Selection of Image Blocks using Genetic Algorithm and Effective Embedding with DCT for Steganography

Shweta Joshi
PG Student
Department of Computer Engineering
St. Francis Institute of Technology
Mumbai-103, India
shweta.joshi.sweet@gmail.com

Dr. Kavita V. Sonawane
Professor, HOD
Department of Computer Engineering
St. Francis Institute of Technology
Mumbai-103, India
kavitavinaysonawane@gmail.com

ABSTRACT
In today's digital world that we live in, security of information is crucial in various communication applications that are widely developed. Steganography is one of the highly secure information hiding techniques. It provides invisible communication and hides the existence of information. This paper focuses on 'before embedding technique' of hiding in image steganography by trying to find suitable places in cover image to embed the secret image. Genetic algorithm (GA) is applied to identify appropriate places in cover image where embedding of secret image will cause minimum distortion. After obtaining these places, embedding is performed using transform domain technique Discrete Cosine Transform (DCT). The secret image is first normalized and then embedded in the lower energy DCT blocks of the selected cover image regions. The experimental results show that the stego images obtained from the proposed method have less visual distortion with satisfactory values in parameters like MSE, PSNR and Correlation used for performance evaluation.

Keywords
Steganography; Genetic Algorithm; DCT; secure communication; Stego Image

1. INTRODUCTION
Now that the internet has become an important and essential tool, the security of information technology and communication has gained utmost importance [1]. Various techniques like cryptography, watermarking, steganography are used to secure confidential information on Internet. Cryptography is the technique of scrambling the intelligible data and making it unintelligible so that people other than the intended recipient will not be able to read the message. In watermarking, the copyright information of a digital media file is identified by inserting pattern of bits into the file [2].

In steganography, the secret information is hidden in seemingly innocent media. The word steganography is the combination of the two Greek words "stegos" meaning "cover" and "grafia" meaning "writing", which defines it as "covered writing" [3]. Steganography hides the secret data inside the cover medium. In steganography, the intended secret message does not attract attention as an item of scrutiny. It is the art and science of sending a secret message by concealing it in such a way that others cannot make out that something is hidden. The main objectives of steganography [4] are:

1. To protect confidential information from intruders
2. To increase data embedding capacity
3. To ensure highest security.

To hide the confidential information there are different types of steganography based on the carriers used like text, image, video, audio and network steganography. Out of these, image steganography is most common due to the advantage that the digital images have high amount of redundant information which can be replaced by secret data without causing much distortion to the image. One of the most important things in designing an image steganographic system is the invisibility factor, which means the human eyes should not be able to distinguish the difference between the cover and stego image.

There are four basic components of image steganography:

1. Cover image: It is the carrier which is available for everyone and the secret data is hidden in it.
2. Secret data: It is the secret information to be embedded into the cover image and can be recovered using the key.
3. Key: It is used for embedding and extraction purpose of the secret data. The key is shared only between the sender and the receiver so that any third party person cannot retrieve the secret data without having the proper key.
4. Stego image: It is the carrier with embedded information. It is sent to the receiver through a communication network.

Two types of techniques can be applied for increasing the robustness of image steganography: 'Techniques before embedding' and 'Techniques after embedding'. In the first phase, care is taken before the embedding is done so that the basic structure of the image is not altered too much. In the second phase, after embedding data, some techniques are used to change the statistical features of pixels' blocks, so that RS is not able to detect the existence of message [5]. According to the previous studies, it has been observed that GA has been largely employed in second phase to increase the robustness of secret messages against steganalysis (detection of covered data in carrier media) [6]. In this paper, we focus on first phase by trying to identify suitable places in cover image for embedding of secret data by employing genetic algorithm for it.

The rest of the paper has been organized as follows: Section 2 presents the literature survey, Section 3 deals with proposed work,

ACM COMPUTE '16, October 21-23, 2016, Gandhinagar, India
 ISBN 978-1-4503-4808-9/16/10..$15.00
DOI: http://dx.doi.org/10.1145/2998476.2998495

in Section 4 experimental results are discussed and analyzed and Section 5 gives the conclusion.

2. LITERATURE SURVEY

This section discusses image steganography techniques and Genetic Algorithm in steganography.

2.1 Types of Image Steganography

Image steganography can be broadly classified into two domains namely spatial domain technique and transform domain technique.

