A Robust Wavelet based Fuzzy-K Means Algorithm in Digital Image Watermarking

Himansu Kumar Jena
Computer Science & Engineering
Centurian University of Technology & Management
Bhubaneswar, India

Manas Ranjan Nayak
Computer Science & Engineering
School of Computer Engineering, KIIT University
Bhubaneswar, India

Nihar Ranjan Hota
Computer Science & Engineering
Einstein Academy of Technology & Management
Bhubaneswar, India

Swati Sucharita Barik
Computer Science & engineering
Centurion University
BBSR, India

Abstract-. **Protection of digital multimedia content such as audio, video, image has become an increasingly important issue for content owners and service providers. Watermarking is the best solution to achieve this copyright protection. Secret digital data can be embedded either in spatial domain or in frequency domain of the cover data. In this paper a Fuzzy K-means algorithm is used to locate the embedding location and Discrete Wavelet transform (DWT) is used for hiding watermark in low frequency band. The quality of the watermarked image and extracted watermark is tested using peak signal to noise ratio (PSNR). Robustness of proposed algorithm is tested under various attacks including salt and pepper noise, Gaussian noise, cropping and JPEG compression. Also it still gives a high quality watermarked image. The watermarking task is performed by MATLAB Program.**

Keywords: ***Digital watermarking, Fuzzy K-means algorithm, Discrete Wavelet transformation (DWT) technique, Mean Square Error(MSE), Peak Signal to Noise Ratio (PSNR).***

I. INTRODUCTION

Now a day's availability of the digital data such as images, audio and video etc. to the public exponentially increases through the internet. At the same time with the rapid improvement of technology, unauthorized manipulation of digital data has lead the industry to look for approaches to implement copyright protection in all sorts of digital multimedia which is done by using digital image watermarking. Among many solutions like Cryptography, Steganography and Watermarking, the Data Hiding capacity of watermarking technique provides one of the best solutions among them. This technique embeds information so that it is not easily perceptible to the others and the watermark should not degrade the quality of the image and should be perceptually invisible to preserve its protective secrecy.

Data hiding scheme for digital image watermarking techniques can be broadly categorized as spatial domain based watermarking techniques and Transform domain based watermarking techniques. Spatial domain methods are less Complex and not robust against various attacks as no transform is used in them and the secret messages are embedded in the image pixels directly. Transform domain methods are robust as compared to spatial domain methods due to the fact that when image is inverse transformed, watermark is spread haphazardly over the entire image, which makes attacker very difficult to read or modify. Although there are various transform domain based techniques available like DFT (Discrete Fourier Transform), DCT (Discrete Cosine Transform), but they have various pitfalls like less robustness (cannot withstand different types of attacks), less imperceptibility (degrades the perceptual quality of the original image) due to which we are considering DWT (Discrete Wavelet Transform) technique.

The watermark technique should reveal many desirable characteristics. That is, a watermarking technique should at least have the following requirements: **Robustness:** means the ability to keep the watermarking information unchanged. and not as an independent document. Please do not revise any of the current designations.

Imperceptibility: It refers to the ability to keep similarity between the original and watermarked versions of cover image.

Security: The watermarking technique should be so secured that no attackers should destroy the watermark.

Complexity: It describes the expenditure to detect and encode the watermark information. It is recommended to design the watermarking procedure and algorithm as complex as possible so that different watermarks can be integrated.

In this paper we propose a Fuzzy K-Means , DWT based watermarking method by means of Haar wavelet transform in which K-means algorithm is used to locate the embedding location so that the watermark can be extracted exactly. Our method gives better simulation result in JPEG compression in comparison with Chang[15] and You[16].

The rest of paper is structured as follows. Basics of Fuzzy-K-means, DWT and review of related works are in Section 2. Section 3 presents the proposed method. Experimental Result is discussed in Section 4. Section 5 concludes of the paper.

II. BASICS OF K-MEANS, DWT AND SOME RELATED WORKS

A. *FUZZY- K-MEANS*

K-means clustering is a classification method which aims to classify data (block) into some similar groups using distance matrices. In our proposed method the image matrix is divided into some non overlapping blocks. K-means clustering aims to groups these blocks into 'k' sets or Groups ($k \leq n$) . From these 'k' groups we are selecting a particular matching group for watermark embedding. In detection procedure it also helps to find that particular group so that we can easily find out the blocks under that groups. The objective function of K-means algorithm is given in Eq(1):

$$J = \sum_{j=1}^{k} \sum_{i=1}^{n} \left\| x_i - y_j \right\|^2 \quad (1)$$

Where $\left\| x_i - y_j \right\|$ is the Euclidian distance between x_i and y_j , 'n' is the number of data points in a block, 'k' is the number of sets or groups.

