Blind Image Watermark Detection Algorithm based on Discrete Shearlet Transform Using Statistical Decision Theory

Baharak Ahmaderaghi, Fatih Kurugollu, Senior Member, IEEE, Jesus Martinez Del Rincon and Ahmed Bouridane, Senior Member, IEEE

Abstract—Blind watermarking targets the challenging recovery of the watermark when the host is not available during the detection stage. This paper proposes Discrete Shearlet Transform (DST) as a new embedding domain for blind image watermarking. Our novel DST blind watermark detection system uses a non-additive scheme based on the statistical decision theory. It first computes the Probability Density Function (PDF) of the DST coefficients modelled as a Laplacian distribution. The resulting likelihood ratio is compared with a decision threshold calculated using Neyman-Pearson criterion to minimise the missed detection subject to a fixed false alarm probability. Our method is evaluated in terms of imperceptibility, robustness and payload against different attacks (Gaussian noise, Blurring, Cropping, Compression and Rotation) using 30 standard grayscale images covering different characteristics (smooth, more complex with a lot of edges and high detail textured regions). The proposed method shows greater windowing flexibility with more sensitive to directional and anisotropic features when compared against Discrete Wavelet and Contourlets.

Index Terms—Digital image watermarking, Frequency domain, DST, Discrete Wavelet Transform (DWT), Contourlet Transform (CT), Laplacian distribution.

I. INTRODUCTION

In the current globally-connected society, where access and distribution of digital multimedia files is ubiquitous and pervasive, virtual opportunities to pirate copyrighted files are in a permanent rise. As a consequence, finding protection methods to block or detect any unauthorized access and keep data transmission safe and secure has become one of the most important challenge during the past decades. Digital watermarking is one method that has been developed in order to protect ownership of data, digital content protection and transaction tracking so that illegal use, modification and distribution of the content can be detected. In this regard, the purpose of digital watermarking is to embed or hide some invisible additional information, called watermark, into another signal such as image, audio or video, known as a host or cover where the visual quality of the embedded host signal should not be significantly degraded. To be effective, watermark detection and extraction should be possible after applying a variety of manipulations and attacks while meeting some criteria in terms of imperceptibility, robustness, security and payload, which are often interdependent.

In general terms, the imperceptibility of a watermark refers to the perceptual similarity between the original and watermarked version of the host data. This is important so as to keep the degradation of host quality to a minimum, so no obvious difference in the fidelity between the original and watermarked hosts can be noticed [1]. Robustness is a measure of the watermarking methods resistance against different types of attacks, for instance, compression, additive noise, etc., are the types of attacks accrue in digital signal processing [1]. Payload refers to the total amount of information that can be hidden within the digital media [2]. The purpose of increasing watermarking payload is to find how transmit more information while satisfying both watermarking robustness and imperceptibility requirements [3]. In particular, the most challenging issue is how to address the trade-off between robustness and imperceptibility, since enhancing robustness implies necessarily increasing the watermark strength and therefore produces a loss of transparency [4]. Finding such an optimized solution still remains a challenge within the watermarking community.

This paper describes a new framework for robust watermarking of image content due to the fact that digital images constitute a major component of digital multimedia files. A watermarking system can be divided into two main processes: embedding and extracting. Current watermarking techniques are broadly classified according to the embedding domain: spatial and transform domains. Although spatial domain based methods are easy to implement, such techniques suffer from some disadvantages, including failure to achieve better robustness against various attacks. For instance, in [5], since the watermark information is embedded in the least significant bits, the effects of simple manipulations like lossy compression, adding noise and filtering are severe and impair the detection of the watermark.

In contrast, imperceptibility and robustness requirements to a variety of attacks can be achieved more efficiently in watermarking systems, based on various transform domains, since watermarking information is spread out over the entire host image [4]. In this regard, watermarking algorithms based on different transform domains such as the DFT (Discrete Fourier Transform) [6], DCT (Discrete Cosine transform) [7], DWT(Discrete Wavelet Transform) [8], Contourlet Transform [9] and others have been proposed [10]. ORuanaidh et al [11] initially proposed in the use of DFT phase for watermarking. In their proposed method, the watermark is embedded in the most significant frequency components of an image where only the DFT phase is used for embedding. Extraction is carried out using a statistical model. Zou et al.[12], developed
a watermarking method based on combining DFT and Hough transform which results in a more robust system that can endure severe attacks such as printing-scanning, scaling and rotating. However, as its main drawback, DFT based schemes suffer against cropping attacks and the watermark cannot survive if aspect ratio changes, since these changes significantly affect the frequency content of the image.

DCT was first applied for watermarking by Koch and Zhao [4]. During the embedding process, some of the host image regions are selected randomly to embed the watermark. These regions are transformed using DCT and then some medium frequency coefficients are modified. In their seminal paper, Cox et al. [13] proposed a spread spectrum based embedding algorithm selecting the most perceptually significant features which is represented by DCT coefficients of the given image. In this algorithm, a Gaussian watermark sequence is embedded into the 1000 highest magnitude DCT coefficients while low frequency regions around the upper-left corner are not used to preserve invisibility. On the other hand, a combination of DCT and single value decomposition (SVD) was proposed for watermarking [14] in order to increase the imperceptibility while obtaining the highest possible robustness. In this method SVD is applied so that the singular values of the watermark are embedded into the DCT coefficients of the original image. The authors argue that better imperceptibility can be achieved by embedding only the singular values of the watermark into the original image. Moreover, better robustness can be obtained by embedding the highest singular values having the highest energy of the watermark into the DC components of the original image. However, the main drawbacks of DCT-based watermarking techniques relate to shortcomings in robustness against high compression levels and have performed poorly for de-synchronization based attacks such as geometric distortions.

DWT transform based schemes were proposed in order to overcome some of the drawbacks of DCT- and DFT-based systems by using multi-resolution techniques. A few watermarking schemes were also proposed based on combining DCT and DWT in order to provide better performance against some attacks [15]. Other works have been carried out to further develop the DWT-based watermarking methods. In [16] SVD was applied to the watermark and original image coefficients in all the frequency bands of DWT. During the embedding stage, the original image was first decomposed into 4 sub-bands using DWT, and then the SVD was applied on each band by modifying their singular values.

