DIGITAL WATERMARKING TECHNIQUES TO SECURE DIGITAL DOCUMENTS

[1]Mr. Desta DanaData ,[2]Akalu Assefa
[1]Assistant Professor in Information Technology Department, School of Informatics, WolaitaSodo University, EthiopiaEmail address –destadanedata@wsu.edu.et
[2] Senior Lecturer in Information Technology Department, School of Informatics, WolaitaSodo University, Ethiopia
Email address – akalu.assefa@gmail.com

ABSTRACT:
The Digital Watermarking is a technology in which identification information is embedded in the data carrier in ways that cannot be easily noticed, and in which the data usage will not be affected. The technology used to often protect the copyright of multimedia data and protects databases and text files. This study majorly focused on types of the document to implement digital watermarking by text image document. The study also includes the embedding and recovering to watermarked and from watermarked documents respectively. Lastly, the study includes sample demonstrations in the document. Therefore, digital watermarking technology can be implemented in many private and organizational digital documents to create secured documents.

Keywords: Digital watermarking, document, DHT, DWT, DCT, embedding, SVD

I. INTRODUCTION:
The success of the Internet, cost-effective and popular digital recording and storage devices, and the promise of higher bandwidth and quality of service (QoS) for both wired and wireless networks has made it possible to create, replicate, transmit, and distribute digital content in an effortless way. The protection and enforcement of intellectual property rights for digital media has become an important issue. In 1998, Congress passed the Digital Millennium Copyright Act (DMCA) which makes it illegal to circumvent any technological measure that protects an owner's intellectual property rights of digital content.[1].
One of the strong data multimedia data protection methods used is digital watermarking mechanisms to secure and hide digital documents in unreliable world. Digital watermarking is the process of hiding digital information in a carrier signal. A digital watermark is basically a digital signature embedded in a noise-tolerant signal such as an audio, video or image data. It is typically used to identify ownership of the file. For example, a digital camera that would use lossless watermarking to embed a biometric identifier together with a cryptographic hash value. A digital signature is a large unique integer generated by encrypting a hash value of a file, image, video, etc. [2]

Application of Digital watermarking in Digital Documents

- **Fingerprinting:-**In order to trace the source of illegal copies the owner can embed different watermarking keys in the copies that are supplied to different customer. For the owner, embedding a unique serial number-like watermark is a good way to detect customers who break their license agreement by copying the protected data and supplying it to a third party.[3], [4]
- **Copyright Protection & Owner identification:-**To protect its intellectual property, the data owner can embed a watermark representing copyright information of his data. This application can be a really helpful tool in settling copyright disputes in court. It is probably the most widely spread use of digital images watermarking and it is also the application we have worked on in the present project.[3]
- **Copy protection:-**The watermarked information can directly control digital recording device. The embedded key can represent a copy-permission bit stream that is detected by the recording device which then decide if the copying procedure should go on (allowed) or not (prohibited).
- **Data Authentication:-**Fragile watermarks are used to detect any corruption of an image or any other type of data. If the watermark is detected, the data is genuine, if not, the data has been corrupted and cannot be considered.
- **Data Hiding (Covert Communications):-** The transmission of private data is probably one of the earliest applications of watermarking. As one would probably have already understood, it consists of implanting a strategic message into an innocuous one in a way that would prevent any unauthorized person to detect it.
- **Medical Safety:-**Embedding the date and patient's name in medical images could increase the confidentiality of medical information as well as the security.[5]
- **Packaging identification:-** by using the barcode systems controlling the ownerships of the products/services
- **Forgery controlling:-** by implementing high encrypted watermarking systems with the strong key the officers can safeguard the original documents.[5],[6],
- **Fact checking/ Fake news identifications:-** by using the metadata of document/images, google image search engines(https://images.google.com/) and other document checking methods can identify either the posted document/images is real or wrong.[7],[8]

