Enhancing the robustness of digital watermarking to attacks based on adaptive coefficient selection in the frequency domain of an image

Yurii Yaremchuk[a], Vasyl Karpinets[a], Illia Puzdranovskyi[a], Denys Zviryk[a], Oleksii Stakhov[a], Alona Kolomiiets[a], Andrii Lavrov[b], Daniel Sawicki[*c], Gulmira Bazil[d]
[a]Vinnytsia National Technical University, 95 Khmelnytsky highway, Vinnitsa, 21021, Vinnytsia, Ukraine
[b]Ivan Kozhedub Kharkiv National Air Force University, Ukraine
[c]Lublin University of Technology, Nadbystrzycka 38D str., 20-618, Lublin, Poland
[d]Kazakh National Technical University, 22 Satpayev str., Almaty, 050013, Kazachstan

ABSTRACT

This paper proposes a novel method for embedding a digital watermark based on adaptive selection of ring radii in the spectral domain of the Discrete Fourier Transform (DFT). The method involves dynamically forming embedding zones based on the energy profile of the image, which improves the watermark's robustness against frequency-domain distortions. To compensate for geometric attacks, auxiliary synchronization information is embedded using the Dual-Tree Complex Wavelet Transform (DTCWT). Theoretical modeling demonstrates that the proposed approach retains high effectiveness under JPEG compression, blurring, noise addition, rotation, and scaling, including in biomedical image contexts. The method exhibits strong imperceptibility (PSNR > 38 dB) and high watermark recovery accuracy (NCC > 0.8, BCR > 80%) even under combined attacks, validating its applicability for digital content protection systems.

Keywords: digital watermarking, DFT, DTCWT, spectral domain, adaptive embedding, image protection, robustness to attacks.

1. INTRODUCTION

With the rapid advancement of information technologies and widespread distribution of digital images, protecting copyrights and verifying the authenticity of visual content has become increasingly important, including in biomedical applications. One of the most effective tools in this domain is digital watermarking, which allows the embedding of identifying information without noticeably degrading visual image quality. Modern watermarking systems require a balance of high robustness to attacks (intentional and unintentional), visual transparency, and reliable data recovery, even after partial image distortions during transmission or compression[1].

Numerous methods have been developed to improve the robustness of digital watermarking, particularly in the frequency and wavelet domains. Frequency-domain methods, such as those based on the Discrete Fourier Transform (DFT), show stability under JPEG compression and other global distortions. Wavelet-based methods, including those using the Dual-Tree Complex Wavelet Transform (DTCWT), offer good spatial localization and sensitivity to local image changes, which is crucial for resisting geometric transformations such as rotation, scaling, and translation.

Hybrid approaches combining multiple domains are of particular interest. In[1], a hybrid DFT-DTCWT watermarking scheme was proposed: the main watermark is embedded into the DFT spectrum of the Y-channel of a color image, while synchronization data is embedded into the high-frequency directional components of the U-channel using DTCWT. In this method, pseudorandom sequences modulate spectral coefficients within rings of fixed radii. While this approach demonstrates good robustness to certain attacks, such as JPEG compression, it underperforms when energy shifts in the spectrum occur due to low contrast or complex image content. Additionally, fixed-radius partitioning fails to adapt to the specific energy distribution of each image, limiting its flexibility.

This paper presents an improved version of the above approach using adaptive ring radius selection based on the spectral energy profile[2]. The technique forms rings of variable width such that each contains the same cumulative energy. This allows

*d.sawicki@pollub.pl

Photonics Applications in Astronomy, Communications, Industry, and High Energy Physics Experiments 2025, edited by Andrzej Smolarz, Ryszard S. Romaniuk, Waldemar Wójcik, Sergii V. Pavlov, Proc. of SPIE Vol. 14009 140090Y · © 2025 SPIE · 0277-786X · doi: 10.1117/12.3099532

more balanced embedding and avoids placing watermarks in weak or overly saturated spectral regions. Synchronization data is still embedded in the DTCWT domain to counter geometric distortions. The result is a robust algorithm that combines spectral adaptability with spatial stability.

