American Scientific Research Journal for Engineering, Technology, and Sciences (ASRJETS)

ISSN (Print) 2313-4410, ISSN (Online) 2313-4402

© Global Society of Scientific Research and Researchers

http://asrjetsjournal.org/

Adaptive Image Watermarking based on K-NN Clustering

Hassan Ouahi[a]*, Abdenbi Mazoul[b]

[a,b]bLaboratory Computing Systems & Vision LabSiv Faculty of Science Ibn Zohr Agadir,Morocco

[a]Email: h.ouahi@uiz.ac.ma

[b]Email: a.mazoul@uiz.ac.ma

Abstract

The key challenge faced by researchers is the rise in the use of social media communication to prove ownership rights to multimedia material such as video, audio, text, graphics, etc. Watermarking is the method of multimedia concealment of digital content that can be used later to prove ownership credentials. The researchers in this field contribute a lot of work, but there is still a need for more robust methods. In this paper, we use the KNN clustering method to find the features in the image, which are then used to embed the content of the watermark. Later, the KNN clustering approach is again used for watermark extraction to classify the characteristics where the watermark is embedded and extraction is performed from those characteristics.

Keywords: Image Watermarking; Discrete Cosine Transform; KNN Clustering.

1. Introduction

Digital images used in social media communication are subject to alterations using different image processing tools. The method of watermarking is used to mask the ownership credentials of digital multimedia files like audio, video, text, graphics and images. Authentication and validation of the integrity of such data is required [1]. In steganography and watermarking techniques, hiding the data may be achieved, but steganography is used to hide secret information that is confidential and not to be exposed to the middle-man. Hidden data in steganography is covered in the source data. Steganography does not care about the source that is used. The only problem in steganography is how data is hidden secretly in the source data. In the other hand, watermarking masks ownership information in the multimedia data that is used to show who the owner.

----

* Corresponding author.

We are more bothered about the source data here in watermarking. Watermarking is used for the proper validation of the source data and to prove confidently who the correct owner is, and also to prove that the content of the data is not corrupted by various attacks [2]. American Scientific Research Journal for Engineering, Technology, and Sciences (ASRJETS) (2021) Volume 00, No 1, pp 00-00 Several methods for embedding the watermark in the multimedia data were proposed [2]. By taking picture as the source, most of the algorithms were developed and embedded content is also primarily image. To make the source content more robust against different attacks, these methods still require more improvements. Clustering is one technique where attackers are unable to find out where precisely the content of the watermark is placed. We will implement more robust clustering approaches in this article. To identify the features in the source image, where the watermark image is embedded and the latter is retrieved from the same features in the watermarked image, we used the KNN clustering process. The rest of this paper is organized as follows. A brief study of related work is done in section 2. The proposed method is discussed in section 3. The results of the proposed algorithm are discussed in section 4. The conclusion is done in section 5.

2. Related Work

A brief analysis of the clustering strategies used by researchers for watermark embedding is performed in this section. In [3] Using the Q-quantization table in JPEG and embedded watermarks in those clusters, FCM clustering was done on the DCT coefficients. The technique is used to extract the watermark from the watermarked image. In [4] used the fuzzy c-means clustering. They embed authenticated data for FCM by creating C distinct clusters where the information is embedded in the image and then using the tamper detection technique to determine if the image is altered anywhere. They made LSBs of pixels in those clusters as null values using the FCM method to form C clusters and embedded the watermark in those pixels. The watermark is embedded here to prove the image's integrity, that is, whether or not the image is tampered in. If the LSBs of the watermark image are altered, this approach does not prove who the right owner is. In [5] to define the features where the watermark should be embedded, Scale Invariant Feature Transformation (SIFT) was used. Using the k-means clustering algorithm, the function regions are then separated into k clusters after defining the features. In one cluster, they inserted one Watermark bit. They split the features into k clusters and were able to embed the watermark's total k-bits. If the watermark image is longer than k-bits, so the number of clusters needs to be increased. In 2012 Lingling An and his colleagues [8] use K-means clustering on the cluster to embed watermark for k number of clusters in the picture and then integer wavelet transformation (IWT) is performed on the cluster. Eventually, they demonstrated that their approach has better performance compared to embedding watermark in the spatial domain. In 2013 Bassem S. Rabil and his colleagues [7] used a function extraction algorithm to extract facial areas and then cluster American Scientific Research Journal for Engineering, Technology, and Sciences (ASRJETS) (2021) Volume 00, No 1, pp 00-00 to create k number of clusters where watermark is embedded. This approach is fragile if there is distortion in the facial region. In [6] used color picture clustering, separating clusters depending on the clustering index and then subdividing them into sub-clusters based on the red, blue and green colors of the pixels, and then embedding the watermark in those clusters. And for extracting the watermarked image, the same procedure is applied. The number of pixels leading to watermark embedding in this process is much lower. In 2016 R.Suganya and his colleagues [9] used k- means clustering approach applied to the LSB bits of the pixels. They separate the clusters on the basis of the number of 0's and 1's in the cluster created by the LSB bits, embedding watermark bits into those clusters. This approach is used to detect whether or not the image has been tampered. The comparison of various algorithms and challenges of those algorithms are listed in the table 1

