HDR image is composed of detailed pixel information in terms of luminance, brightness, and texture features. The HDR images are compressed to JPEG format by a mobile camera. The existing color image watermarking system will not be efficient in handling HDR-compressed JPEG images. As a solution, quaternion DTCWT and SVD transform model. Apart from embedding a watermark in an optimal region without influencing

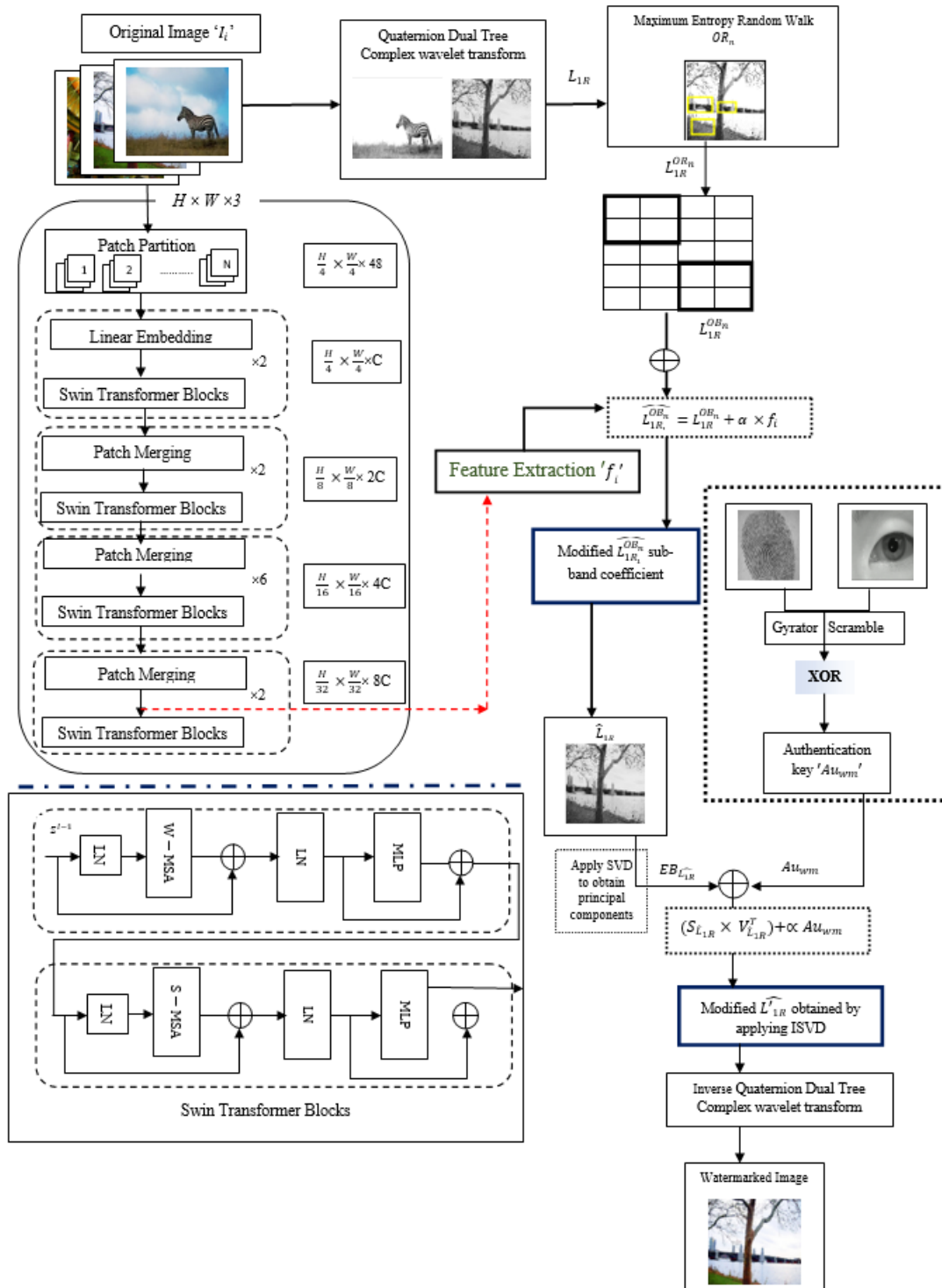


FIGURE 2. Detailed Embedding process.

imperceptibility, generating a robust watermark is also a major criterion in the watermarking technique. The proposed model attempted the Swin base transformer model which is highly invariant to translation and produces dense global watermark features of an original image, that enhance the

system to attain high robustness against various attacks than the existing traditional watermark generation techniques. Apart from a robust watermark, embedding the watermark in an optimal region is determined by the maximum entropy random walk (MERW) region.

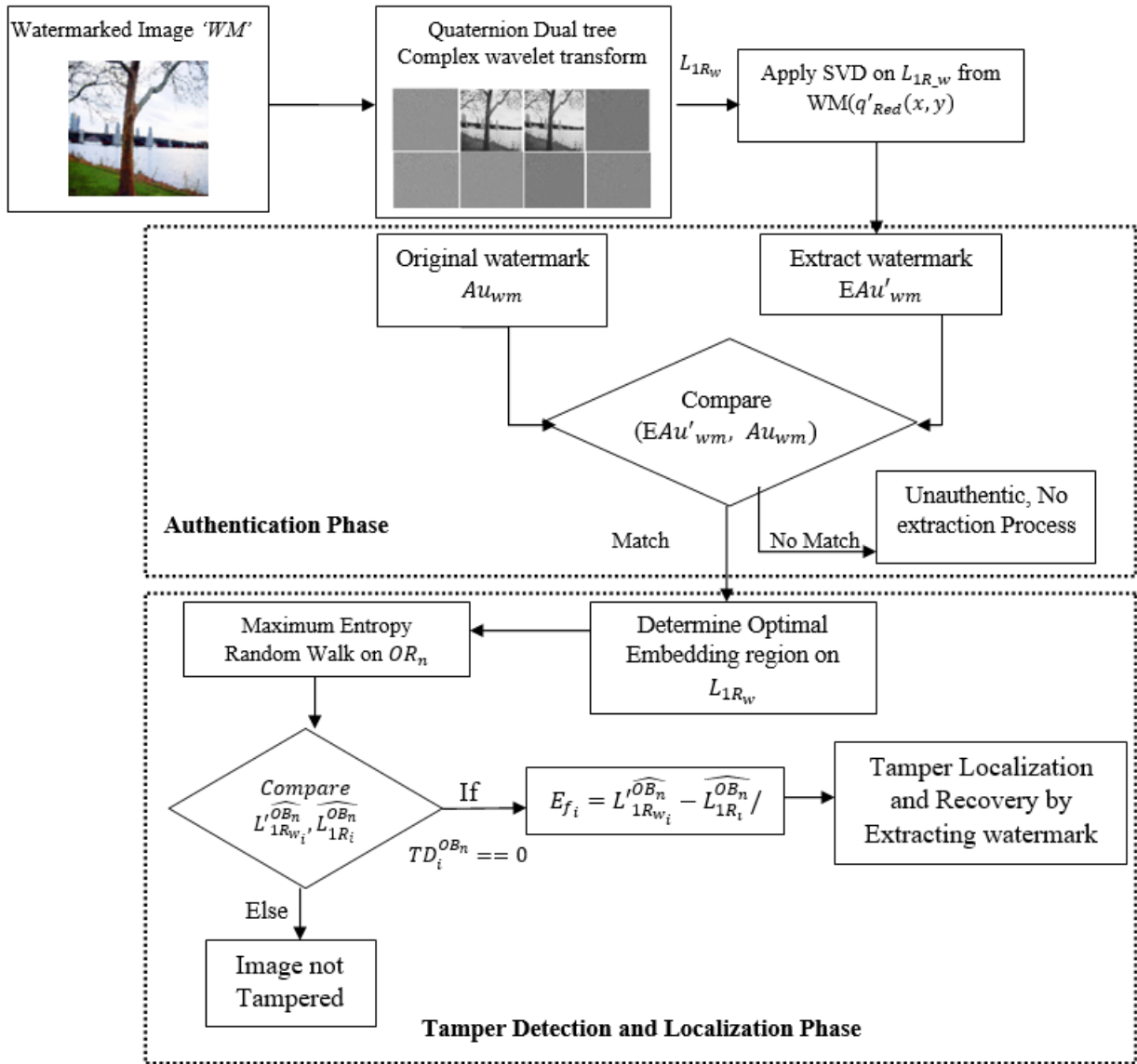


FIGURE 3. Extraction process.