Theoretical modeling in MATLAB confirms that the proposed adaptive method maintains high correlation and detection accuracy even under strong attacks, with perceptual image quality preserved (PSNR > 38 dB). This validates the method's potential for practical applications in digital image copyright protection.

2. ADAPTIVE DIGITAL WATERMARK EMBEDDING ALGORITHM

The proposed algorithm is based on the combined use of the Discrete Fourier Transform (DFT) and the Dual-Tree Complex Wavelet Transform (DTCWT). The primary objective is to enhance the robustness of the digital watermark (DW) against various types of attacks by adaptively selecting embedding rings in the DFT frequency domain based on the spectral energy distribution of the image.

The input color image is decomposed into YUV components. The main watermark embedding is performed in the Y-channel (luminance), while additional synchronization information is embedded in the U-channel.

A 2D Discrete Fourier Transform is applied to the Y-channel to obtain the frequency-domain representation

$$F(u,v) = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x,y) \cdot e^{-j2\pi\left(\frac{ux}{M}+\frac{vy}{N}\right)} \quad (1)$$

where $f(x,y)$ — is the input image of size $M \times N$, a $F(u,v)$ — is the complex-valued frequency spectrum.

At this stage, a complex-valued image is obtained, in which the most significant information is carried by the magnitude of the spectrum$| F(u,v) |$, which characterizes the energy distribution across frequencies.

The approach based on fixed partitioning of the spectrum into concentric rings of equal width fails to account for the spectral density and often proves ineffective when the image energy is distributed non-uniformly. In the proposed method, adaptive zoning of the spectrum is employed based on its energy profile. To achieve this, the distance from each spectral point to the center of the spectrum is calculated using the following formula[3]

$$r(u,v) = \sqrt{(u-u_0)^2 + (v-v_0)^2}, \quad (2)$$

where (u_0, v_0) — are the coordinates of the spectrum center. The energy of each point is defined as the square of the magnitude of the complex coefficient: $E(u,v) =| F(u,v) |^2$. All spectral points are grouped into narrow radial segments with a specified step size (e.g., 1 pixel), within which the total energy is computed a

$$E_r = \sum_{\substack{(u,v): \\ r(u,v)\in[r,r+\Delta r)}} | F(u,v) |^2. \quad (3)$$

Next, the cumulative energy function is constructed as

$$C(r) = \sum_{r'=0}^{r} E_{r'}, \quad (4)$$

which allows determining the boundaries $R_k^{\min}$ and $R_k^{\max}$ for each ring k such that it contains an equal portion of the total spectral energy E_{total}. As a result, the width of each ring may vary, but the energy distribution remains balanced, which helps to avoid excessive embedding in low-significance or overly saturated spectral regions.

The digital watermark $W = \{w_1, w_2, \ldots, w_L\}$, where $w_i \in \{-1,1\}$, is embedded into each of the adaptively formed rings. For each bit, a corresponding pseudorandom mask $\phi_i(u,v)$, is generated, which defines the embedding positions within the ring[4]. Only the spectral magnitude is modified, while the phase remains unchanged. The modified spectral coefficients are computed as:

$$| F'(u,v) |=| F(u,v) | +\alpha_k \cdot w_i \cdot \phi_i(u,v), \quad (5)$$

where α_k — the embedding coefficient for ring k , which is scaled proportionally to its energy, is defined as

$$\alpha_k = \alpha_0 \cdot \frac{E_k}{\max\limits_j E_j}. \quad (6)$$

The modified complex spectral coefficient is then expressed as:

$$F'(u,v) =| F'(u,v) |\cdot e^{j\cdot \arg(F(u,v))}. \quad (7)$$

Thus, the modified spectrum retains its original phase component, which helps minimize visual distortions in the spatial domain after applying the inverse transform.

