

Watermarked and Noisy Images identification Based on Statistical Evaluation Parameters



Sattar B. Sadkhan^{*}, Nidaa A. Abbas^{**}

^{*} Information Networks Dept., College of IT- University of Babylon, Iraq, drengsattar@ieee.org

^{**} Software Dept., College of IT- University of Babylon, Iraq, drnidaa_muhsin@ieee.org

Abstract:

A watermark scheme is an important technique for copyright protection of digital images. Digital watermarking is the process of computer-aided information hiding in a carrier signal. The main interest of this paper is copyright protection, and it takes into consideration four important aspects: (i) Implementation the images watermarking by Least Significant Bit method (LSB) for JPEG gray images using invisible watermark, (ii) Evaluation the watermarking images using different statistical parameters, (iii) Identifying watermark images from noisy images by showing that the difference in results using open set identification, (iv) Proposing threshold equations that can be used to differentiate among noisy and watermarked images based on the used statistical parameters of the tested images. By comparing the image quality, obtained by the proposed method with the calculated statistical metrics like Variance, Standard Deviation, Kurtosis and Skewness. The results are promising and give us a great indication to differentiate between the images of watermarking and noisy images.

Keywords: Watermarking image; copyright; Identification; Statistical Metrics

I. INTRODUCTION :

Fundamentally, watermarking can be described as a method for embedding information into another signal. In case of digital images, the embedded information can be either visible or invisible from the user.

Digital images are subject to a wide variety of distortions during acquisition, processing, compression, storage, transmission and reproduction, any of which may result in a degradation of visual quality. For applications in which images are ultimately to be viewed by human beings, the only "correct" method of quantifying visual image quality is through subjective evaluation. In practice, however, subjective evaluation is usually too inconvenient, time-consuming and expensive. The goal of research in objective image quality assessment is to

develop quantitative measures that can automatically predict image quality [1].

An objective image quality metric can play a variety of roles in image processing applications. Most existing approaches are known as full-reference, meaning that a complete reference image is assumed to be known. In many practical applications, however, the reference image is not available, and a no-reference or "blind" quality assessment approach is desirable [2, 3].

The simplest and most widely used full-reference quality metric is the mean squared error (MSE), computed by averaging the squared intensity differences of the distorted and reference image pixels, along with the related quantity of Peak Signal-to-Noise Ratio (PSNR).

This method, as well as all the statistics based measures, is simple to calculate,

have clear physical meanings, and are mathematically convenient in the context of optimization.

Identification codes for noisy channels were introduced by R. Ahlswede and G. Dueck for the situation in which the receiver needs to identify whether the coming message equals a specified one. If not, then they don't care what it are [4]. It turned out that this weaker requirement dramatically increases the sizes of message sets which could be handled: double exponential grown in the block lengths of codes.

Y. Steinberg and N. Merhav notice that in most cases people check watermarks in order to identify them (e.g. Copyright) rather than recognize them and so they introduced identification codes to watermarking models [5]. In their models the attack channels are single memoryless channels. That means the attacker's random strategy is known by information hider (encoder) and the decoder. They notice that the assumption is not robust and so suggested to study more robust models. As to the resources shared by encoders and decoders they consider two cases, the decoder either completely knows the covertext or he knows nothing about it. (In all cases the attacker must not know the covertext because otherwise there would be no safe watermarking).

In this paper, we will concentrate on invisible watermarks, and the aims of this paper are (i) implementation the image watermarking by Least Significant Bit method (LSB) for a JPEG grey image using invisible watermark, (ii) evaluates the watermarking image using statistical parameters, (iii) Identify watermark image from noisy image by showing that the difference in results using open set identification, (iv) proposing threshold equations that can be used to differentiate among noisy and watermarked images

based on the used statistical parameters of the tested images.