A. DETAILED PROCESS

The proposed model embedding process is clearly illustrated in Fig. (2) and the extraction process is depicted in Fig. (3). The original image is pre-processed by converting the pixel to quaternion form which can maintain the correlation among the color channels, followed by that Dual-tree complex wavelet transform is applied on the quaternion vector coefficient to produce two low frequency (L_{1R}, L_{1Im}) and ($H_1, H_2, H_3, H_4, H_5, H_6$) six high-frequency sub-bands. To embed the watermark in an optimal region, a novel maximum entropy random walk algorithm is employed to determine the low-sensitive region. The random walker ' R_n ' search the maximum entropy region by the transition probability score of each pixel. Once the high MERW determines the optimal region ' OR_n ' through stochastic

distribution score, those regions are bounded to further embed the watermark ' OB_n '. The robust global deep feature watermark is generated from the Swin transformer model where the dataset is trained, to learn the feature. From the pre-trained model, the original image features are obtained through class labels. Before embedding the watermark, QSVD is used as in [23], where the Swin features are partitioned according to the optimal block size and embedded in the principal component of QSVD transform using the scaling factor. In the proposed work $\alpha = 0.063$ is found to be an appropriate embedding factor. Inverse SVD is performed on the modified low-frequency sub-band L'_{1R} . Following that, the secured content authentication watermark is obtained from the dual gyrator scrambled image which is embedded in the imaginary low frequency sub-bands.

The extraction process in the reverse of embedding process where the gyrator scrambled watermark is recovered from the watermarked. The regenerated authentication watermark is compared to the extracted authentication watermark; if they match, the extraction operation can continue; otherwise, extraction process not allowed.

To detect the tampering of the watermarked image, the principal component EB''_{OB_n} is extracted from low frequency sub-band of QDTCWT from the tampered image and the original principal components EB'_{OB_n} are extracted from the watermarked image. By comparing both applied to extract the principal components tamper localization is detected and tampered region is recovered using swin features.

Algorithm

Process 1: Watermark embedding process

- 1) Apply quaternion dual-tree complex wavelet transform on input image I_i , and the red channel quaternion dual-tree complex wavelet transform is expressed as,

$$[L_1R, L_1Img, (H_11R, H_11Img, H_12R, H_12Img, H_13R, H_13Img, H_14R, H_14Img, H_15R, H_15Img, H_16R, H_16Img)] = QDTCWT(q_R(m, n)) \quad (33)$$

where L_1, H_1 refers low and high frequency sub-bands, R and Img refers real and imaginary part coefficient

- 2) Determination of suitable embedding region using MERW on each channel of LL sub-band:

- a) *Graph Construction:* Construct a graph representation for L_1 sub-band image, $G(L_{1R}) = V, E$ represent the graph, where $V = v_0, v_1, v_2 \dots v_n, v_n \in I(m,n), E = E_0, E_1 \dots E_n, E_n \in V$
- b) *Determine the energy value:* Energy value referred as $E(v_i)$ computed by Eq. (10) where $E(v_i)$ represent scalar function that map node v_i to its energy value in the graph.
- c) Compute the transition probabilities ' P_{ij} ' of each node v_i computed using Eq. (12).
- d) Normalize the transition probabilities ' P_{ij} ' and construct the transition matrix T, $P_{ij} \in I(m,n)$

$$T(i, j) = \frac{P_{ij}}{\Sigma} P(v_i, A_{ij}(v_i)) \quad (34)$$

here A_{ij} denotes the adjacent matrix value node of v_i . 'T' matrix size depends on the number of nodes in the graph.

- e) Initialize random walker R_n and perform random walk process from a seed pixel or set of seed pixels.
- f) The stationary probability $p_i(t)$ from node v_0 to node v_n is determined as explained in II(B).
- g) After several iterations of the random walk, the MERW scores for each pixel.
- h) Let $S(v_i)$ represent the MERW score of node v_n . The MERW score can be calculated using the

probability values obtained from the random walk process:

$$S(v_i) = -\log(T(v_i)) \quad (35)$$

The negative logarithm of the probability represents the entropy or uncertainty associated with being at node v_i .

- i) Apply a threshold to the MERW scores to determine the block boundaries. Pixels with high MERW scores are considered part of the optimal region OR_b , while pixels with low scores are considered high sensitive region.

$$OR_n = M_S(v_i, v_j) \geq t, i, j \in I(m, n) \quad (36)$$

where t represent average threshold values of MERW score

- j) Embedding blocks are selected as ' OB_n ' from the optimal region ' OR_n '
- 3) Robust swin watermark feature generation
 - a) Train the dataset images I_i on the Swin transformer.
 - b) Divide the image into an even number of n-sized patches.
 - c) Perform linear embeddings by flattening the image patches.
 - d) The linear input layer is inputted to $stage_1$ Swin transformer module.
 - e) In stage 2, all the patches are combined in patch merging layers followed by swin module.
 - f) Hierarchical four stages of features are learnt and robust feature maps are extracted from the last layer f_i from the pre-trained model.
- 4) Select the optimal embedding block OB_n and embed the watermark f_i in selected optimal embedding block OB_n coefficients,

$$\widehat{L_{1R_i}^{OB_n}} = \widehat{L_{1R}^{OB_n}} + \alpha \times f_i \quad (37)$$

where i represent number of selected blocks where the features blocks are embedded accordingly.

- 5) Preprocessing watermark dual biometric image b_1 and b_2
- 6) Dual owner biometric image b_1, b_2 scrambled with gyrator scrambling method.

$$Scramble_1 = gyrator(b_1) \quad (38)$$

$$Scramble_2 = gyrator(b_2) \quad (39)$$

- 7) Final authentication watermark is generated using XOR method.

$$Au_wm = XOR(Scramble_1, Scramble_2) \quad (40)$$

- 8) The authentication watermark is embedded only in the red channel L_{1R} sub-band principal component coefficient of SVD transform

$$[U_{\widehat{L_{1R}}}, S_{\widehat{L_{1R}}}, V_{\widehat{L_{1R}}}^T] = SVD(\widehat{L_{1R}}) \quad (41)$$

- 9) Modified L'_{1R} coefficient is obtained by embedding the watermark 'Au_{wm}' in low frequency imaginary sub-band and inverse SVD transform applied as expressed as,

$$EB'_{L'_{1R}} = (S_{L'_{1R}} \times V_{L'_{1R}}^T) + \alpha Au_{wm} \quad (42)$$

$$L'_{1R} = ISVD[U_{L'_{1R}}, EB'_{L'_{1R}}, V_{L'_{1R}}^T] \quad (43)$$

- 10) The embedding process from step (1 – 4) is carried out for all the three channels and the modified QDTCWT is obtained.

$$IQDTCWT[(\widehat{L'_{1R}}, L_{1Img}, (H_{11R}, H_{11Img}, H_{12R}, H_{12Img}, H_{13R}, H_{13Img}, H_{14R}, H_{14Img}, H_{15R}, H_{15Img}, H_{16R}, H_{16Img}))] = ModifiedCoefficient(q'_R(m, n)) \quad (44)$$

- 11) Finally, the watermarked image 'WM' is obtained using inversing the QDTCWT as expressed in Eq (32).

Process 2: Watermark extraction process

- 1) Apply QDTCWT on the received watermarked image 'WM' as performed similarly in the embedding process, where the below Eq. (45) express the QDTCWT of red channel,

$$[L_{1Rw}, L_{1Imgw}, (H_{11Rw}, H_{11Imgw}, H_{12Rw}, H_{12Imgw}, H_{13Rw}, H_{13Imgw}, H_{14Rw}, H_{14Imgw}, H_{15Rw}, H_{15Imgw}, H_{16Rw}, H_{16Imgw})] = QDTCWT(WM(q_{Red}(x, y))) \quad (45)$$

- 2) To verify the content authentication by extracting the watermark from the principal component of low frequency imaginary sub-band L'_{1Img}

$$[U'_{L'_{1Rw}}, S'_{L'_{1Rw}}, V_{L'_{1Rw}}^T] = SVD(L_{1Rw}) \quad (46)$$

$$EAu'_{wm} = S'_{L'_{1Rw}} \times V_{L'_{1Rw}}^T \quad (47)$$

- 3) Original authentication watermark is regenerated from the owner's original biometric image by repeating the embedding process steps from (5)-(7)

$$A = Compare(EAu'_{wm}, Au_{wm}) \quad (48)$$

If A == true
 Content authenticated
 Do
 Extraction process
 Else
 Watermarked image is unauthenticated and no extraction process

- 4) To extract the tamper detection and recovery watermark, the optimal region 'OB_n' is determined by performing MERW algorithm in 'L_{1Rw}' sub-band as in step 2 of embedding process algorithm.