Table 1: Features and challenges of image watermarking based on clustering methods

Author	Used Method	Features	Challenges
Wei-che and his colleagues	FCM	• Used for tamper detection • Watermark image is imperceptible	• Restoration of tampered regions cannot be done • If all LSBs are modified we cannot prove the ownership
Huawei Tian and his colleagues	KM	• Algorithm is semi-blind • Algorithm is invariant to geometric distortions	• Embedded watermark size is very less • If watermark size is more then it take more time to increase the number of clusters
Jianzhen Wi	FCM and HVS	• Used q-quantization method for FCM clustering • More Robust	• It require cover image to verify the watermark
R. Suganya and his colleagues	KM on LSB	• Embedding watermark is easy • It is Blind Watermarking method	• Does not prove ownership if the image is tampered • Not Robust
Mohamed Tahar Ben Othman	Pixel clustering	• Used for color images •Each cluster is subdivided into sub clusters if needed	• Embedding watermark capacity is very less

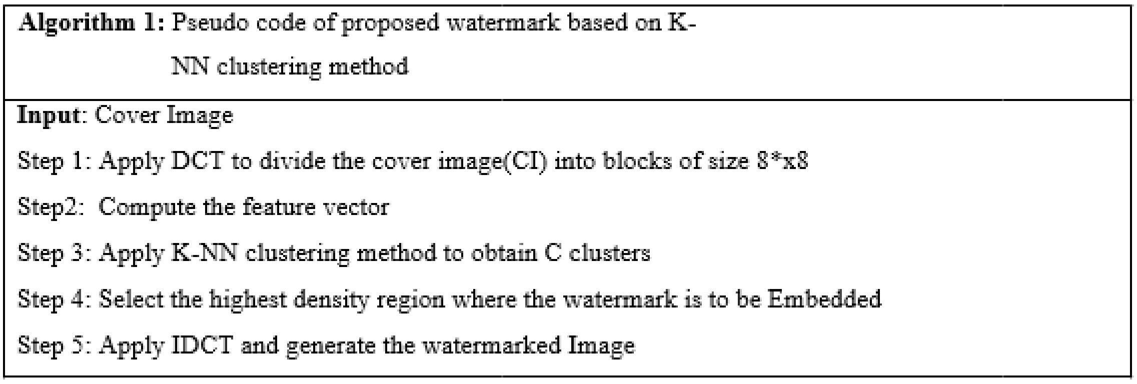
3. Architectural view of proposed adaptive image watermarking based on K-NN clustering

3.1. Proposed Methodology

Although several watermarking methods have been developed, more robust methods are expected to be used to show ownership credentials. There is a need for algorithms that support different watermarking attacks, such as collusion attacks, etc. [10]. The technique here suggests a more robust algorithm for embedding watermark in in transformation domain by constructing clusters using the clustering algorithm based on K-NN as seen in Figure 1. In algorithm 1, the steps used in the proposed image watermarking algorithm are compiled.

The source image (CI) here is divided into blocks of size 8x8 using DCT. In addition, each block is divided again into 2x 2 blocks from which vector characteristics are extracted by scanning both horizontally and vertically. The K-NN clustering algorithm is applied after computing the feature vector for all blocks of the source image and appropriate features for embedding watermarks are identified. These features are embedded with the binary watermark (WA). Then the Inverse DCT is applied on the resultant image which is the intended watermark Image (WI). The key benefit of this approach is that the technique is more robust to collusion attacks.