In [17] the coefficients of the original image are quantized in the wavelet domain and the binary watermark is embedded into the wavelet-blocks that can be obtained by grouping four coefficients at different sub-bands at corresponding coordinates. The method has shown promising results against various types of attack, including the geometric and non-geometric attacks. In spite of the success of DWT and its different variants, such as the dual tree complex wavelets transform (DTCWT) [18], and the non-redundant complex wavelets transform (NRCWT) [19], multi-resolution transforms based on DWT suffer of limited directionality in their filtering structure [19].

Images to be watermarked usually contain sharp transitions between objects in the scene such as lines, edges and corners or textural regions. These structures are formed in multiple and fine grain directions and orientations. The coefficients in DWT based transforms cannot accurately represent these structures because of their limited directionality. Although DTCWT-based methods exhibit relevant advantages in comparison with the previous transform domains in this regard by having improved directionality with more orientations and approximate Shift Invariance, it is difficult to design it with perfect reconstruction properties and good filter characteristics to solve line-like edges discontinuities across curves (curve singularities) and geometrical smoothness issues [18].

To overcome this limitation, a variety of transforms such as Ridgelets, Curvelets [20] and Contourlet (CT) [21] have been deployed to provide a better framework for capturing the directionality and the geometry of the scene using multiresolution decomposition. Curvelets and ridgelets, same as DWT, their construction is not associated with a multi-resolution analysis. This and other issues make the discrete implementation of curvelets very challenging as claimed in [22], therefore two different implementations of it have been suggested [20] and [23]. In an attempt to provide a better discrete implementation, The Contourlet transform was developed as an improvement over wavelet and Curvelet and ridgelets [21].

Zaboli and Moin [24] proposed a CT based watermarking using human visual system characteristics. In their method, the host image is first decomposed using CT into four levels. In order to add the watermark, a binary logo is scrambled through a well-known PN sequence in order to enhance the system security and provides a random distribution of original image. A more recent research is carried out based on a combination of SVD and CT [25], where the eigenvalues of a QR watermark matrix are embedded into the eigenvalues of the original images coefficients in the Contourlet domain. This method has shown an improved robustness against various types of attacks such as scaling, compression and filtering. Moreover, the proposed method has better imperceptibility when compared with other Contourlet based watermarking techniques. Although Contourlet aims to better capture the directionality of the image features, this is still insufficient and causing visual artifacts into the host image, which is not a desirable property in applications such as watermarking [23].

Watermarking techniques also can be classified based on the usage of the original image during extraction process. If during the extracting procedure the original image is required this is called non-blind watermarking [13], whereas a technique is called blind if it works under the assumption that the original image will not be available at extraction. In this paper, the main focus is on blind digital image watermarking. In a blind schema the watermark extraction can be obtained by applying statistical methods. Cheng and Huang [26] pointed out that the watermark detection problem can be viewed as a statistical hypothesis testing problem. Therefore, this type of detection requires a suitable modelling for PDF of the host image. Barni et al.[27], applied Weibull PDF in order to model the magnitude of a set of full-frame discrete Fourier transform coefficients. The DWT coefficients modelled using Generalizes Gaussian (GG) [28] or Laplacian PDF [29].
In this paper we propose a new transform domain using DST to the problem of image blind watermarking. The DST shows promising results in image processing applications such as edge detection [22] and image denoising [30], in compare with other transforms such as the DWT and CT, both visually and with respect to the Peak Signal-to-Noise Ratio (PSNR). This leads us to conclude that its directional properties have potential in watermarking. As previously explained, complex structures present in images, such as curves, edges and textural regions are not easy to capture. The Shearlet transform has the ability to capture image features more precisely. For example, edges can be more accurately captured due to the efficient multi-resolution filter which produces more specific directional localization for a higher number of directional components. This means that some features that might remain undetected in one resolution can be spotted in another resolution. This can potentially increase the data embedding capacity for watermarking while preserving the imperceptibility requirements and providing higher robustness. This can be achieved by embedding more information in the edges of the image as the human visual system is less sensitive to changes near the edges. The DST transform offers a plethora of advantages for watermarking problems, namely; (i) it captures directional features more precisely, (ii) it has no restrictions on the number of directions and no constraints on the size of the supports for the shearing. There are also no constraints on the size of the supports for the shearing, unlike, for instance, directional filter banks [22] where using a small window size would result in a performance loss. Therefore, the Shearlet transform is designed to deal with directional and anisotropic features, typically present in images, and has the ability to effectively capture the geometric information of edges.

In our earlier works, we already proposed DST as the transform domain for a non-blind watermarking framework based on spread spectrum [32] and a further refinement of it using perceptual models based on the human visual system [33]. While these earlier works showed the potential of DST for watermarking, they were both limited for their non-blind nature, requiring the original image during extraction. To overcome this limitation, this paper proposed a new framework on blind watermarking using DST.

The novel contributions of this work can be summarized as follows:

- Novel use of DST for blind watermarking applications. This transform has not been used in watermarking applications before, according to the best knowledge of the authors. The only exceptions are its use in our conference papers [32] for a basic non-blind watermarking framework as part of our preliminary work and conference papers [33] for a basic non-blind perceptual watermarking model.
- A fully new framework on blind watermarking using DST. This method is derived based on the statistical decision theory, Bayes decision theory, the Neyman-Pearson criterion, and the distribution of the DST coefficients in the case of grey scale images.
- PDF of the DST coefficients is estimated as a Laplacian distribution. This approach is evaluated against all different attacks using a variety of images (30 images) having different image content and characteristics.

The rest of the paper is organized as follows: Section II provides a brief description of the DST. The proposed watermarking system is described in Section III. Section IV covers the implementation details of the proposed method, where results and comparative evaluations against different attacks are given. Finally, Section V concludes the paper.

II. BACKGROUND: THE DISCRETE SHEARLET TRANSFORM

Shearlet transform is an affine function containing a single mother Shearlet function that is parameterized by scaling, shear and translation parameters with the shear parameter capturing the direction of the singularities [31]. An important advantage of this transform over other transforms is due to the fact that there are no restrictions on the number of directions for the shearing. There are also no constraints on the size of the supports for the shearing, unlike, for instance, directional filter banks [22] where using a small window size would result in a performance loss. Therefore, the Shearlet transform is designed to deal with directional and anisotropic features, typically present in images, and has the ability to effectively capture the geometric information of edges.