II. METHODOLOGY
There are different types of watermarking methods implemented on digital images or documents. some of those are mentioned in below:

I. Text Document Watermarking Much of the early work on recognizing the potential problems with intellectual property rights of digital content and addressing these issues with early watermarking techniques was in the area of document watermarking. These techniques were devised for watermarking electronic versions of text documents which are in some formatted version such as postscript or PDF.[9]

II. Image Watermarking Many techniques have been developed for the watermarking of still image data.[10] For grey-level or color-image watermarking, watermark embedding techniques are designed to insert the watermark directly into the original image data, such as the luminance or color components or into some transformed version of the original data to take advantage of perceptual properties or robustness to particular signal manipulations. Requirements for image watermarking include imperceptibility, robustness to common signal processing operations, and capacity. Common signal processing operations which the watermark should survive include compression (such as JPEG), filtering, rescaling, cropping, A/D and D/A conversion, geometric distortions, and additive noise. Capacity refers to the amount of information (or payload) that can be hidden in the host image and detected reliably under normal operating conditions.[11],[12],[4],[13], [14]

To get those applications by implementing the different algorithms of Digital watermarking in our document or images by using the scenarios of the following diagrams:

III. EXPERIMENTAL IMPLEMENTATION SCENARIO

(a) Watermarking the document i.e. encrypting the original document into watermarked at the sides of the Sender

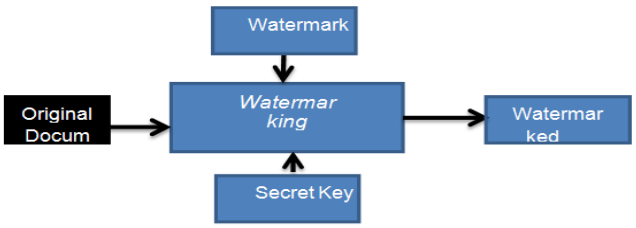

Fig-1- Digital Watermarking embedding

(b) Recovering from The watermarked Document which is applicable at receiver sides.

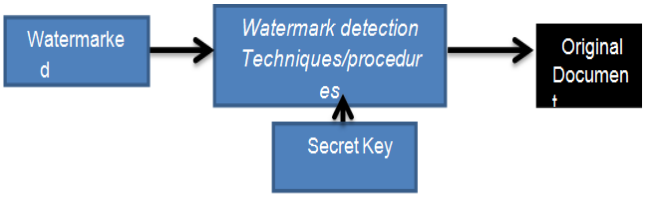

Fig-2- Digital watermarking extraction

Implementation of Algorithms of Digital Watermarking:

A. ***DWT:*** The DWT (Discrete Wavelet Transform) is a powerful and useful multi-resolution decomposition method in digital watermarking. It is often applied on image processing, and has been applied to such as noise reduction, edge detection, and data compression. It is consistent with the visual perception process of human eyes. DWT can localize the signal in spatiotemporal, it is a new signal analytic theory but has already been widely used. DWT uses discrete wavelet transform to decompose the original image into four sub-bands LL1, HL1, LH1, and HH1, which can be separate into lower frequency sub-bands and higher frequency sub-bands. And the low frequency sub-band LL1 which stands for the coarse level coefficients can be further decomposed into four sub-bands LL2, HL2, LH2, and HH2. [11],[15],[4],[16],[17]