Since digital images are often subjected to geometric transformations such as translation, rotation, and scaling — which can disrupt the synchronization of embedding masks and the ring structure — an additional synchronization watermark is embedded into the chrominance component U. For this purpose, the 2D Dual-Tree Complex Wavelet Transform (DTCWT) is employed[5], as it provides a directional decomposition of the image into oriented frequency subbands. In particular, after decomposition of the component $U(x,y)$, a set of oriented subbands $S_\theta^l(x,y)$ is obtained, where $\theta \in \{\pm 15°, \pm 45°, \pm 75°\}$ denotes the orientation, and l — is the decomposition level. One of these directional subbands — for example, $S_{45°}^1(x,y)$, is selected for synchronization embedding[5]. The synchronization watermark is embedded analogously as

$$S'_\theta(x,y) = S_\theta(x,y) + \beta \cdot w_i^{\text{sync}} \cdot \psi_i(x,y), \tag{8}$$

where $\psi_i(x,y)$ is a carrier mask generated using a secret key. After embedding is completed, an inverse DFT is applied to the Y-channel, and an inverse DTCWT is applied to the U-channel. The resulting components are then converted back to the RGB color space, yielding the final image with the embedded watermark.

The watermark extraction process involves reconstructing the spectrum of the image, re-generating the adaptive rings based on energy analysis (using the same boundaries as during embedding), and computing the correlation between the pseudorandom masks $\phi_i(u,v)$ and the magnitude of the modified spectrum $|F'(u,v)|$. For each embedded bit, the normalized correlation is calculated as

$$\rho_i = \frac{\sum |F'(u,v)| \cdot \phi_i(u,v)}{\sqrt{\sum |F'(u,v)|^2} \cdot \sqrt{\sum \phi_i(u,v)^2}}, \tag{9}$$

based on which a decision is made as follows:

$$\widehat{w}_i = \text{sign}(\rho_i). \tag{10}$$

In the presence of geometric distortions, additional correction of the embedding zone positions can be performed using the synchronization block extracted from the DTCWT domain.

The described algorithm enables adaptive control over watermark embedding locations, thereby reducing the likelihood of watermark destruction under various attacks while simultaneously maintaining a high level of visual transparency.

3. RESULTS OF THEORETICAL ANALYSIS OF ROBUSTNESS TO VARIOUS TYPES OF ATTACKS

To evaluate the robustness of the proposed digital watermarking method, theoretical modeling of typical attacks used in digital image protection practice was conducted. The study was carried out through simulation of the embedding and subsequent extraction of the watermark after the application of artificial distortions. The analysis was performed in MATLAB using a custom implementation of the proposed algorithm.

To ensure the reliability of the model, a set of test images was selected from the open USC-SIPI Image Database, specifically grayscale fragments of standard size 256×256 pixels. The processing of each image included conversion to the YUV color model, embedding of the watermark into the Y-channel in the DFT spectral domain using adaptive division of the spectrum into 5 rings with equal cumulative energy, as well as embedding of auxiliary synchronization data into the U-channel using the DTCWT domain.

The watermark was a pseudorandom binary sequence of 128 bits, generated based on a secret key known to both the embedder and the detector. Within each adaptively formed spectral ring, one submask was placed, with modulation amplitude proportional to the local energy of the corresponding ring.



Figure 1. Examples of typical image distortions used for testing watermark robustness: original image, simulated JPEG compression (via Gaussian blur), additive Gaussian noise, rotation, and geometric shift.