The Shearlet transform is implemented by applying a Laplacian pyramid scheme and directional filtering [22]. Shearlets are formed by dilating, shearing and translating the mother function $\psi \in L^2(\mathbb{R}^2)$ [34]. DST is obtained by sampling continuous Shearlet transform on a discrete subset of the Shearlet group $S$, which are associated to an orthonormal basis for $L^2(\mathbb{R}^2)$ [31]. The DST for a mother function $\psi$ is defined as below:

$$SH\{\psi_{j,k,l}\} = 2^{j/2}\psi(B_k A_{2j} - L) : j, k \in Z, L \in Z^2 \}$$

where $j, k, l$ are the scale, orientation and location indexes and

$$A = \begin{pmatrix} 4 & 0 \\ 0 & 2 \end{pmatrix}, B = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

are the dilation matrix and the shear matrix respectively. For a given image $f(N_{rows} \times N_{columns})$, the DST can be expressed as [34]:

$$<f, \psi_{j,l,m}^d> = 2^{3/2} \int_{\mathbb{R}^2} f(\xi) \left\{ \hat{V}(2^{-j}\xi) \hat{w}_{j,l,m}(\xi) e^{2\pi i A_{2j}^{-1} B_{2j}^{-1} m} \right\}$$

where

$$V(\xi_1, \xi_2) = \hat{\psi}_1(\xi_1) X D_0(\xi_1, \xi_2) + \hat{\psi}_1(\xi_2) X D_1(\xi_1, \xi_2)$$

and $X$ denotes the indicator function of the set $D$. $D_0$ and $D_1$ are the horizontal and vertical trapezoids, respectively, $d \in \{0, 1\}$, $\xi = (\xi_1, \xi_2) \in \mathbb{R}^2$, $\xi \geq 0$, $l = (-2^j \ldots, 2^j - 1)$ is the junction of the horizontal trapezoids, $w_{j,l,m}^d(\xi)$ is a window function localized on a pair of trapezoids and $V$ is the pseudo-polar coordinates.

$$D_0 = \{ (\xi_1, \xi_2) \in \mathbb{R}^2 : |\xi_1| \geq 1, \frac{2\xi_2}{\xi_1} \leq 1 \}$$

$$D_1 = \{ (\xi_1, \xi_2) \in \mathbb{R}^2 : |\xi_2| \geq 1, \frac{\xi_1}{\xi_2} \leq 1 \}$$
Thus the Shearlet coefficients can be obtained as

\[ X = \int \int 2^{-3/2} g_j(u, v) \left( w(2^j v - l) \right) \exp \left( 2\pi i \left( \frac{n_1 + n_2}{4} \xi_1 + \frac{n_2}{2} \xi_2 \right) \right) d\xi_1 d\xi_2. \]  

(5)

where \( g_j(u, v) \left( w(2^j v - l) \right) = \hat{f} (\xi) \left\{ \hat{V} (2^{-2j} \xi) \hat{w}_{d,l}(\xi) \right\}, \) is the discrete samples on a pseudo-polar grid. \( \hat{W} \) is a window function localized on a pair of trapezoids, \( g_j(n_1, n_2) \) are the values of the DFT on a pseudo-polar grid, \( n_1 \) and \( n_2 \) are finite sequence of values for a given image \( N_{rows} \times N_{columns} \) and \( u, v \) are the pseudo-polar coordinates \((u, v) \in \mathbb{R}^2\) as follows:

\[ (u, v) = \left( \xi_1, \xi_2 \right) \text{if} \left( \xi_1, \xi_2 \right) \in D_0 \]

\[ (u, v) = \left( \xi_2, \xi_1 \right) \text{if} \left( \xi_1, \xi_2 \right) \in D_1. \]  

(6)

In other words, a DST applies filtering to a given image using the Laplacian pyramid algorithm [35], which is implemented in the spatial domain. This is accomplished in the multiscale partition by decomposing an image into a low-pass and a high-pass filtered image and then downsampling the result by 4. In order to extract the frequency components of the input image, directional localization for different directional components is obtained by translating a window function \( W \). Depending on the chosen shearing filter size, the first level decomposition generates 4 or 8 sub-bands. An illustration of the frequency-domain implemented Shearlet support for 4 scales is shown in Figure 1. Figure 2 shows the structure of the orientations corresponding to each DST sub-bands and the corresponding coefficients for an example image. It is worth noticing that different sub-images have the same size, however; for illustrative purposes in Figure 2(b) they are shown with the different sizes.

### III. DST-BASED BLIND WATERMARKING

In relation to its application for image watermarking, the DST ability to better represent directional features as claimed in [36], may allow watermark embedding to adapt to the diagonal features in the host image more efficiently. In this section, a new DST-based watermarking framework for blind watermarking is developed in order to explore the possible improvements on DST performance against signal processing, geometric and compression based attacks. In addition, this proposed new blind watermark detection scheme for DST coefficients is optimal for non-additive schemes relying on the statistical decision theory.
A. Digital Image Statistical Watermark Detection Based On Discrete Shearlet Transform Domain

Non blind watermarking systems, such as the one proposed in [32], are limited in their application field, since they require access to the host image during the detection process. However, this is not always the case for some applications such as image authentication [4]. As alternative blind watermarking, targets the recovery of the watermark when the host (in this case an image) is not available during the detection stage. This makes blind watermarking systems more complicated, but more practical since the original image is not required in the receiver side. In order to reconstruct the watermark, blind schemas assume that original and watermarked coefficients are strongly correlated [37]. Under this assumption, the watermark detection problem can be viewed as a statistical hypothesis testing problem [37]. Thus, the statistical behaviour of the noisy transformed coefficients can be used to derive a decision rule which decides whether a candidate watermark is actually embedded in the data (hypothesis $H_1$) or not (hypothesis $H_0$). In this section a new blind watermark detection scheme for DST coefficients is proposed as optimal for non-additive schemes relying on the statistical decision theory. The proposed method is derived according to the Bayes decision theory, the Neyman-Pearson criterion which is used to minimize the missing detection probability subject to a fixed false alarm probability ($P_{FA}$) [38], and PDF distribution of the DST coefficients.