- 5) Swin features are extracted from the optimal region blocks $L'_{1Ri}{}^{OB_n}$ from the low frequency sub-band L_{1Rw} ,

$$L'_{1Rw_i}{}^{OB_n} = MERW(L_{1Rw}) \quad (49)$$

$$TD_i^{OB_n} = Compare(L'_{1Rw_i}{}^{OB_n}, \widehat{L'_{1Ri}}{}^{OB_n}) \quad (50)$$

If $TD_i^{OB_n} == 0$
 Image Not tampered, extraction process not required
 Else
 Tamper detected and extract watermark

- 6) Extract the watermark from the watermarked low frequency band to localize

$$E_{fi} = L'_{1Rw_i}{}^{OB_n} - \widehat{L'_{1Ri}}{}^{OB_n} / \alpha \quad (51)$$

- 7) Extracted swin features are combined to produce extracted watermark which has the ability to recover the tampered region.

IV. EXPERIMENTAL RESULTS

The performance of the proposed algorithm against various attacks where the watermarking characteristics efficiency is evaluated in terms of PSNR, SSIM, HDR-VDP, and NCC. This section deals with experimental setup, dataset description, evaluation metrics, performance analysis and comparative analysis.

A. EXPERIMENTAL SETUP

The experiment is carried out in Windows 11th Gen Intel(R) Core(TM) i5-1155G7 with processor speed @ 2.50GHz 2.40 GHz and 8GB capacity of RAM with 64-bit processor. The software used for simulation purpose is MATLAB R2021a. In order to generate watermark, the transformer model is trained in google colab platform with T4 GPU hardware accelerator.

B. DATASET DESCRIPTION

To evaluate the performance of the proposed algorithm, the experiment is conducted on the selected dataset LVZ-TMO (HDR images) [56], CASIA [57] and benchmark dataset (Color images). So, the LVZ-TMO dataset [56] contains 457 HDR images with four classes namely, indoors with 71 images, nature with 173 images, nighttime with 80 images, and river-side sunset with 133 images, with dimension 1024 × 1024 in compressed JPEG format, CASIA₁ dataset has 1721 images belongs to two classes with image dimension 384 × 256 and the benchmark images like crown, boat, lena, aeroplane, and Peppers with dimension of 256 × 256, and 512 × 512. The original images and the augmented images totally 10,990 images are inputted, in order to train and to extract features from the swin transformer model.

To authenticate the watermarked image, biometric images are utilized from FVC2002 [58] and MMU iris



FIGURE 4. Sample Test Images (A) LVZ- TMO HDR images (B) Benchmark Images (C) CASIA dataset images.

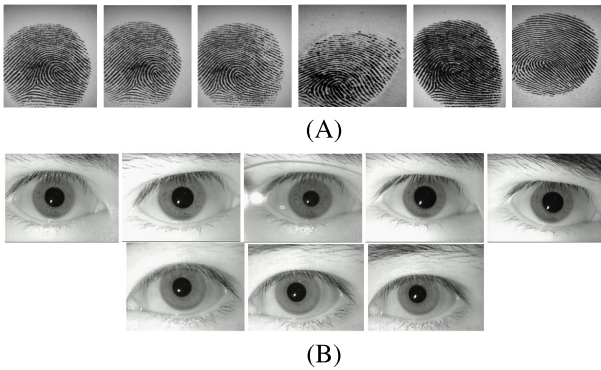


FIGURE 5. Owners Sample biometric images (A) FVC-DB2004 (B) MMU Iris dataset.

dataset [59] (Biometric images). The FVC2002 dataset contains 880 images of size 288×384 (108 Kpixels), and iris datasets named Multimedia University (MMU1) [59] contains merely 460 images with dimension 320×240 in '.bmp' format respectively. The sample test images from LVZ-TMO and CASIA₁ dataset is shown in Figure (4). Figure (5)(A) shows a sample biometric image from FVC2002-DB_{2B} and MMU iris dataset that are listed in Figure (5)(B).

C. EVALUATION METRICS

To protect HDR and RGB images from tampering, the optimal semi-fragile watermarking technique is designed using the swin transformer model. The performance analysis was conducted with respect to robustness, imperceptibility,



FIGURE 6. Performance evaluation of Zero attack watermarked test images using PSNR, SSIM and HDR-VDP.

and authenticity. Common metrics to assess image quality in terms of imperceptibility are the Peak Signal-to-Noise Ratio (PSNR) and the Structural Similarity Index Measure (SSIM). Bit Error Rate (BER) and Normalised Correlation Coefficient (NCC) are the most widely used measures to quantify robustness.

1) IMPERCEPTIBILITY

Generally, the embedded watermark shouldn't degrade the quality of the original image, which is why it has imperceptibility properties. In assessing the imperceptibility The metrics used to assess the degree of resemblance between the original and watermarked image are the peak-signal-to-noise ratio and the Structure resemblance Index (SSIM).

- 1) Peak signal to noise ratio (PSNR) [31]: The PSNR metric's performance is determined by the degree of similarity between the original and watermarked images, with a minimum value of 30 dB, indicating a significant range. Lesser than 30 dB is considered as less imperceptible which means the embedded watermark has influenced the watermarked image more. Whereas SSIM range between 0 to 1, closer to 1 is of acceptable range. The PSNR and SSIM metrics are computed through the below given equations,

$$PSNR = 10 \log_{10} \left(\frac{255^2}{\left[\sum_{M,N} (I(m,n) - WM(m,n))^2 \right] \frac{1}{M*N}} \right) \quad (52)$$

where M, N are the image dimension, original image represented as 'I' and watermarked image as 'WM'

- 2) Structure similarity index Measure (SSIM) [31]: The imperceptibility performance is determined by luminance, contrast and correlation between original and watermarked image which ranges between 0 to 1. The value range near to 1 is considered to be of acceptable quality which is represented as,

$$SSIM = \frac{(2\mu_I \mu_{WM}) (2\sigma_{I,WM})}{(\mu_I^2 + \mu_{WM}^2) (\sigma_I^2 + \sigma_{WM}^2)} \quad (53)$$

where μ_I and μ_{WI} represent the mean of I and WI (i.e., image luminance), " σ_I " and " σ_{WM} " represent the standard deviation of I and WI (i.e. image contrast), " $\sigma_{I,WM}$ " represent the correlation coefficient of original 'I' and watermarked image 'WM'

Attacks	Tampered Image	Recovered image and performance measures	Tampered Image	Recovered image and performance measures	Tampered Image	Recovered image and performance measures
Salt and Pepper	HDR Image					
	Density = 0.02	PSNR = 67 dB HDR-VDP=89.61	Density = 0.05	PSNR = 67.02 dB HDR-VDP=90.38	Density = 0.1	PSNR = 66.81 dB HDR-VDP=88.24
	RGB Image					
	Density = 0.02	PSNR = 67.99 dB SSIM = 0.9990	Density = 0.05	PSNR = 65.7 dB SSIM = 0.9929	Density = 0.1	PSNR = 63.72 dB SSIM = 0.9982
Gaussian Noise	HDR Image					
	Variance = 0.02	PSNR = 66.92 dB HDR-VDP=88.32	Variance = 0.05	PSNR = 65.98 dB HDR-VDP=87.09	Density = 0.1	PSNR = 65.27 dB HDR-VDP=87.018
	RGB Image					
	Variance = 0.02	PSNR = 66.98 dB SSIM = 0.9993	Variance = 0.05	PSNR = 66.82 dB SSIM = 0.9919	Density = 0.1	PSNR = 64.97 dB SSIM = 0.9803
Speckle Noise	HDR Image					
	Density=0.02	PSNR =60.82 dB HDR-VDP=85.91	Density=0.05	PSNR =59.91 dB HDR-VDP=82.54	Density = 0.1	PSNR =59 dB HDR-VDP=82.28
	RGB Image					
	Density =0.02	PSNR =59.75 dB SSIM =0.9981	Density=0.05	PSNR =59.72 dB SSIM =0.9921	Density = 0.1	PSNR=59.23 dB SSIM =0.9921

FIGURE 7. Performance evaluation of Noise attack on watermarked test images using PSNR, SSIM and HDR-VDP.

