ISSN 2278-3091

Volume 9, No.4, July – August 2020

International Journal of Advanced Trends in Computer Science and Engineering

Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse363942020.pdf

https://doi.org/10.30534/ijatcse/2020/363942020

Reversible steganographic method based on interpolation and genetic algorithm

Amine Benhfid[1], El Bachir Ameur[2], Youssef Taouil[3], Ismail Kich[4]

[1-4]Research Team MSISI - LaRIT. Faculty of Sciences, Ibn Tofail University Kenitra, Morocco,

aminebenhfid@gmail.com[1], ameurelbachir@yahoo.fr[2], taouilysf@gmail.com[3], kichsma@gmail.com[4]

ABSTRACT

Recently, Information security is amongst the major challenges in the world of information communication. Hence the necessity of Steganography which is one of the solutions to fulfill this need; it is the technique of hiding a secret information into a seemingly harmless digital file to ensure successful extraction by the intended receiver. In this paper, a new reversible steganographic method based on the genetic algorithms (GA) combined with the interpolation by linear box-splines on the three directional mesh is proposed. The GA helps to choose among the interpolated pixels the suitable location where to dissimulate the secret information. The population that the GA is operating on is paths that indicate which pixel shall be used for the dissimulation. From a generation to the next one, the selected path provides a lesser Mean Square Error, which helps to obtain a good imperceptibility. As for the capacity, the interpolation technique used in this work provides a large space to conceal data. These hypotheses were corroborated by intensive experimental results that were accomplished on a wide set of test images. We obtained a good trade-off between imperceptibility and capacity.

Key words: Steganography, Data hiding, Interpolation, Box-spline, Optimal Pixel Adjustment Procedure, Message Adaptive Error.

1. INTRODUCTION

The principle of steganography is the protection of the information integrity by dissimulating it into a harmless digital file; the secret data can be a text, an image, a sound, a video or even a software in an object (image, sound, video…etc). This method allows exchanging ``secret'' messages with your correspondent through the Internet, social networks, storage devices or any other media that will go unnoticed by others. Steganography is mostly applied on images because they are shared massively every day in social networks and because the human visual system is very weakly sensitive to minor variations within image's intensity. The ultimate objective of a steganographic model is to embed the largest amount of secret data in an image while preventing third parties from suspecting anything. Therefore, it is imperative to decrease visually and statistically the modifications generated by the dissimulation algorithm. In general, there are two approaches to conceive a steganographic model, the frequency domain and the spatial domain. In the first one, data is hidden in the high frequency coefficients of a frequency transform of the host image. The Discrete Cosine Transform and Discrete Wavelet Transform are the most used transforms in this approach. Generally, the frequency domain-based techniques are characterized by high imperceptibility but low capacity. In the spatial domain-based algorithms, data is hidden in the image pixels. This approach is used for high capacity and acceptable imperceptibility purpose. The earliest technique in this approach is the Least Significant (LSB) method, it consists of replacing the LSB of each of the image's pixels by a bit if the secret data; it is very simple and provides a good imperceptibility; however, it leaves detectable statistical traces making it exposed to attacks. Later, other techniques were proposed to resolve this problem. Other techniques were prosed to take into consideration the host image's texture, like Pixel Value Differencing (PVD) and Interpolation. In PVD techniques, the discrepancy between adjacent pixels indicates the number of data bits to be inserted. In Interpolation-Based Steganography (IBS), down sampling is performed to choose the interpolating pixels that will serve to compute the interpolated pixels; the error of interpolation is an important factor that decides the amount of data to be hidden in each interpolated pixel. One of the most useful features of IBS is its reversibility; we can retrieve both secret data and host image in the extraction phase. In IBS, the adopted interpolation method and the calculus of the interpolation error are important factors since they have direct influence on the capacity of embedding. In [1], authors used the neighbor mean interpolation (NMI) hiding method; the interpolated pixels are the mean values of the two adjacent pixels in rows, in columns and the three adjacent pixels in both directions. The NMI provides an average capacity with modest values of PSNR. Later, the Interpolation by Neighboring Pixels (INP) was proposed in [2]; it ameliorated the capacity of embedding of the NMI. Instead of considering the mean values, it takes 75% from the precedent pixel and 25% from the following one; and this is performed in rows and columns.

In [3], authors proposed a chaos-based steganographic scheme; the number of bits embedded in every pixel is calculated by chaos maps that do the choice randomly.

Furthermore, data is not hidden sequentially in pixels; the GA chooses the sequence to follow during the dissimulation. This work reaches a PSNR of 54 dB for 0,25 bit per pixel; for a non-interpolation-based steganography, the compromise between capacity and imperceptibility still has to improve. In [4], authors proposed a reversible steganographic scheme based on difference expansion; they used multiple layer embedding while employing the difference expansion of quad to decrease the difference value in pixel. The findings attest that they could reach a large capacity of embedding. However, the obtained compromise capacity-imperceptibility is not very good. In [5], a reversible data hiding scheme is proposed; authors utilize the histogram shifting to embed the secret data. To help minimizing the effects on the host image caused by the side information, authors thought of using the GA. They could obtain a good trade-off between capacity and imperceptibility. Later, they proposed an amelioration of their findings in [6]; to create an additional space to hide the secret data, they proposed to shift the peak bins towards right direction then towards left direction in a symmetrical way, the shifting does leave space to dissimulate data.

The optimization algorithms can indeed minimize the distortion caused by the dissimulation; but this minimization does cost an important part of the capacity. Hence, in this paper, we use an optimization algorithm only after the interpolation because the optimization is more efficient when we have a great error that needs a minimization; and that is the case for the interpolation. The error of interpolation becomes great when the host image has plenty of texture; on the one hand, this provides a large space of hiding. The error minimization by the GA is so remarkable, hence a good imperceptibility for a large capacity. GA is a type of artificial intelligence, a model based on Darwin's theory of evolution, thus the name. It is part of the order of evolutionary algorithms, which have the particularity of being inspired by the adaptation of nature. In [19] the authors suggested an image steganography method, which provides security in three levels.

In this paper, a new steganographic method based on interpolation is proposed. The host image is down-sampled by factor 2; then, we compute the interpolated pixels using bi-variate box-spline on three directional mesh. The error of interpolation is calculatedusing a formula that enlarges the dissimulation capacity. At this point, we apply the GA [18] to construct several paths data can be hidden according to. The fitness function that controls the choice is the Mean Square Error (MSE) between the host image and the stego image. After refining these paths in several generations, we choose the most suited path, the one that provides the smallest MSE value. Afterwards, we append an identifier of the selected path to the secret data, the identifier is necessary to retrieve data in the extraction phase. Experiment results on a set of variety of images has been realized to assess the performance of the suggested work; the GA ameliorates greatly the imperceptibility without requiring any decrease of the capacity. The PSNR reaches values that rival frequency domain techniques which are usually characterized withhigher PSNR than spatial domain techniques. Compared to literature, we obtain at the same time greater PSNR and larger capacity.