Table 2



```
Algorithm 1: Pseudo code of proposed watermark based on K-
             NN clustering method
Input: Cover Image
Step 1: Apply DCT to divide the cover image(CI) into blocks of size 8*x8
Step2:  Compute the feature vector
Step 3: Apply K-NN clustering method to obtain C clusters
Step 4: Select the highest density region where the watermark is to be Embedded
Step 5: Apply IDCT and generate the watermarked Image
```

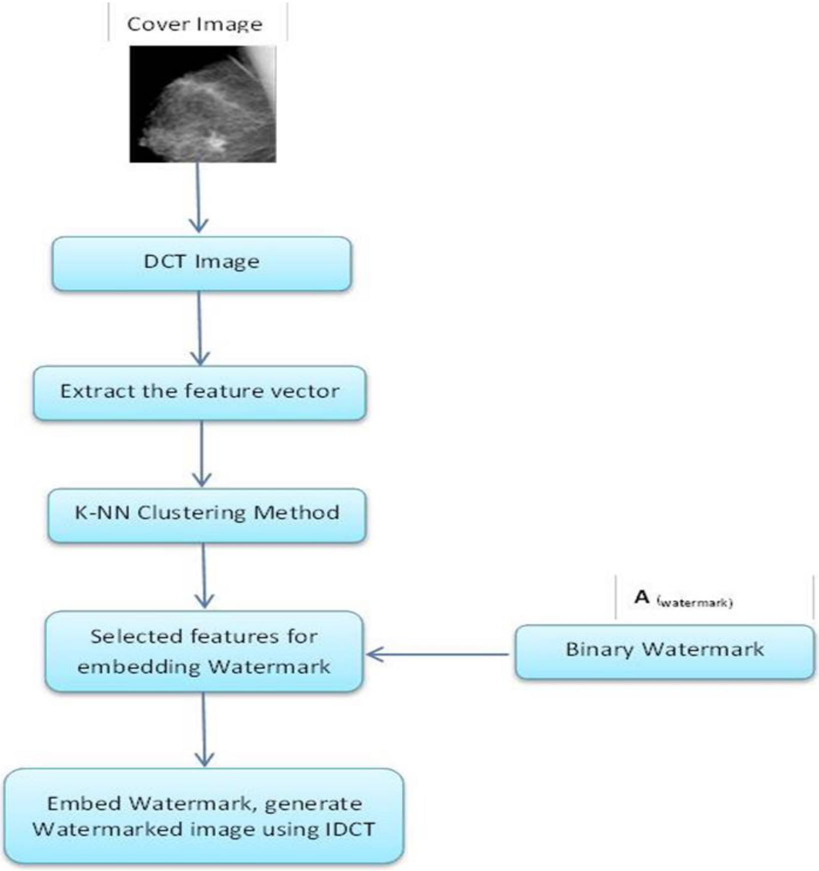

Cover Image
DCT Image
Extract the feature vector
K-NN Clustering Method
A (watermark)
Selected features for embedding Watermark
Binary Watermark
Embed Watermark, generate Watermarked image using IDCT

Figure1: Proposed method of K-NN clustering based image watermarking

3.2. Objective of the Proposed Method

The key objective of the proposed process is to minimize MSE and find more bits to embed watermark. Mean Squared Error (MSE) is a metric that checks whether two images are similar. It is worked out using this equation.

$$MSE = \frac{1}{MeNe}\sum_{x=1}^{Me}\sum_{y=1}^{Ne}\left[I_O(x,y) - I_W(x,y)\right]^2 \qquad (1)$$

Here I0(x,y) is the Cover image and Iw(x,y)is the watermarked image.