In clustering distance measure will determine in which cluster a data elements should go and it will influence the shape of the clusters. The distance measure can be performed by Euclidian distance and Manhattan distance and is given by the formula in Equation (2) and (3).

(Euclidian)
$$d(x,y) = \sum_{i=1}^{p} |x_i - y_i| \quad (2)$$

(Manhattan)
$$d(x,y) = \sqrt{\sum_{i=1}^{p} |x_i - y_i|^2} \quad (3)$$

Where $x_i - y_i$ gives distance between two points x_i and y_i.

Suppose a data vector consisting of variances and we want to group them 'k' groups where k=1, 2…n.

Now K-means algorithm groups each variance to a group according to the following logic

Initialize K cluster $C_k^{(a)}$, iteration $i = 0$. Assign each data Xj to the cluster with the nearest cluster $C_k^{(i)}$.

Currently we use the ordinary Euclidean distance metric $\left\| C_k^{(i)} - X_j \right\|$. Set new cluster centers $C_k^{(i+1)}$ to the center of each cluster:

$$C_k^{(i+1)} = E\{X_j\}_{j \in C_k^{(i)}}$$

Repeat these steps for all data.

1.1. Discrete Wavelet Transform (DWT)

- Use a zero before decimal points: "0.25," not ".25." Use "cm3," not "cc." (*bullet list*)

A. *Equations*

The equations are an exception to the prescribed specifications of this template. You will need to determine whether or not your equation should be typed using either the Times New Roman or the Symbol font (please no other font). To create multileveled equations, it may be necessary to treat the equation as a graphic and insert it into the text after your paper is styled.

Number equations consecutively. Equation numbers, within parentheses, are to position flush right, as in (1), using a right tab stop. To make your equations more compact, you may use the solidus (/), the exp function, or appropriate exponents. Italicize Roman symbols for quantities and variables, but not Greek symbols. Use a long dash rather than a hyphen for a minus sign. Punctuate equations with commas or periods when they are part of a sentence, as in

$$a + b = \gamma \quad (1)$$

Note that the equation is centered using a center tab stop. Be sure that the symbols in your equation have been defined before or immediately following the equation. Use "(1)," not "Eq. (1)" or "equation (1)," except at the beginning of a sentence: "Equation (1) is ..."

B. *Some Common Mistakes*

- The word "data" is plural, not singular.
- The subscript for the permeability of vacuum μ_0, and other common scientific constants, is zero with subscript formatting, not a lowercase letter "o."
- In American English, commas, semi-/colons, periods, question and exclamation marks are located within quotation marks only when a complete thought or name is cited, such as a title or full quotation. When quotation marks are used, instead of a bold or italic typeface, to highlight a word or phrase, punctuation should appear outside of the quotation marks. A parenthetical phrase or statement at the end of a sentence is punctuated outside of the closing parenthesis (like this). (A parenthetical sentence is punctuated within the parentheses.)
- A graph within a graph is an "inset," not an "insert." The word alternatively is preferred to the word "alternately" (unless you really mean something that alternates).
- Do not use the word "essentially" to mean "approximately" or "effectively."
- In your paper title, if the words "that uses" can accurately replace the word using, capitalize the "u"; if not, keep using lower-cased.
- Be aware of the different meanings of the homophones "affect" and "effect," "complement" and "compliment," "discreet" and "discrete," "principal" and "principle."
- Do not confuse "imply" and "infer."

- The prefix "non" is not a word; it should be joined to the word it modifies, usually without a hyphen.
- There is no period after the "et" in the Latin abbreviation "et al."
- The abbreviation "i.e." means "that is," and the abbreviation "e.g." means "for example."

An excellent style manual for science writers is [7].

Discrete Wavelet transform (DWT) is a mathematical tool for hierarchically decomposing an image. The DWT separates image into four spatial directions lower resolution image (LL), and horizontal (HL), vertical (LH) and diagonal (HH) detail components as shown below. Magnitude of DWT coefficients is larger in the lowest bands (LL) at each level of decomposition and is smaller for other bands (HH, LH, and HL). The block diagram of 1 level DWT decomposition is given below as shown in fig 1.