The remaining parts of the document arestructured as follows: in section 2, a review on related works is presented. In section 3, the suggested algorithm is explained. In section 4, we present and discuss our results and compare them with previous works. Section 5 concludes the paper.

2. PROPOSED WORK

To ensure a high quality of the image, the suggested approach takes into account the complexity of the reduced pixels to adapt the length of the hidden message. The long messages are hidden in high complexity zones, while short messages are hidden in smooth zones. This variance is used to evaluate whether a zone is smooth or complex.

In this part, we introduce the interpolation method by bivariate linear box-spline on three directional mesh, which guaranties that the interpolation error E has an approximation order of O(h^2) where h is the mesh step. Based on this interpolation method, a new reversible steganographic approach that hides the secret data in the error between the cover and the interpolated pixels is proposed. To improve the imperceptibility by finding the nearest value to the cover pixel that can dissimulate data, we use optimal pixel adjustment procedure (OPAP) and message adaptive error (GA) methods.

2.1.Interpolation by linear box spline on three directional mesh.

Let Δ be the three directional mesh, which is the uniform triangulation of the plane whose set of vertices is $\mathbb{Z}^2$ and whose edges are parallel to the three directions$e1 = (1,0)$, $e2 = (0,1)$ and $e3 = (1,1)$ as shown in Figure 1. Let us denote $S_n^r(\Delta)$ the space of piecewise polynomial functions of degree n and class C^r defined on Δ. A function fis said to be of class C^rif the derivatives $f', \dots, f^{(r)}$exist and are continuous.

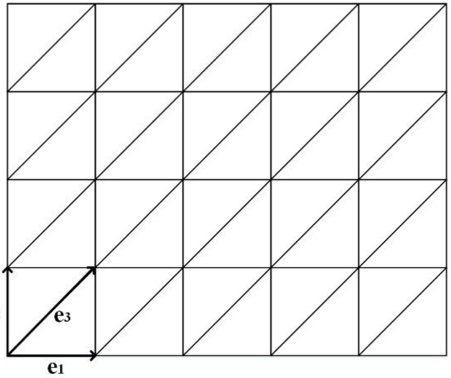


Figure 1: Three directional mesh Δ.

The interpolation problem in the space $S_n^r(\Delta)$ consists of finding the interpolate operator:

$$I = \sum_{\alpha\in\mathbb{Z}^2} \boldsymbol{c}_\alpha \mathbb{B}(. - \alpha)$$

Which interpolates a given function f on the vertices of the triangulation Δ

$$I(i,j) = f(i,j) \qquad (1)$$

Where $\mathbb{B}$ is the B-spline in $S_n^r(\Delta)$ of compactly support on Δ. This problem is equivalent to determine the coefficients$\{c_\alpha\}$, see [7]-[8]-[9]-[10]. In this part, we are studying the spatial interpolation problem$S_1^0(\Delta)$ of bivariate linear spline function on the three directional mesh Δ. For this purpose, we first define the box-spline, which constitutes the foundation of this space.

The Haar bivariate box-spline is defined by the following equation:

$$\mathbb{B}_1(x,y) \coloneqq \chi[0,1)^2(x,y) = \begin{cases} 1 \; if\,(x,y) \in [0,1)^2 \\ 0 \; if \; (x,y) \in \mathbb{R}^2[0,1)^2 \end{cases}$$

From which we see that $\mathbb{B}_1$is a piecewise constant polynomial function; its support is shown in Figure 2(a) It is discontinuous around the boundary of its support and assumes the constant value 1 in the interior of its support $(0,1)^2$ as presented in Figure 2(B)

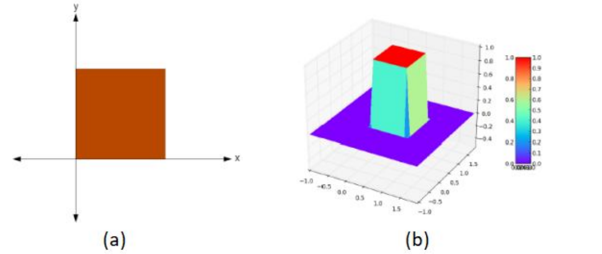


Figure 2:(a): The support of $\mathbb{B}_1$, (b): The Haar box spline $\mathbb{B}_1$.

The hat form depicted in Figure 3(b) is obtained, whereas the support is provided in Figure 3(a).

$$\mathbb{B}_2(x,y) \coloneqq \int_0^1 \mathbb{B}_1(x-t, y-t)dt, \forall (x,y) \in \mathbb{R}^2 (2).$$

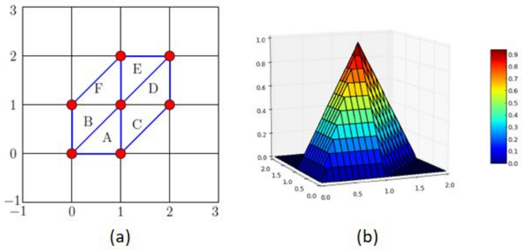


Figure 3:(a): The support of $\mathbb{B}_2$, (b): The Haar box spline$\mathbb{B}_2$.

An explicit computation of the equation above leads to:

$$\mathbb{B}_2(x,y) = \begin{cases} y & if\,(x,y) \in A \\ x & if\,(x,y) \in B \\ 1+y-x & if\,(x,y) \in C \\ 2-x & if\,(x,y) \in D \\ 2-y & if\,(x,y) \in E \\ 1+x-y & if\,(x,y) \in F \\ 0 & otherwise \end{cases} \qquad (3)$$

The family $\mathbb{B}_2(.-\alpha), \alpha \in \mathbb{Z}^2$ constitutes a basement of the space $S_1^0(\Delta)$ of piecewise linear polynomial on the three directional mesh Δ.

Let$h \in \mathbb{R}_+^*, p, q \in \mathbb{N}\{0,1\}$ and let Δ_{pq} be the restriction of the uniform triangulation hΔ on a rectangular domain $\Omega = [0,ph] \times [0,qh]$. Let $V = \{v_{i,j}\} = \frac{(ih,jh)}{i} = 0,\dots,\frac{p}{j} = 0,\dots,q$ be the set of (p+1) × (q+1) interpolation points in the rectangular domain Ω.