2.1.1 Spatial Domain Techniques

In spatial domain, the cover image is modified to hide the secret data by directly replacing the values of its pixels. This technique is widely used in data hiding for its simplicity, high payload capacity and high efficiency. The common technique in this domain is Least Significant Bit (LSB) [7]. In LSB, the secret message bits are directly substituted in the LSBs of the cover image. Other spatial domain techniques are PVD, GLM, OPAP, EMD and DE. In Pixel Value Differencing (PVD) [8] method, the image is divided into non-overlapping blocks after which the variance between the pixel values is computed and then modification to the variance is made for embedding data. The Gray Level Modification (GLM) [9] is used to map data by modifying the gray level of the image pixels. The Optimal Pixel Adjustment Process (OPAP) [10] enhances the stego image quality after the data embedding is done. The methods based on pixel pair matching are Exploiting Modification Direction (EMD) [11] and Diamond Encoding (DE) [12]. DE is the extension of EMD embedding scheme.

2.1.2 Transform Domain Techniques

In transform domain, the cover image is converted from spatial domain form to transform domain form using certain transforms. Discrete Fourier Transform (DFT), Discrete Cosine Transform (DCT), Discrete Wavelet Transform (DWT), Lempel–Ziv–Welch (LZW) Transform and Contourlet Transform are some widely used transform techniques. Transform domain techniques are stronger than spatial domain techniques. In this approach, the magnitude of the transform coefficients is changed to hide the secret data. In DFT, secret data is hidden by altering the values of Fourier coefficients. DWT uses wavelets for embedding the data. LZW is a lossless data compression algorithm and is widely used in GIF and TIFF image format [13]. Contourlet Transform is a two-dimensional transform method for image representation in which the contourlet coefficients that have larger magnitude correspond to the edges in the image [14].

2.2 Genetic Algorithm in Steganography

The Genetic Algorithm (GA) is a population-based search algorithm. The process of natural selection is imitated by GA. GA is the application of principle of survival of the fittest on a population of potential solutions. GA is an iterative procedure in which better approximations to a solution are produced successively. During each generation (generation is the temporal increment), the fitness of population (as domain solutions) is calculated. Based on these evaluations, a new population of candidate solutions is generated. Specific genetic operators like selective reproduction, crossover and mutation are applied to generate the new population.

Ronak Karimi *et al.* [5] proposed an approach which focuses on the before embedding technique. This approach uses GA to find suitable places in cover image for embedding the secret message based on LSB substitution that will cause lesser changes in original cover image. Shen Wang *et al*. [15] propose a steganography method which is based on genetic algorithm for ensuring the security against RS analysis. Stego image is secured by modifying the pixel values using GA to keep their statistic characters as it is. P. M. Siva Raja and E. Baburaj [16] have proposed LSB Matching Revisited (LSBMR) using Genetic Algorithm. In this, the embedding regions are selected according to the size of the secret message using the GA. The GA is also used to optimize the threshold value of the selected image regions. Elham Ghasemi *et al.* [17] have employed a GA based mapping function to embed data in discrete wavelet transform coefficients in 4*4 blocks of the cover image. After embedding the message, optimal pixel adjustment process (OPAP) is applied to minimize the error between the cover and stego image. H. R. Kanan and Bahram Nazeri [18] present a GA based approach to find best starting point in host image for hiding secret image such that the Peak Signal-to-Noise Ratio (PSNR) of the stego image is maximized.

3. PROPOSED WORK

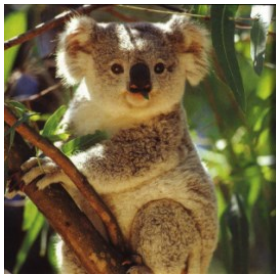
The method presented in this study tries to find out proper places in the cover image using GA where embedding of secret image will cause least changes in the cover image. The embedding is done in lower energy blocks of transformed (DCT) cover image. Low energy blocks have high frequency. Embedding of secret data in high frequency regions will lead to less distortion in cover image. High frequency patches in cover image suitable for embedding could be found directly by applying transform domain techniques. But application of GA helps to identify and analyze the fine image details and contributes to the quality of stego image produced.

The role of genetic algorithm in cover image block selection, embedding algorithm and extracting algorithm are described below.