B. DST coefficient probability distribution function

In order to apply the decision theory and derive the optimum behaviour of the ML (Maximum-likelihood) detector, a suitable distribution model for PDF of the DST coefficients is required as a first step. We have estimated PDF of the DST coefficients for thirty images (see experimental section for more details on the data) for all five resolutions and 49 sub-bands.

It can be noticed, as shown in Figure 3, that the statistical model of the Shearlet approximates a Laplacian distribution, therefore this model was chosen can be better modelled as a Normal Inverse Gaussian (NIG). The Laplacian distribution is defined as follows:

$$f(\chi) = \frac{\lambda}{2} \exp(-\lambda|\chi|)$$  \hspace{1cm} (7)

The Laplacian is symmetrical about zero, and it can be readily matched to the sample DST distribution by finding the appropriate parameter for $\lambda$. It is also worth to notice that the statistical model of the Shearlet coefficients can be modelled as a Normal Inverse Gaussian (NIG)\(^1\) [39]. However, in our case, this distribution is not best choice due to have high computational cost caused by complexity of the mathematical structure of this distribution (it contains four variables that need to be estimated simultaneously) which leads to difficulty in order to apply the central limit theorem [40]. This is required in our blind watermarking framework, in order to calculate $P_{FA}$. Therefore, and as the most suited distribution model, the Laplacian distribution, was chosen. An example is shown in Figure 4 where the DST coefficients distribution averaged for both all thirty images and all the fourth level sub-bands are illustrated and compared with a Laplacian, Gaussian distribution and NIG approximations.

In order to validate the previous findings, the similarity between the real DST coefficient distribution and the hypothetical distribution models using NIG, Laplacian and Gaussian, are estimated using Relative Entropy (Kullback-Leibler divergence). The Relative Entropy, $D$, measured how well our hypothetical distribution $Q$ fills the observation of the real distribution $P$ between the DST coefficients and the estimated one $Q$, and is obtained as below, where achieving smaller value for $D$ implies greater similarity between two distributions, being $D \geq 0$.

$$D_{KL}(P \parallel Q) = \sum_x p(x) \log_2 \left( \frac{p(x)}{Q(x)} \right).$$  \hspace{1cm} (8)

The $D$ value obtained 12, 17 and 25 for NIG, Laplacian and Gaussian distribution, respectively. These results confirm that NIG is the nearest distribution to the real one while the Laplacian distribution remains a good approximation to the NIG model.

C. Hypothesis Testing Problem and Formulation

Given an image $I$, the aim is to verify whether the image $I$ contains the watermark $W^*$ (chosen from the sequence of possible watermark $W$) or not. By applying statistical detection theory the following hypotheses are under consideration [40]:

\(^1\)Please note that this assumption is not proven theoretically, but it is the result of observation by curve fitting on our 30 images dataset. Please also note this assumption is not critical in our theoretical development where Laplacian model is applied.
Hypothesis $H_0$:

- **Case 1**: The DST coefficients, $Y$, do not contain any watermark.
- **Case 2**: The DST coefficients, $Y$, contain a watermark other than $W^*$. For notation purpose, we will denote that the DST coefficients $Y$ contain a watermark $w_0$, where $w_0$ is another random watermark selected from a set $W$ of watermarks different from $W^*$.

Hypothesis $H_1$:

- The DST coefficients, $Y$, contain the watermark $W^*$.

The embedding rule adopted in this paper is multiplicative (non-additive) embedding due to its adaptation with frequency domain and the fact that it fulfils invisibility constraints thus increasing system security [41]:

\[ y_i = x_i(1 + \alpha w_i^*) \]  

(9)

where $x = (x_1,...,x_N)$ is a sequence of the original DST coefficients of image $I$, $w^* = (w_1^*,...,w_N^*)$ is the watermark sequence that is uniformly distributed in [-1, 1], $\alpha$ is a gain factor controlling the watermark strength, and $(y_1,...,y_N)$ is the sequence of watermarked DST coefficients of the watermarked image, $I'$. By relying on the decision theory, the observation variables are the vector $Y$ of possibly marked coefficients. The likelihood ratio of these coefficients to be watermarked $l(Y)$ is obtained as:

\[ l(Y) = \frac{f_y(Y|w^*)}{f_y(Y|w_0)} \leq T \]  

(10)

where $f_y(y|w)$ is PDF of the vector $Y$ conditioned to $w$ and $T$ is the decision threshold. Note that, for Hypothesis $H_0$, Case 1 and Case 2 can be treated together under the assumption that $w_0$ is allowed to include the null sequence.

As we deal with image watermarking in this paper, therefore following assumptions are read for sake of mathematical calculation.

**Lemma 1**: The components of $Y$ are independent of each other and $Y$ satisfies $f_y(Y|w_0) > 0$ by considering Hypotheses $H_0$ and $H_1$ and equation (15), it can be shown that:

\[ H_0 = \text{case 1}: y_i = x_i \]  

(11)

\[ H_0 = \text{case 2}: y_i = x_i(1 + \alpha w_0i) \Rightarrow x_i = \frac{y_i}{1 + \alpha w_0i} \]  

(12)
To further calculate the likelihood ratio, PDF of the DST coefficients is required. By assuming the previously justified Laplacian distribution as PDF of the DST coefficients:

\[ f(x_i) = \frac{\sqrt{2}}{2\sigma_i} \exp\left(-\frac{\sqrt{2}}{\sigma_i} |x_i - \mu_i| \right) \]  

which is equivalent to the following expression when using \( \frac{\sqrt{2}}{\sigma_i} = \lambda \)

\[ f(x_i) = \frac{\lambda}{2} \exp(-\lambda |x_i - \mu_i|) \]

where \( \mu_i \) and \( \sigma_i^2 \) are the mean and variance of the sub-band to which the coefficients belong.