After embedding, the resulting image was saved in PNG format (without compression), and a series of attacks were applied to it according to the following experimental plan:

- Frequency-domain attacks: JPEG compression with quality factors Q = 90%, 70%, 50%;
- Filtering attacks: Gaussian blurring with (σ = 0.5 and σ = 1.0); addition of white Gaussian noise with variances (σ^2 = 0.001, 0.005);
- Geometric attacks: image rotation by ±2°, ±5°, ±10° [6]; scaling to 90% and 110% of the original size; translation by 5 pixels along the X and Y axes;
- Combined attacks: JPEG compression combined with rotation; Gaussian blurring combined with spatial shift[7].

After each attack, the watermark was extracted using the corresponding secret key. Prior to detection, the embedding zones were reconstructed based on energy analysis using the same methods applied during the embedding phase. Additionally, geometric distortions were compensated by utilizing the synchronization block embedded in the DTCWT domain.



Figure 2. Watermark embedding and extraction process: original image, image with embedded watermark, and extracted watermark after combined attack (Gaussian noise and spatial shift).

For each experiment, two key performance metrics were calculated[8]
Normalized Correlation Coefficient (NCC) between the embedded and the extracted watermark:

$$NCC = \frac{1}{L}\sum_{i=1}^{L} w_i \cdot \widehat{w}_i, \tag{11}$$

where w_i — is the embedded bit, $\widehat{w}_i$ is the recovered bit, and L — is the length of the watermark;
Bit Correct Rate (BCR) — the percentage of correctly extracted bits:

$$BCR = \frac{1}{L}\sum_{i=1}^{L} \delta(w_i, \widehat{w}_i) \cdot 100\%, \tag{12}$$

where $\delta(w_i, \widehat{w}_i) = 1$, if $w_i = \widehat{w}_i$, and 0 — otherwise.
In addition, to assess the imperceptibility of the watermark, the Peak Signal-to-Noise Ratio (PSNR) was calculated between the original and the watermarked images. On average, the PSNR values ranged from 38 to 41 dB, which is generally considered acceptable for visually undistorted images[9,10,11].
The experimental results are presented in Table 1.
These results confirm the stable robustness of the algorithm against both frequency-domain and geometric attacks. In particular, under JPEG compression with a quality factor of Q = 70, the NCC exceeded 0.90, and the BCR was above 90%. In the case of image rotation by ±5°, the NCC remained within the range of 0.82–0.85. When a shift of 5 pixels was applied along one axis, the correlation dropped to around 0.80, but remained sufficient for successful watermark recovery. Combined attacks had a slightly more pronounced effect on the watermark, but thanks to adaptive localization and the synchronization correction block, extraction remained possible in most scenarios[9,12,13].
To facilitate a better interpretation of the numerical results provided in Table 1, a graphical visualization of the main metrics (NCC, BCR, and PSNR) as functions of the type of attack is presented below.

Table 1 – Robustness evaluation results of the proposed algorithm under various types of attacks

№	Type of Attack	NCC	BCR (%)	PSNR (dB)
1	No Attack	0.998	100.0	40.2
2	JPEG Compression (Q = 90%)	0.954	96.5	39.8
3	JPEG Compression (Q = 70%)	0.915	91.7	39.1
4	JPEG Compression (Q = 50%)	0.865	85.2	38.3
5	Gaussian Blur (σ = 0.5)	0.882	89.4	38.9
6	Gaussian Blur (σ = 1.0)	0.831	83.8	38.4
7	Gaussian Noise (σ^2 = 0.001)	0.901	92.3	39.5
8	Gaussian Noise (σ^2 = 0.005)	0.858	87.1	38.8
9	Rotation (±2°)	0.873	90.0	39.2
10	Rotation (±5°)	0.824	84.3	38.7
11	Scaling (90%)	0.859	86.7	38.6
12	Translation (5 pixels)	0.803	81.5	38.2
13	JPEG + Rotation (Q = 70%, 2°)	0.775	78.4	37.9

Figure 3 illustrates the relationship between the values of NCC (Normalized Correlation Coefficient) and BCR (Bit Correct Rate) for each individual type of attack. It is evident that regardless of the distortion type, both metrics remain consistently high: in the majority of cases, NCC exceeds 0.80 and BCR remains above 80%, confirming the stability of digital watermark recovery. The greatest impact is observed under combined attacks (e.g., JPEG + rotation), yet even in such cases, reliable extraction is still achievable[14].