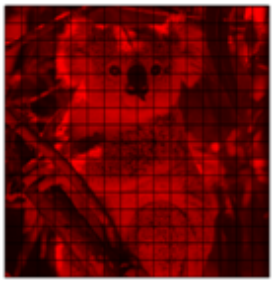
3.1 Role of Genetic Algorithm in Proposed System for Cover Image Block Selection

GA is an optimization algorithm which tries to find the optimal solution for a problem in hand. The aim of the proposed method is to find best optimal blocks in cover image for effective embedding of secret image. The fitness function in the proposed method is designed in such a way that image quality attribute like texture variations are analyzed before embedding to select the image blocks where embedding of secret information will generate good quality stego image. The GA applied for identifying optimal cover image blocks is as below:

Take the cover image and separate it into R-G-B planes. Consider one plane at a time for GA execution. Let's consider an example. The sample cover image is as shown in Figure 1(a). Now, consider R-Plane of the cover image, divide it into 16 × 16 non-overlapping blocks as shown in Figure 1(b). Thus, 256 blocks each of 16 × 16 pixels are obtained as shown in Figure 1(c).Then, chromosomes are encoded as an array of 256 genes containing pixel numbers of each 16 × 16 pixel cover image R-Plane blocks. Hence, we get total 256 chromosomes as our initial population for GA. Our GA aims at minimizing the standard deviation [19] of chromosomes (image blocks) because embedding in lower standard deviation blocks will not lead to much distortion. Thus, the definition of fitness function will be as shown in equation (1),



(a)



(b)

1	2	3	4	.....	14	15	16
17	18	19	20	.....	30	31	32
33	34	35	36	.....	46	47	48
49	50	51	52	.....	62	63	64
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
209	210	211	212	.....	222	223	224
225	226	227	228	.....	238	239	240
241	242	243	244	.....	254	255	256

(c)

Figure 1: (a) Sample Cover Image (b) 16 × 16 Blocks of R-Plane (c) Numbered Grid of 256 Blocks of R-Plane

$$\text{Standard Deviation } (\sigma) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2} \quad (1)$$

where $\{x_1, x_2, \ldots, x_N\}$ are the pixel values in the cover image blocks and μ is the mean value of the pixels. N is the size of the chromosome.

The next step is associated with formation of the second generation based on selection processes via genetic operators like selective reproduction, crossover and mutation. A pair of best parents is selected for reproduction of every individual based on their fitness. The contents of these two chromosomes are interacted for production of two newborn chromosomes. This interaction process is called crossover where the contents of the two parent chromosomes are exchanged to produce the newborn. In addition, mutation is applied during each process to maintain genetic diversity in population. Mutation refers to alteration of one or more gene values in a chromosome from its initial state. After the execution of Genetic Algorithm is completed, a population of optimal R-Plane image blocks is obtained. These blocks are then sorted from low to high standard deviation values. The first 150 blocks of lower standard deviation (STD) from these sorted blocks are selected for further computation. DCT is applied to these 150 blocks and their energy is computed. Energy of each block is computed as the summation of square of coefficients within that block. Energy [20] of a DCT block is computed as shown in equation (2),

$$E = \sum_{i=1}^{n} d^2 \quad (2)$$

where d is the DCT coefficient value and n is the number of DCT coefficients in the block.

DCT blocks are then arranged from low to high energy. Then, the lower energy DCT blocks are considered for secret image embedding. For example, if secret image is of size 128 × 128, then 64 lower energy blocks will be selected in which embedding will be done. The number of blocks to be selected can be considered according to the secret image size. The same procedure is applied for G-Plane and B-Plane to select the cover image blocks for embedding.

3.2 Embedding Algorithm

1. GA is applied on each plane (R-G-B) of the cover image as described in Section 3.1 to obtain the suitable blocks of cover image where the secret image can be embedded securely.
2. Separate secret image into R-G-B planes. Normalize each plane of secret image (find the maximum pixel value of the secret image and then divide each value of secret image by its maximum pixel value). Divide these normalized planes into non-overlapping blocks. The size of secret image blocks is same as cover image blocks.
3. Replace the DCT coefficients of selected lower energy blocks of transformed cover image with the normalized secret image blocks. The R-plane, G-plane and B-plane of normalized secret image are embedded into R-plane, G-plane and B-plane of transformed cover image respectively.
4. Apply inverse transform (inverse DCT) to the blocks and index the blocks.
5. Reconstruct the image using R-G-B planes with embedded secret message to get back the cover image. This gives the stego image.

3.3 Extracting Algorithm

1. Separate the stego image into R-G-B planes. Then divide each plane into non-overlapping blocks.
2. Based on indexing, access the key image blocks.
3. Convert these key image blocks of each plane into transform domain form (DCT).
4. Extract the data from the transformed stego image blocks of each plane.
5. Reconstruct each plane of secret image with the help of extracted data. De-normalize each plane (Multiply each value of the extracted secret data by its maximum pixel value).
6. Finally, reconstruct the R-G-B planes to retrieve the secret image.

