EAI Endorsed Transactions
on _____ Research Article EAI.EU

Text Steganography in Statistically Clustered Iris Image

Irtefaa A. Neamah[1,*] and Hind Rustum Mohammed [2]

[1]Assistant Professor, Department of Mathematics, Faculty of Computer Science and Mathematics, University of Kufa, Iraq.
[2]Professor, Department of Computer Science, Faculty of Computer Science and Mathematics, University of Kufa, Iraq.
irtefaa.radhi@uokufa.edu.iq, hindrustum.shaaban@uokufa.edu.iq

Abstract

The hiding text within the iris to increase the data protection method is discussed in this work. It is impossible to distinguish between the iris image before and after concealment, and the difference between the two images only after using statistical measures such as PSNR and MSR to compare them. The proposed database consists of 500 images with different formats (tif, gif, png, jpg, .bmp) selected for analysis. The proposed method is shown with more accurate results, stronger image encoding, and high-efficiency text protection using performance evaluation factors to assess business standards. The success of hiding high-text ratios proved successful. Experimental results were shown based on a statistical strategy, and that the text was converted into two random variables, X and Y, which were distributed to Asia. Then, the random variables' data were included in the iris segment, cut-off, and iris' clustered image. It appears that the use of our proposed scheme can include sufficient data in the image of the iris that maintains the accuracy of the identification.

Keywords: Steganography, PNSR, MSE, Clustered Image, Iris Image.

Received on 28 August 2020, accepted on 04 November 2020, published on 18 November 2020

doi: 10.4108/_____

1. Introduction

Data hiding is one of the key ways to protect your privacy. The objective of hiding biometric data is to include enough personal data in the maintenance and biometric templates performance recognition. Methods of hiding current dynamic data usually include data in an area that does not contain basic attributes for dynamic measurements. In the template data of iris is only included in the blue channel. However, these perform schemes reasonably, well, to hide biometric data. How to minimize the effect, which remains unanswered, is embedded in biometrics recognition [1].

Digital data has one of the great advantages that it can be reproduced without quality losing. It can also be easily modified and created for authorized parties that want to prevent illegal distribution of secret documents in many contexts; as video security regulations and legal evidence, i.e., image, audio, or video [2].

A single image is composed of a group of pixels. There are many categories of pixels. Therefore, those pixels that belong to the same category have similar values and must be different from other categories. So, within the same cluster, a group of pixels is combined. Then cluster values are calculated based on feature selections. The method is called k-means or (Lloyd's algorithm). After the cluster value calculation, three histograms have been used in RGB. Peak calculation is used for each histogram and calculates the peak value in Red histogram, Green histogram, and Blue histogram [3].

To accurately divide the iris areas into ideal images, a new technique is proposed, which uses the detection of statistical distribution mechanisms to compensate for iris image detection errors resulting from the detection of color fragmentation and how to protect the text and hide it within the iris after the adoption of fragmentation [4].

The remaining sections show Section 2: Statistical cluster analysis and Statistical Proposed Methodology approach discussed in Section 3, The results shown in Section 4 and performance analysis discussed under section 5, The paper concluded with future work in Section 6.

*Corresponding author. Email: irtefaa.radhi@uokufa.edu.iq

2. Statistical Cluster Analysis

Clustering is an unattended automated learning method that means there is no information about the output. Data that contains instances are divided—then described by their features or attributes. It is then divided into separate subtotals by aggregation of the algorithms. Instances should be in a similar set with some criteria. The instances of different groups should be as different as possible [5]. The assembly is used in science, medicine, economics, astronomy, web intelligence, management, security, etc. Over the past decades, many assembly techniques have been developed, improved, or modified to solve many problems. Some known assembly algorithms are k-mean, DBSCAN, hierarchical assembly, and so on. One of the simplest methods is K-mean, and it aims is to reduce the total distance between the instances and their centroids, which are represented as the mean of all cases in corresponding groups [6].

It is probably impossible to aggregate algorithms because hundreds of themes in literature can be found. One reason why many clustering algorithms exist is the fact that the cluster cannot be defined accurately [7]. Based on an understanding of the cluster model, a measure of similarity and difference, clustering algorithms can be divided into several categories. The connection models divide the data into groups depending on the distance connection. The hierarchical assembly is one example of connection forms. Creation of distribution models groups using statistical distributions [8].