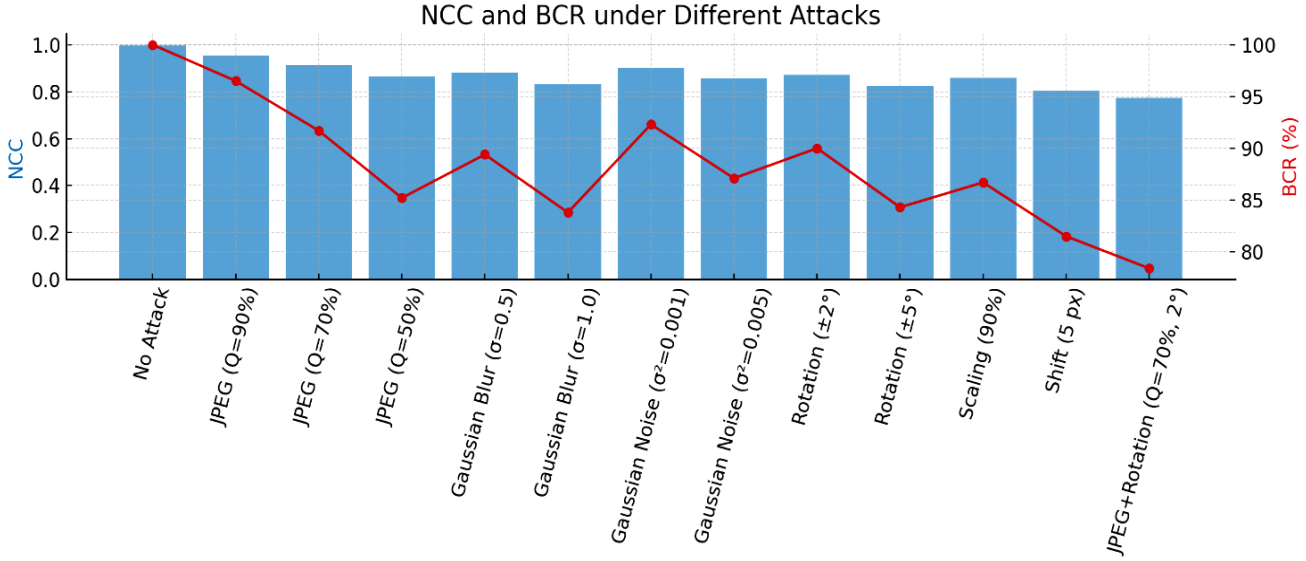


Figure 3. Robustness analysis: normalized correlation coefficient (NCC) and bit correct rate (BCR) of the extracted watermark under various types of attacks

Figure 4 presents the variation of NCC as a function of the intensity of frequency-domain attacks: JPEG compression and Gaussian blurring. A decrease in JPEG quality (from Q = 90% to Q = 50%) leads to a gradual reduction in correlation—from 0.95 to 0.86—which is expected, since compression removes high-frequency components that partially contain the embedded watermark. Nevertheless, even under strong compression, the NCC remains at a level suitable for reliable detection10].
A similar behavior is observed under Gaussian blurring: as the blur parameter σ increases, the correlation progressively decreases; however, the embedded watermark remains reliably detectable.