3) Visual Dynamic Predictor (VDP) [11]: In addition to the above metrics, the performance of the HDR image in terms of imperceptibility is determined by VDP which is defined as:

$$\begin{aligned}
 & \text{HDR-VDP} \\
 &= \frac{1}{F \cdot O} \sum_{f=1}^F \sum_{o=1}^O w_f \log \left(\frac{1}{M} \sum_{m=1}^M D_p^2[f, o](m) + \varepsilon \right) \quad (54)
 \end{aligned}$$

where w_f represent vector of per-band pooling weights determined by maximizing correlations with subjective opinion scores, m is the pixel index, M is the total number of pixels, D_p denotes the noise-normalized difference the f^{th} spatial frequency ($f = 1$ to F) band and ($f = 1$ to F) band and o^{th} orientation ($o = 1$ to O) of the steerable pyramid of the original and watermarked image, $\varepsilon = 10^{-5}$ is a constant to avoid singularities when D_p is close to 0.















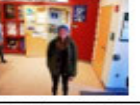
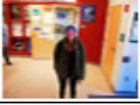













Watermarked Image	Tampered Image	Recovered Image and Performance measure	Tampered Image	Recovered Image and Performance measure
Gaussian Filter	HDR Image			
	Filter = (3 × 3)	PSNR= 65 dB HDR-VDP=84.06	Filter = (5 × 5)	PSNR=64.40 dB HDR-VDP=82.28
				
	Filter = (7 × 7)	PSNR=64 dB HDR-VDP=80.34	Filter = (9 × 9)	PSNR=64.002 dB HDR-VDP=80.63
				
	RGB Image			
	Filter = (3 × 3)	PSNR=64.91 dB SSIM=0.9999	Filter = (5 × 5)	PSNR= 65.03 dB SSIM=0.9962
				
	Filter = (7 × 7)	PSNR=65.807 dB SSIM=0.9982	Filter = (9 × 9)	PSNR=63.112 dB SSIM=0.9899
				
Averaging Filter	HDR Image			
	Filter = (3 × 3)	PSNR=65.92 dB HDR-VDP=85.1	Filter = (5 × 5)	PSNR= 65.86 dB HDR-VDP=85
				
	Filter = (7 × 7)	PSNR= 63.824dB HDR-VDP=84	Filter = (9 × 9)	PSNR= 64.92 dB HDR-VDP=84.62
				
	RGB Image			
	Filter = (3 × 3)	PSNR=64.2 dB SSIM =0.9871	Filter = (5 × 5)	PSNR= 62.58 dB SSIM=0.9820
				
Filter = (7 × 7)	PSNR = 58.02 dB SSIM =0.9801	Filter = (9 × 9)	PSNR=57.2 dB SSIM =0.9761	
				

FIGURE 8. Performance measured by PSNR, HDR-VDP and SSIM metric for filter attack on sample test images.

2) ROBUSTNESS

Using Normalised Correlation Coefficient (NCC) and Bit Error Rate (BER), which have a range of 0 to 1, are used to calculate the robustness performance of the embedded watermark.

- 1) Normalized Correlation coefficient [31] This metric measures the correlation among the extracted watermark and the original watermark. When the values are closer to 1, then the images are highly correlated

(i.e., robust) else, it is uncorrelated (i.e., Less robust). NCC metric is defined as:

$$NCC = \frac{(\sum_{i=1}^{nL} \sum_{j=1}^{nK} (|f_i + E_{f_i}| / 2))}{n_{MN}} \tag{55}$$

where W and EW represents the original and extracted watermark and n_{MN} dimension, (i, j) represent the position of a pixel in the image.












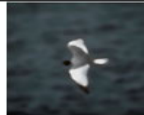



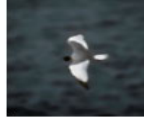
Median Filter	HDR Image			
	Filter = (3 × 3)	PSNR=64.5 dB HDR-VDP=85	Filter = (5 × 5)	PSNR=63.9 dB HDR-VDP=84.96
				
	Filter = (7 × 7)	PSNR=59.72 dB HDR-VDP=84.64	Filter = (9 × 9)	PSNR=59.99 dB HDR-VDP=84.052
				
	RGB Image			
	Filter = (3 × 3)	PSNR=62.87 dB SSIM =0.9982	Filter = (5 × 5)	PSNR=62 dB SSIM =0.9982
				
	Filter = (7 × 7)	PSNR=58.7 dB SSIM =0.9882	Filter = (9 × 9)	PSNR=56 dB SSIM =0.9812
				

FIGURE 8. (Continued.) Performance measured by PSNR, HDR-VDP and SSIM metric for filter attack on sample test images.

2) Bit error rate (BER) [31]: The BER performance is calculated by dividing the total number of embedded watermark bits by the number of watermark bits that were extracted incorrectly. There won't be an error if it ranges closer to 0, otherwise numbers will be closer to 1 which is represented as,

$$BER = \frac{N_{Err}}{N_{Bit}} \times 100 \tag{56}$$

D. PERFORMANCE ANALYSIS

Unauthorized users may intentionally or unintentionally tamper with the watermarked image 'WM' at the receiver side during an unsecured network transmission. The quality degradation between the recovered image and the watermarked image on the sample test images for various attacks, such as (i) Zero or No attack, (ii) Unintentional attack, such as Noise attacks, geometrical attacks, and image processing operations like histogram equalization, sharpening attack, and JPEG compression, is used to assess the watermark features ability of the Swin transformer-based watermark. (iii) Copy-move forgeries, image splicing, and content removal attacks are instances of intentional attacks.

1) ZERO ATTACKS OR NO ATTACKS

The imperceptibility metric is used to calculate the efficiency of the suggested approach in terms of quality deterioration when the watermark is embedded for the zero attacked

watermarked image. The watermarked images are measured using PSNR and SSIM metric, and also the HDR watermarked image similarity is additionally measured using HDR-VDP2 (High dynamic range-Visual Difference Predictor). Figure. (6) clears that the highest PSNR value achieved is 68.75 dB, the maximum SSIM value obtained is 0.999, whereas maximum VDP-quality achieved is 94.76.

2) IMAGE PROCESSING ATTACKS

The performance of the proposed approach for HDR and RGB watermarked images evaluated for various image processing attacks in order to estimate imperceptibility and robustness efficiency for various image tampering ratios.

Image processing attacks are namely: (1) Noise attacks, (2) Filtering attacks, and (3) Common attacks. The noise attacks are categorized as salt and pepper noise (S& P), Gaussian noise (GN), and Speckle noise (SN) tested for various density and variance values of 0.02, 0.05, 0.1. Filtering attacks such as Gaussian filter, Median filter, and average filter are evaluated for various filter size of 3 × 3, 5 × 5, 7 × 7 and 9 × 9. The common attacks with various parameters namely sharpening attack with tampering ratio of 10%, 20%, 40%, and 50%, histogram attacks with tampering ratio of 10%, 50% and the JPEG compression attacks with varying quality factor of 10%, 50%, 70%, and 90%. The imperceptibility performance of the proposed algorithm for unintentional attack is evaluated by PSNR, SSIM, and HDR-VDP metrics

TABLE 1. BERs comparison between proposed Vs existing algorithm for color images.

Attacks & Index	Image	BER (in %)					
		Ying Sun et al. [33]	Aberna et al. [38]	Junlin O et al. [15]	Awasthi D et al. [25]	Proposed	
Salt and Pepper (density=0.02)	3	0.0031	0.072	0.161	0.002	0.016	
Salt and Pepper (density=0.05)		0.061	0.068	0.106	0.219	0.0065	
Gaussian noise (variance=0.02)		0.0281	0.81	0.141	0.015	0.082	
Gaussian noise (variance=0.05)		0.0682	0.126	0.019	0.085	0.082	
Speckle Noise (density=0.02)		0.112	0.012	0.120	0.1108	0.102	
Speckle Noise (density=0.05)		0.013	0.167	0.1142	0.22	0.241	
Median filter size =3*3		0.072	0.03	0.133	0.146	0.0129	
Median filter Size=5*5		0.0186	0.28	0.170	0.191	0.0891	
Average filter size = 3*3		0.150	1.02	0.148	0.221	0.0183	
Average filter size = 5*5		0.342	1.03	0.0127	0.075	0.159	
Gaussian filter size = 3*3		0.618	0.812	0.074	0.028	0.0216	
Gaussian filter size = 5*5		1.024	0.09	0.16	0.0059	0.240	
JPEG with QF 20%		7	0.341	0.0254	0.0463	0.18	1.09
Rotation _{10⁰}			1.062	0.64	0.084	0.12	0.053
Copy-move _{20%}	0.24		0.607	0.022	0.0259	0.051	
Scaling _{25%}	0.36		0.093	0.0124	0.0821	0.0124	
Image Splicing _{20%}	0.012		0.580	0.146	0.1406	0.098	
Content Removal 40%	0.0286		0.35	0.33	0.144	0.1012	

TABLE 2. NCCs comparison between proposed Vs existing watermarking algorithm for color images.