**Lemma 2:** Barni and Bartolini [40] formulated that, under the assumption of an imperceptible watermark, i.e. when the embedding strength is set to be much smaller than one (\( \alpha \ll 1 \)), then:

\[ P(y|w) \approx P(y|0) \]  

In this case the integral is very small and centered at \( y_i \), therefore the component can be linearly approximated using Taylors theorem. By applying the previous change of notation and a new **Lemma 2**, \( l(y) \) is defined as follows:

\[ l(y) = \prod_{i=1}^{N} \left( \frac{\lambda e^{-\lambda(y_i - \mu_{i,x} - \mu_{i,w_i^*})}}{\lambda^N e^{-\lambda(y_{i} - \mu_{i,x} - \mu_{i,w_i^*})}} \right) \]

\[ T_2 = \left( \frac{1}{2} \right)^N \]

The detector decide \( H_1 \) if \( \text{Ln}l(y) > \text{Ln}T_2 \)
The detector decide \( H_0 \) if \( \text{Ln}l(y) < \text{Ln}T_2 \)

The likelihood ratio is obtained as follows:

\[ \frac{\sum_{i=1}^{N} (|y_i - \mu_{i,x} - |1 + \alpha_i w_i^*|^{-1}|y_i - \mu_{i,x} - \mu_{i,w_i^*}|) \leq \frac{1}{\lambda} \text{Ln}(T_2) \]

where \( \lambda = \frac{\sqrt{2}}{\sigma_i} \)

\[ \frac{\sum_{i=1}^{N} (|y_i - \mu_{i,x} - |1 + \alpha_i w_i^*|^{-1}|y_i - \mu_{i,x} - \mu_{i,w_i^*}|) \leq \frac{\sigma_i}{\sqrt{2}} \text{Ln}(T_2) \]

\[ T_3 = \frac{\sigma_i}{\sqrt{2}} \text{Ln}(T_2) \]

By simplifying, \( g_i = (|y_i - \mu_{i,x} - |1 + \alpha_i w_i^*|^{-1}|y_i - \mu_{i,x} - \mu_{i,w_i^*}|) \) the decision rule is obtained as follows:

\[ Z(y) = \sum_{i=1}^{N} g_i > T_3 \]  

**D. Decision Threshold**

By analysing the decision rule obtained from the previous section, it can be seen that the detector operates by comparing the likelihood ratio against a detection threshold:

\[ T = \frac{p_0(l \mid H_0)}{p_1(l \mid H_1)} \]

where \( p_0 \) and \( p_1 \) are the prior probability of hypotheses \( H_0 \) and \( H_1 \), respectively. In a desirable system, the threshold should be set to minimize the overall error probability \( P_e \). This can be achieved by setting the missed detection probability \( P_m \) (failure to detect the presence of the watermark in an image that contains one) and the false alarm probability \( P_{FA} \) (detection of watermark in an image when it does not actually contain one) to be equal. However, in the case of an attack, the threshold selected to minimize the error probability \( P_e \) will not be suitable since the missed detection probability \( P_m \) becomes higher than the false alarm probability \( P_{FA} \).

In order to address this issue, the Neyman-Pearson criterion can be used to obtain the threshold \( T \) in such a way that the missed detection probability is minimized, subject to a fixed false alarm probability [38].

\[ D = (H_1 | R = H_0) \]

\[ D = (H_1 | R = H_0) \]

\[ P_{FA} = P(D) \]

\[ = P(Z(y) > T | w_0) = P(Z(y) > T) \]

\[ = \int_{T}^{\infty} f_z(Z)dz \]

where

\[ Z(x) = Z(y) \]

\[ = P(Z(y) > T | w_0) = P(Z(y) > T) \]

\[ = \frac{\sqrt{2}}{\sigma_i} (|x_i - \mu_{i,x} - |1 + \alpha_i w_i^*|^{-1}|x_i - \mu_{i,x} - \mu_{i,w_i^*}|) \]

(21)

By applying the central limit theorem, PDF of the \( Z(x) \) can be assumed to be a normal distribution [38] with mean and variance as follows:

The mean can be derived as:

\[ \mu_{Z(x)} = E[Z(x)] \]

\[ = E\left( \frac{\sqrt{2}}{\sigma_i} (|x_i - \mu_{i,x} - |1 + \alpha_i w_i^*|^{-1}|x_i - \mu_{i,x} - \mu_{i,w_i^*}|) \right) \]

\[ = E[|x_i - \mu_{i,x} - |1 + \alpha_i w_i^*|^{-1}|x_i - \mu_{i,x} - \mu_{i,w_i^*}|)] \]

\[ = \sum_{i=1}^{N} \left[ 1 - |1 + \alpha_i w_i^*|^{-1}(\lambda | \mu_{i,x} - \mu_{i,w_i^*}| + \frac{1}{\lambda} \exp(-\lambda(\mu_{i,x} - \mu_{i,w_i^*})) \right) \]

(22)

Similarly for calculating variance:

\[ \sigma_{Z(x)}^2 = E[Z^2(x)] \]
\[
\sigma_z^2(x) = \sum_{i=1}^{N} \left[1 + |1 + \alpha_i w_i^*|^{-2} \right] \left(2 - \exp(-2\lambda|\mu_i\alpha_i w_i^*|)\right) - 2|1 + \alpha_i w_i^*|^{-1} \exp(-\lambda(\mu_i\alpha_i w_i^*)) - \lambda|\mu_i\alpha_i w_i^*| \exp(-\lambda(\mu_i\alpha_i w_i^*)^2) \right])
\] (24)

Then, false alarm probability can be calculated:

\[
P_{FA} = \int_{T}^{\infty} f_z Z(x) dz \]
\[
= \int_{T}^{\infty} \frac{1}{\sqrt{2\pi\sigma_z^2}} \exp\left(-\frac{|Z_x - \mu_z|^2}{2\sigma_z^2}\right) dz(x)
\]
\[
= Q\left(\frac{T - \mu_z}{\sigma_z}\right)
\] (25)

where \(Q\) is the Q-function or tail probability of the standard normal distribution of \(Z(x)\):

\[
P_{FA} = Q\left(\frac{T - \mu_z}{\sigma_z}\right) \Rightarrow Q^{-1}(P_{FA}) = \frac{T - \mu_z}{\sigma_z}
\] (26)

Finally, the threshold will be obtained as below:

\[
T = \sigma_z Q^{-1}(P_{FA}) + \mu_z.
\] (27)

The embedding and detecting framework proposed for our blind watermarking system, is depicted in Figure 5. During the embedding process [Upper block], first the original image \(N \times N\), is decomposed to 5 levels using DST, then the watermark consists of a sequence of random real numbers uniformly distributed in the range [-1, 1] of length \(N\) is generated and embedded into the original image \(I\). Once the watermark is embedded into the Discrete Shearlet coefficients, the image is recomposed to create the watermarked image \(I'\). The watermarked image is then passed through the attack channel [lower block] where some distortions are applied in order to remove the watermark. This produces the attacked image \(I''\) that is then passed to the detecting stage. It is important to remember that in this blind schema, the original image is not available during the detection stage. Instead, a statistical model is used during the decision stage and calculated directly from the watermarked and possibly attacked images.