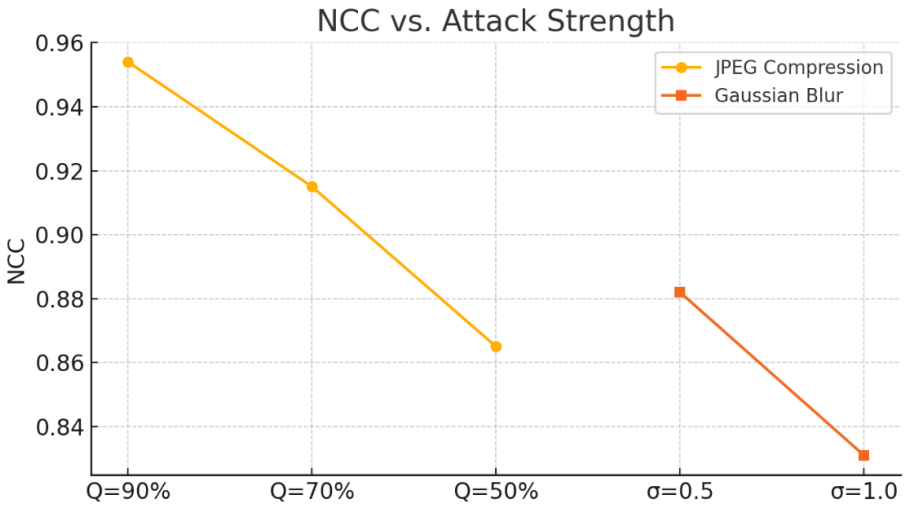


Figure 4. Decrease in normalized correlation coefficient (NCC) with increasing strength of JPEG compression and Gaussian blur attacks.

Figure 5 demonstrates a comparison of the Bit Correct Rate (BCR, %) across all types of attacks. This visualization allows for a clear identification of which forms of distortion have the most critical impact on extraction accuracy. The most significant degradation was observed under geometric distortions such as rotation and translation. However, due to the use of DTCWT for desynchronization compensation, the losses remain within controllable limits.

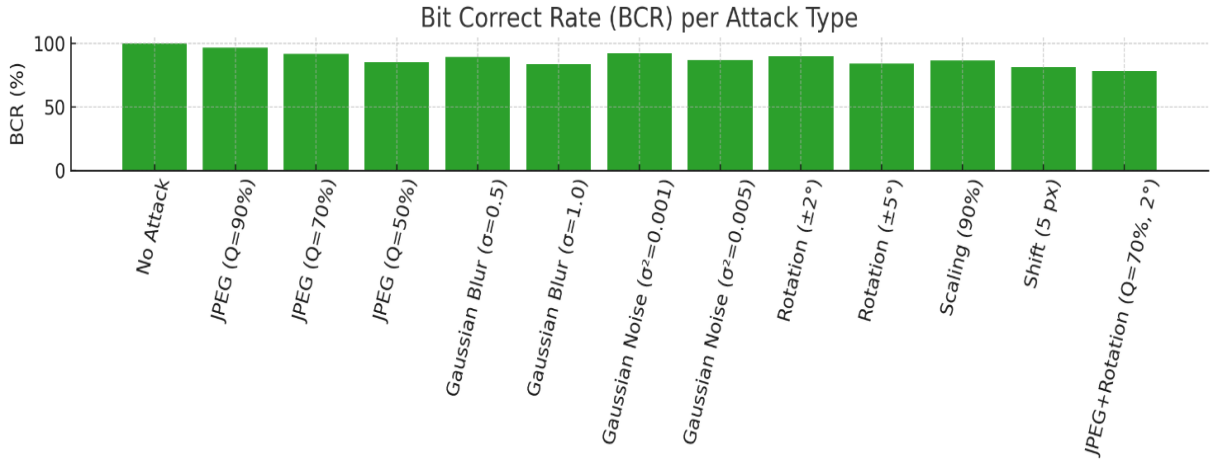


Figure 5. Bit Correct Rate (BCR) of the extracted watermark under different types of attacks

Figure 6 illustrates the variation of PSNR between the original and the modified (watermarked and attacked) images. Despite the embedding of the watermark and the application of subsequent distortions, the average PSNR values remain within the range of 38–41 dB. This indicates a high level of visual transparency of the proposed algorithm and the absence of perceptible artifacts to the human observer.