4. EXPERIMENTAL RESULTS AND DISCUSSION

Thirty RGB images of size 256 × 256 are used as cover images (C) and five RGB images of different sizes are taken as secret images (S) = {64 × 64, 128 × 128, 160 × 160, 176 × 176, 192 ×

192}. All the secret images are embedded separately in all thirty cover images. The stego image produced by secret image of size 128 × 128 for sample cover image 'Koala' and the retrieved secret image are shown in Figure 2. Similarly, Figure 3 shows stego image produced by secret image of size 176 × 176 and the retrieved secret image. The quality of stego images produced by the proposed method has been tested based on various performance evaluation metrics like MSE, PSNR and Correlation. Table 1 displays the MSE, PSNR and Correlation values for stego images produced by cover image Koala with all five secret images. Figure 4 displays the graph of MSE and PSNR values for all thirty stego images obtained by secret image of size 176 × 176.

Mean Square Error (MSE) between two images can be computed as shown in equation (3),

$$MSE = \frac{1}{NM}\sum_{m=0}^{M-1}\sum_{n=0}^{N-1}[x(m,n) - y(m,n)]^2 \quad (3)$$

where M and N: Number of rows and columns in the cover image respectively, x(m,n): Cover image, y(m,n): Stego image.

Peak Signal-to-Noise Ratio (PSNR) [21] is computed as shown in equation (4),

$$PSNR = 10\log_{10}\left(\frac{S^2}{MSE}\right) \quad (4)$$

where S is the maximum possible pixel value of the image. For example, if it has an 8-bit unsigned integer data type, S is 255.

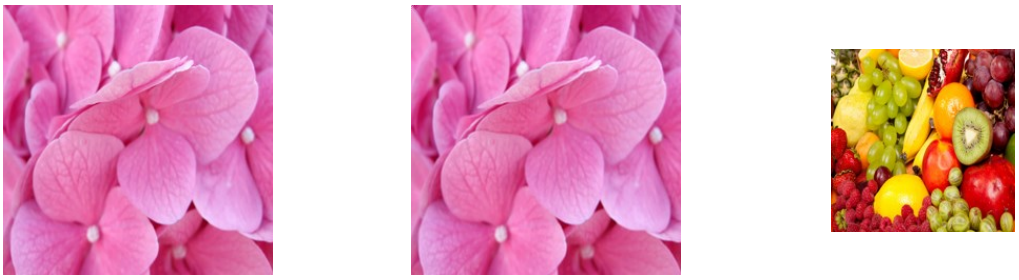
The Pearson's Correlation Coefficient [22, 23] is computed as shown in equation (5),

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x}).(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2.\sum_{i=1}^{n}(y_i - \bar{y})^2}} \quad (5)$$

where x_i and y_i are the cover image positions and x bar and y bar are stego image positions.

(a) (b) (c)

Figure 2: (a) Cover Image Koala (b) Stego Image (c) Retrieved Secret Image (128 × 128)



(a) (b) (c)

Figure 3: (a) Cover Image (b) Stego Image (c) Retrieved Secret Image (176 × 176)

Table 1. Performance Metric Values for Cover Image Koala with Different Sizes of Secret Images

Size of Secret Image	MSE	PSNR	Correlation
64 × 64	0.3428	52.7810	0.9999
128 × 128	3.5261	42.6579	0.9984
160 × 160	6.6337	39.9133	0.9968
176 × 176	6.2244	40.1898	0.9970
192 × 192	8.7725	38.6996	0.9956

4.1 Observations and Analysis:

From Table 1 it is observed that when smaller sizes of secret images are embedded in a cover image, MSE between cover and stego images is lesser and as the size of secret images increase, MSE is more. Figure 2 and 3 show that there is less visual distortion in the stego images and the secret images are retrieved properly. The ideal value for MSE is zero and Cover Image 1 for secret image of size 176 × 176 has the least MSE (MSE = 0.4355) which is close to zero, among all the thirty cover images considered, as observed in the Figure 4. PSNR is inversely proportional to MSE, hence, lesser the MSE, more will be the PSNR. If PSNR has value above 36 dB then the visibility between the cover and stego image looks the same i.e. quality of stego image is good [24]. The proposed method gives PSNR above 36