**IV. PERFORMANCE EVALUATION**

To verify the effectiveness of the proposed algorithm, a series of experiments were conducted.

**A. Dataset**

In our experiments, thirty \(512 \times 512\) sized well-known grayscale images were used as host images. A set of standard test images which are used frequently in the literature were selected from a wide range of image processing databases[42] to represent different image features (Figure 6). Some of these images are smooth with a lack of detailed features, others are more complex with a lot of edges and some textured regions. The rest contains high detail textured regions. This set is selected from the following references [10],[19].

**B. Blind Watermarking**

In this section, the performance of the blind statistical detector described in Section III.B is tested on the thirty standard greyscale images \(512 \times 512\) images (Figure 6). The original image is not available during the detection stage. Instead, a statistical model is used during the decision stage and calculated directly from watermarked and possibly corrupted images. Each image is transformed using DST and the watermark consists of a sequence of random real numbers uniformly distributed in the range [-1, 1]. The watermark is embedded in the most significant coefficient through all DST levels at the 5th level of resolution and sub-bands of the host image. The watermark detection is performed in the transform domain using maximum-likelihood detection, whereby the decision threshold is calculated using the Neyman-Pearson criterion. In order to investigate the performance of our proposed method, the results from all the different method are compared under the same conditions. DWT coefficients were selected from the
3rd level of resolution as suggested by [43]. CT coefficients were selected from the sub-band on the 3rd levels of resolution as suggested in [3] to optimise imperceptibility and robustness. Three performance metrics were taken into account during this analysis: The imperceptibility of the watermark by using PSNR, the Root-mean squared error (RMSE), and Structural similarity (SSIM) as fidelity measurements, the probability of false alarm and the probability of missed detection and the robustness of the watermark against a number of commonly used attacks. In particular SSIM measures the quality of the image using an initial distortion-free image as reference. SSIM is designed to improve traditional methods such as PSNR and MSE which have been proved to be inconsistent with the human eye perception [44]. The resulting SSIM index is a decimal value between -1 and 1, where 1 is only reachable in the case of two identical sets of data.

By comparing the results using SSIM (Figure 7) and RMSE (Figure 8) as metrics, it is concluded that the proposed algorithm based on DST has a better imperceptibility as reflected in having smaller RMSE (which indicates that the watermarked image is close to the original one on a pixel-by-pixel basis) and higher similarity SSIM, where, more closer to 1 indicates the watermarked image is more similar to original one. Among the reasons for this improved imperceptibility, we can cite: the smaller sizes of the shearing filters represented by (eq.2) in comparison with the directional filters used by DWT and CT [22], having the greater windowing flexibility as represented by (eq.3,4) as claimed in [22] that can be utilized and makes possible Incorporating sub-sampling and providing additional directional information. On the other word, by choosing smaller size of filters we can represent edges more precisely and by having greater windowing flexibility we can develop a variety of alternative implementations. This is more noticeable by considering each transform reaction based on image characteristic. For example, DST is more adapted with images having a lot of edges and textured regions (Barbara). For images having smooth areas with a lack of detailed features (Bunny), DST adaptation is still better than CT and DWT. DST also adapted perfectly for images contain mostly high detail textured regions such as Baboon.

1) Robustness: To investigate the effects of attacks on the blind watermarking algorithm, different tests were carried out to evaluate its performance. The results are compared against an equivalent DWT and CT blind watermarking schemas, as it was shown that the DWT and CT coefficient distributions can be also expressed using Laplacian model [43]. In order to ensure a fair comparison, given that every method has a different imperceptibility/robustness balance, all the methods were tuned to provide an approximately 43dB PSNR value before the attack [43]. In this regard, the alpha value is set to 0.25 for DWT, 0.2 for CT and 0.2 for DST. During blind detection, the parameters of the proposed model are directly estimated from the DST, CT and DWT coefficients of the watermarked image to fulfil the assumption that watermarked
image is close to the original one if strength parameter $\alpha$, is much less than 1 ($\alpha << 1$). It is to be noted that, in practice, our chosen strength parameter values 0.2 and 0.25 will be acceptable [43] under this approximation while providing acceptable levels of robustness. The embedding was performed in all the coefficients obtained from the 3rd level of decomposition for DST and 3rd level for DWT and CT in order to provide the better resolution and therefore the biggest payload that each method allows.

Table I shows the average performance of the False alarm rate (FA), which represents the watermarks that were detected when that watermark was not actually embedded, Missed detection (MD), embedded watermarks that were not detected and True positive (TP) which represents the embedded watermarks that each method allows.

In the first attack, Gaussian noise is added to the watermarked images. To extract the watermark, the missing part(s) of the images are cropped by cutting off 25%, 50% and 75% of some random part of the image. To apply Gaussian noise with standard deviations 0.01, 0.05 and 0.09 on the watermarked image, the missing part(s) of the image will become out of focus which would not valuable for the attacker. From these experimental results represented in Table II, it is found that DST provides comparable robustness to its counterpart in terms of AWGN attack, consistently better than CT and DWT.

In the second attack, Gaussian low pass filter is applied to the watermarked image to analyse the effect of blurring using standard deviation varied from 0.3, 0.5 and 0.8 and $3 \times 3$ spatial filter. From these experimental results represented in Table III, it is found that DST also performs better against blurring attacks in when compared against DWT and CT counterpart.