The K-mean modified function algorithm was used in many applications, for example, image fragmentation. K-algorithm means an iterative algorithm with two operations performed at each frequency. The algorithm begins with randomly generated k centers, i.e., the center points. In each generation, each object (instance) is assigned to the nearest centroid. The most commonly used distance scale in the k-mean in Euclidean space. After the appointment step, the locations of the central countries are updated [9].

$$C_i = \frac{1}{|S_i|} \Sigma X_j \qquad (1)$$

Where |Si | is the number of instances in the cluster Si, i = 1, 2. . . k. The quality estimation is an important part of all clustering algorithms; in the k-mean algorithm, the aim is to reduce the sum of variations between cases and corresponding raster devices [10].

3. Statistical Proposed Methodology approach

The section at beginning built as testing the system by [11] interval database, when the image determined before enhancement and then measure the segmentation and hide text rate of the system when the determination process performed after enhancement the images, figure1 showed the sample of the database in this paper [12].

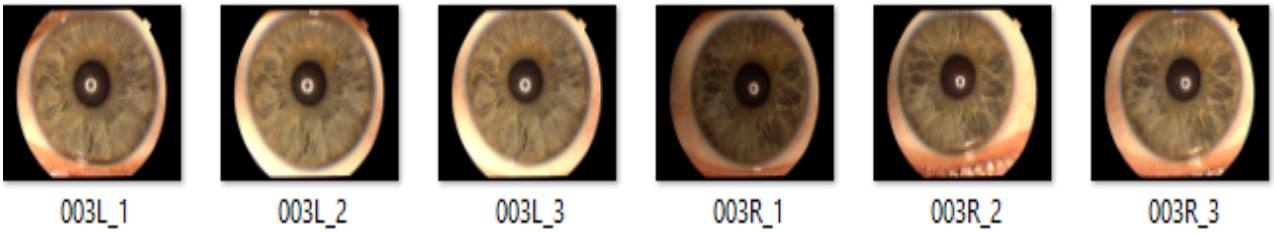

Figure 1. Sample of the images

The idea is to hide the largest possible amount of text data within the smallest possible storage space. One of the most important biometric characteristics is the iris, which is used to increase information security. Therefore, the iris selection and segmentation provide us with less size than the cover image to hide large-sized text information. After including important personal information within the iris's smallest spot, it can retrieve the iris' image as it was and send it with hidden information without notice to those who object to sending the image [13].

This paper used to convert the text to be hidden into statistical data that follows binary exponential distribution [14]. We assumed that the length of the text is n and by dividing it into two parts. The first part represents the variable X, and the second part represents the variable Y if the length of the text is an even number. If the length of the text is an odd number, we add an optional character for the text as in the formula below

$$L(x,y) = \begin{cases} \frac{n}{2} & n \quad Even \\ \frac{n+c}{2} & n \quad Odd \\ & c \quad any \quad Character \end{cases} \qquad (2)$$

L (x, y) is the length of random variables x and y, n is the original text's length.

After converting the text into two random variables, we can include these random variables as digital data inside the iris image after cutting and fining the edges accurately and classifying them. Thus, the iris's Image after chipping containing hidden text can be read according to the algorithm shown in figure 2. It is worth noting that this method can be applied to any image with different extensions and sizes.