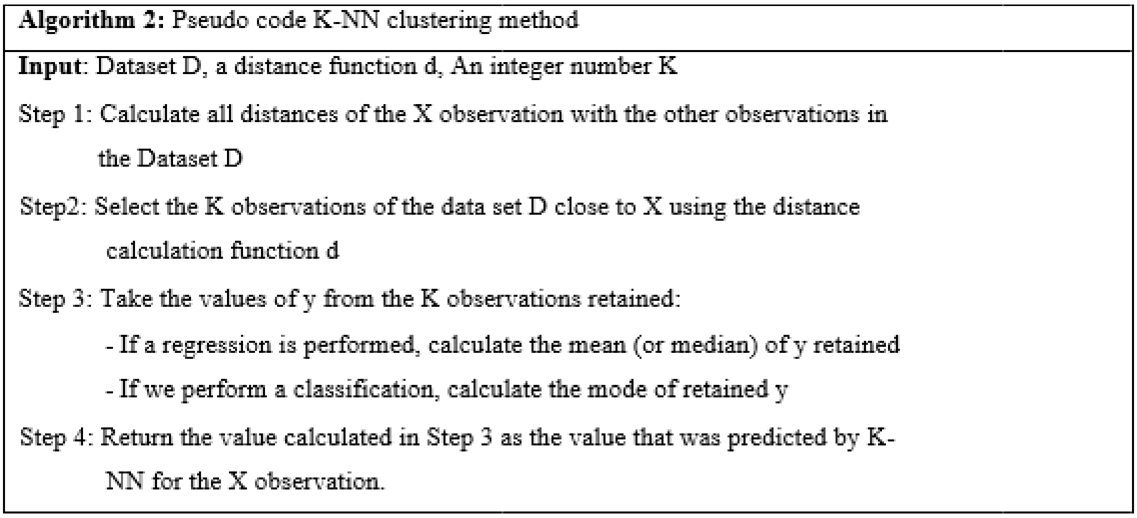
3.3. K-Nearest Neighbor Clustering Method

The K-NN (K-nearest neighbors) algorithm is a supervised learning method. It can be used both for regression and classification. Its operation can be likened to the following analogy "tell me who your neighbors are, I'll tell you who you are". To make a prediction, the K-NN algorithm will not compute a predictive model from a Training Set as is the case for linear regression. Indeed, K-NN does not need to build a predictive model. Thus, for K-NN there is no training phase as such. This is why it is sometimes categorized in Lazy Learning. In order to make a prediction, K-NN uses the dataset to produce a result.To make a prediction, the K-NN algorithm will base itself on the entire dataset. For an observation, which is not part of the dataset, which we want to predict, the algorithm will look for the K instances of the dataset closest to our observation. Then for these K neighbors, the algorithm will use their output variable y to calculate the value of the variable y of the observation that we want to predict. In order to find the K Nearest Neighbor to a data to be classified, we can choose the Euclidean distance. Given two data represented by two vectors xi and xj, the distance between these two data is given by:

$$d(\mathbf{x}_i, \mathbf{x}_j) = \sqrt{\sum_{k=1}^{d}(x_{ik} - x_{jk})^2} \qquad (2)$$

The pseudo code of K-NN clustering is demonstrated in algorithm 2.

Table 3



```
Algorithm 2: Pseudo code K-NN clustering method
Input: Dataset D, a distance function d, An integer number K
Step 1: Calculate all distances of the X observation with the other observations in
        the Dataset D
Step2: Select the K observations of the data set D close to X using the distance
        calculation function d
Step 3: Take the values of y from the K observations retained:
        - If a regression is performed, calculate the mean (or median) of y retained
        - If we perform a classification, calculate the mode of retained y
Step 4: Return the value calculated in Step 3 as the value that was predicted by K-
        NN for the X observation.
```

The choice of the K-value to be used to make a prediction with K-NN, varies depending on the dataset. As a general rule, the fewer neighbors (a small K number) are used, the more underfitting will occur. On the other hand, the more neighbors (a high K number) are used; our prediction will be more reliable. However, if we use K number of neighbors with K=N and N being the number of observations, we risk having overfitting and consequently a model that does not generalize well on observations it has not yet observed. K-NN is a fairly simple algorithm to learn. Mainly because it does not need a model to make a prediction, the important cost is that he has to remember all the observations in order to be able to make his prediction. Thus it is necessary to be careful about the size of the training set.

The working principle of the algorithm is shown in Figure 2.