Since the Courant Hat box spline $\mathbb{B}_2$ satisfies $\mathbb{B}_2(1,1) = 1$ and $\mathbb{B}_2(i,j) = 0$for all other vertices of Δ, then we obtain an explicit formula of the interpolation operatorI_h.

$$I_h = \sum_{\alpha \in \Delta_{p,q}} f(\alpha)\mathbb{B}_2(.h - \alpha + (1,1)) \qquad (4)$$

Which interpolates a given function f defined on Ω at the points of the set$V$, i.e.

$$I_h(v_{i,j}) = f(v_{i,j}), \quad for\ i = 0,\dots,p\ and\ j = 0,\dots,q. \ (7)$$

Moreover, the interpolation error verifies $||I_h - f||_\infty \leq Kh^2$ where $K$is a constant and$||.||_\infty$is infinity norm (see [8]).Let $X \coloneqq \left\{x_{ij} = \frac{v_{ij}+v_{i+1,j}}{2}, i \neq p\right\}, Y \coloneqq \left\{y_{ij} = \frac{v_{ij}+v_{i,j+1}}{2}, j \neq q\right\}$be the set of midpoints of horizontal and vertical edges respectively and let$Z \coloneqq \{z_{ij} = \frac{v_{ij}+v_{i+1,j+1}}{2}, i \neq p, j \neq q\}$ be the set of the center of all squares in the uniform triangulation Δ_{pq}, as shown in Figure 4.

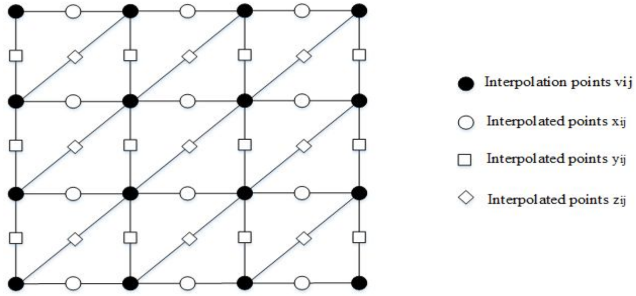


Figure 4: Points and vertices pixels.

By using the equations 3 and 4, we obtain for $i = 0,\dots,p and j = 0,..,q$

$$\begin{cases} I_h(x_{i,j}) = \dfrac{f(v_{i,j}) + f(v_{i+1,j})}{2} \\ I_h(y_{i,j}) = \dfrac{f(v_{i,j}) + f(v_{i,j+1})}{2} \\ I_h(z_{i,j}) = \dfrac{f(v_{i,j}) + f(v_{i+1,j+1})}{2} \end{cases} \qquad (5)$$

In our method, the cover image C represents the signal, which we interpolate$C_{i,j} = f(v_{i,j})$. By using the equation (8), the virtual predicted pixel $I_{i,j}$is obtained by the followingequation:

$$I_{(i,j)} = \begin{cases} \left\lfloor \dfrac{O_{i,j-1} + O_{i,j+1}}{2} \right\rfloor & if\ i = 2p, j = 2q+1 \\ \left\lfloor \dfrac{O_{i-1,j} + O_{i+1,j}}{2} \right\rfloor & if\ i = 2p+1, j = 2q \\ \left\lfloor \dfrac{O_{i-1,j-1} + O_{i+1,j+1}}{2} \right\rfloor & otherwise \end{cases} \qquad (6)$$

2.2 Interpolation error computation

Tang et al. proposed in [11] (2014) a high-capacity reversible steganography (CRS) scheme. They improved the INP scheme by using two values C_{min} and C_{max} in the calculation of the predicted values$I$. These two values are computed as follows:

$$C_{max} = \max\{C_{i,j}, C_{i+1,j}, C_{i,j+1}, C_{i+1,j+1}\}$$
$$C_{min} = \min\{C_{i,j}, C_{i+1,j}, C_{i,j+1}, C_{i+1,j+1}\}$$

The maximum value C_{max} and the minimum value $C_{min}$are used to compute the reference value AD as can be seen in the following equation:

$$AD = \left\lfloor \frac{3 \times C_{min} + C_{max}}{4} \right\rfloor \quad (7)$$

The predicted value$I_{(i,j)}$is calculated using the neighboring pixels and the value AD defined in equation (7)

$$I_{(i,j)} = \begin{cases} \left\lfloor \frac{AD + (O_{i,j-1} + O_{i,j+1})/2}{2} \right\rfloor & if\ i = 2m, j = 2n + 1 \\ \left\lfloor \frac{AD + (O_{i-1,j} + O_{i+1,j})/2}{2} \right\rfloor & if\ i = 2m + 1, j = 2n \\ \left\lfloor \frac{O_{i-1,j-1} + I_{i-1,j} + I_{i,j-1}}{3} \right\rfloor & otherwise \end{cases} \quad (8$$

The differences $E_k(i,j)$ are computed by the comparison between the virtual predicted value $I$and the mean value of $C_{max}$and C_{min}. The formula is given as follows:

$$E_1(i,j) = \begin{cases} C_{max} - I_{(i,j+1)} & if\ I_{(i,j+1)} < \frac{C_{min} + C_{max}}{2} \\ I_{(i,j+1)} - C_{min} & if\ I_{(i,j+1)} \geq \frac{C_{min} + C_{max}}{2} \end{cases}$$

$$E_2(i,j) = \begin{cases} C_{max} - I_{(i+1,j)} & if\ I_{(i+1,j)} < \frac{C_{min} + C_{max}}{2} \\ I_{(i+1,j)} - C_{min} & if\ I_{(i+1,j)} \geq \frac{C_{min} + C_{max}}{2} \end{cases} \quad (9)$$

$$E_3(i,j) = \begin{cases} C_{max} - I_{(i+1,j+1)} & if\ I_{(i+,j+1)} < \frac{C_{min} + C_{max}}{2} \\ I_{(i+1,j+1)} - C_{min} & if\ I_{(i+1,j+1)} \geq \frac{C_{min} + C_{max}}{2} \end{cases}$$

2.3 Genetic algorithm

The GA is a research algorithm based on the mechanisms of natural selection and genetics. The individuals are selected by a "survival of the strongest" strategy with a random but structured exchange of information. This group of individuals is called the population. The characteristics (or variables to be determined) are then used in gene sequences that will be combined with other genes to form chromosomes and then new individuals. Each solution is associated with an individual, and this individual is assessed and classified according to its similarity to the optimal option to the problem.

As in constrained biological systems, the best members from the population are those who have better chanceto reproduce and transfer elements of their genetic heritage to the next generation.

A new population, or generation, is then created by combining the parents' genes. It is expected that some individuals of the new generation will have the best characteristics of both parents, and therefore will be a better solution to the problem. The new group (the new generation) is then subject to the same selection criteria, and then generates its own offspring. This process is repeated several times, until all individuals have the same genetic heritage. Members of this last generation, who are usually very different from their ancestors, have genetic information that corresponds to the optimal solution to the problem.