Attacks & Index	Image	NCC					
		Ying Sun et al. [33]	Aberna et al. [38]	Junlin O et al. [15]	Awasthi D et al. [25]	Proposed	
Salt and Pepper (density=0.02)	3	0.9958	0.999	0.9992	0.9926	0.9999	
Salt and Pepper (density=0.05)		0.9711	0.9998	0.9899	0.9884	0.9923	
Gaussian noise (variance=0.02)		0.9621	0.99	0.9882	0.9795	0.9996	
Gaussian noise (variance=0.05)		0.9479	0.9981	0.9806	0.9681	0.9902	
Speckle Noise (density=0.02)		0.9845	0.986	0.9782	0.9601	0.9969	
Speckle Noise (density=0.05)		0.9120	0.972	0.9786	0.9542	0.9892	
Median filter size =3*3		0.9918	0.973	0.9726	0.9654	0.9899	
Median filter Size=5*5		0.8788	0.97	0.9733	0.9542	0.9810	
Average filter size = 3*3		0.9917	0.98	0.9814	0.9588	0.9827	
Average filter size = 5*5		0.9482	0.98	0.9891	0.9782	0.9791	
Gaussian filter size = 3*3		0.9925	0.9854	0.9790	0.9706	0.9848	
Gaussian filter size = 5*5		0.9867	0.9802	0.9709	0.9786	0.9809	
JPEG with QF 20%		7	0.9735	0.9781	0.99811	0.9980	0.9857
Rotation _{10⁰}			0.9975	0.98	0.9789	0.9706	0.9932
Copy-move _{20%}	0.9587		0.991	0.9621	0.9647	0.9981	
Scaling _{25%}	0.9969		0.9862	0.9601	0.9768	0.9862	
Image Splicing _{20%}	0.9862		0.9813	0.9962	0.9752	0.9847	
Content Removal 40%	0.9721		0.9887	0.9893	0.9626	0.9821	

TABLE 3. BER comparison between proposed Vs existing systems on HDR images.

Index	Attacks	Image 3			Image 5			Image 7		
		[8]	[11]	Proposed	[8]	[11]	Proposed	[8]	[11]	Proposed
[1]	Salt and Pepper (density=0.02)	0.35	0.082	0.006	0.26	0.068	0.012	0.47	0.0921	0.0182
[3]	Gaussian noise (variance=0.02)	0.91	0.0629	0.0028	0	0.0436	0.028	0.68	0.0726	0.0355
[5]	Speckle Noise (density=0.02)	0.026	0.105	0.021	2.81	0.0242	0.0281	0.21	0.0802	0.0926
[7]	Median filter size =3*3	1.18	0.011	0.0129	3.41	0.019	0.016	0.94	0.079	0.056
[9]	Average filter size = 3*3	0.98	0.121	0.0183	0.71	0.13	0.019	0.65	0.093	0.034
[11]	Gaussian filter size = 3*3	0.78	0.160	0.211	1.04	0.056	0.092	3.82	0.069	0.102
[13]	JPEG with QF 20%	0.7	0.142	1.09	2.79	0.1702	1.009	0.8	0.18	0.189
[16]	Rotation (20 ⁰)	0.66	0.0689	0.056	1.13	0.0569	0.049	0.09	0.069	0.0759
[19]	Scaling (50%)	2.97	0.0094	0.0128	1.72	0.0183	0.0093	2.07	0.0742	0.0261
[21]	Translation (Tx=-25,Ty= 20)	0.47	0.0897	0.0142	0.71	0.0249	0.0136	1.19	0.028	0.0138
[23]	Copy-move (20%)	0.23	0.7056	0.8204	0.35	0.9321	0.81	2.42	0.83	0.829
[25]	Image Splicing (20%)	0.28	0.0567	0.0114	0.62	0.0142	0.914	0.92	0.1114	0.95
[27]	Content Removal (20%)	0.6	0.027	0.416	0.18	0.087	0.426	0.48	0.0917	0.421

(between Watermarked and Recovered image) and the results are shown in Figures (7), (8) and (9).

It is observed from the Fig. (7), (8) and (9) that, the performance of noise attack for color image as

Tampered Image	Recovered image and performance measures	Tampered Image	Recovered image and performance measures
Histogram equalization			
10%	PSNR=64.78 dB SSIM =0.9982	50%	PSNR=64.15 dB SSIM =0.999
JPEG Compression			
Quality factor -10	PSNR =63.71 dB SSIM =0.9991	Quality factor- 50	PSNR =62.07 dB SSIM =0.999
Quality factor -70	PSNR =59.82 dB HDR-VDP=82.29 SSIM =0.9981	Quality factor- 90	PSNR =56.27 dB SSIM =0.9956
Sharpening			
10%	PSNR =60.63 dB HDR-VDP=84.32 SSIM =0.9897	20%	PSNR =58.08 dB HDR-VDP=82.92 SSIM =0.9828
40%	PSNR =55.88 dB SSIM =0.9686	50%	PSNR =52.28 dB SSIM =0.9617

FIGURE 9. Performance evaluation of common attack on watermarked test images using PSNR, SSIM and HDR-VDP.

achieved good result than the HDR images with a maximum PSNR value of 67.99 dB and HDR-VDP resulting in 90.38. Whereas the filtering, histogram, sharpening and JPEG compression attack is evaluated on sample test image via various filter sizes, yielding high fidelity with average PSNR values of 64 dB and HDR-VDP of 85.

3) GEOMETRIC ATTACKS

The proposed systems performance against RST attacks is examined, to evaluate the robustness of the watermark in terms of image recovery. Intentional geometrical attacks are namely rotational attacks with tampering ratios of 10°, 20°, 40^{circ}, and 60°, scaling attack with tampering ratios of 25%, 50%, 75%, and 125%, and translation attacks with


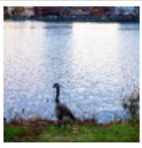




















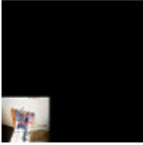

Geometric Attacks				
Attacks	Attacked Image	Recovered Image and Performance measure	Attacked Image	Recovered Image and Performance measure
Rotation	10°	PSNR = 65.59 dB HDR-VDP=92.32 SSIM =0.9969	20°	PSNR= 65.1 dB HDR-VDP = 93 SSIM =0.9950
				
	40°	PSNR = 64.92 dB SSIM =0.9942	60°	PSNR = 63.81 dB SSIM =0.9899
				
Scaling	25%	PSNR= 64.14 dB HDR-VDP=95.74 SSIM =0.9928	50%	PSNR= 63.50 dB HDR-VDP = 95.25 SSIM =0.9892
				
	75%	PSNR= 63.21 dB SSIM =0.9829	125%	PSNR = 62.02 dB SSIM =0.9808
				
Translation	Tx=-25,Ty= 20	PSNR= 66.12 dB SSIM = 0.9951	Tx=50,Ty= -50	PSNR = 65.28 SSIM = 0.992
				
	Tx=45,Ty= -25	PSNR=64.18 dB HDR-VDP=93.25 SSIM =0.9905	Tx= -50,Ty=50	PSNR = 61 dB HDR-VDP=89.01 SSIM =0.9725
				

FIGURE 10. Performance measured by PSNR, HDR-VDP and SSIM metric for geometric attack on sample test images.

different axis ranges of $T_x = -25, T_y = 20, T_x = 50, T_y = -50, T_x = 45, T_y = -25,$ and $T_x = -50, T_y = 50$ respectively. The watermarked image is tampered intentionally using MATLAB command, where the performance is computed using multiple metrics and the results are depicted in Fig. (10). The system shown better imperceptibility performance with the highest PSNR values of 63.5 dB, SSIM value of 0.999, and an average HDR-VDP of 92.

4) INTENTIONAL ATTACKS

Tamper detection and localization efficiency for the proposed model is evaluated by intentionally tampering the watermarked image. The intentional attacks are namely content removal, copy-move forgery, content removal, and Image Splicing. From Fig. (11) it is inferred that the proposed work is highly efficient in tamper localization and the deep global swin features maps are highly capable in recovering

Intentional attacks						
Attacks	Attacked Image	Tamper detection	Recovered Image and Performance measure	Attacked Image	Tamper detection	Recovered Image and Performance measure
Content removal	10%		PSNR=65.002 dB VDP=92.62	20%		PSNR=65.25 dB VDP=93.08
	40%		PSNR=64.92 dB SSIM=0.9997	50%		PSNR=61.69 dB SSIM=0.9889
Image Splicing	10%		PSNR=67.82 dB VDP=97.62	20%		PSNR=62 dB VDP=89.19
	30%		PSNR=59.37 dB SSIM=0.9781	40%		PSNR=58.98dB SSIM=0.9732
Copy-move	5%		PSNR=66.84 dB VDP=94.16	10%		PSNR=66.01 dB VDP=93.09
	20%		PSNR=65 dB SSIM=0.9921	40%		PSNR=63.1 dB SSIM=0.9939

FIGURE 11. Performance evaluation of intentional attack on watermarked test images using PSNR, SSIM and HDR-VDP.

the image. The proposed system performed better for content removal and copy-move attack. For image splicing attack, the fidelity performance of HDR image achieved better result when compared with RGB image. From figure (6) - (11), it is inferred that the proposed algorithm significantly outperforms for various attacks.