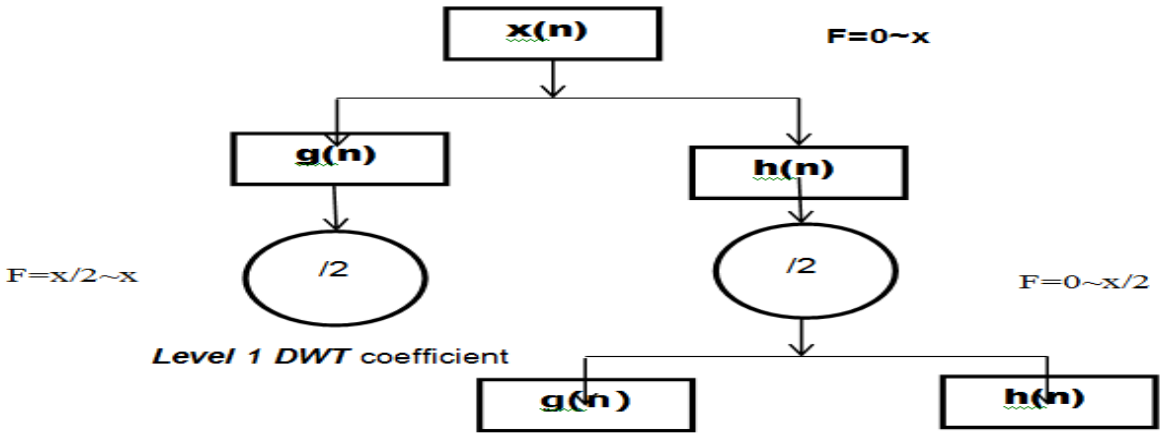

Fig.-3-Single Level DWT decomposition

Fig.1 is process of the DWT decomposition, it is computed by successive low and high pass filtering of the discrete time domain signal. The x(n) in the Fig.1

denotes the signal. The g(n) is the low pass filter and the h(n) stands for the high pass filter. A single level of decomposition can be expressed as in Eq.1

$$Y_{\text{high}}[k] = |_n x(\text{n})(\text{g } 2\text{k} - \text{n})$$
$$Y_{\text{low}}[k] = |_n x(\text{n})(\text{h } 2\text{k} - \text{n})$$

Eq.1 a single level of decomposition

The decimation and filtering process is continued until we reach the desired level which depends on the length of signal. Then we concatenate all the coefficients, start from the last level of decomposition to get the DWT of the original signal.

Below are briefly introduction of some novel digital watermarking algorithms based on DWT. First is a binary image algorithm, the algorithm enhances the security of embedding watermarks by using the Logistic chaotic sequences and generalized cat mapping to scramble the watermarks. The algorithm is used in a copyright protection zero-watermarking scheme. The algorithm decomposes the original image into the appropriate levels by the DWT, and divides the obtained approximation image into non-overlapping blocks, then use SVD (singular value decomposition) to get the singular values. Finally, the algorithm uses XOR operation between the first singular value of each block and pixel value of the actual binary character watermarking sequentially. This algorithm can ensure the quality of original image, and has very good robustness against the common image processing attacks.[18],[19],[20]

SVD Implementation steps:

1. Input image is gray scale.
2. Partition the image into blocks of n×n pixels
3. Apply SVD transformation to each partitioned block
4. Calculate the number of non-zero co-efficient in the D component of each block. This is calculated to determine the complexity of the block.
5. Select greater complexity blocks using PRNG [pseudo random number generator] and also using the feature of D component.
6. For each selected greater complexity block, in the first column of U, magnitude difference between the neighbouring coefficients is calculated.
7. First, if the magnitude difference matches the embedding watermark (e.g. positive relationship matching a bit value of 1 or negative relationship matching a bit value of 0), the coefficients are retained. Second, if the magnitude difference does not match the embedding watermark, the coefficient must be modified.
8. To retain the image quality and provide a stronger robustness of a watermarking scheme, the difference value is first checked to be above certain threshold.