- *Genetic operators*

The basic GA has three simple operations that are no more complicated than algebraic operations:

- Selection: depending on the quality of the individuals, each is assigned a percentage of chance of being chosen for reproduction, which matches with the relative importance of the quality of the individual in relation to the whole quality of the population.

- Crossover: consists of combining any two individuals (called parents) to produce two other individuals (called children) not necessarily better than the parents; there are several variants of this operator, but in general it consists in cutting two individuals (at the same places in both individuals) into one or more points and exchanging the parts located between these points, the latter are randomly generated.

- Mutation: This operator acts on an individual. It consists in randomly selecting one or more genes (bits) and modifying their values. It is used with a Pm rate, which must be relative to the size of the population.

- *The parameters of the Genetic Algorithm (GA)*

As with pseudo-random methods, the effectiveness of GA depends strongly on their parameters. We briefly present the main parameters of these algorithms as follows:

- The mutation rate or the mutation operator: it is applied with a probability Pm ; if this rate is high, then the research becomes purely random, the population is diversified and the GA loses its effectiveness. If, on the other hand, this rate is low, the population is less diversified and there is also a risk of stagnation. Empirical studies recommend a frequency around one mutation every 1000 bits for good results. Notice that these parameters depend closely on the type of problem to be solved.

-The crossover rate:it is the probability in which a set of individuals is chosen to apply the crossover operator to them. Its choice is generally experimental and its value is very often taken between 0.3 and 0.7. The more the rate is high, the more the population undergoes significant changes. Thus, convergence will be very quick if the crossing rate is close to 1.

- The size of the population: Convergence conditions change with the size of the population; when the population is large, its diversity increases, which reduces convergence towards a local optimum, the execution time of each generation increases, research may be carried out in a redundant way and the overall efficiency of the algorithm is globally affected. Moreover, if the population size is small, then the probability of lingering on local minima is high. The size, depending on the case, is between 25 and 100.

- Stopping condition: Generally, a GA ends after a certain number of generations, but the algorithm can also be terminated when a certain condition is reached, for example when the quality of an individual exceeds a certain threshold

2.4 Proposed steganography Algorithm

In this work, the hidden messageis divided into blocks; the size of each block is calculated based on the interpolation error whose calculus was explained in section 2.2. In most works, data is hidden in the normal path following the interpolated pixels sequentially as they are. However, one can choose smartly a good path that can hide data with the lowest error possible. Finding the best path is defined as an NP-hard optimization problem; the larger are the image's size and the number of blocks, the harder the problem becomes. In this case, using an advanced search algorithm such as GA is a suitable solution. In this proposed research, the role of GA is to find the best path. The fitness function that determines the survival of population is the MSE between the original image (OI) and the stego image. Figure 5depicts in detail the proposed steganography model based on GA.

- *Embedding Algorithm*

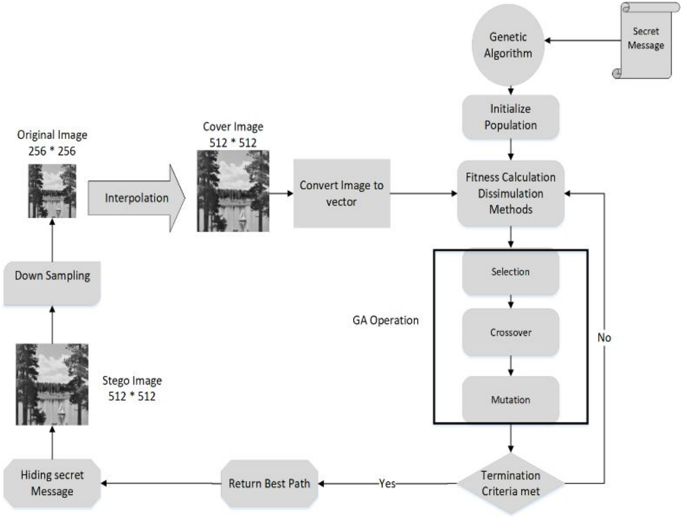


Figure 5: Proposed embedding Algorithm.

The steganographic model developed in this paper can be described in the following steps:

Input: OI (Original Image), Secret Data (Message to embedding)

Output: S (Stego Image)

Begin:

Step 1: Read the OI as a matrix.

Step 2: Using our interpolation method, the image C is constructed, its size is the double of the OI's.

Step 3: Transform the imageC into a vector so that we can easily browse paths, the browsing is done in only one dimension instead of two dimensions.

Step 4: Generate a population of individuals of size : $x_1, x_2, \ldots, x_N$. The individuals are paths that indicate the order we can follow to dissimulate data into the interpolated pixels. Therefore, N is chosen so that there are enough paths, thus we can have a very good one to choose.

Step 5:Calculate the chances of survival of all individuals x_i. Each path has its own error of dissimulation MSE; the smaller the MSE of a path (individual) x_i, the bigger is his chance of survival.

Step 6: We keep half the population that have the best survival chances.

Step 7: Choose a pair of individuals for reproduction (according to each individual's chances of survival).

Step 8: According to the probabilities associated with each genetic operator, apply these operators.

Step 9: Place the produced individuals in the new population.

Step 10: Check if the size of the new population is correct. If not, return to step 7.

Step 11: Replace the old population of individuals with the new one.

Step 12: Return to step 5.

Step 13: When the best path is found, associate it to a key which is appended to the secret message.

Step 14: Hide the secret message following the best path found to create stego image.

Step 15: Increase imperceptibility of image with OPAP.

End

- *Extraction Algorithm*

In order to retrieve the hidden message, let us take as input the image-stego.We will use the corresponding extracting process, which is described as follows:

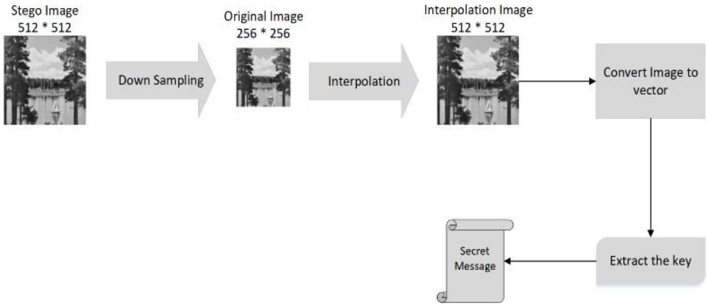


Figure 6: Proposed Extracting Algorithm.

Input: S (Stego Image)

Output: Secret Data

Begin:

Step 1: Read the stego image as a matrix.

Step 2: using our Interpolation Method, we will construct the image C.