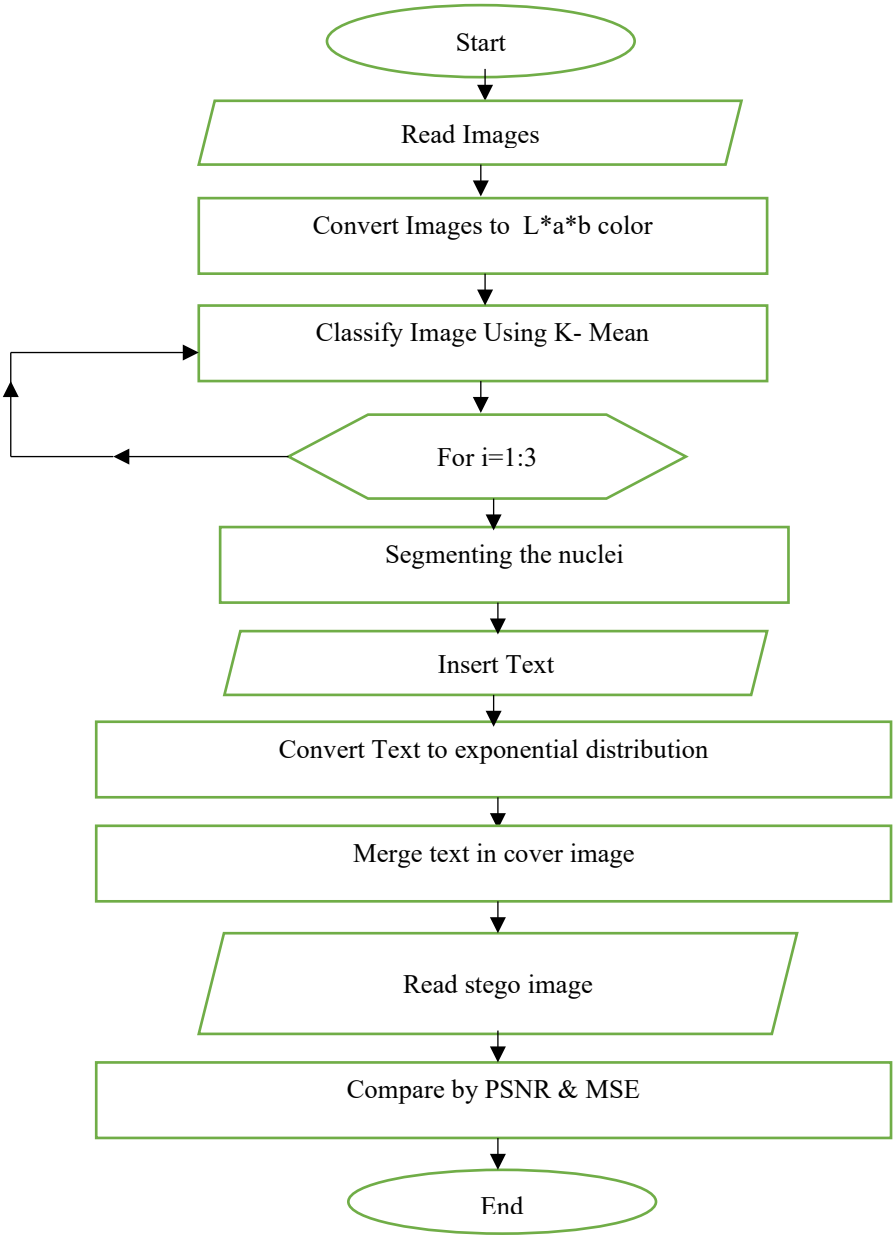

Figure 2. Proposed system flowchart

4. Experimental Results

The flowchart procedures proposed in this paper have been applied to two types of iris images; the first (figure 3) is iris colored images from the database. The second group (figure 4) consisted of taking colored iris pictures taken from electronic sites. The number of pictures of the two groups was 500 pictures. As an example, the results of the proposed approach were presented. The results were presented for only four pictures representing the two eyes (left and right) Shown in the following figure:

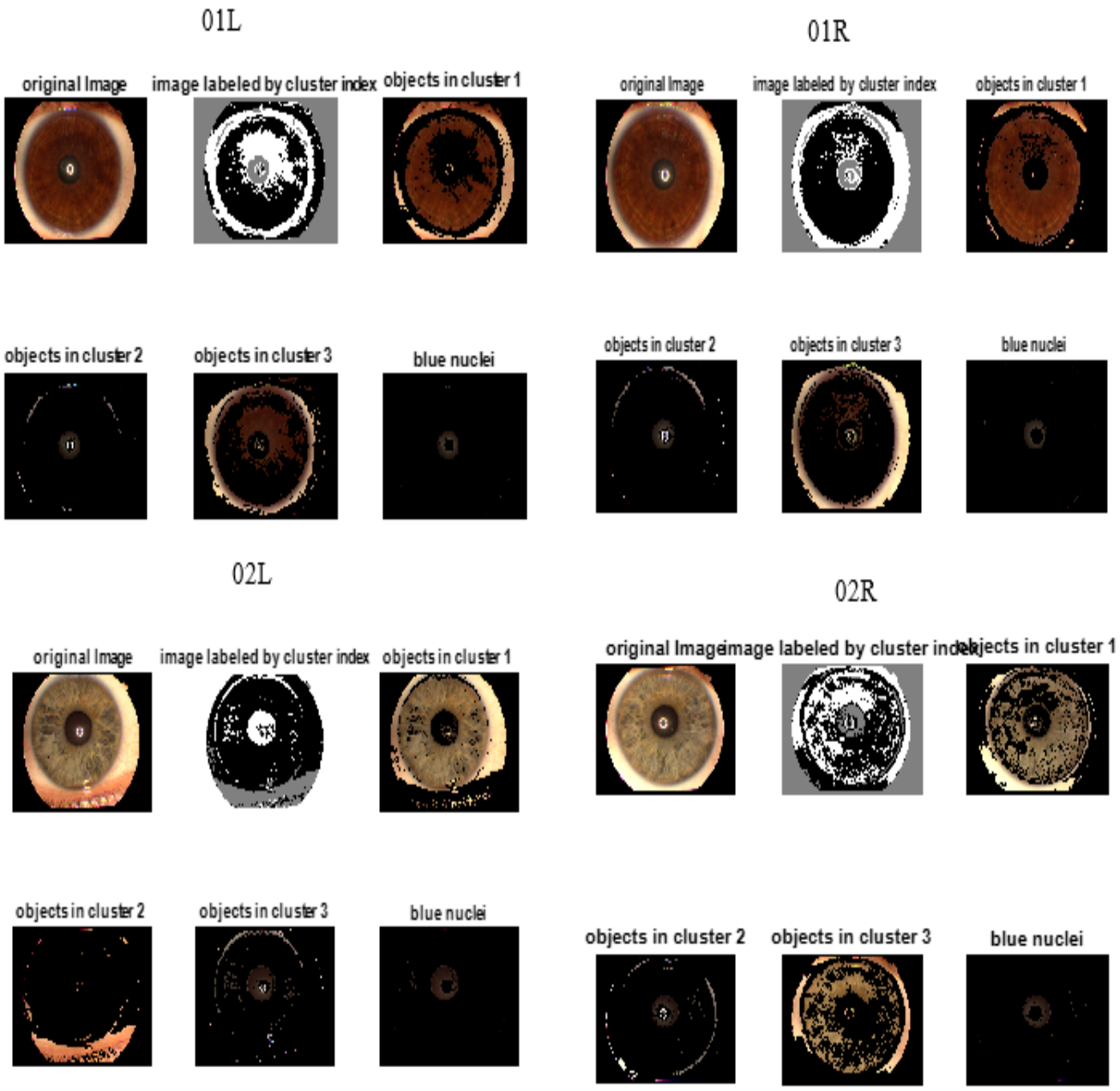

Figure 3. Clustering steps (left and right iris, respectively)

The final stage in the application of the proposed method, which represents the most accurate part of the iris (chosen in the process of detection edges and identification, the integration of the text to hide in the selected part and we will use the following text as an example:

The message will hide in the cover image is:

"The book of nature is written in the language of Mathematics" –Galileo

To determine the accuracy, reliability, and validity of the proposed method's results, a comparison between the iris's original image after the summer, without the hidden text and between the iris's image after the summer and the hidden text. Note that in Fig. 4, the naked eye shows no differences between the two images; for this, we must use the two images' data and find the differences more accurately.