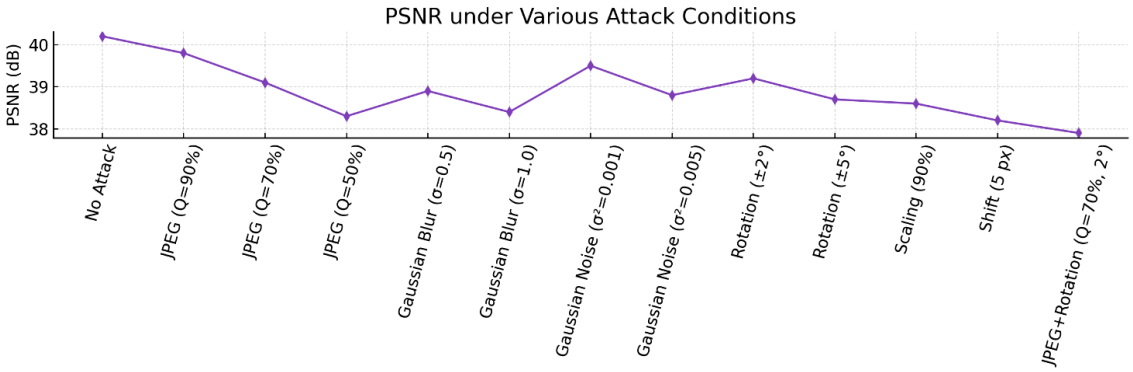


Figure 6. Peak Signal-to-Noise Ratio (PSNR) of the marked image under various attack conditions, reflecting visual quality preservation

The analysis of the presented graphs and numerical results confirms that the proposed algorithm demonstrates robustness against a wide range of attacks, including frequency-domain distortions (JPEG compression, blurring), additive noise, and geometric transformations (rotation, translation, scaling). In the majority of cases, the digital watermark is successfully extracted from the image even after combined attacks, as evidenced by consistently high NCC and BCR values.

At the same time, the visual quality of the watermarked image remains suitable for practical use, with PSNR values exceeding 38 dB. Thus, the theoretical study has shown that the proposed algorithm ensures strong resilience to various types of attacks while maintaining a high level of imperceptibility, making it well-suited for practical deployment in digital image protection applications.

4. CONCLUSIONS

In this work, a method for embedding a digital watermark using adaptive selection of spectral ring radii in the Discrete Fourier Transform (DFT) domain was proposed and theoretically substantiated. Unlike classical approaches with fixed zoning of the frequency domain, the developed algorithm enables dynamic formation of embedding rings based on the image's energy profile, allowing the watermark to be placed in spectrally stable regions.

Additional embedding of synchronization information in the DTCWT domain provides compensation for possible geometric distortions such as translation, rotation, or scaling. Thus, the proposed approach combines the advantages of localized embedding, high visual transparency, and self-synchronization capability.

The results of theoretical modeling, conducted on a set of representative images — including medical images — and a wide range of attacks (JPEG compression, blurring, noise addition, geometric transformations), confirmed the high robustness of the method. Even under severe distortions, the watermark remains detectable, with performance metrics showing NCC > 0.8 and BCR > 80%, indicating the practical effectiveness of the algorithm

REFERENCES

[1] Ying, Q., Lin, J., Qian, Z., Xu, H., and Zhang, X., "Robust digital watermarking for color images in combined DFT and DT-CWT domains", Mathematical Biosciences and Engineering, vol. 16, no. 5, 4788–4801, (2019). doi: 10.3934/mbe.2019241