In the next Section a brief description of watermarking implantation functions. In Section 3, the watermark identification and their types, is described. In Section 4, classifications of watermarking, like visible and invisible and their categories are presented. Section 5, fundamental steps of Least Significant Bit algorithm and its implementation in image watermark. In Section 6, hypothesis testing is proposed to provide the statistical certainty for the watermark identification. Finally, in Section 7, simulation experiments of particular algorithm are presented indicating its performance.

II. WATERMARKING IMPLEMENTATION FUNCTIONS:

A watermarking system is usually divided into three distinct steps, embedding, attack and detection. In embedding, an algorithm accepts the host and the data to be embedded and produces a watermarked signal. The watermarked signal is then transmitted or stored, usually transmitted to another person. If this person makes a modification, this is called an attack. There are many possible attacks [2]. Detection is an algorithm which is applied to the attacked signal to attempt to extract the watermark from it. If the signal was not modified during transmission, then the watermark is still present and it can be extracted.

If the signal is copied, then the information is also carried in the copy. The embedding takes place by manipulating the content of the digital data, which means the information is not embedded in the frame around the data, it is carried by the signal itself. Fig. 1 shows general digital watermark life-cycle phases with embedding, attacking, and detection and retrieval functions.

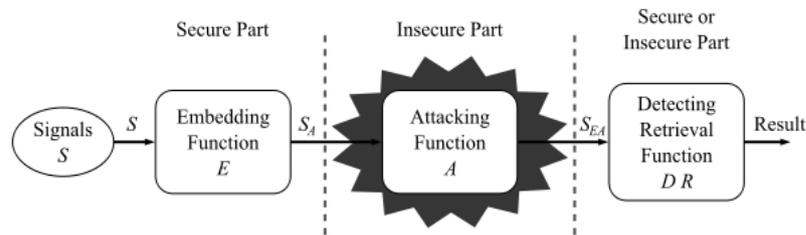


Fig. 1 General digital watermark life-cycle phases with embedding, attacking, and detection and retrieval functions.

The information to be embedded in a signal is called a digital watermark, although in some contexts the phrase digital watermark means the difference between the watermarked signal and the cover signal. The signal where the watermark is to be embedded is called the host signal. A watermarking system is usually divided into three distinct steps, embedding, attack, and detection. In embedding, an algorithm accepts the host and the data to be embedded, and produces a watermarked signal.

A. Embedding Function

The watermark embedding scheme can either embed the watermark directly into the host data or to a transformed version of the host data. Some common transform domain watermarking for image data can be in the frequency domain like Discrete Cosine Transform (DCT) based [6], [7], and references therein or wavelet based [8] or in spatial domain like Least Significant Bit method (LSB).

Some of the “watermarking techniques” described in the literatures are simple additive watermarking schemes expressed as:

$$X = S + W \quad (1)$$

Where S is the original host signal, X is the watermarked signal, and W is the watermark signal.

B. Attack Function

Digital watermarking is not as secure as data encryption. Therefore, digital watermarking is not immune to hacker attacks.

Watermarking attacks are broadly divided into the following categories:

1. Removal Attacks
2. Geometrical Attacks
3. Cryptographic Attack
4. Protocol Attacks

In basic attack, the attacker takes advantage of the limitations in design of the embedding technique [9, 10, 11].

C. Detection Function

Watermark detection is the most important part of the watermark algorithm. Detection or verification refers to the process of making a binary decision at the decoder—whether a specific watermark is or is not present in the received data [2].

III. WATERMARK IDENTIFICATION:

Identification refers to the process of being able to decode one of N possible choices (messages) at the receiver. An application for this includes copyright protection where multiple copies of the same content get a unique label so that misuse of one of the copies can be traced back to its owner. Identification problems can be categorized as “open set” or “closed set.” Open set identification refers to the possibility that one of N or no

watermark exists in the data. Closed set refers to problems where one of N possible watermarks is known to be in the received data and the detector has to pick the most likely one.