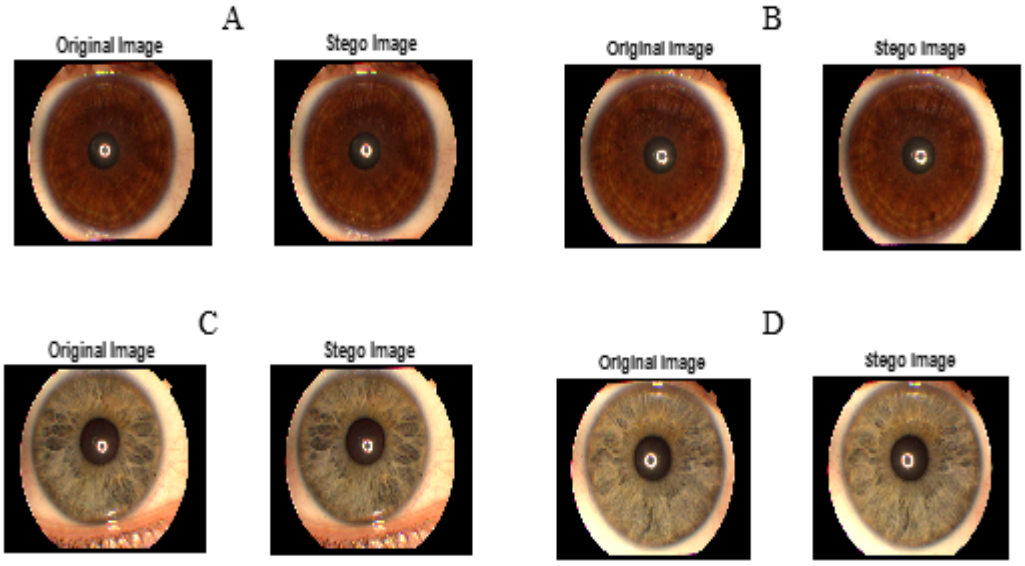

Figure 4. Original and steganography image

A, B, C, D, respectively, in figure 5. Each part contains the original image and cover image, which contains hidden text. As shown, there is no sign of the difference between the two images at a time in figure 6.

A set of criteria for evaluating the proposed work performance were taken, which are the following statistical criteria (PNSR, SNR, and MSE).