dB for all the images as seen from the Table 1 and Figure 4. Cover Image 30 for secret image of size 176 × 176 has the least PSNR (PSNR=39.3721) amongst the thirty cover images (as displayed by the graph in Figure 4) and is above 36 dB. The correlation coefficient can range between values −1 and +1. Higher correlation value i.e. correlation close to one states that good linear relationship exists between cover and stego images and that there is less difference between the two images [25]. The average correlation for all thirty cover images with secret image of size 176 × 176 is 0.9989, which means the overall performance of the proposed method is good. The correlation values displayed by Table 1 are also satisfactory. The embedding capacity of the proposed method is 68.75% (i.e. secret image up to size 176 × 176 can be embedded into 256 × 256 cover image without causing much visual distortion and maintaining quality of stego image). Based on analysis with respect to the poor results where error is more, it is observed that there are more discontinuities i.e. texture variations in those cover images. Adding secret information into the image adds to more discontinuities which comes as noise and leads to more distortion and error in the stego image. But, the overall performance of the proposed method is satisfactory with acceptable values of MSE, PSNR and Correlation as desired.

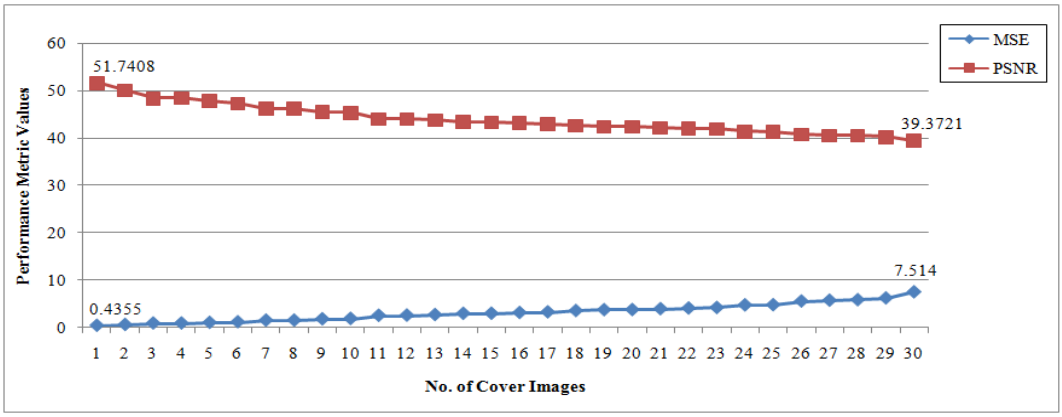


Figure 4: MSE and PSNR Values for Secret Image of Size 176 × 176

5. CONCLUSION

In this paper, GA-based approach is introduced for carrier in image steganography. Our proposed method focused on the before embedding technique which tries to identify appropriate places in cover image using GA for embedding secret image with minimum distortion. For this, the cover image was divided into non-overlapping blocks and these image blocks were considered as the population in GA execution. This paper focused and proved the proposed effective use of GA with new fitness function i.e. standard deviation which actually refers to the image texture while selecting the appropriate image blocks for embedding. Further, DCT was applied to these selected image blocks that helps to filter out high energy blocks and select low energy blocks for embedding. As DCT was applied only to the selected blocks of image, the computational complexity of the algorithm was reduced. These above two steps played an important role in the generation of quality stego image. The experimental results exhibit that a good quality stego image is obtained from the proposed method. The genetic algorithm in our method increases the strength of stego image and DCT allows effective embedding of secret image. The values of image quality metrics show that the stego images are secure and less prone to detection of secret data. Few of the stego images obtained by the proposed method have higher MSE values which gives scope for further improvement. In future, our plan is to implement this method in other frequency domain techniques such as wavelet transforms and to improve the embedding capacity.

6. ACKNOWLEDGMENTS

I would like to thank my respected guide Dr. Kavita V. Sonawane for her encouragement and support throughout. She has been a constant source of inspiration.

We are also grateful to the Institute and Department of Computer Engineering for providing us the necessary infrastructure and facilities.

7. REFERENCES

[1] Jan HP Eloff, Martin S. Olivier and Tayana Morkel. An overview of image steganography. In *Proceedings of the Fifth Annual Information Security South Africa Conference (ISSA)*, 2005.

[2] Bagheri Baba Ahmadi. Image Watermarking: Blind Linear Correlation Technique. *World Applied Programming*, pages 93-100, 2015.