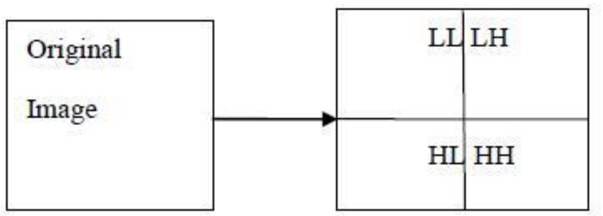

Fig. 1: DWT Decomposition of 1Level Transforms

Furthermore, from these DWT coefficients, the original image can be reconstructed. This reconstruction process is called the inverse DWT (IDWT). If C[m, n] represents an image, the DWT and IDWT for C[m, n] can similarly be defined by implementing the DWT and IDWT on each dimension and separately (Perumal, Kumar ,Sumalatha, and Kumar, 2009).

Advantages using DWT:

a) The larger the magnitude of wavelet coefficient, the more significant it is.
b) b) Watermark detection at lower resolutions is computationally effective because at every successive resolution level, less no. of frequency bands are involved.
c) High resolution sub bands help to easily locate edge and textures patterns in an image.
d) Due to its great frequency component separation properties, the DWT, in contrast to DCT, is very useful to identify the coefficients to be watermarked [3].
e) Watermarking in the wavelet domain is compatible with the JPEG 2000 compression standards.

III. RELATED WORKS

Hui - Yu Huang [2] proposes a technique lossless data hiding method for a DWT. Using the quantization factors for DWT and was not robust to some attacks like JPEG compression, our proposed approach can offer high hiding capacity and preserve the image quality of stego-images. The original image can be recovered losslessly when the secret data had been extracted from stego - images. Barni et al. [4] proposed a wavelet domain based method which exploits the characteristics of human visual system (HVS). Based on the texture and the luminance content of all image sub-bands, a mask is accomplished pixel by pixel. This method embedded the watermark image in HH sub-band of the first level, and this subband was not robust to attack like JPEG compression.

Reddy et al. [5] proposed a method in which the authors used a gray scale logo as watermark. To embed watermark, HVS characteristics were used to select the significant coefficients and watermark is added to these selected coefficients. Further, they used the model of Barni et al. [4] to calculate the weight factors for wavelet coefficients of the host image. They extracted watermark from the distorted image by taking into consideration the distortion caused by the attacks. Lin et al.[6] proposed a wavelet –tree based blind image watermarking scheme using distance vector of binary cluster. The embedding is done by comparing the statistical difference and distance vector of wavelet tree to decide about embedding bit 0 or 1. This method preserves strong robustness against filtering attacks but cannot resist JPEG compression attacks.

A. Proposed Method of Watermark Embedding Extraction and Verification

The proposed method embeds watermark by decomposing the host image using discrete wavelet transform by means of Haar wavelet transform. The algorithm is described in this section by outlining the major steps in its two, conversely, if there are not at least two sub-topics, then no subheads should be introduced. Styles named "Heading 1," "Heading 2," "Heading 3," and "Heading 4" are prescribed.

procedures; the watermark embedding procedure and the watermark extraction procedure. The host is a 512×512 bit gray scale Lena image and the watermark is a binary image having either 0 or 1values.

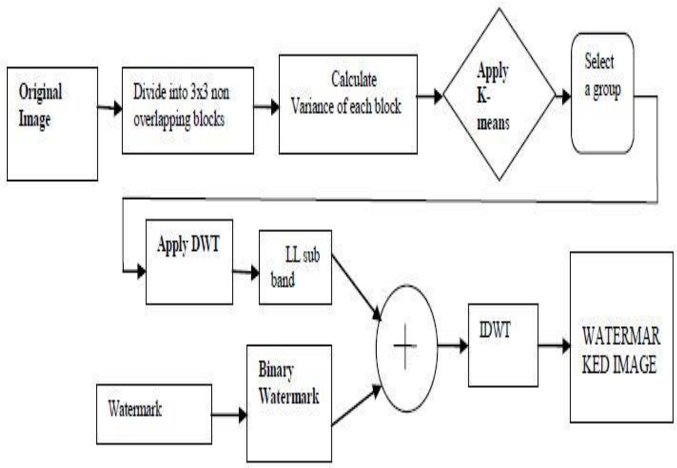

Figure 2: Block Diagram of Watermark Embedding Process

The Embedding Algorithm:

The blocks of the embedding algorithm is explained in the following steps:

Step 1: Read the cover image file into a two dimensional decimal array to handle the file data more easily.
Step 2: Divide the cover image into 3x3 non overlapping blocks. By this division each 3x3 block can be categorized as a smooth or complex block.
Step 3: Calculate Variance of each block.
Step 4: Apply Fuzzy K-means algorithm to group the similar variance of blocks.
Step 5: Select a suitable block and transform it to the transform domain using 1D Haar integer wavelet transform resulting LL, LH, HL and HH. [LL, LH, HL, HH] = DWT(C) where c is the selected group.
Step 6: Embed the message into LL sub-band and apply inverse wavelet to get the watermarked image.