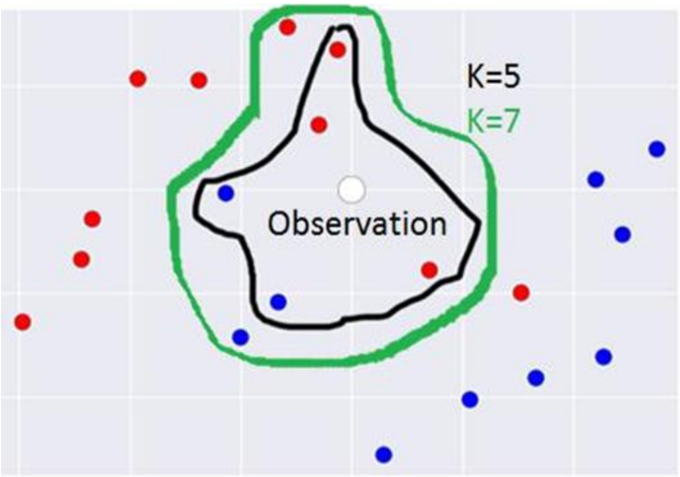
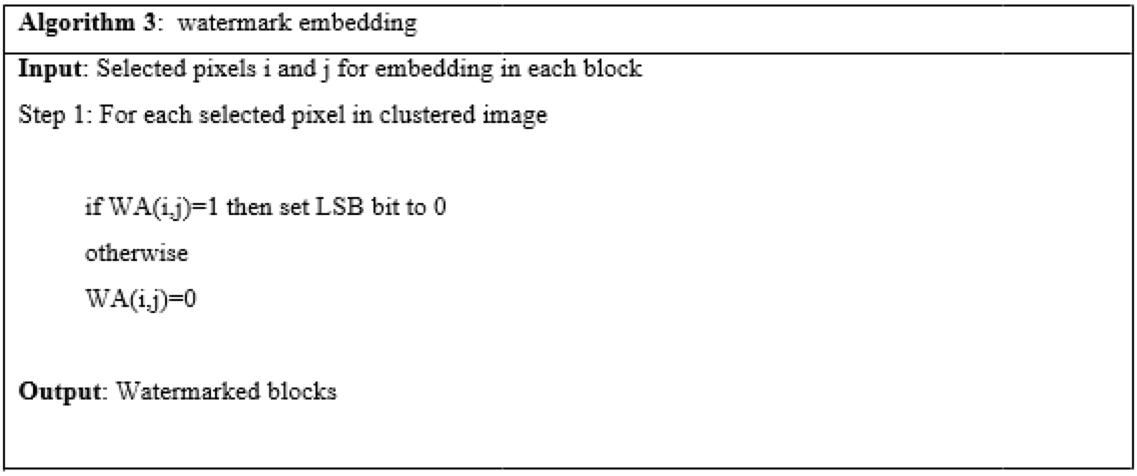


Figure 2: K-NN method

3.4. Watermark Embedding

The watermark image (WA) is transformed to a binary representation of a 1-d vector for embedding watermarks in a clustered image. Set the LSB bit to 1 in the chosen WA pixels if WA is 0, otherwise 0.

Table 4



```
Algorithm 3:  watermark embedding
Input: Selected pixels i and j for embedding in each block
Step 1: For each selected pixel in clustered image

        if WA(i,j)=1 then set LSB bit to 0
        otherwise
        WA(i,j)=0

Output: Watermarked blocks
```

3.5. Extraction

To extract the watermark from watermarked image (WI), we apply the reverse methodology. The process is shown in Figure3.