Step 3: Transform the image C into vector to easily retrieve the best path chosen in embedding phase.

Step 4:Extract the key to find the path in order to correctly extract the secret data.

Step 5: Extract the data following the path.

End

- *Example of embedding and extracting*

In this section, we present an example to explain the proposed steganography algorithm. LetImgbe the cover Image of size 4x4 and let msg be the secret message to be hidden inImg.

Input:

Img=

186	184	186	201
175	193	205	175
157	175	175	193
178	184	165	175

We convert the image Img to vector C as follows:

C= {186, 175, 157,178, 184, 193, 175, 184, 186, 205, 175, 165, 201, 175, 193, 175}

Msg=

0	1	0	0	1	0	1	1

Step 1: The identifiers of the individuals (paths) are {8, 7, 6, 5, 4, 3, 2, 1}

Step 2: We choose randomly 4 individuals {7, 4, 2, 8}.

Index 7 ={14, 8, 1, 9, 15, 13, 3, 10, 7, 6, 11, 4, 2, 12, 16, 5}
Index 4={8, 15, 12, 6, 9, 10, 2, 16, 5, 4, 11, 13, 1, 3, 7, 14}
Index 2= {2, 13, 15, 7, 10, 9, 6, 5, 4, 1, 14, 12, 3, 8, 11, 16}
Index 8= {6, 5, 8, 16, 7, 10, 9, 4, 12, 11, 15, 14, 13, 3, 1, 2}

Step 3:We hide msg in C using the chosen individuals. Then, we compute the corresponding MSE to obtain the survival chance of everyone. We obtained the following:

Stego 7= {186, 175, 157,178, 184, 193, 175, 184, **187**, 205, 175, 165, 201, **173**, 193, 175}
Stego 4= {186, 175, 157,178, 184, **195**, 175, **185**, 186, 205, 175, **166**, 201, 175, **192**, 175}
Stego 2= {186, **173**, 157,178, 184, 193, 175, 184, 186, 205, 175, 165, **200**, 175, **194**, 175}
Stego 8= {186, 175, 157,178, 184, 193, 175, **186**, 186, 205, 175, 165, 201, 175, 193, 175}

Index	MSE
7	5
2	6
4	7
8	4

Step 4: To produce the next generation, we use one of the GA operators (for example: Mutation) on individuals 7 and 4; the child (resulting individual) will replace individual 4. We do the same to individuals 8 and 2 using for example Crossover this time. Hence, we obtain the following generation

New generation
8
7
f(8,2)=child
F(7,4)=child

Step 5: If at least one individual meets the termination criteria (an individual provides an MSE ≤ Threshold or if we are in the last iteration), we stop and obtain the elected individual; else, we return to the step 3.

2.5 Fitness function

The objective of applying (OPAP) is to minimizethe error between the OI and the stego image S. To dissimulate data, we calculate the error of interpolation as in CRS which was defined in the equation (9). When a message of $L_k(i,j)$ bit size is hidden by substitution into the interpolated pixel $I_{i,j}$, the message occupies the $L_k(i,j)$ least significant bits of the resulting pixel that we call $S_{i,j}^{(1)}$. The idea is to edit the bits of $S_{i,j}^{(1)}$ that do not include the data (weights above $L_k(i,j)$) to produce a closer value to the initial pixel $O_{i,j}$. This can be achieved by adding $\pm 2^{L_k(i,j)}$ to $S_{i,j}^{(1)}$; we obtain two new pixels named$S_{i,j}^{(2)}$ and $S_{i,j}^{(3)}$, they are calculated as follows: $S_{i,j}^{(2)} = S_{i,j}^{(1)} + 2^{L_k(i,j)}$ and $S_{i,j}^{(3)} = S_{i,j}^{(1)} - 2^{L_k(i,j)}$ The corresponding stego pixel $S_{i,j}$ is determined among $S_{i,j}^{(1)}, S_{i,j}^{(2)} andS_{i,j}^{(3)}$ by choosing the closest one to $O_{i,j}$.

$$S_{i,j} = S_{i,j}^{(r)} withr = \text{argmin}_{k=1,2,3}(|O_{i,j} - S_{i,j}^{(k)}|)$$

3. EXPERIMENT RESULTS

3.1 Evaluation of the proposed embedding Algorithm

In this section, we use the standard images shown in Figure 7. They are 512 × 512 8-bit gray-scale images, which are mostly used in tests. We conceal texts of different sizes into these images; then, we evaluate the performance based on the metrics: Capacity and Imperceptibility.

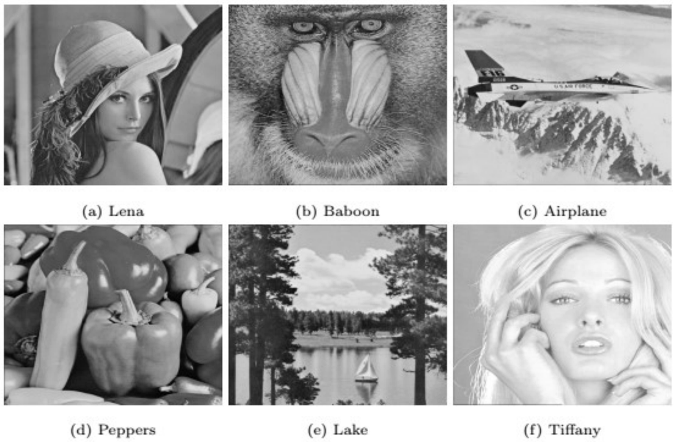

(a) Lena (b) Baboon (c) Airplane
(d) Peppers (e) Lake (f) Tiffany

Figure 7: Six 512×512 grayscale testing images.

- *Imperceptibility and capacity test*

The capacity of embedding is the amount of secret information that we can embed into the C with no harmful modifications. It is computed by the number of bits we can dissimulate in one pixel (bit per pixel bpp).

The imperceptibility is when we are unable to differentiate between the C and the stego image. It is evaluated by several parameters, such as the Peak Signal to Noise Ratio (PSNR), Image Fidelity, Normalized Correlation Coefficient and Q-index.