B. WATERMARK EXTRACTION

The Algorithm for watermark extracting is given below.
Step 1: Read the watermarked image into a two dimensional decimal array.
Step 2: Read the cover image into a two dimensional decimal array.
Step 3: Divide both two images into 3x3 non overlapping blocks. By this division each 3x3 block can be categorized as a smooth or complex block.
Step 4: Calculate Variance of each block of two images watermarked and cover.
Step 5: Apply Fuzzy K-means algorithm to group the similar variance of blocks .
Step 6: Select a suitable block and transform it to the transform domain using 1D Haar integer wavelet transform resulting LL, LH, HL and HH. [LL, LH, HL,HH] = DWT(C) where ‘C’ is the selected group.
Step 7: Subtract the LL sub band value of cover image from watermarked image and get the watermark
Step 8: calculate MSE and PSNR.

The Procedure of water mark Extraction is shown in fig.3

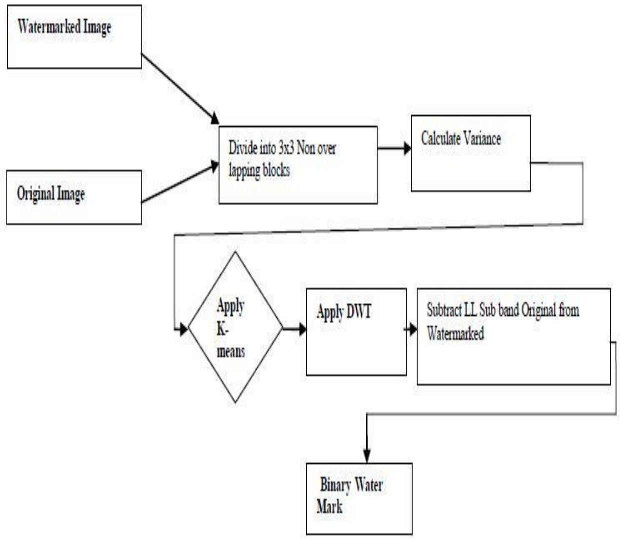

Figure3: Block Diagram of Watermark Extraction

One commonly employed measure to evaluate the imperceptibility of the watermarked image is the peak signal-to-noise ratio (PSNR).Assuming that the original image X and the watermarked image X' both have image sizes M x N. The mean square error (MSE) between the original and the watermarked images can be represented by

$$MSE = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}\left(X(i,j)-X'(i,j)\right)^2$$

Consequently, the PSNR in decibel (dB) can be calculated by

$$PSNR = 10\log_{10}\left(MAX^2/MSE\right) \quad \text{(db)}$$

where MAX is the maximum possible pixel value of the image.

IV. EXPERIMENTAL RESULT

The Original image used is an 512×512 gray scale Lena image and the watermark is a string. First we convert the watermark to a binary array and embed it to the Original image. Figurer below shows cover image Lena, and the watermark, with PSNR value. At the watermark extraction process the binary array of the watermark is accurately extracted . In order to demonstrate, the robustness of the proposed watermarking algorithm, the watermarked image is attacked by a variety of attacks namely JPEG Compression, Gaussian noise addition, Cropping, Salt and peeper noise addition and Poisson noise addition . After these attacks on the watermarked image, the extracted Binary

Watermark is compared with the original one. The PSNR value is given in fig 4 and after adding different noise the PSNR value is given in fig 5.

	**NIHAR RANJAN HOTA**		Binary Watermark of the String retrieved accurately with PSNR given below.
Original lena Image(512x512)	watermark	Watermarked Image	PSNR= 52.008

Figure4: from left Original Image , Watermark, Watermarked Image and extracted watermark.

Adding Different Noises to Image and their PSNR values:

		
JPEG Compression(Q=100) (h)	JPEG Compression(Q=80) (i)	JPEG Compression(Q=60) (j)
PSNR=51.590	PSNR=42.466	PSNR= 40.540

Figure 5: Adding different Noise and their PSNR values.