E. ROBUSTNESS ANALYSIS

The swin watermark robustness efficiency is evaluated for various attacks where the performance of the extracted watermark from the tampered image is calculated using

Normalized correlation coefficient (NCC) and Bit error rate (BER) metric and the results obtained is mentioned in Fig. (12) and (13). In the proposed work, swin features maps are embedded as watermark in quaternion transform of RGB channel which not only guarantee the imperceptibility of watermarked image and also achieve higher robustness. Moreover, in the adaptive less sensitive region the watermark information are effectively dispersed in all the three RGB channels of the host image, thus the robustness as well as the payload of the proposed model is improved.

Watermarked Images	Embedded Watermark Features	Extracted Watermark Features After Attacks	
		Salt & Pepper Noise (Density=0.02)	Salt & Pepper Noise (Density=0.05)
		 NCC = 0.999 BER = 0.12	 NCC = 0.998 BER = 0.042
		 NCC = 0.9998 BER = 0.28	 NCC = 0.9973 BER = 0.0312
		 NCC = 0.9923 BER = 0.101	 NCC = 0.9897 BER = 0.916
		 NCC = 0.9802 BER = 0.124	 NCC = 0.952 BER = 0.1064
		 NCC = 0.9732 BER = 1.91	 NCC = 0.9689 BER = 2.12
		 NCC = 0.999 BER = 0.18	 NCC = 0.9927 BER = 0.094
		 NCC = 0.9879 BER = 0.0011	 NCC = 0.9823 BER = 0.0289
		 NCC = 0.9812 BER = 0.79	 NCC = 0.9791 BER = 0.107
		 NCC = 0.9818 BER = 0.0191	 NCC = 0.9881 BER = 0.023
		 NCC = 0.9742 BER = 0.067	 NCC = 0.9626 BER = 0.247
		 NCC = 0.9902 BER = 0.009	 NCC = 0.9893 BER = 0.025

FIGURE 12. Robustness performance in terms of NCC and BER against various attacks on HDR Images.

V. COMPARATIVE ANALYSIS

This section conducts a comparative analysis between the suggested algorithm and the existing reversible

multi-watermark [33], semi-blind deep learning features [38], semi-fragile reversible [15] and a hybrid transform algorithm [25] that has been proposed, in order to demonstrate





































Watermarked Image	Embedded Watermark Feature	Extracted Watermark Features After Attack	
		Salt & Pepper Noise (Density=0.02)	Salt & Pepper Noise (Density=0.05)
		 NCC = 0.999 BER = 0.016	 NCC = 0.9992 BER = 0.06
		 NCC = 0.9999 BER = 0.78	 NCC = 0.9996 BER = 0.84
		 NCC = 0.9929 BER = 0.81	 NCC = 0.9899 BER = 0.0271
		 NCC = 0.9829 BER = 0.0138	 NCC = 0.9799 BER = 0.056
		 NCC = 0.9824 BER = 0.0387	 NCC = 0.9782 BER = 0.812
		 NCC = 0.9994 BER = 0.045	 NCC = 0.9993 BER = 0.62
		 NCC = 0.9823 BER = 0.298	 NCC = 0.973 BER = 0.098
		 NCC = 0.9897 BER = 0.0628	 NCC = 0.97899 BER = 0.092
		 NCC = 0.991 BER = 0.052	 NCC = 0.9864 BER = 0.21

FIGURE 13. Robustness performance in terms of NCC and BER against various attacks on RGB Images.

the endurance of the proposed algorithm. Embedding the grayscale invariance watermark in multi-level and multi-feature regions [33]. The multi-level watermark embedding is carried out on three levels of non-overlapping blocks, and

multi-region embedding is carried out by utilizing the SURF key feature points. The semi-blind watermarking approach based on the 2-level DWT-SVD model [23] embeds the convolution attention features as watermark to effectively

TABLE 4. Proposed model Vs existing model imperceptibility performance for payload.

Images	Ying sun et al., [33] Max 17,000 bits	Abernal et al., [38] Nearly 65,000 bits	Awasthi D et al., [25] Nearly 65,000 bits	Proposed Nearly 786,432 bits
1	38.8817	45.5728	60.32	64.72
2	36.0018	42.15	61.09	63.18
3	37.881	43.817	59.98	65.94
4	34.2896	39.98	62.01	64.801
5	35.872	43.0124	62.41	64.92
6	36.8859	42.897	60.106	63.64
12	36.461	36.292	59.806	62.07
13	34.6567	44.87	58.82	63.81
14	33.2016	38.6015	623.3	63.21
15	34.5178	40.869	61	66.12
16	34.102	40.415	60.9899	64.84
17	34.8523	42.783	61.29	62.013
18	33.7719	43.689	60.77	61.69

detect and recover the tampered region. The semi-fragile colour image watermarking algorithm in scheme [15] uses quaternion discrete fourier transform (QDFT) to embed watermark. The authentication binary watermark followed by that synchronized watermark is embedded in the QDFT coefficients. The quaternion Hadamard transform (QHT) with schur decomposition technique is suggested by Li et al. in [60] where the watermark embedding is carried out in Q matrix. In order to be robust against geometric attack, and to detect geometric distortion Zernike moment is induced. Whereas in [18] effective tamper detection system is proposed by extracting invariant features from Vision transformer model. multi-watermark are generated using Schur decomposition, SVD and vision transformer and each watermark is embedded in the LSB bit of each blocks in LH, LL sub-band of DWT and curvelet coefficients to produce watermarked image. In order to verify the efficiency of the suggested approach, a balanced comparison is conducted by replicating the existing algorithm on sample images from (4) for different attacks. The obtained results are compared with the existing system is presented in Table (1). When compared to conventional watermarking systems, a greater performance is attained in terms of NCC and BER values because of the adaptive embedding region and a global deep watermark feature. Table (2) shows that the suggested work's NCC robustness efficiency on RGB images outperformed the performance of the all other existing system [17], [18], [23] but the proposed system lag robustness for few attack than the system [60]. On the other hand, the suggested work on RGB's BER performance outperformed than all the existing state-of-art systems in terms of noise attack outcomes depicted in Table (1). However, the findings of [18] and [23] performance were superior when it came to filtering attacks.

Table (3) depicts the robustness performance of HDR image against existing system [8], [11] in terms of BER. From table (3) the proposed system shows better BER performance than the state-of-art methods.

Additionally, in order to validate the efficiency of the suggested work on HDR images, a fair comparison is conducted using benchmark HDR images by replicating the existing

Tucker decomposition [11] technique for various intentional and unintentional attacks. The obtained results are compared with the existing system which is tabulated in Table (2). In [11] extensive evaluations on several HDR images made by encoding two widely-used transfer function (TFs) confirm the strong HVS-imperceptibility capabilities of the method, as well as the robustness of the embedded watermarks to tone mapping, lossy compression, and common signal processing operations. The error rate performance on several test images is evaluated in order to assess the HDR image robustness performance. Based on the outcome, it is deduced that the BER of the suggested system achieved better than the existing system.

The watermarking system payload characteristics of the proposed and the existing algorithm is measured as number of watermark bits embedded in the original image pixel. In addition to that the trade-off between imperceptibility and robustness for varying payload system is measured by PSNR and NCC metric that is tabulated in Table (4). It is inferred from table (4) that the performance of proposed system shows better results.

VI. CONCLUSION AND FUTURE WORK

The growth of photographic technology has increased in real-world scenarios that resulted in high-quality images. Also, the uncontrollable multimedia image manipulation insists on the need for secured content authentication and tamper detection systems. Much research suggested a successful model for color image content authentication and tamper detection systems, but limited research has been carried out on HDR images for tampering attacks. Considering the above-mentioned difficulties, we proposed an adaptive semi-blind watermarking system for HDR and RGB images using quaternion Dual-Tree complex wavelet transform for content authentication, proof of ownership, and tamper detection application using the Swin transformer model. This work focuses mainly on a robust watermark and an optimal secured authentic system for various intentional and unintentional attacks. For the first time, we have

evaluated the Swin transformer-based feature generation model for the first time on the HDR and RGB image datasets.