B. ***DCT :-***(Discrete Cosine Transformation) is a Fourier-related transform, it only uses real numbers. DCT is roughly twice long than DFT (discrete fourier transform), operating at a finite number of real discrete data points. Like other Fourier-related transforms, DCT uses different frequencies and amplitudes to get a sum of sinusoids and then uses that sum to indicate a function or a signal. Compared with DFT or other Fourier-related transforms, DCT has different boundary conditions and only uses cosine functions. DCT is an invertible, linear function. It makes RN into RN or an N x N square matrix where R expresses the set of real numbers.[21],[22],[23],[13],[24] The most popular form of DCT is the DCT-II which is often referred to as "the DCT". The formulation of DCT-II is given as Eq.2.

$$X_k = \sum_{n=1}^{N-1} \text{X}_n \cos\left[\frac{\pi}{N}\left(\text{n} + \frac{1}{2}\right)k\right] \qquad k = 0, \ldots\ldots, \text{N} - 1.$$

Eq.2 DCT-II formulation

In DCT-II, Xn is even around n=-1/2 and even around n=N-1/2, Xk is even around k=0 and odd around k=N, which means the boundary conditions.

DCT Implementation steps:

(1) Image is segmented into non-overlapping blocks of 8x8.
(2) Forward DCT is applied to these blocks.
(3) Block selection criteria like HVS are applied.
(4) Coefficient selection criterialike host images odd or even
(5) Selected Co-efficient is modified for embedding watermark.
(6) Inverse DCT transform is applied on each block.

C. ***DHT :-***(Discrete Hadamard Transformation) is a non-sinusoidal orthogonal transformation. A signal is decomposed into a set of orthogonal rectangular waveforms which are called Hadamard functions. Since the amplitude of Hadamard functions has only two values +1 or -1, the transformation is real and has no multipliers. The Hadamard matrix is a square array of plus and minus ones whose columns and rows are orthogonal to one another. The product of an N x N Hadamard matrix and its transpose is the identity matrix. The 2D-Hadamard transform is given as Eq.3.

$$[\text{V}] = (H_n[U]H_n)/N$$

Eq.3 2D-Hadamard transform

Where [U] is the original image and [V] is the transformed image, H is an N x N Hadamard matrix, N=2n, n is an integer. The elements of the transform matrix Hn are binary real numbers. The inverse 2D-Hadamard transform is given as Eq.4.

$$[\text{U}] = (H_n^{-1}[UV]H_n^*) = (H_n[U]H_n)/N$$

Eq.4 inverse 2D-Hadamard transform

Below are briefly introduction of some novel digital watermarking algorithms based

IV. RESULT BY SAMPLE DEMONSTRATIONS:

A. Text Document Watermarking samples

i. Properties in the document: (sample used for the proposed work)

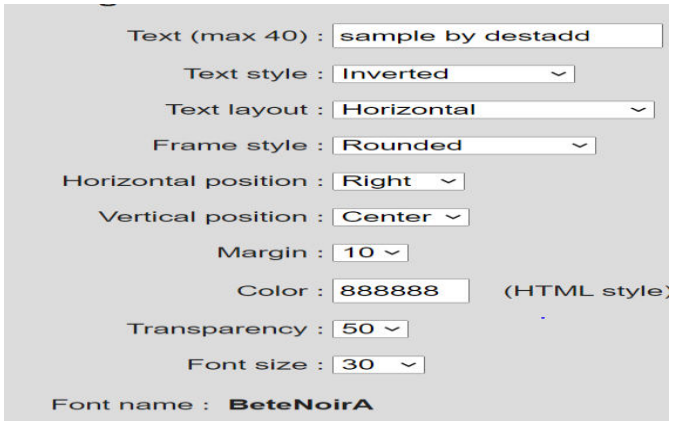

Fig-4-: properties in text watermarking

ii. Metadata Information on document:

The organization in secretly can use the secret information/ metadata to safeguard the organization secrets. The metadata use in QR or different text document:

The lists of Metadata in information hiding as listed in follows

- The system current date.
- The organization logo
- The CEO signatures.
- Establishment date of the organizations/company
- The full information of product and services in QR.
- Barcode of the identifications.