V. CONCLUSION

			
Crop of topmost left side (a)	Crop at middle (b)	Crop at down left side (c)	Crop at topmost right (d)
PSNR=37.935	PSNR= 52.008	PSNR=22.383	PSNR= 38.121
			
Adding Gaussian noise (e)	Adding Poisson noise (f)	Salt and pepper(0.02) (g)	Speckle(.01) (h)
PSNR=20.43	PSNR=27.129	PSNR=21.369	25.168

The proposed technique of blind digital image watermarking has been performed in single-level DWT. Robustness of this method is carried out by a variety of attacks. The PSNR gives more accurate value in JPEG Compression in comparison to other attacks. The proposed scheme has resulted in an efficient watermarking scheme for effective copyright protection of images. The algorithm has been tested with 10 host images and its performance has been seen. The quality of the watermarked image is good in terms of perceptibility and PSNR. Which happens to be (52.008 dB) for Lena image, (48.327 dB) for boat image, (51.002 dB) for cameraman image, (52.008 dB) for baboon image, and so on. It can be noticed that all PSNR values are higher than 40 dB which is quite acceptable for the human eye, with almost no sign of watermark existence. The proposed scheme has satisfied both the requirements of effective copyright protection scheme: imperceptibility and robustness. The security of the proposed method lies in the reference image. Since no attacker can extract the data without accessing the reference image. If any intruder tries to remove the watermark then the watermark is removed by degrading the image quality. Hence, the quality of the image degradation is directly proportional to the quality of the extracted watermark.

REFERENCES

[1] Han, J. and Kamber, M. Data Mining: Concepts and Techniques , (Morgan Kaufmann, San Francisco, California) 2000.
[2] Hui-Yu Huang & Shih-Hsu Chang “A lossless data hiding based on discrete Haar wavelet transform”,10th IEEE International Conference on Computer and Information Technology, 2010.
[3] R. Dugad, K. Ratakonda, N. Ahuja, “A New Wavelet-Based Scheme for Watermarking
[4] Images”, International Conference on Image Processing (ICIP 1998), Vol. 2, pp. 419-423, 1998.
[5] M. Barni, F. Bartolini, A. Piva, Improved wavelet based watermarking through pixel wise masking, IEEE Transactions on Image Processing, vol-10, pp.783-791,2001
[6] A.A. Reddy, B.N. Chatterjii, A new wavelet based logo–watermarking scheme, Pattern Recognition Letters 26 ,pp. 1019–1027,2005.
[7] W.H.Lin,Y.R.Wang and S.J.Horng, A wavelet –tree-based watermarking method using distance vector of binary cluster,Expert Systems with Applications, vol-36,no.6,pp.9869-9878,2009
[8] Mohammad Reza Soheili “A Robust Digital image Watermarking Scheme Based on DWT” Journal of Advances in Computer Research, pp75-82.1,2010.
[9] Souvik Bhattacharyya and Gautam Sanyal “Data Hiding in Images in Discrete Wavelet Domain Using PMM”, World Academy of Science, Engineering and Technology, 2010.
[10] M.Ouhsain and A.B.hamza image watermarking scheme using nonnegative matrix factorization and wavelet transform, Expert systems with application,vol-36,no.2,pp-2123-2129,2009.
[11] Victor V., Guzman, Meana, “Analysis of a Wavelet-based Watermarking Algorithm”, IEEE Proceedings of the International Conference on Electronics, Communications and Computer, pp. 283-287,2004.
[12] Wan Adnan, W.A.; Hitam, S.; Abdul-Karim, S., Tamjis, M.R., “A review of image watermarking”, Research and Development, SCORED, pp. 381-384,2003.
[13] F. Kumgollu, A Bouridane, M A Roula and S Boussaktd, “Comparison of Different Wavelet Transforms for Fusion Based Watermarking Applications”, IEEE Transactions, Vol.3, pp. 1188-1191,2003.
[14] Cong Jin, Shu-Wei Jin, “Wavelet Packets-Based Robust Blind Digital Watermark Scheme”, IEEE International Conference on Wavelet Analysis and Pattern Recognition, pp. 724-728,Aug 2008.
[15] Vijay R Ayangar, S. N. Talbar, “A Novel DWT-SVD Based Watermarking Scheme”, MCIT, pp. 105- 108,2010.
[16] C.C. Chang, P.Y. Lin, and J.S. Yeh, “Preserving Robustness and Removability for Digital Watermarks Using Subsampling and Difference Correlation,” Information Sciences, vol. 179,no. 13, pp. 2283–2293,2009.
[17] X. You et al., “A Blind Watermarking Scheme Using New Nontensor Product Wavelet Filter Banks,” IEEE Trans. Image Processing, vol. 19, no. 12, pp. 3271–3284,2010.