With the use of a pixel correlation-based graph entropy system, the optimal embedding region is identified using the Maximal Entropy Random Walk approach, which strikes a balance between robustness and imperceptibility. The less sensitive Quaternion Dual-Tree Complex Wavelet Transform coefficient in all three channels, which not only demonstrated high robustness because of its shift-invariance and strong directionality property but also preserved the pixel correlation, was another way to improve system performance. Furthermore, the novel Quaternion Dual-Tree Complex Wavelet Transform (QGBT) yields superior tamper detection and localization results because of its strong perceptual features. In addition, to confirm the content authenticity, the dual scrambled biometric watermark is integrated into the principal components of the singular value decomposition matrix. Utilising a gyrator transform to jumble the owner's biometric image, the protected authentication system is achieved while preventing outside parties from detecting or extracting the watermark from the image. Extraction of the watermark is possible, only when the owner's dual biometric image matches with the extracted watermark. To authenticate and detect tampering owner's biometric image and the original watermarked image are required which is known to be a semi-blind watermarking system.

The suggested algorithm's performance is assessed for a range of deliberate and inadvertent attacks. The imperceptibility and robustness performance of the proposed algorithm for each attack with varying tampering ratios is evaluated by various quality metrics like PSNR, NCC, BER, SSIM, and VDP. With a maximum PSNR of approximately 65 dB and an NCC of 0.999, it is evident from the result analysis that the suggested model outperformed the others in terms of imperceptibility and robustness. The VDP metric was also used to measure the HDR quality visualisation, and it achieved the highest value of 94.76, respectively. An adaptive, content authentication, tamper detection, tamper localization and recovery, and proof of ownership application were all successfully obtained by the suggested system. Also, the trade-off between watermarking properties is well balanced by an adaptable system. In the future, this system can be extended for high dynamic range videos.

REFERENCES

- [1] M. Z. Konyar and S. Solak, "Efficient data hiding method for videos based on adaptive inverted LSB332 and secure frame selection with enhanced vigenere cipher," *J. Inf. Secur. Appl.*, vol. 63, Dec. 2021, Art. no. 103037.
- [2] S. Jiao, C. Zhou, Y. Shi, W. Zou, and X. Li, "Review on optical image hiding and watermarking techniques," *Opt. Laser Technol.*, vol. 109, pp. 370–380, Jan. 2019.
- [3] M. Z. Konyar and S. Öztürk, "Reed Solomon coding-based medical image data hiding method against salt and pepper noise," *Symmetry*, vol. 12, no. 6, p. 899, Jun. 2020.
- [4] C.-M. Yu, K.-C. Wu, and C.-M. Wang, "A distortion-free data hiding scheme for high dynamic range images," *Displays*, vol. 32, no. 5, pp. 225–236, Dec. 2011.
- [5] L. Agilandeewari and K. Ganesan, "A robust color video watermarking scheme based on hybrid embedding techniques," *Multimedia Tools Appl.*, vol. 75, no. 14, pp. 8745–8780, Jul. 2016.
- [6] X. Xue, M. Okuda, and S. Goto, "Bilateral filtering based watermarking for high dynamic range image," in *Proc. Int. Symp. Intell. Signal Process. Commun. Syst. (ISPACS)*, Dec. 2011, pp. 1–5.
- [7] C.-C. Chang, T.-S. Nguyen, and C.-C. Lin, "A new distortion-free data embedding scheme for high-dynamic range images," *Multimedia Tools Appl.*, vol. 75, no. 1, pp. 145–163, Jan. 2016.
- [8] F. Yazdan Bakhsh and M. E. Moghaddam, "A robust HDR images watermarking method using artificial bee colony algorithm," *J. Inf. Secur. Appl.*, vol. 41, pp. 12–27, Aug. 2018.
- [9] Y.-T. Lin, C.-M. Wang, W.-S. Chen, F.-P. Lin, and W. Lin, "A novel data hiding algorithm for high dynamic range images," *IEEE Trans. Multimedia*, vol. 19, no. 1, pp. 196–211, Jan. 2017.
- [10] K. R. Perez-Daniel, F. Garcia-Ugalde, and V. Sanchez, "Watermarking of HDR images in the spatial domain with HVS-imperceptibility," *IEEE Access*, vol. 8, pp. 156801–156817, 2020.
- [11] M. Yu, Y. Wang, G. Jiang, Y. Bai, and T. Luo, "High dynamic range image watermarking based on tucker decomposition," *IEEE Access*, vol. 7, pp. 113053–113064, 2019.
- [12] A. Khan, M. Kuribayashi, K. Wong, and V. M. Baskaran, "HDR image watermarking using saliency detection and quantization index modulation," 2023, *arXiv:2302.11361*.
- [13] J. Ouyang, H. Shu, X. Wen, J. Wu, F. Liao, and G. Coatrieux, "A blind robust color image watermarking method using quaternion Fourier transform," in *Proc. 6th Int. Congr. Image Signal Process. (CISP)*, vol. 1, Dec. 2013, pp. 485–489.
- [14] S. Kaçar, M. Z. Konyar, and Ü. Çavuşoğlu, "4D chaotic system-based secure data hiding method to improve robustness and embedding capacity of videos," *J. Inf. Secur. Appl.*, vol. 71, Dec. 2022, Art. no. 103369.
- [15] J. Ouyang, J. Huang, and X. Wen, "A semi-fragile reversible watermarking method based on qdft and tamper ranking," *Multimedia Tools Appl.*, pp. 1–24, Oct. 2023.
- [16] J. Ouyang, J. Huang, X. Wen, and Z. Shao, "A semi-fragile watermarking tamper localization method based on QDFT and multi-view fusion," *Multimedia Tools Appl.*, vol. 82, no. 10, pp. 15113–15141, Apr. 2023.
- [17] A. M. Cheema, S. M. Adnan, and Z. Mehmood, "A novel optimized semi-blind scheme for color image watermarking," *IEEE Access*, vol. 8, pp. 169525–169547, 2020.
- [18] P. Aberna, L. Agilandeewari, and A. Bansal, "Vision transformer-based watermark generation for authentication and tamper detection using Schur decomposition and hybrid transforms," *Int. J. Comput. Inf. Syst. Ind. Manag. Appl.*, vol. 15, pp. 107–121, 2023.
- [19] A. Palani and A. Loganathan, "Multi-image feature map-based watermarking techniques using transformer," *Int. J. Electr. Electron. Res.*, vol. 11, no. 2, pp. 339–344, May 2023.
- [20] L. Agilandeewari, M. Prabukumar, and F. A. Alenizi, "A robust semi-fragile watermarking system using pseudo-zernike moments and dual tree complex wavelet transform for social media content authentication," *Multimedia Tools Appl.*, vol. 82, no. 28, pp. 43367–43419, Nov. 2023.
- [21] S. I. Lee, J. Y. Park, and I. K. Eom, "CNN-based copy-move forgery detection using rotation-invariant wavelet feature," *IEEE Access*, vol. 10, pp. 106217–106229, 2022.
- [22] Y. Liu and M. Lapata, "Text summarization with pretrained encoders," 2019, *arXiv:1908.08345*.
- [23] Y. Chen, Z. Jia, Y. Peng, and Y. Peng, "Robust dual-color watermarking based on quaternion singular value decomposition," *IEEE Access*, vol. 8, pp. 30628–30642, 2020.
- [24] L. Agilandeewari and K. Muralibabu, "A robust video watermarking algorithm for content authentication using discrete wavelet transform (DWT) and singular value decomposition (SVD)," *Int. J. Secur. Appl.*, vol. 7, no. 4, pp. 145–158, 2013.
- [25] D. Awasthi and V. K. Srivastava, "LWT-DCT-SVD and DWT-DCT-SVD based watermarking schemes with their performance enhancement using Jaya and particle swarm optimization and comparison of results under various attacks," *Multimedia Tools Appl.*, vol. 81, no. 18, pp. 25075–25099, Jul. 2022.
- [26] L. Agilandeewari and K. Ganesan, "An efficient Hilbert and integer wavelet transform based video watermarking," *J. Eng. Sci. Technol.*, vol. 11, no. 3, pp. 327–345, 2016.