B. **PDF Document Watermarking:-**by Watermarking your PDF document can signify whether the PDF document is a copy or an original, who the company that authorizes a form is, even the kind of PDF document, such as an estimate, invoice, office letters, personal documents.. and so on. In PDF watermarking there are different ways to hides the secured information either text/image. There are some of the properties used in PDF watermarking as mentioned in below:

(1) Text(Secured message)
(2) Images/pictures
(3) Logo
(4) Personal biometric data like fingerprint,
(5) Transparency
(6) Positioning
(7) Styles/fonts..

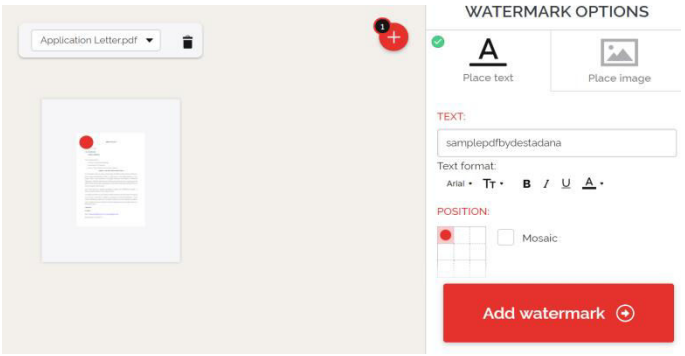

Fig-5-: PDF watermarking tools working demo

C. Image Digital watermarking

i. QR, Copyright, texted and other image

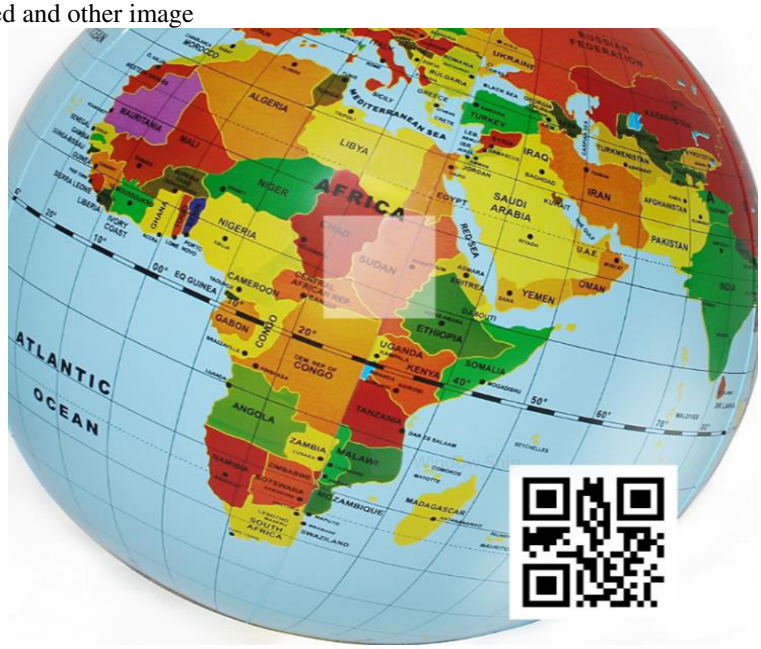

Fig-6-: Quick Response (QR) watermarked image.