IV. CLASSIFICATIONS OF WATERMARKING:

1. Visible

The watermark is visible when a text or a logo used to identify the owner. Any text or logo to verify or hide content can be expressed as follows:

$$F_w = (1-\alpha)F + \alpha*W \quad (2)$$

Where F_w is Watermarked Image, α is a constant; $0 \leq \alpha \leq 1$, IF $\alpha=0$ No watermark, if $\alpha=1$ watermark present, F is the original image and W is a watermark

2. Invisible

The watermark is embedded into the image in such a way that it cannot be perceived by the human eye. It is used to protect the image authentication and prevent it from being copied.

V. LEAST SIGNIFICANT BIT (LSB):

LSB coding is one of the earliest methods in watermarking and steganography. It can be applied to any form of watermarking. In this method the LSB of the carrier signal is substituted with the watermark. The bits are embedded in a sequence which acts as the key. In order to retrieve it back this sequence should be known. The watermark encoder first selects a subset of pixel values on which the watermark has to be embedded. It then embeds the information on the LSBs of the pixels from this subset. LSB coding is a very simple technique but the robustness of the watermark will be too low. With LSB is coding almost always the watermark cannot be retrieved without a noise component [5].

VI. PERFORMANCE EVALUATION METRICS:

To measure the quality of the watermarked image statistical analysis is used.

A. Pearson Correlation Coefficient

Pearson's correlation coefficient, r , is widely used in statistical analysis, pattern recognition, and image processing [12]. Applications include comparing two images for the purposes of image registration, object recognition, and disparity measurement. For monochrome digital images, the Pearson correlation coefficient is defined as [13]:

$$r = \frac{\sum_i (x_i - x_m)(y_i - y_m)}{\sqrt{\sum_i (x_i - x_m)^2} \sqrt{\sum_i (y_i - y_m)^2}} \quad (3)$$

Where x_i is the intensity of the i th pixel in image 1, y_i is the intensity of the i th pixel in image 2, x_m is the mean intensity of image 1, and y_m is the mean intensity of image 2.

The correlation coefficient has the value $r = 1$ if the two images are absolutely identical, $r = 0$ if they are completely uncorrelated, and $r = -1$ if they are completely anti-correlated, for example, if one image is the negative of the other.

B. Mean

We can think of $r \times c$ matrix (image) as a set of c column vectors, each having r elements. Often, with matrices, we want to compute mean scores separately within columns, consistent with the equation below.

$$X_c = \sum X_{ic} / r \quad (4)$$

Where X_c is the mean of a set of r scores from column c , $\sum X_{ic}$ is the sum of elements from column c .

C. Variance

Variance is a measure of the variability or spread in a set of data. Mathematically, it is the average squared deviation from the mean value. We use the following formula to compute variance.

$$\text{Var}(X) = \frac{\sum (X_i - \bar{X})^2}{N} = \frac{\sum x_i^2}{N} \quad (5)$$

Where N is the number of scores in a set of scores X is the mean of the N scores. X_i is the i th raw score in the set of scores x_i is the i th deviation score in the set of scores $\text{Var}(X)$ is the variance of all the scores in the set

D. Standard Deviation

The standard deviation shows how much variation or "dispersion" exists for the average (mean, or expected value). A low standard deviation indicates that the data points tend to be very close to the mean, whereas high standard deviation indicates that the data points are spread out over a large range of values.

The standard deviation of any matrix can be expressed in the following way:

$$\sigma = \sqrt{\frac{i}{N} \sum_{i=1}^N (x_i - \bar{x})^2} \quad (6)$$

Where N is the total number of elements in a column of that matrix and x_i are the matrix's elements in column i .

E. Kurtosis

The classical measure of nonGaussianity is Kurtosis or the fourth-order cumulant. The Kurtosis of y is classically defined by [14]:

$$\text{kurt}(y) = E\{y^4\} - 3(E\{y^2\})^2 \quad (7)$$

Kurtosis can be either positive or negative. Random variables that have a

negative Kurtosis are called subGaussian, and those with positive Kurtosis are called superGaussian, and zero for Gaussian.