- [27] A. Loganathan and G. Kaliyaperumal, "An adaptive HVS based video watermarking scheme for multiple watermarks using BAM neural networks and fuzzy inference system," *Expert Syst. Appl.*, vol. 63, pp. 412–434, Nov. 2016.
- [28] L. Li, R. Bai, J. Lu, S. Zhang, and C.-C. Chang, "A watermarking scheme for color image using quaternion discrete Fourier transform and tensor decomposition," *Appl. Sci.*, vol. 11, no. 11, p. 5006, May 2021.
- [29] L.-Y. Hsu and H.-T. Hu, "QDCT-based blind color image watermarking with aid of GWO and DnCNN for performance improvement," *IEEE Access*, vol. 9, pp. 155138–155152, 2021.
- [30] M. Yin, W. Liu, J. Shui, and J. Wu, "Quaternion wavelet analysis and application in image denoising," *Math. Problems Eng.*, vol. 2012, pp. 1–21, Oct. 2012.
- [31] L. Agilandeewari and K. Ganesan, "RST invariant robust video watermarking algorithm using quaternion curvelet transform," *Multimedia Tools Appl.*, vol. 77, no. 19, pp. 25431–25474, Oct. 2018.
- [32] K. Prabha, M. J. Vaishnavi, and I. S. Sam, "Quaternion Hadamard transform and QR decomposition based robust color image watermarking," in *Proc. 3rd Int. Conf. Trends Electron. Informat. (ICOEI)*, Apr. 2019, pp. 101–106.
- [33] Y. Sun, X. Yuan, X. Wang, and J. Li, "Reversible multi-watermarking for color images with grayscale invariance," *Multimedia Tools Appl.*, vol. 82, no. 11, pp. 16323–16342, May 2023.
- [34] P. Aberna and L. Agilandeewari, "Digital image and video watermarking: Methodologies, attacks, applications, and future directions," *Multimedia Tools Appl.*, vol. 83, no. 2, pp. 5531–5591, Jan. 2024.
- [35] L. Yu, Y. Zhang, H. Han, L. Zhang, and F. Wu, "Robust median filtering forensics by CNN-based multiple residuals learning," *IEEE Access*, vol. 7, pp. 120594–120602, 2019.
- [36] Y. Rao and J. Ni, "A deep learning approach to detection of splicing and copy-move forgeries in images," in *Proc. IEEE Int. Workshop Inf. Forensics Secur. (WIFS)*, Dec. 2016, pp. 1–6.
- [37] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 1–7.
- [38] A. Palani and A. Loganathan, "Semi-blind watermarking using convolutional attention-based turtle shell matrix for tamper detection and recovery of medical images," *Expert Syst. Appl.*, vol. 238, Mar. 2024, Art. no. 121903.
- [39] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16×16 words: Transformers for image recognition at scale," 2020, *arXiv:2010.11929*.
- [40] M. Yang, "Visual transformer for object detection," 2022, *arXiv:2206.06323*.
- [41] B. Zhang, "SegViT: Semantic segmentation with plain vision transformers," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 35, 2022, pp. 4971–4982.
- [42] J. Li, Y. Yan, S. Liao, X. Yang, and L. Shao, "Local-to-global self-attention in vision transformers," 2021, *arXiv:2107.04735*.
- [43] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 9992–10002.
- [44] B. Han, H. Wang, D. Qiao, J. Xu, and T. Yan, "Application of zero-watermarking scheme based on Swin transformer for securing the metaverse healthcare data," *IEEE J. Biomed. Health Informat.*, early access, Mar. 15, 2023, doi: 10.1109/JBHI.2023.3257340.
- [45] F. Verdoja and M. Grangotto, "Graph Laplacian for image anomaly detection," *Mach. Vis. Appl.*, vol. 31, nos. 1–2, p. 11, Feb. 2020.
- [46] R. Mihalcea, "Graph-based ranking algorithms for sentence extraction, applied to text summarization," in *Proc. ACL Interact. Poster Demonstration Sessions*, 2004, pp. 170–173.
- [47] W. Ju, D. Xiang, B. Zhang, L. Wang, I. Kopriva, and X. Chen, "Random walk and graph cut for co-segmentation of lung tumor on PET-CT images," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 5854–5867, Dec. 2015.
- [48] R.-H. Li, J. X. Yu, and J. Liu, "Link prediction: The power of maximal entropy random walk," in *Proc. 20th ACM Int. Conf. Inf. Knowl. Manage.*, Oct. 2011, pp. 1147–1156.
- [49] V. Gopalakrishnan, Y. Hu, and D. Rajan, "Random walks on graphs for salient object detection in images," *IEEE Trans. Image Process.*, vol. 19, no. 12, pp. 3232–3242, Dec. 2010.
- [50] L. Wang, J. Zhao, X. Hu, and J. Lu, "Weakly supervised object localization via maximal entropy random walk," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2014, pp. 1614–1617.
- [51] J.-G. Yu, J. Zhao, J. Tian, and Y. Tan, "Maximal entropy random walk for region-based visual saliency," *IEEE Trans. Cybern.*, vol. 44, no. 9, pp. 1661–1672, Sep. 2014.
- [52] P. Korus and J. Huang, "Improved tampering localization in digital image forensics based on maximal entropy random walk," *IEEE Signal Process. Lett.*, vol. 23, no. 1, pp. 169–173, Jan. 2016.
- [53] C. Huang, J. Li, and G. Gao, "Review of quaternion-based color image processing methods," *Mathematics*, vol. 11, no. 9, p. 2056, Apr. 2023.
- [54] N. Kingsbury, "Complex wavelets for shift invariant analysis and filtering of signals," *Appl. Comput. Harmon. Anal.*, vol. 10, no. 3, pp. 234–253, May 2001.
- [55] B. Chen, G. Coatrieux, G. Chen, X. Sun, J. L. Coatrieux, and H. Shu, "Full 4-D quaternion discrete Fourier transform based watermarking for color images," *Digit. Signal Process.*, vol. 28, pp. 106–119, May 2014.
- [56] K. Panetta, L. Kezebou, V. Oludare, S. Agaian, and Z. Xia, "TMO-Net: A parameter-free tone mapping operator using generative adversarial network, and performance benchmarking on large scale HDR dataset," *IEEE Access*, vol. 9, pp. 39500–39517, 2021.
- [57] (2023). *Casia Dataset*. [Online]. Available: <https://www.kaggle.com/datasets/sophatvathana/casia-dataset>
- [58] (2023). *FVC2002–Second International Fingerprint Verification Competition*. [Online]. Available: <http://bias.csr.unibo.it/fvc2002/databases.asp>
- [59] (2023). *MMU IRIS Dataset*. [Online]. Available: <https://www.kaggle.com/datasets/naureenmohammad/mmu-iris-dataset>
- [60] J. Li, C. Yu, B. B. Gupta, and X. Ren, "Color image watermarking scheme based on quaternion Hadamard transform and schur decomposition," *Multimedia Tools Appl.*, vol. 77, no. 4, pp. 4545–4561, Feb. 2018.



P. ABERNA received the M.Tech. degree in software engineering from Vellore Institute of Technology, Vellore, in 2018, where she is currently pursuing the Ph.D. degree with the School of Information Technology and Engineering. Her research interests include digital image watermarking, image and video processing, multimedia security, digital forensics, machine learning, and deep learning.



L. AGILANDEESWARI received the bachelor's degree in information technology and the master's degree in computer science and engineering from Anna University, in 2005 and 2009, respectively, and the Ph.D. degree. She is currently a Professor with the Department of Software Systems and Engineering, School of Computer Science Engineering and Information Systems (SCORE), VIT, Vellore. She has around 19 years of teaching experience and published more than 60 articles in peer-reviewed reputed journals. Her reputed publications include research articles in peer-reviewed journals, namely *Expert Systems with Applications*, *Multimedia Tools and Applications*, *International Journal of Remote Sensing*, *IEEE ACCESS*, *Journal of Ambient Intelligence and Humanized Computing*, and *Journal of Applied Remote Sensing* indexing at Thomson Reuters with an average impact factor of five. She produced two Ph.D. candidates. She also published more than 20 engineering books as per Anna University Syllabus. Her areas of research interests include image and video watermarking, image and video processing, neural networks, cryptography, fuzzy logic, machine learning, the IoT, information-centric networks, and remote sensing. She is a Lifetime Member of the Computer Society of India. She is a Peer Reviewer in journals, including *Neural Computing and Applications*, *IEEE ACCESS*, *Pattern Recognition*, *International Journal of Remote Sensing*, *Array*, *Artificial Intelligence Review*, *Informatics in Medicine Unlocked*, *Neurocomputing*, *Computers*, *Electrical Engineering*, *Journal of King Saud University: Computer and Information Sciences*, and *Journal of Engineering Science and Technology* (JESTEC).