[3] Mansi S. Subhedara and Vijay H. Mankarb. Current status and key issues in image steganography: A survey. *Computer Science Review*, October 2014.

[4] S. G. Shelke and S. K. Jagtap. A Novel Approach: Pixel Matching Based Image Steganography. In *International Conference on Pervasive Computing (ICPC)*, pages 1-4, 2015.

[5] Ali Hanani, Masoud Nosrati and Ronak Karimi. Steganography in Image Segments using Genetic Algorithm. In *Fifth IEEE International Conference on Advanced Computing & Communication Technologies (ACCT)*, pages 102–107, 2015.

[6] M. Nosrati and R. Karimi. A Survey on Usage of Genetic Algorithms in Recent Steganography Researches. *World Applied Programming*, pages 206-210, 2012.

[7] L. H. Chen and Y. K. Lee. High capacity image steganographic model. In *IEEE Proceedings-Vision, Image and Signal Processing*, 147(3):288–294, June 2000.

[8] D. C. Wu and W. H. Tsai. A steganographic method for images by pixel value differencing. *Pattern Recognition Letters*, pages 1613–1626, 2003.

[9] E. Chang and V. Potdar. Gray level modification steganography for secret communication. In *2nd IEEE International Conference on Industrial Informatics*, pages 355–368, May 2004.
[10] Chi-Kwong Chan and L. M. Cheng. Improved hiding data in images by optimal moderately-significant-bit replacement. *Electronics Letters*, pages 1017-1018, 2001.
[11] S. Wang and X. Zhang. Efficient Steganographic embedding by exploiting modification direction. *IEEE Communications Letters*, 10(11):781-783, November 2006.
[12] C. Lee, H. Wu, R. Chao and Y. Chu. A novel image data hiding scheme with diamond encoding. *EURASIP Journal on Information Security*, 2009.
[13] "Lempel-Ziv-Welch (LZW) Compression," Source: http://www.fileformat.info/mirror/egff/ch09_04.htm
[14] Kolsoom Shahryari and Mehrdad Gholami. High Capacity Secure Image Steganography Based on Contourlet Transform. *Advances in Computer Science: an International Journal*, 2(4):62-65, September 2013.
[15] Bian Yang, Shen Wang and Xiamu Niu. A Secure Steganography Method based on Genetic Algorithm. *Journal of Information Hiding and Multimedia Signal Processing*, 1(1):28-35, January 2010.
[16] E. Baburaj and P.M. Siva Raja. Data Hiding Scheme For Digital Images Based on Genetic Algorithms with LSBMR. *International Journal of Computer Applications*, 59(5):8-15, December 2012.
[17] Elham Ghasemi, Jamshid Shanbehzadeh and Nima Fassihi. High Capacity Image Steganography based on Genetic Algorithm and Wavelet Transform. *Intelligent Control and Innovative Computing*, Springer US, pages 395-404, 2012.
[18] Bahram Nazeri and Hamidreza Rashidy Kanan. A novel image steganography scheme with high embedding capacity and tunable visual image quality based on a genetic algorithm. *Expert Systems with Applications*, pages 6123-6130, October 2014.
[19] "Standard Deviation Formulas," Source: https://www.mathsisfun.com/data/standard-deviation-formulas.html
[20] H. B. Kekre, Rekha Vig and Tanuja Sarode. Multi-resolution Analysis of Multi-spectral Palmprints using Hybrid Wavelets for Identification. *International Journal of Advanced Computer Science and Applications (IJACSA)*, 2013.
[21] Dr. Der Chen Soong and Yusra A. Y. Al-Najjar. Comparison of Image Quality Assessment: PSNR, HVS, SSIM, UIQI. *International Journal of Scientific & Engineering Research*, August 2012.
[22] Joseph Lee Rodgers and W. Alan Nicewander. Thirteen ways to look at the correlation coefficient. *The American Statistician*, 42(1):59-66, February 1988.
[23] Son, S.-W., C. Christensen, P. Grassberger, and M. Paczuski. PageRank and rank-reversal dependence on the damping factor. *Physical Review E*, 2012.
[24] Gautam Sanyal, Indradip Banerjee and Souvik Bhattacharyya. Robust image steganography with pixel factor mapping (PFM) technique. In *IEEE International Conference on Computing for Sustainable Global Development (INDIACom)*, pages 692-698, 2014.
[25] "Interpret the key results for Correlation," Source: http://support.minitab.com/en-us/minitab-express/1/help-and-how-to/modeling-statistics/regression/how-to/correlation/interpret-the-results/