F. Skewness

Skewness is a measure of the asymmetry of the probability distribution of a real-valued random variable. The skewness value can be positive or negative, or even undefined. Qualitatively, a negative skew indicates that the tail on the left side of the probability density function lies longer than the right side and the bulk of the values to the right of the mean. A positive skew indicates that the tail on the right side is longer than the left side and the bulk of the values lying to the left of the mean. A zero value indicates that the values are relatively evenly distributed on both sides of the mean, typically but not necessarily implying a symmetric distribution. Mathematically, skewness is calculated from [15]:

$$K_3(x) = \frac{E[(x-\mu)^3]}{\sigma^3} \quad (8)$$

Where μ and σ are the mean and standard deviation of a random variable x , respectively and $E []$ is the mathematic expectancy.

VII. PROPOSED SYSTEM FOR IDENTIFICATION OF IMAGES WATERMARKING:

The proposed system is implemented under Dell Laptop, with O.S. Windows 7, Processor Core 2 Duo and RAM 2.00 GB using the programming facilities of MATLAB. In this section we proposed an identification system using statistical evaluation parameters, Fig. 2, represents the proposed system.

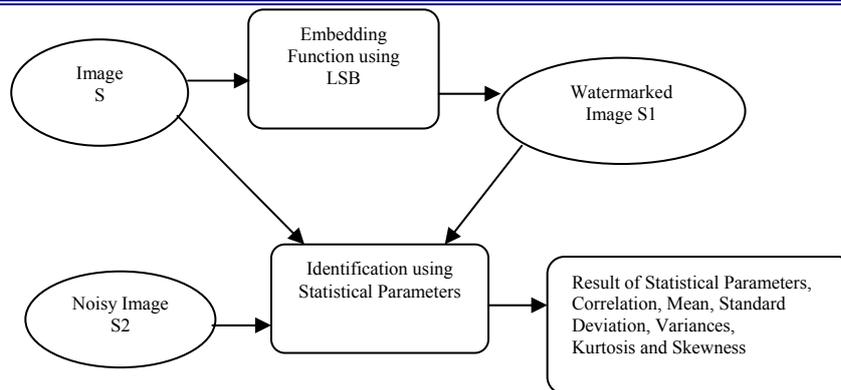


Fig. 2 the Proposed System

The steps of the proposed system are: -

1. Read grey Image JPG type
2. Embed copyright image using LSB algorithm
3. Compare between original, watermarked, and noisy image using statistical parameters
4. Identify between three images using open set

A. Proposed system in Detail

In step 1 is the input to the system by reading the image JPG type the result is the matrix of two dimensional that representation of the given image. Step 2 represents the embedding function which embeds the copyright image represented by Fig. 4 using Least Significant Bit (LSB) algorithm, Least significant bit Watermarking) Steps are

1. A raw bitmap image 'A' will be selected from the set of standard test images. Let this be the base image on which the watermark will be added.
2. A raw bitmap image 'B' will be selected from the set of standard test images. This will be the watermark image which will be added to the base image.
3. The most significant bit henceforth will be mentioned as an MSB, of watermark image 'B' will be read and these will be written on the Least Significant Bit,

henceforth will be mentioned as LSB, of the base image 'A'.

Thus, 'A' will be watermarked with 'B' resulting in a combined image 'C'. 'C' therefore will now contain an image 'A' which has its LSBs replaced with the MSBs of 'B'. The technique used will be LSB technique which is a form of spatial domain technique. This technique is used to add invisible and visible watermarks in the image

Step 3, identification function that compares the original images with watermarked image which result from embedding function and noisy image using statistical parameters represented by correlation, Mean, Standard Deviation, Variances, Kurtosis and Skewness.