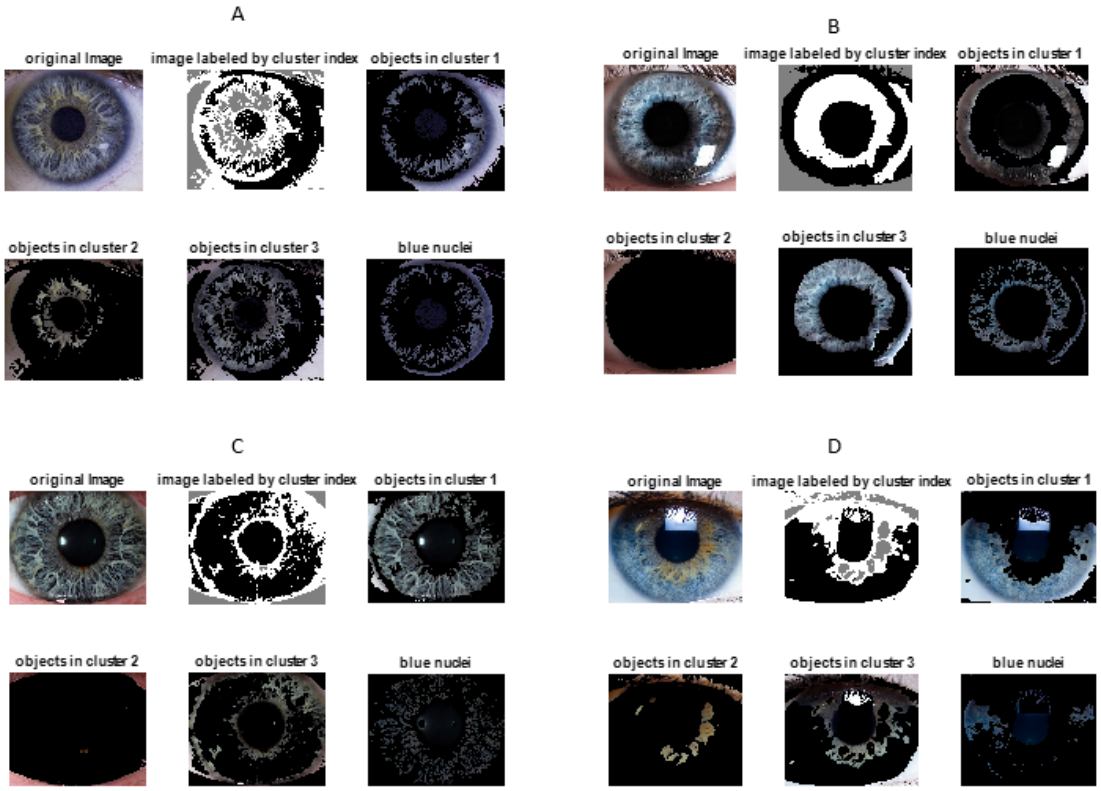

Figure 5. Clustering steps

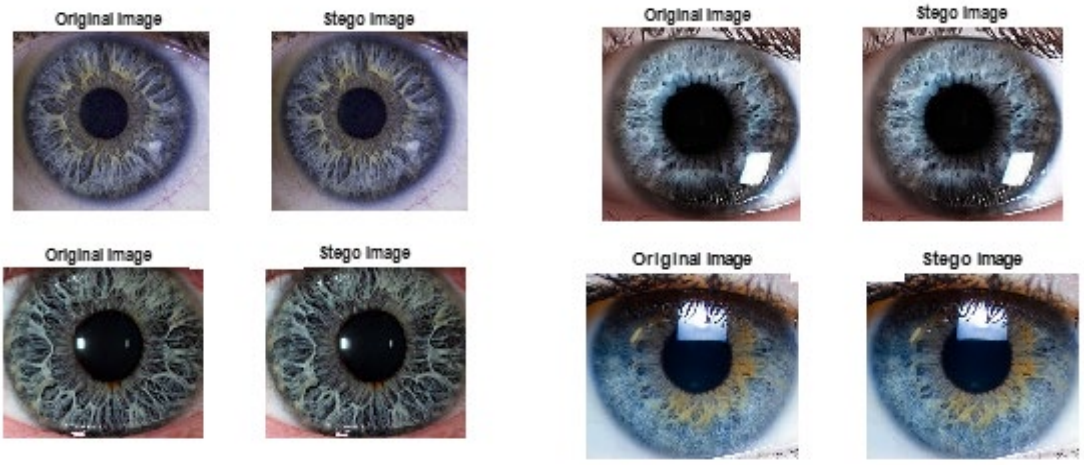

Figure 6. Original and steganography

5. Performance

MSE (mean-squared error) is one way to quantify the difference between values implied by an estimator and the true values of the estimated quantity. MSE is calculated using the following equation:

$$MSE = \frac{\sum [I_1(m,n) - I_2(m.n)]^2}{M * N} \quad (3)$$

In the above equation, the number of rows and columns in the entered images are M and N, respectively. PSNR (Peak Signal to Noise Ratio) is a standard measurement technology used to hide information to test stego image quality. If the PSNR value is higher, that means the better the stego image quality. PSNR is calculated as the following equation:

$$PNSR = 10 \log_{10} \frac{R^2}{MSE} \quad (4)$$

R is the maximum fluctuation in the input image data type in the previous equation. For example, if the input image has a double-precision floating-point data type, R is 1. If R is 255, R contains an 8-bit integer data type, etc. [9]. After the standards were applied to measure the performance of the algorithm, we obtained the results that we will show in the following table 1 and 2:

Table 1. Implementation of the algorithm for dataset color iris images with size 576x768

No.	Image	PSNR	SNR	MSE
1	01L	61.7364	28.3690	0.0436
2	01R	67.7570	34.6219	0.0109
3	02L	61.7364	30.7113	0.0436
4	02R	67.7570	24.7661	0.0109

Table (1) shows that the comparison criteria between the original and stego image are almost identical. The reason is that the images from the same database and with the same sizes. Note that the right iris of most images in the same database is equal to most other right irises. The same is true for the left iris. This is what we see in the measure of performance with PNSR and MSE.

The standard scale SNR has varied values between all images. This explains to us more than the cover's image, despite its overlap with the concealment image now that the digital measurements show their difference.