### Table I

<table>
<thead>
<tr>
<th>Laplacian Model</th>
<th>$P_{FA} = 10^{-3}$</th>
<th>DWT</th>
<th>CT</th>
<th>DST</th>
</tr>
</thead>
<tbody>
<tr>
<td>FA</td>
<td>23</td>
<td>17</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>MD</td>
<td>7</td>
<td>5</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>TP</td>
<td>92</td>
<td>96</td>
<td>99</td>
<td></td>
</tr>
</tbody>
</table>

The effect of five attacks, including Additive White Gaussian Noise (AWGN), Compression, Blurring, Cropping and Rotation are tested on the all 30 watermarked images. For each attack, the detector responses were related to the actual embedded watermark. Table II-VI contains the number of successful detection for most commonly used attacks on each individual watermarked image as well as the global average and average of False Alarm rate and Missed detection. It is worth noting these results were obtained based on 3000 trials.

In the first attack, Gaussian noise is added to the watermarked image with zero mean and standard deviations 0.01, 0.05 and 0.09. It can be possible to add bigger values than 0.09, this will make the image out of focus which would make the quality of the image so distorted that the image would not valuable for the attacker. From these experimental results represented in Table II, it is found that DST provides comparable robustness to its counterpart in terms of AWGN attack, consistently better than CT and DWT.

In the second attack, Gaussian low pass filter is applied to the watermarked image to analyse the effect of blurring using standard deviation varied from 0.3, 0.5 and 0.8 and $3 \times 3$ spatial filter. From these experimental results represented in Table III, it is found that DST also performs better against blurring attacks in when compared against DWT and CT counterpart.

In the third attack, the watermarked images are cropped by cutting off 25%, 50% and 75% of some random part of the images. To extract the watermark, the missing part(s) of the image should be replaced with those parts of the original non...
In the fourth attack, the watermarked image is compressed to provide an output quality of 50%, 70% and 90% of the original images. No smoothing is applied. According to Table V, it can be concluded that DST performs very well against JPEG compression in comparison with DWT, but in terms of the severe compression attack CT provides slightly better results than DST.

Finally, the watermarked image is slightly rotated and cropped to discard areas of the image that contain less useful information, such as black areas resulting from the rotation by applying 1, 2, 5 and 7 degrees rotation in a counter clockwise direction. According to Table VI, it can be concluded that DST provides very good robustness against rotation attacks in comparison with DWT and CT. More precisely, this is due to DST improved property to capture more directions and having shift-invariant structure that allows Shearlet to capture the image more efficiently.

Table IV NUMBER OF SUCCESSFUL DETECTIONS (TRUE POSITIVES) AND AVERAGE OF FA AND MD FOR ALL 30 IMAGES AFTER APPLYING CROPPING ATTACK BY CUTTING OFF 25%, 50% AND 75% FOR DWT, CT AND DST FOR 3000 TRIALS

<table>
<thead>
<tr>
<th></th>
<th>DWT</th>
<th>CT</th>
<th>DST</th>
</tr>
</thead>
<tbody>
<tr>
<td>Image</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Barrel</td>
<td>97</td>
<td>96</td>
<td>97</td>
</tr>
<tr>
<td>Barbera</td>
<td>99</td>
<td>86</td>
<td>96</td>
</tr>
<tr>
<td>Boat</td>
<td>98</td>
<td>94</td>
<td>77</td>
</tr>
<tr>
<td>Bunny</td>
<td>98</td>
<td>75</td>
<td>99</td>
</tr>
<tr>
<td>Cameraman</td>
<td>98</td>
<td>89</td>
<td>89</td>
</tr>
<tr>
<td>Clock</td>
<td>95</td>
<td>87</td>
<td>77</td>
</tr>
<tr>
<td>Elaine</td>
<td>97</td>
<td>88</td>
<td>91</td>
</tr>
<tr>
<td>F16</td>
<td>100</td>
<td>90</td>
<td>86</td>
</tr>
<tr>
<td>Fingerprint</td>
<td>100</td>
<td>91</td>
<td>83</td>
</tr>
<tr>
<td>Flinstone</td>
<td>100</td>
<td>97</td>
<td>85</td>
</tr>
<tr>
<td>Flower</td>
<td>100</td>
<td>86</td>
<td>80</td>
</tr>
<tr>
<td>Girl</td>
<td>96</td>
<td>80</td>
<td>81</td>
</tr>
<tr>
<td>House</td>
<td>97</td>
<td>95</td>
<td>78</td>
</tr>
<tr>
<td>Jelly Beans</td>
<td>94</td>
<td>92</td>
<td>93</td>
</tr>
<tr>
<td>Lake</td>
<td>97</td>
<td>72</td>
<td>100</td>
</tr>
<tr>
<td>Lena</td>
<td>100</td>
<td>97</td>
<td>86</td>
</tr>
<tr>
<td>Living room</td>
<td>100</td>
<td>92</td>
<td>84</td>
</tr>
<tr>
<td>Moon surface</td>
<td>96</td>
<td>92</td>
<td>98</td>
</tr>
<tr>
<td>Peppers</td>
<td>100</td>
<td>97</td>
<td>71</td>
</tr>
<tr>
<td>Pirate</td>
<td>97</td>
<td>83</td>
<td>100</td>
</tr>
<tr>
<td>Scientist</td>
<td>97</td>
<td>94</td>
<td>89</td>
</tr>
<tr>
<td>Splash</td>
<td>97</td>
<td>87</td>
<td>95</td>
</tr>
<tr>
<td>Straw</td>
<td>99</td>
<td>87</td>
<td>94</td>
</tr>
<tr>
<td>Tree</td>
<td>100</td>
<td>96</td>
<td>73</td>
</tr>
<tr>
<td>Truck</td>
<td>100</td>
<td>92</td>
<td>78</td>
</tr>
<tr>
<td>Walk bridge</td>
<td>100</td>
<td>95</td>
<td>79</td>
</tr>
<tr>
<td>Woman-blonde</td>
<td>98</td>
<td>85</td>
<td>87</td>
</tr>
<tr>
<td>Woman dark hair</td>
<td>99</td>
<td>97</td>
<td>98</td>
</tr>
<tr>
<td>zebra</td>
<td>97</td>
<td>87</td>
<td>95</td>
</tr>
<tr>
<td>Average</td>
<td>98.53</td>
<td>94.9</td>
<td>76.4</td>
</tr>
</tbody>
</table>