Fig. 3 shows by plotting the differences between three images using Statistical Parameters. The results of the five measures (Mean, Standard Deviation, Variances, Kurtosis and Skewness) are used for the original, watermarked and noisy images as shown in Table I. Fig. 4 represents the watermark image size of (50×50) . In other hand Fig.5 (a) show the original image size (128×128) , Fig.5 (b) represent watermarked images, and Fig. 5 (c) show noisy images with Gaussian noise, mean 0 and variance 0.01

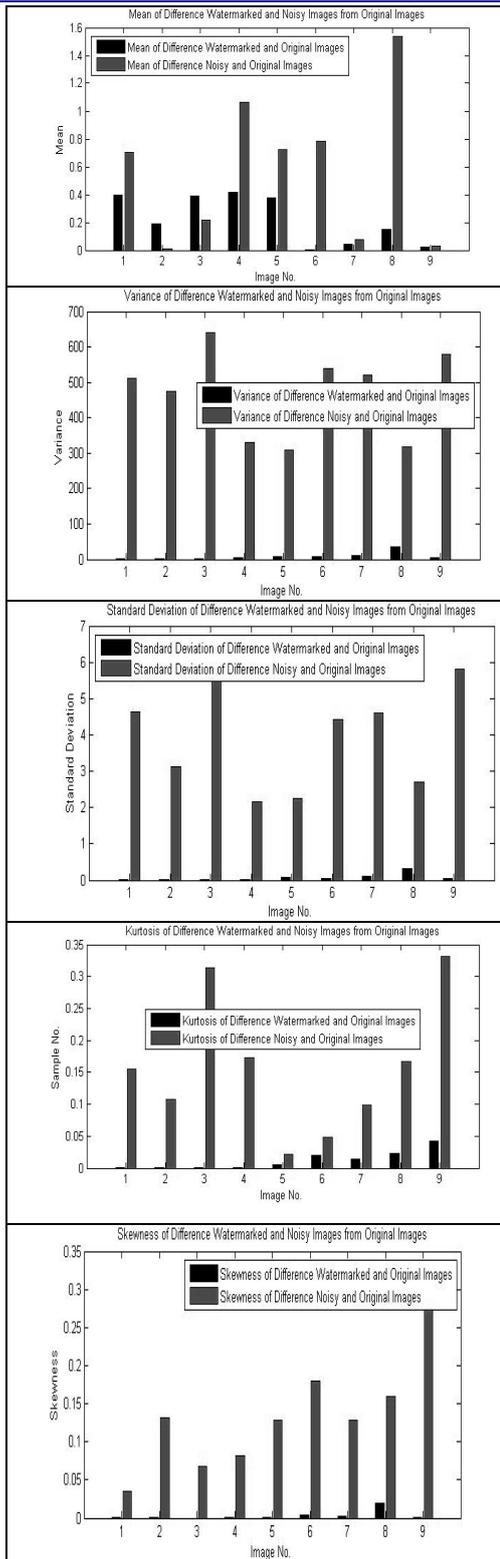


Fig. 3 Comparison between three images using Statistical Parameters



Fig. 4 Image to be Watermarked, Size (50 × 50)