[2] Yaremchuk, Y.Y., Karpinets, V.V., Puzdranovskyi, I.V., Zviryk, D.S., "Using adaptive selection of the radius of coefficient rings to enhance the robustness of DWT to attacks," in Proceedings of the International Conference PHOTONICS–ODS 2025, Vinnytsia National Technical University, Vinnytsia, Ukraine (2025). https://conferences.vntu.edu.ua/index.php/ods/ods_2025/paper/view/24706/

[3] Hongyan, Xu, “Digital media zero watermark copyright protection algorithm based on embedded intelligent edge computing detection”, Mathematical Biosciences and Engineering, 18(5), 6771-6789 (2021). doi: 10.3934/mbe.2021336

[4] Haoyu, Lu, Daofu, Gong, Fenlin, Liu, Hui, Liu, Jinghua, Qu. “A batch copyright scheme for digital image based on deep neural network”, Mathematical Biosciences and Engineering, 16(5), 6121-6133 (2019). doi: 10.3934/mbe.2019306

[5] Shanqing, Zhang, Xiaoyun, Guo, Xianghua, Xu, Li, Li, Chin-Chen, Chang, “A video watermark algorithm based on tensor decomposition”, Mathematical Biosciences and Engineering, 16(5), 3435-3449 (2019). doi: 10.3934/mbe.2019172

[6] Wenfa, Qi, Wei, Guo, Tong, Zhang, Yuxin, Liu, Zongming, Guo, Xifeng, Fang, “Robust authentication for paper-based text documents based on text watermarking technology”, Mathematical Biosciences and Engineering, 16(4): 2233-2249 (2019). doi: 10.3934/mbe.2019110

[7] Chuanda, Cai, Changgen, Peng, Jin, Niu, Weijie, Tan, Hanlin, Tang, “Low distortion reversible database watermarking based on hybrid intelligent algorithm”, Mathematical Biosciences and Engineering, 20(12), 21315-21336 (2023). doi: 10.3934/mbe.2023943

[8] Xuehu, Yan, Xuan, Zhou, Yuliang, Lu, Jingju, Liu, Guozheng, Yang. “Image inpainting-based behavior image secret sharing”, Mathematical Biosciences and Engineering, 17(4), 2950-2966 (2020). doi: 10.3934/mbe.2020166

[9] Shaozhang, Xiao, Xingyuan, Zuo, Zhengwei, Zhang, Fenfen, Li, “Large-capacity reversible image watermarking based on improved DE”, Mathematical Biosciences and Engineering, 19(2), 1108-1127 (2022). doi: 10.3934/mbe.2022051

[10] Vyatkin, S. I., Romanyuk, O. N., Pavlov, S. V., Snigur, A. V., Komada, P., Smailova, S., Yeraliyeva, B., "A function-based approach to real-time visualization using graphics processing units", Proc. SPIE 11581, 115810E (2020).

[11] Bisikalo, O., Kharchenko, V., Kovtun V., Krak, I., Pavlov, S., “Parameterization of the Stochastic Model for Evaluating Variable Small Data in the Shannon Entropy Basis. Entropy. 25(2), 184 (2023). https://doi.org/10.3390/e25020184

[12] Qiuling, Wu, Dandan, Huang, Jiangchun, Wei, Wenhui, Chen, “Adaptive and blind audio watermarking algorithm based on dither modulation and butterfly optimization algorithm”, Mathematical Biosciences and Engineering, 20(6): 11482-11501 (2023). doi: 10.3934/mbe.2023509.

[13] Vasilevskyi, O.M., Sevastianov, V.M., Ovchynnykov, K.V., Didych, V.M., and Burlaka, S.A., “Accuracy of Potentiometric Methods for Measuring Ion Activity in Solutions,” Proceedings of Seventh International Congress on Information and Communication Technology, 181–189, (2022).

[14] Romanyuk, O. N., Pavlov, S. V., Melnyk, O. V., Romanyuk, S. O., Smolarz, A., Bazarova, M. Z., "Method of anti-aliasing with the use of the new pixel model", Proc. SPIE 9816, 981617 (2015).