Table V NUMBER OF SUCCESSFUL DETECTIONS (TRUE POSITIVES) AND AVERAGE OF FA AND MD FOR ALL 30 IMAGES AFTER APPLYING JPEG ATTACK USING QUALITY OF 50%, 70% AND 90% FOR DWT, CT AND DST FOR 3000 TRIALS

<table>
<thead>
<tr>
<th></th>
<th>DWT</th>
<th>CT</th>
<th>DST</th>
</tr>
</thead>
<tbody>
<tr>
<td>Image</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Barrel</td>
<td>97</td>
<td>96</td>
<td>97</td>
</tr>
<tr>
<td>Barbera</td>
<td>99</td>
<td>86</td>
<td>96</td>
</tr>
<tr>
<td>Boat</td>
<td>98</td>
<td>94</td>
<td>77</td>
</tr>
<tr>
<td>Bunny</td>
<td>98</td>
<td>75</td>
<td>99</td>
</tr>
<tr>
<td>Cameraman</td>
<td>98</td>
<td>89</td>
<td>89</td>
</tr>
<tr>
<td>Clock</td>
<td>95</td>
<td>87</td>
<td>77</td>
</tr>
<tr>
<td>Elaine</td>
<td>97</td>
<td>88</td>
<td>91</td>
</tr>
<tr>
<td>F16</td>
<td>100</td>
<td>90</td>
<td>86</td>
</tr>
<tr>
<td>Fingerprint</td>
<td>100</td>
<td>91</td>
<td>83</td>
</tr>
<tr>
<td>Flinstone</td>
<td>100</td>
<td>97</td>
<td>85</td>
</tr>
<tr>
<td>Flower</td>
<td>100</td>
<td>86</td>
<td>80</td>
</tr>
<tr>
<td>Girl</td>
<td>96</td>
<td>80</td>
<td>81</td>
</tr>
<tr>
<td>House</td>
<td>97</td>
<td>95</td>
<td>78</td>
</tr>
<tr>
<td>Jelly Beans</td>
<td>94</td>
<td>92</td>
<td>93</td>
</tr>
<tr>
<td>Lake</td>
<td>97</td>
<td>72</td>
<td>100</td>
</tr>
<tr>
<td>Lena</td>
<td>100</td>
<td>97</td>
<td>86</td>
</tr>
<tr>
<td>Living room</td>
<td>100</td>
<td>92</td>
<td>84</td>
</tr>
<tr>
<td>Moon surface</td>
<td>96</td>
<td>92</td>
<td>98</td>
</tr>
<tr>
<td>Peppers</td>
<td>100</td>
<td>97</td>
<td>71</td>
</tr>
<tr>
<td>Pirate</td>
<td>97</td>
<td>83</td>
<td>100</td>
</tr>
<tr>
<td>Scientist</td>
<td>97</td>
<td>94</td>
<td>89</td>
</tr>
<tr>
<td>Splash</td>
<td>97</td>
<td>87</td>
<td>95</td>
</tr>
<tr>
<td>Straw</td>
<td>99</td>
<td>87</td>
<td>94</td>
</tr>
<tr>
<td>Tree</td>
<td>100</td>
<td>96</td>
<td>73</td>
</tr>
<tr>
<td>Truck</td>
<td>100</td>
<td>92</td>
<td>78</td>
</tr>
<tr>
<td>Walk bridge</td>
<td>100</td>
<td>95</td>
<td>79</td>
</tr>
<tr>
<td>Woman-blonde</td>
<td>98</td>
<td>85</td>
<td>87</td>
</tr>
<tr>
<td>Woman dark hair</td>
<td>99</td>
<td>97</td>
<td>98</td>
</tr>
<tr>
<td>zebra</td>
<td>97</td>
<td>87</td>
<td>95</td>
</tr>
<tr>
<td>Average</td>
<td>98.53</td>
<td>94.9</td>
<td>76.4</td>
</tr>
</tbody>
</table>

Based on the results obtained as described above, it can be concluded that the proposed DST blind watermarking method provides better results, in terms of robustness, compared to DWT and CT watermarking based techniques using the same statistical model (Laplacian). This is due to the fact that DST has a greater windowing flexibility that can be utilized to capture the image characteristics like curve and edges. This is more noticeable by considering each transform reaction based on image characteristic. For example, DST is more adapted with images having a lot of edges and textured regions (Barbera). For images having smooth areas with a lack of detailed features (Bunny), DST adaptation is still better than DWT and CT. DST also adapted perfectly for images contain mostly high detail textured regions such as Baboon.

V. Conclusion

In this paper we have proposed a novel blind watermarking framework based on the Discrete Shearlet Transform for blind image watermarking. This idea is justified through its
structure and potential to provide higher payload and better imperceptibility. A blind system framework was implemented to test the suitability of DST for watermarking based on decision theory. This system presents theoretical novelties in the latter structure and the probabilistic model in order to allow DST to be integrated. As a main advantage this blind watermarking method does not require the transmission of the original clean image. To achieve this, the distribution of the DST coefficients for different sub-bands and resolutions are investigated. Thus, PDF obtained from DST coefficients is modeled using a Laplacian channel. This model has proved to be effective and simpler, allowing the corresponding mathematical description of the full framework. Finally, a maximum likelihood detection scheme based on Laplacian modelling of the DST coefficients is implemented under a hypothesis condition using detection rules based on the Neyman-Pearson criterion in order to improve the robustness as well as adapting the watermark strength to the host image by considering the visual sensitivity. The proposed method is less sensitive to fine parameter tuning in comparison with non-blind methods [33], i.e. parameters can remain unchanged even under different attacks and the original image is not required during the detection stage. From the experimental results it is found that the DST based embedding provides a good imperceptibility and an improved payload as predicted. In terms of robustness, the results demonstrate superior robustness against common image processing manipulations compared to DWT and CT. This is more obvious in compression, noise and rotation attacks.

REFERENCES