When the PSNR is beyond 30 dB, the human eye is unable to differentiate between the cover and stego images. This metric is calculated as follows:

$$PSNR = 10\log_{10}\frac{(255)^2}{MSE}\ db$$

Where $MSE = \frac{1}{512^2}\sum_{i=1}^{512}\sum_{i=1}^{512}(C_{i,j} - S_{i,j})^2$

The (IF) measures the ratio of the error energy to the C energy; its value is between 0 and 1, the more it is close to 1, the better is the imperceptibility. It is calculated as follows:

$$IF = 1 - \frac{\sum_{i=1}^{M}\sum_{j=1}^{N}\left(S_{i,j} - C_{i,j}\right)^2}{\sum_{i=1}^{M}\sum_{j=1}^{N}C_{i,j}^2}$$

The Normalized Correlation Coefficient (NCC) and the Q-index are used to measure the similarity of the cover and stego image. These two parameters are always between -1 and 1. The closest they are to 1, the more similar are the images. When they are close to 0, the images are said uncorrelated. They are calculated as follows:

$$NCC = \frac{\theta_{cs}}{\theta_c\theta_s}, Q = 4\ \frac{\theta_{cs}}{{\theta_c}^2 + {\theta_s}^2}\frac{\eta_{cs}}{{\eta_c}^2 + {\eta_s}^2} \quad (10)$$

The parameters η_c and η_s are the mean values of the cover andstego images respectively, andθ_c, θ_s and $\theta_{cs}$are given by the following expressions

$$\theta_c = \sqrt{\sum_{i=1}^{M-1}\sum_{j=1}^{N-1}\left(C_{i,j}-\eta_c\right)^2}\ ,\theta_s = \sqrt{\sum_{i=1}^{M-1}\sum_{j=1}^{N-1}\left(S_{i,j}-\eta_s\right)^2}$$

$$\theta_{cs} = \sum_{i=1}^{M-1}\sum_{j=1}^{N-1}\left(C_{i,j}-\eta_c\right)\left(S_{i,j}-\eta_s\right)$$

We accomplish a comparison to the NMI, INP and CRS methods, the earliest interpolation methods used in steganography. Tables 1 and 2 show the values of the capacity and the metrics of the imperceptibility by the NMI [1], INP [2], CRS [11],IBLBS [12] and the proposed GA. We used the images Lena(a), Mandrill(b), Airplane(c), Peppers(d), Lake(e) and Tiffany(f), they have different level of texture which is a crucial parameter in steganographybased on interpolation.

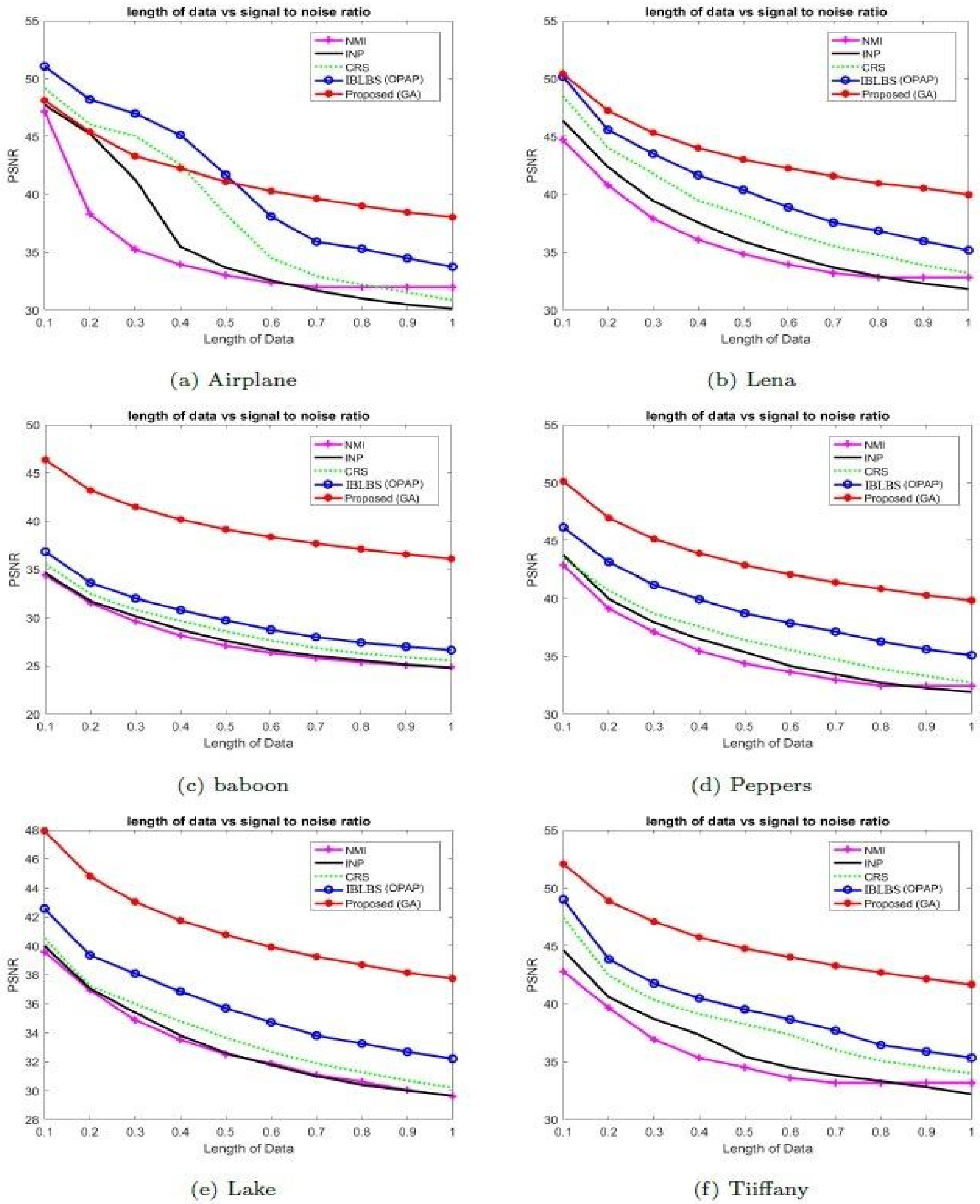
Table 1: The comparison of capacity (bpp) and PSNR between NMI, INP, CRS, IBLBS and proposed approach based on (GA).

Image	Metric	NMI	INP	CRS	Proposed	
					IBLBS	GA
a	PSNR Bpp	34.84 0.76	35.94 1.27	38.25 1.76	40.38 1.81	43.01 1.81
b	PSNR Bpp	27.08 1.62	27.60 2.32	28.58 2.89	29.71 2.94	39.14 2.94
c	PSNR Bpp	33.02 0.68	33.66 1.12	38.26 1.58	41.68 1.64	41.10 1.64
d	PSNR Bpp	34.38 0.74	35.37 1.27	36.38 1.78	38.71 1.84	42.88 1.84
e	PSNR Bpp	32.50 1.04	32.59 1.64	33.65 2.17	35.70 2.22	40.77 2.22
f	PSNR Bpp	34.50 0.64	35.44 1.09	38.23 1.57	39.52 1.64	44.76 1.64

Table 2: The comparison of the IF, NCC and Q-index between NMI, INP, CRS, IBLBS and proposed approach based on (GA).