Table 2. Implementation of the algorithm for general color iris images

No.	Image	Size	PSNR	SNR	MSE
1	00A	**383 436 3**	57.5054	40.9965	0.1155
2	00B	**214 227 3**	52.1429	35.2866	0.3970
3	00C	**264 319 3**	60.5531	40.0364	0.0572
4	00D	**354 422 3**	57.0217	35.6313	0.1291

When applying the same performance criteria for iris images taken from the Internet, we observe that performance standards are clearer. This means that all the measurement values are different, and it is very clear in the PNSR scale, SNR, and MSE, respectively. Different sizes of images and specifications of each image affect the performance of the standards. Finally, the results obtained confirm the difference between the two images shown in each of the forms, such as in Figure (4) and Figure (6)

6. Conclusions

After applying the proposed method of hiding the text data inside the iris image after identifying and classification the selected area of the iris accurately to reduce the effect of identification of the iris that carries data. The proposed method is to include data more confidentially to protect users' data. This study's experimental results showed us that we could include the largest volume of data within the iris's smallest spot while maintaining high accuracy in secrecy to protect it from discrimination by attacks in the transmission of information. We can use this technique to hide large size data (any data size) and embed it in very small images. For example, the smallest part of the footprint, or the smallest part we get from a distorted image. Also, we convert the segmented part from the original image and convert it to the same size as the original image and use it to hide. The work's future scope focused on big data by hiding the image's large digital images using wavelet theory.

References

[1] A.E. Hassanien "Hiding iris data for authentication of digital images using wavelet theory." Pattern Recognition. Vol. 16, no.5, pp. 637–643, 2006.

[2] S. Basar, A. Adnan, & N.H. Khan "Color Image Segmentation Using K- Means Classification on RGB Histogram." Recent Advances in Telecommunications, Informatics and Educational Technologies, vol.23, no.6, pp. 257-262, 2014.

[3] N. Dhanachandra, K.Manglem, & Y. Chanu "Image segmentation using k-means clustering algorithm and subtractive clustering algorithm" Procedia Computer Science, Elsevier, pp.764–771, 2015.

[4] P. Kruus, C. Scace, M. Heyman, & M. Mundy, "A survey of steganographic techniques for image files" Advanced Security Research Journal, vol.1, no.1, pp. 41-52, 2003.

[5] R. Subramani and C. Vijayalakshmi, "Augmented Lagrangian Algorithm for Hydrothermal Scheduling," EAI Endorsed Transactions on Energy Web, vol. 5, no. 18, p. 154815, Jun. 2018.

[6] S. P and D. B, "Improvised_XgBoost Machine learning Algorithm for Customer Churn Prediction," EAI Endorsed Transactions on Energy Web, p. 164854, Jul. 2018.

[7] A. Bobryakov, V. Borisov, A. Gavrilov, and E. Tikhonova, "Compositional fuzzy modeling of energy- and resource-saving in socio-technical systems," EAI Endorsed Transactions on Energy Web, vol. 0, no. 0, p. 155863, Jul. 2018.

[8] E. Tuba, D. Dolicanin-Djekic, R. Jovanovic, D. Simian, & M. Tuba, "Combined Elephant Herding Optimization Algorithm with K-means for Data Clustering." Information and Communication Technology for Intelligent Systems, vol.17, no.12, pp. 665-672, 2019.

[9] G. O. Young, "Synthetic structure of Industrial Plastics" In J. Peters, Edition Plastics McGraw-Hill, vol.4, no.4 pp. 15-64, 1964.

[10] S. Arhun, V. Migal, A. Hnatov, S. Ponikarovska, A. Hnatova, and S. Novichonok, "Determining the Quality of Electric Motors by Vibro-Diagnostic Characteristics," EAI Endorsed Transactions on Energy Web, p. 164101, Jul. 2018.

[11] D. Devikanniga, A. Ramu, and A. Haldorai, "Efficient Diagnosis of Liver Disease using Support Vector Machine Optimized with Crows Search Algorithm," EAI Endorsed Transactions on Energy Web, p. 164177, Jul. 2018.

[12] E. Lee, M. Schmidt & J. Wright "Improved and simplified inapproximability for k-means" Information Processing Letters, vol.14, no,4, pp. 40-43, 2017.

[13] D. Michal, D, & M. Libor, M. "Iris Database" http://phoenix.inf.upol.cz/iris/ Accessed by May 24, 2020.

[14] L.Sheng, C. Xin, W. Zichi, Q. Zhenxing, & Z. Xinpeng "Data Hiding in Iris Image for Privacy Protection." IETE Technical Review, vol.23, no.9, pp. 34-41, 2018.