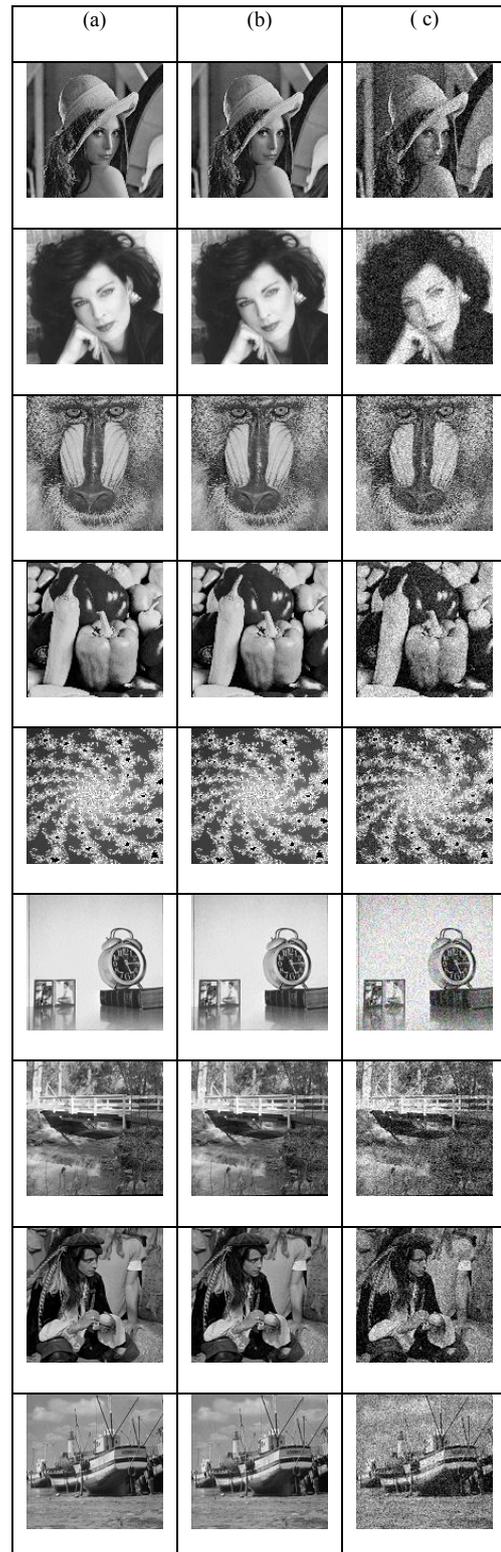


Fig. 5 (a) Original Image size (128×128), (b) watermarked Images, and (c) Noisy Images with Gaussian noise, mean 0 and variance 0.01

VIII. CONCLUSION:

We have presented in this work an objective quality metric based on statistical parameters, and tested its performances regarding five distinct quality assessment tasks. The testing aspect within the proposed system based on Mean, Standard Deviation, Variances, Kurtosis and Skewness. We notice that the correlation parameter is ineffective in the results therefore is not mentioned.

The experimental results showed good performances of the metrics Standard Deviation, Variances as identification parameters. We can conclude general formulas for these statistical parameters.

From the used testing parameters we can summarize the followings:

- 1) The used mean value parameter is oscillating in its results from images to others. Hence we couldn't based on its behavior as one of the identification parameters set
- 2) The other four statistical parameters provided good results to identify the noise images from watermarked images, and we raced to the proposed equation for each parameter, as follows:

For the standard deviation and throughout the calculation of the differences between the standard deviation of watermarked images and original images, and that corresponding to the difference between noisy images and originals, used for the all nine tested different images, we can give the following equations which can be

considered as identification threshold equation to recognize if the image is noisy or it is a watermarked image (the number like 2 which is taken is try and error):

$$\text{if } sdv(N) - Sdv(O) \geq 2 \begin{cases} \text{yes} & \text{noisy} \\ \text{No} & \text{watermarked} \end{cases}$$

- 3) The same concepts were followed for the other parameters, and we reached to the following equations that can be considered as important threshold equations to enhance the decision making about the nature of testing images.

$$\text{if } \text{var}(N) - \text{var}(O) \geq 300 \begin{cases} \text{yes} & \text{noisy} \\ \text{No} & \text{watermarked} \end{cases}$$

$$\text{if } \text{Kur}(N) - \text{Kur}(O) \geq 0.02 \begin{cases} \text{yes} & \text{noisy} \\ \text{No} & \text{watermarked} \end{cases}$$

$$\text{if } \text{Skw}(N) - \text{Skw}(O) \geq 0.03 \begin{cases} \text{yes} & \text{noisy} \\ \text{No} & \text{watermarked} \end{cases}$$

- 4) For all tested different images, the proposed equations are well satisfied the identification aim.