Image	Metric	NMI	INP	CRS	Proposed	
					IBLBS	GA
a	IF NCC Q-index	0.99881 0.99545 0.99543	0.99908 0.99644 0.99644	0.99946 0.99796 0.99794	0.99967 0.99873 0.99873	0.99963 0.99858 0.99858
b	IF NCC Q-index	0.9931 0.96386 0.96371	0.99381 0.96803 0.96801	0.99504 0.97490 0.97484	0.99619 0.98053 0.9805	0.99913 0.99517 0.99516
c	IF NCC Q-index	0.99881 0.99061 0.99060	0.99888 0.99117 0.99116	0.99953 0.99633 0.99633	0.99977 0.99821 0.99821	0.9997 0.99774 0.99773
d	IF NCC Q-index	0.99767 0.99307 0.99305	0.99758 0.99281 0.9928	0.99810 0.99442 0.99440	0.99897 0.99696 0.99695	0.99962 0.99887 0.99887
e	IF NCC Q-index	0.99744 0.99408 0.99408	0.99728 0.99374 0.99373	0.99782 0.99504 0.99501	0.99867 0.99695 0.99694	0.99945 0.99872 0.99872
f	IF NCC Q-index	0.99942 0.98465 0.98464	0.99949 0.98672 0.98670	0.99972 0.99281 0.99280	0.99982 0.99527 0.99527	0.99990 0.99794 0.99794

Results in tables 1 and 2 indicate that the proposed approaches provide the larger capacity; and at the same time, they have the highest imperceptibility. We proposed to interpolate by the average of the neighboring pixels horizontally, vertically and diagonally to minimize the error between the interpolated and the original pixels; this gives a better image quality. Furthermore, the GA finds the best path to follow during the dissimulation; it allows hiding each fragment of data in the most suited interpolated pixel in order to achieve a minimal error, thus a better imperceptibility. This explains how the GA could ameliorate remarkably the PSNR values of our precedent work [12]. The values of IF, NCC and Q-index in table 2 obtained by the proposed approaches are closer to their optimal value 1 than the three other methods. For thecapacity, we calculate the error of interpolation the same way as CRS. But, still the capacity of the proposed work surpasses CRS's; this difference is due to the utilized interpolation technique. The interpolation in CRS is based on the minimum of the four reference pixels; while in the suggested approach, the interpolation is the average in the three directions, which provides a greater error of interpolation, thus a larger space to hide data.



(a) Airplane

(b) Lena

(c) baboon

(d) Peppers

(e) Lake

(f) Tiiffany

Figure 8: Comparison of all interpolation methods with the two proposed approaches over six standard testing images.

Figure 8 shows the evolution of the PSNR as we increase the size of the hidden message from 0.1 bpp to 1bpp for the six test images shown before in Figure 7. As can be seen in Figure 8, the proposed work using GA offers better trade-off between the capacity and PSNR. When the C has a lot of texture, the resulting interpolation error is consequently big; Therefore, the correction by GA is more important. This explains why the difference of PSNR in images Mandrill and Lake are very big, these two images contain a lot of contours.

- *Security test*
-

To evaluate the security of the proposed steganographic scheme, we use the statistical attack χ^2[13]. It is based on the hypothesis that the dissimulation process does not change significantly the sumof the frequencies of each pair of consecutive pixels $(2i, 2i+1)$ in the histogram; and the frequency of $2i$ and $2i+1$ tends to the average of their initial values. Let $\nu_{2i}$and ν_{2i+1}be the frequencies of $2i$ and $2i+1$ in the histogram respectively. Their average is $\eta_i = \frac{\nu_{2i}+\nu_{2i+1}}{2}$the χ_n^2 statistic with n degree of freedom is then calculated as follows:$\chi_n^2 = \sum_{i=0}^{n+1} \frac{(\nu_{2i}-\eta_i)^2}{\eta_i}$

The degree of freedom is the number of pairs $(2i, 2i + 1)$having a sum of frequenciesgreater than 4, i.e.

$$n = card\{(2i, 2i + 1); 0 \leq i \leq 127; v_{2i} + v_{2i+1} > 4$$

Then, the probability of attackingis computed as follows:

$$p = 1 - \frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)}\int_0^{\chi_n^2} e^{-\frac{t}{2}} t^{\frac{n}{2}-1}\,dt$$

While Γ is defined by$\Gamma(x) = \int_0^{+\infty} t^{x-1}e^{-t}dt$

Table 3 shows the PSNR and χ^2 values of the same six images. To measure the contribution of the GA, only classicLSB substitution is used, as it's the simplest way of insertion; we mask data by substituting the LSBs of the interpolated pixels. Results are shown in the column LSB for both PSNR and χ^2 test. Clearly, the PSNR of the LSB substitution is the lowest, followed by the PSNR of the procedure OPAP, this procedure ameliorates the PSNR of LSB. However, the PSNR of the proposed algorithm is remarkably higher; this proves the effectiveness of the proposed GA and IBLBS; the GA elects the best path that allows to obtain a small MSE.

Table 3: χ^2results on six usual images after using OPAP and GA.

Metric	**PSNR**			χ^2**[13]**		
Image	LSB	GA	OPAP	LSB	GA	OPAP
a	37.2916	43.0120	40.3762	4.1481e-10	0	6.5503e-15
b	27.2397	39.144	29.7128	0.5239	0	2.2922e-04
c	38.6339	41.0968	41.6837	0	0	0
d	35.5388	42.877	38.7154	6.0263e-13	0	0
e	32.7401	40.774	35.6969	4.8950e-13	0	0
f	36.7134	44.757	39.5246	0	0	0

The χ^2 results in table 3show a very low rate of detectability; the proposed work is not detected by this attack. The changes of the histogram are not predictable by χ^2.

3.2 Comparison to literature

In table 4, the proposed work is compared to schemes developed in [14], [15], [16] and [17]. The proposed algorithm provides the largest capacity of embedding with the highest values of PSNR.

Table 4: Comparison of our proposed work with eight data hiding methods basing on image quality and capacity.