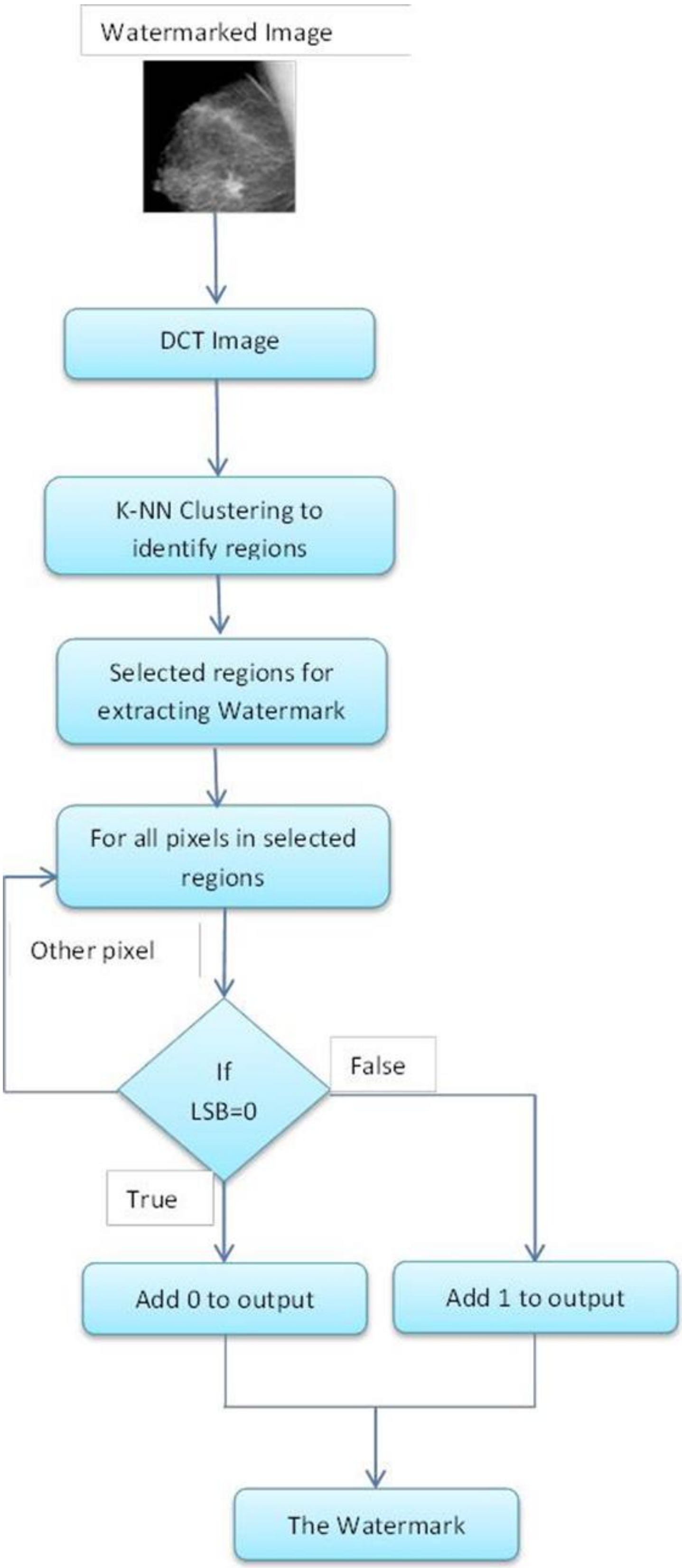

Watermarked Image
DCT Image
K-NN Clustering to identify regions
Selected regions for extracting Watermark
For all pixels in selected regions
Other pixel
If LSB=0
False
True
Add 0 to output
Add 1 to output
The Watermark

Figure 3: Watermarking Extraction

4. Results and discussions

In this section, evaluation of the proposed adaptive watermarking approach based on K-NN clustering is done. The algorithm is applied to a number of different Image categories such as smooth, classic, pattern pictures, medical etc. We use a medical image grayscale 256*256 A_0002_1.LEFT_CC extracted from the DDSM database and the watermark is the logo with the size 20*20, Figure 4.

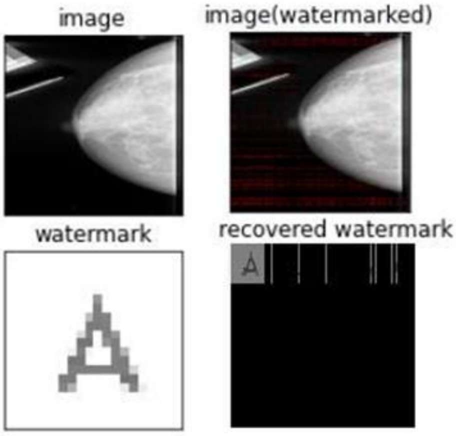


Figure 4: Image watermarking

The proposed algorithm has been implemented using python 3.8(jupyter) language programming. The Peak Signal to Noise Ratio (PSNR) is used as distortion measurement between the original and a watermarked image. It is define as

$$PNSR = 10\log_{10}\left(\frac{255^{2}}{\frac{1}{N}\sum_{n=1}^{N}(w_n)^{2}}\right)$$

$$w_n = s_n - y_n$$

Where N is the total number of the pixels, s represents the original image, and y is the watermarked image.