ii. Image digital Watermarking

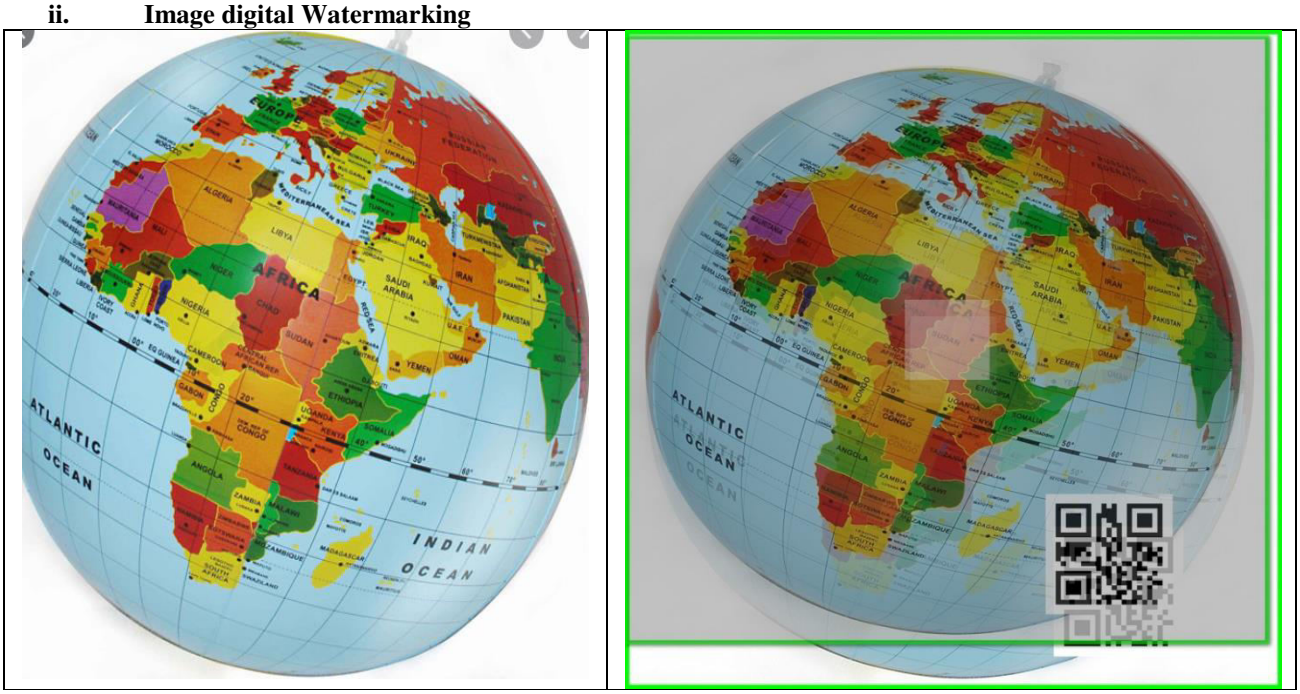

Fig-7- Original Image with Digital watermarked image

V. CONCLUSIONS

One of the strong data multimedia data protection methods used is digital watermarking mechanisms to secure and hide digital documents in an unreliable world. Digital watermarking is the process of hiding digital information in a carrier signal. The digital watermarking technique is embedded in all forms of media content, including digital text, images, audio, video, and even certain objects. The study mostly focused on how the embedding and finding the original documents after the watermarking process. To conduct the study we majorly include text and image multimedia data. The tools and software are available for embedding imperceptible information via subtle changes to the data of the original digital content. And digital watermarks can be easily detected and read by computers, networks, and a variety of digital devices, validating the original content and/or initiating actions. Finally, we have used different online, free open source digital watermarking tools to change different data types like text, images, audio, and video media to watermarked and secured hidden digital documents.

REFERENCES

[1] C. I. Podilchuk and E. J. Delp, "Digital watermarking: algorithms and applications," IEEE Signal Process. Mag., vol. 18, no. 4, pp. 33–46, 2001.

[2] C. Honsinger, "Digital watermarking," J. Electron. Imaging, vol. 11, no. 3, p. 414, 2002.

[3] K. Deb, M. S. Al-Seraj, M. M. Hoque, and M. I. H. Sarkar, "Combined DWT-DCT based digital image watermarking technique for copyright protection," in 2012 7th International Conference on Electrical and Computer Engineering, 2012, pp. 458–461.

[4] A. Al-Haj, "Combined DWT-DCT digital image watermarking," J. Comput. Sci., vol. 3, no. 9, pp. 740–746, 2007.

[5] S. Castelo et al., "A topic-agnostic approach for identifying fake news pages," in Companion Proceedings of The 2019 World Wide Web Conference, 2019, pp. 975–980.