Methods	Metric	Lena	Mandrill
Ou et al. [14]	PSNR Capacity	35.3729 1.0954	38.9982 0.5321
Wang et al. [16]	PSNR Capacity	48.6747 0.2716	48.9441 0.0952
Pei et al. [15]	PSNR Capacity	26.8483 1.8388	21.4183 1.2496
Jie et al. [17]	PSNR Capacity	34.6967 1.6945	30.1452 1.8483
Arham et al.[4]	PSNR Capacity	42.06 0.74	31.68 0.72
wang et al.[5]	PSNR Capacity	42.89 0.52	32.48 0.49
Benhfid et al.[12]	PSNR	40.3762	29.7128
	Capacity	1.8109	2.9432
wang et al. [6]	PSNR	32.01	31.33
	Capacity	1.77	0.61
Proposed (GA)	PSNR	43.0120	39.144
	Capacity	1.8109	2.9432

The proposed work surpasses the work developed by Ou et al. [14] in both PSNR and the capacity. The difference in PSNR is significant for the images Lena and Airplane, respectively 8 and 7 dB. Meanwhile, the capacity of the given algorithm is respectively 1.6 times and 1.3 times larger. While the PSNR values are fairly equal, the capacity of the image Mandrill is nearly 6 times higher. The PSNR obtained by Wang et al. [16] is constant for the three images with a significantly small capacity of hiding. It is higher than the PSNR we obtain with 5 dB, 9 dB and 7 dB but our capacity is respectively 6, 30 and 5 times greater than theirs. In comparison to Pei et al. [15], we obtained a comparable capacity for Lena and Airplane; but the difference in PSNR is very big, 16 dB and 14 dB. For Mandrill, the difference in PSNR is 18 dB and the capacity is more than twice greater. In Jie et al. [17], while the capacity increases with only 0.2 bpp from results, the PSNR decreases with 4 dB. In [4]-[5], the PSNR value is almost the same for the image Lena, but our capacity is around 2,5 times larger than theirs; for the image Mandrill, both results are bigger. In [6], for both images, both the PSNR and capacity of the proposed work are greatly larger. On the contrary of these works, the proposed algorithm exploits the regularity of the images with a lot of texture. The difference in the capacity for the image Mandrill is very big. The more texture, the greater is the error of interpolation; therefore, the available space to conceal data becomes larger. Results in this table indicate that the proposed work gives better compromise between the PSNR and the capacity of hiding than the other works. We obtain around 40 dB in PSNR for almost 2bpp capacity; and even for 3 bpp, the PSNR is still high (39 dB).

4. CONCLUSION

In this paper, we introduced new data hiding method that allowed us to increase the capacity while keeping the imperceptibility at a good level. Depending on the image's complexity, the adopted interpolation technique and the method utilized to calculate the interpolation error enabled us to obtain a good capacity. Results show that the capability and PSNR of the proposed approaches whether using OPAP or

GA are greater than NMI, INP and CRS algorithms. Moreover, the χ^2test results prove the undetectability of the proposed algorithm. Therefore, these substantial performance improvements demonstrate the efficiency of the suggested approach. Compared to steganographic schemes based on wavelet transform, wavelets prohibit the concealment of data in the approximation bloc; it reduces the hiding space by a quarter. However, the imperceptibility is good. The suggested method provides a reduced imperceptibility than wavelet-based solution but better capacity. In our future works, we will focus on strengthening the robustness of our work and study its resistance to attacks that try to destroy the information hidden inside the stego image.

REFERENCES

1. K.Jung, K.Yoo. **Data hiding method using image interpolation**, Computer Standards and Interfaces. 31, 465-470, 2009.
2. C.F. Lee,Y.L. Huang. **An efficient image interpolation increasing payload in reversible data hiding**, Expert Syst Appl. vol. 39, 6712-6719, 2012.
3. S. Dogan, **ArtifIntel**l Rev, 46: 129. 2016. https://doi.org/10.1007/s10462-016-9459-9.
4. Arham, Aulia, H. A. Nugroho, and T. B. Adji. **Multiple layer data hiding scheme based on difference expansion of quad**.Signal Processing 137:52-62, 2017. doi: 10.1016/j.sigpro.2017.02.001.
5. J. Wang, J. Ni, X.Zhang, and YQ. Shi. **Rate and Distortion Optimization for Reversible Data Hiding Using Multiple Histogram Shifting**. IEEE Transactions on Cybernetics 47.2:315.2017.
6. W. Wang, J. Ye, T. Wang and W. Wang. **A high capacity reversible data hiding scheme based on right-left shift**, Signal Processing, 2018. doi: 10.1016/j.sigpro.2018.04.008.
7. M. Unser, Splines. **A perfect fit for signal and image processing**, IEEE Signal Processing Mag, Nov. vol. 16. no. 6, pp. 22-38. 1999.
8. C. De Boor, K.Höllig, S.Riemenschneider, **Box Splines**, Springer-Verlag, New York, Berlin. 1993.
9. C. De Boor, K. Höllig, S. Riemenschneider, **Bivariate Cardinal Interpolation by Splines on a Three-Direction Mesh**, Illinois journal of mathematics, vol. 29, no. 4, pp. 533–566. 1983.
10. C.K. Chui, **Multivariate Splines**, Soc. for Indust. & Appl. Math, Philadelphia. 1988.
11. M.W. Tang, J. Hu, W. Song, **A high-capacity image steganography using multi-layer embedding**, Optik. 125, 3972-3976. 2014.
12. A.Benhfid, et al. **Reversible steganographic method based on interpolation by bivariate linear box-spline on the three directional mesh**. Journal of King Saud University – Computer and Information Sciences, 2018. https://doi.org/10.1016/j.jksuci.2018.09.016.
13. A.Westfeld and A.Pfitzmann. **Attacks on Steganographic Systems**, Lecture Notes in Computer Science in International Workshop on Information Hiding. 1768, 61-76. 1999.
14. B. Ou, Y. Zhao, R. Ni, **Reversible watermarking using optional prediction error histogram modification**, Neurocomputing. 93, 67-76. 2012.
15. Q. Pei, X. Wang, Y. Li, H. Li, **Adaptive reversible watermarking with improved embedding capacity**, J SystSoftw. 86, 2841-2848. 2013.
16. X.T. Wang, C.C. Chang, T.S. Nguyen, M.C. Li, **Reversible data hiding for high quality images exploiting interpolation and direction order mechanism**, Digital Signal Process. 23, 569-577. 2013.
17. J. Hu, T. Li, **Reversible steganography using extended image interpolation technique**, Computers and Electrical Engineering. vol. 46. 2015.
18. N. Ahmad@Mohd Din, R. H. Salimin, I. Musirin, N. M.S. Honnoon, N. Aminuddin.**Solving Economic Load Dispatch for Power Generation Using Genetic Algorithm Techniques**.International Journal of Advanced Trends in Computer Science and Engineering. Volume 8, No.1.3, 2019 https://doi.org/10.30534/ijatcse/2019/6181.32019
19. Marilou O. Espina, Arnel C. Fajardo, Bobby D. Gerardo, and Ruji P. Medina. **Multiple Level Information Security Using Image Steganography and Authentication**, International Journal ofAdvanced Trends in Computer Science and Engineering, Vol 8, No.6, November – December 2019. https://doi.org/10.30534/ijatcse/2019/100862019.