Where Sdv, var, Kur and Skw are Standard Deviation, Variance, Kurtosis and Skewness respectively. N and O represent Noisy and Original images successively.

Some ideas for future works could be the use of spread spectrum techniques instead of the LSB method in addition to the Principal Component Analysis (PCA) as extra parameters. Also we will use frequency domain to show the performance of the proposed method.

TABLE I: Testing images using statistical parameters, noisy images with Gaussian noise mean 0 and variance 0.01, image size (128 × 128)

Image No.	Image Type	Correlation between two images	Mean	Standard Deviation	Variances	Kurtosis	Skewness
1	O	0.9999 (O, W)	98.9850	53.0589	2.8152e+003	2.2896	0.2153
	W		99.3814	53.0567	2.8150e+003	2.2891	0.2155
	N	0.9030 (O, N)	99.6864	57.6827	3.3273e+003	2.4444	0.2508
	(W-O)		0.3964	0.0022	0.19	0.0005	0.0002
	(N-O)		0.7014	4.6238	512.10	0.1548	0.0355
2	O	1.0000 (O, W)	105.3309	74.1127	5.4927e+003	1.8436	0.6856
	W		105.5219	74.1215	5.4940e+003	1.8442	0.6859
	N	0.9481 (O, N)	105.3441	77.2375	5.9656e+003	1.9516	0.5538
	(W-O)		0.1910	0.0088	1.30	0.0006	0.0003
	(N-O)		0.0132	3.1248	472.90	0.1080	0.1318
3	O	0.9999 (O, W)	129.2408	48.2662	2.3296e+003	1.9958	-0.1058
	W		129.6336	48.2729	2.3303e+003	1.9952	-0.1058
	N	0.8831 (O, N)	129.0246	54.4815	2.9682e+003	2.3097	-0.0383
	(W-O)		0.3928	0.0067	0.70	0.0006	0
	(N-O)		0.2162	6.2153	638.60	0.3139	0.0675
4	O	1.0000 (O, W)	114.1697	75.4484	5.6925e+003	1.4417	-0.1020
	W		114.5842	75.4201	5.6882e+003	1.4403	-0.1006
	N	0.9510 (O, N)	115.2332	77.6138	6.0239e+003	1.6140	-0.0208
	(W-O)		0.4145	0.0283	4.30	0.0014	0.0014
	(N-O)		1.0635	2.1654	331.40	0.1723	0.0812
5	O	1.0000 (O, W)	136.4821	67.2970	4.5289e+003	1.9951	0.2602
	W		136.8589	67.2288	4.5197e+003	1.9906	0.2589
	N	0.9382 (O, N)	135.7562	69.5570	4.8382e+003	1.9729	0.1322
	(W-O)		0.3768	0.0682	9.19	0.0045	0.0013
	(N-O)		0.7259	2.2600	309.30	0.0222	0.1280
6	O	0.9951 (O, W)	185.4709	58.5410	3.4271e+003	3.1938	-1.2126
	W		185.4674	58.4848	3.4205e+003	3.2131	-1.2167
	N	0.9238 (O, N)	184.6835	62.9715	3.9654e+003	3.1454	-1.0330
	(W-O)		0.0026	0.0562	6.59	0.0193	0.0040
	(N-O)		0.7865	4.4305	538.30	0.0484	0.1796
7	O	0.9862 (O, W)	112.9603	54.1226	2.9293e+003	2.6359	0.4897
	W		112.9143	54.0124	2.9173e+003	2.6506	0.4867
	N	0.9049 (O, N)	112.8809	58.7367	3.4500e+003	2.5375	0.3606
	(W-O)		0.0460	0.1102	12	0.0147	0.0030
	(N-O)		0.0794	4.6141	520.70	0.0984	0.1291
8	O	0.9905 (O, W)	88.3239	57.4678	3.3025e+003	1.9209	-0.0052
	W		88.4742	57.1509	3.2662e+003	1.9446	0.0142
	N	0.9193 (O, N)	89.8584	60.1564	3.6188e+003	2.0878	0.1541
	(W-O)		0.1503	0.3169	36.30	0.0237	0.0194
	(N-O)		1.5345	2.6886	316.30	0.1669	0.1593
9	O	0.9881 (O, W)	128.9118	46.9142	2.2009e+003	3.2934	-0.8725
	W		128.9384	46.8678	2.1966e+003	3.3363	-0.8714
	N	0.8812 (O, N)	128.8824	52.7159	2.7790e+003	2.9624	-0.5398
	(W-O)		0.0266	0.0464	4.30	0.0429	0.0011
	(N-O)		0.0294	5.8017	578.10	0.3310	0.3327