The watermark (WA) is inserted in the Cover Image (CI) with and create watermarked image (WI) after embedding. Later WA is extracted from WI. The performance of algorithms using performance measurements such as PSNR is shown. The algorithm proposed has better performance. Using DWT and DCT-DWT algorithms, the watermark extraction is better for medical image there are more bits without any importance for medical image that we can use for embedding Figure 5,

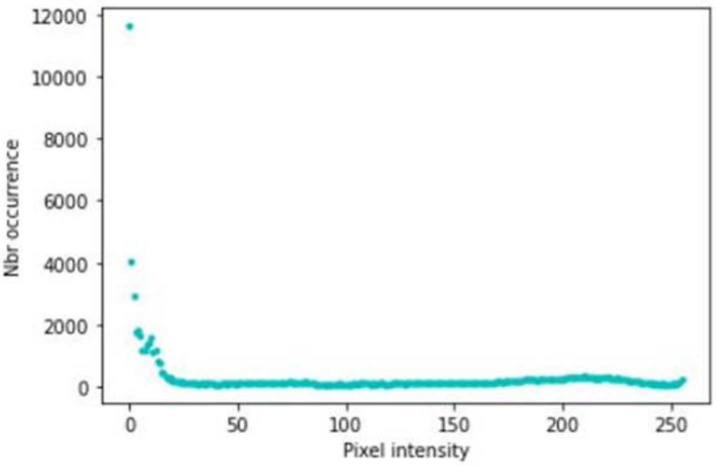


Figure 5: bits for embedding

For this image we have 11647 1 pixels with value 0 and 4061 with value 1, these pixels have no importance in medical images so we can use there LSB bit for Insertion.

The watermarking used in this paper has this parameter:

Average PSNR = 46.15

Average SSIM = 0.997

SSIM is the Structural Similarity

5. Recommendations

In our study, a watermarking application is presented for inserting patient information into medical images. Being a current problem, the large number of images has started to take place in almost all stages of daily life. Due to the growing interest in medical applications, these medical applications and devices have become accessible on the internet. Thus, our next goal is to find a solution to detect tumor evolution for patients. For this purpose we recommend the use of texture and watermarking techniques.

6. Conclusions

We have proposed an adaptive K-NN clustering technique in this paper to embed the watermark in the cover image. We see that the performance of the watermark image proposed is higher than traditional methods. Our adaptive process of finding the best parameter K compared to other clustering approaches, make the mechanism more stable. In future work we will cluster the watermark as well and embed the each cluster of watermark in different cluster of the image.

References

[1]. Chu-Hsing Lin,Jung-Chun Liu,Chih-Hsiong Shih,Yan-Wei Lee "A Robust Watermark Scheme for Copyright Protection" 2008 International Conference on Multimedia and Ubiquitous Engineering

[2]. Kavitha Soppari , N.Subhash Chandra "Study of Digital Watermarking Algorithms for Digital Rights Management and their Attacks", International Journal of Computer Trends and Technology (IJCTT) – Volume 67 Issue 1 - Jan 2019.

[3]. Jianzhen Wi andJianyig Xie"ADAPTIVEIMAGEWATERMAFUNG SCHEME BASED ON HVSAND FUZZY CLUSTERINGTHEORY"IEEE Int. Conf.

[4]. Wei-Che Chen, Ming-Shi Wang "A fuzzy cmeans clustering-based fragile watermarking scheme for image authentication" Expert Systems with Applications 36 (2009) 1300 1307, Science Direct

[5]. Huawei Tian, Yao Zhao, Rongrong Ni, and Jeng-Shyang Pan "Geometrically Invariant Image Watermarking Using Scale-Invariant Feature Transform and K-Means Clustering" ICCCI 2010, Part I, LNAI 6421, pp. 128–135, 2010. ©Springer-Verlag Berlin Heidelberg2010

[6]. Mohamed Tahar Ben Othman "Digital Image Watermarking based on image clustering"

[7]. Bassem S. Rabil • Robert Sabourin • Eric Granger. "Rapid blockwise multi-resolution clustering of facial images for intelligent watermarking" Machine Vision and Applications DOI 10.1007/s00138-013-0493-1

[8]. Lingling An, Xinbo Gao, Xuelong Li,Dacheng Tao, Cheng Deng, Jie Li "Robust Reversible Watermarking via Clustering and Enhanced Pixel-Wise Masking" IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 21, NO. 8, AUGUST 2012

[9]. R. Suganya1 ,R Kanagavalli "TamperDetection using Watermarking Scheme and KMean Clustering for Bio Medical Images",International Journal for Modern Trends in Science and Technology Volume: 02, Issue No: 11, November 2016

[10]. Lin ZHANG, Fengyong QIAN, Yi GAO, Yuesheng ZHU "A New Integration Scheme of Robust and Fragile for Secured Digital Watermarking" 2008 ISECS International Colloquium on Computing, Communication, Control, and Management