[6] K. Shu, D. Mahudeswaran, S. Wang, D. Lee, and H. Liu, "Fakenewsnet: A data repository with news content, social context and dynamic information for studying fake news on social media," ArXiv Prepr. ArXiv180901286, vol. 8, 2018.

[7] A. Piva, M. Barni, F. Bartolini, and V. Cappellini, "DCT-based watermark recovering without resorting to the uncorrupted original image," in Proceedings of international conference on image processing, 1997, vol. 1, pp. 520–523.

[8] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, "Fake news detection on social media: A data mining perspective," ACM SIGKDD Explor. Newsl., vol. 19, no. 1, pp. 22–36, 2017.

[9] M. Jiansheng, L. Sukang, and T. Xiaomei, "A digital watermarking algorithm based on DCT and DWT," in Proceedings. The 2009 International Symposium on Web Information Systems and Applications (WISA 2009), 2009, p. 104.

[10] "Search Results for medicine - Technical Bulletin." https://vsearch.nlm.nih.gov/vivisimo/cgi-bin/query-meta?query=medicine&v%3Aproject=technical-bulletin (accessed May 03, 2018).

[11] P.-W. Chan and M. R. Lyu, "A DWT-based digital video watermarking scheme with error correcting code," in International Conference on Information and Communications Security, 2003, pp. 202–213.

[12] W. C. Chu, "DCT-based image watermarking using subsampling," IEEE Trans. Multimed., vol. 5, no. 1, pp. 34–38, 2003.

[13] M.-J. Tsai and H.-Y. Hung, "DCT and DWT-based image watermarking by using subsampling," in 24th International Conference on Distributed Computing Systems Workshops, 2004. Proceedings., 2004, pp. 184–189.

[14] A. G. Bors and I. Pitas, "Image watermarking using DCT domain constraints," in Proceedings of 3rd IEEE International Conference on Image Processing, 1996, vol. 3, pp. 231–234.

[15] H. Daren, L. Jiufen, H. Jiwu, and L. Hongmei, "A DWT-based image watermarking algorithm," in IEEE International Conference on Multimedia and Expo, 2001. ICME 2001., 2001, pp. 80–80.

[16] A. Al-Haj, A. A. Mohammad, and L. Bata, "DWT-based audio watermarking.," Int Arab J Inf Technol, vol. 8, no. 3, pp. 326–333, 2011.

[17] G. Sun and Y. Yu, "DWT based watermarking algorithm of color images," in 2007 2nd IEEE Conference on Industrial Electronics and Applications, 2007, pp. 1823–1826.

[19] K. A. Navas, M. C. Ajay, M. Lekshmi, T. S. Archana, and M. Sasikumar, "Dwt-dct-svd based watermarking," in 2008 3rd

International Conference on Communication Systems Software and Middleware and Workshops (COMSWARE'08), 2008, pp. 271–274.

[20] F. Huang and Z.-H. Guan, "A hybrid SVD-DCT watermarking method based on LPSNR," Pattern Recognit. Lett., vol. 25, no. 15, pp. 1769–1775, 2004.

[21] M. Barni, F. Bartolini, V. Cappellini, and A. Piva, "A DCT-domain system for robust image watermarking," Signal Process., vol. 66, no. 3, pp. 357–372, 1998.

[22] A. Al-Haj, "Combined DWT-DCT digital image watermarking," J. Comput. Sci., vol. 3, no. 9, pp. 740–746, 2007.

[23] J. R. Hernandez, M. Amado, and F. Perez-Gonzalez, "DCT-domain watermarking techniques for still images: Detector performance analysis and a new structure," IEEE Trans. Image Process., vol. 9, no. 1, pp. 55–68, 2000.

[24] G. C. Langelaar and R. L. Lagendijk, "Optimal differential energy watermarking of DCT encoded images and video," IEEE Trans. Image Process., vol. 10, no. 1, pp. 148–158, 2001.