References

- [1] Zhou Wang, Alan Conrad Bovik, Hamid Rahim Sheikh and Eero P. Simoncelli, "Image Quality Assessment: From Error Visibility to Structural Similarity", IEEE TRANSACTIONS ON IMAGE PROCESSING, 4(13) , APRIL (2004).
- [2] Christine I. Podilchuk and Edward J. Delp, "Digital Watermarking: Algorithms and Applications", IEEE SIGNAL PROCESSING MAGAZINE, JULY (2001)
- [3] R. Ahlswede and N. Cai , "Watermarking Identification Codes with Related Topics on Common Randomness", Information Transfer and Combinatorics, LNCS 4123, pp. 107–153, Springer-Verlag Berlin Heidelberg (2006).
- [4] R. Ahlswede and G. Dueck, Identification via channels, IEEE Trans. Inform. Theory, 1(35), pp.15-29, (1989).
- [5] Y. Steinberg and N. Merhav, "Identification in the presence of side information with application to watermarking", IEEE Trans. Inform. Theory, (47), pp. 1410–1422, (2001).
- [6] A.G.Borsand I.Pitas, "Image watermarking using DCT domain constraints," in IEEE Proc. Int. Conf. on Image Processing, Lausanne, Switzerland, Sept., (3), pp. 231-234, (1996).
- [7] Ahmed A. Abdulfetah, Xingming Sun, Hengfu Yang and Nur Mohammad, "Robust Adaptive Image Watermarking using Visual Models in DWT and DCT Domain", Information Technology Journal 9 (3) pp. 460-466, (2010).
- [8] Dhandapani Samiappan and Krishnan Ammasi , "Robust Image Watermarking Using Discrete Wavelet Transform", *Journal of Computer Science* DOI: 10.3844/jcssp.2011.1.5, 1(7), pp. 1-5
- [9] S. Voloshynovskiy, S. Pereira and T. Pun, "Watermark attacks", Erlangen Watermarking Workshop 99, Oct., (1999).
- [10] Neil F. Johnson, "An Introduction to Watermark Recovery from Images", SANS Intrusion Detection and Response Conference (IDR'99) held in San Diego, CA, February 9-13, (1999).
- [11] Mir Shahriar Emami and Ghazali Bin Sulong, "Set Removal Attack: A New Geometric Watermarking Attack", International Conference on Future Information Technology, IPCSIT, IACSIT Press, Singapore, (13), (2011).
- [12] Eugene K. Yen and Roger G. Johnston, "The Ineffectiveness of the Correlation Coefficient for Image Comparisons", <http://jps.anl.gov/vol.2/3-Correlation.pdf>
- [13] Joseph Lee Rodgers and W. Alan Nicewander, "Thirteen Ways to Look at the Correlation Coefficient", *The American Statistician*, 1(42), pp. 59-66, (1988).
- [14] Dr. Eng. Sattar B. Sadkhan, Dr. Nidaa A. Abbas , "Performance Evaluation of Speech Scrambling Methods Based on Statistical Approach", FONDAZIONE GIORGIO RONCHI, 5, Oct., Italy, (2011).
- [15] A. Azzalini and A. D. Valle, "The multivariate skew-normal distribution," *Biometrika*, (83), pp. 715-726, December 1, (1996).