International Journal of Research in Engineering and Science (IJRES)
ISSN (Online): 2320-9364, ISSN (Print): 2320-9356
www.ijres.org Volume 11 Issue 11 ǁ November 2023 ǁ PP. 174-183

A Hybrid GLCM and Statistical Moments System for Digital Image Forgery Detection

M.E.ElAlami, W.K.ElSaid, Samra.Mamdouh
Moh_ElAlmi@mans.edu.eg
prof_wessam@mans.edu.eg
Sit_el_banat9@yahoo.com

Abstract
Image forgery detection systems have recently become one of the most essential trends in the field of software engineering. In order to verify the originality of a digital image, numerous organizations and institutions are now using digital image forgery detection systems. This scientific research paper provides an image forgery detection system that can detect forged images and static video frames. The proposed system compares the efficiency of the three major methods: GLCM, Statistical Moments and the combination of GLCM and Statistical Moments. The main features produced from GLCM approach include: contrast, energy, homogeneity and correlation. While, mean, median, standard deviation, kurtosis and skewness are the major statistical metrics obtained from Statistical Moments approach. The system evaluation procedures are divided into two important phases. The first is the validity test, in which its usability in real-world context has been validated. The second is the efficiency test, in which its ability to detect tampering in digital images has been verified. Luckily, the results of both tests were very encouraging.

Keywords: *Image Forgery Detection, GLCM, Statistical Moments, Validity Test, Efficiency Test.*

Date of Submission: 06-11-2023 Date of Acceptance: 20-11-2023

I. Introduction

Before 2000, most images were printed on conventional paper with low quality. Then at the beginning of the year 2000, most editorial offices began to shift from paper-printed images to digital image formats, which became indispensable in our daily lives, as it is now no longer acceptable to use images printed on paper in any operations [1].

Technological progress has played a major role in the production of digital images, as there is now no longer a need to use special machines, as simply the camera of any digital device such as laptops, tablets and smartphones can be used to create high resolution. Digital images are available for use in various fields such as medicine, engineering, arts, ... etc. [2].

A digital image is a two-dimensional matrix of a limited set of cell elements known as image elements (pixels) in mathematics. Each pixel has a color value that varies depending on the image type; in a binary image, each pixel has a color value that is either zero or one, whereas in a grayscale image, each pixel has a color value ranging from 0 to 255, and in a color image, each pixel has three color values ranging from 0 to 255, with the first value representing the Red color, the second representing the Green color, and the third representing the Blue color [3].

With the cheapness and widespread of computer devices, many computer applications have spread dramatically around the world, giving rise to a new branch of computer science called "digital image processing", which allows image data to be captured and manipulated electronically to be suitable for use in several fields such as physics, chemistry, medicine, engineering, statistics, ... etc [4].

Image processing is a technique for enhancing the raw images received from the input devices -like cameras and sensors- as well as retrieving the maximum amount of image features and information [5], through a series of common steps namely image enhancement, noise reduction, segmentation, restoration, encoding, compression and features extraction [6].

Digital image processing has had an impact on several fields such as, Satellite and spacecraft remote sensing, image transmission and storage for business purposes, medical processing, radio detection and ranging (RADAR), sound navigation and ranging (SONAR), and acoustic image processing [7].

In fact, one of the most significant processes in digital image processing is feature extraction, which seeks to extract the most relevant information from the original image data and represents it in a smaller dimensionality space [8, 9]. There are numerous feature extraction approaches that can be categorized as either generic characteristics like color, texture, and shape or domain specialized features [10, 11].

Recently, GLCM has emerged as one of the most important image processing algorithms for extracting textural picture features [12, 13] by calculating how frequently pairs of pixels with given values and in a specified spatial relationship appear in an image [14]. GLCM-derived statistical measures include contrast, correlation, energy, homogeneity, correlation, and so on [15].

Unfortunately, with the rapid development of image processing techniques another bad aspect of image processing has emerged, as manipulation of digital images content has become much easier, not only for professional fakers, but also for the inexperienced counterfeiters through powerful software for editing images, like Microsoft Paint, Adobe Photoshop, Adobe Illustrator and Gimp [16].

As a result, a new area has emerged in the field of processing images named "digital image forensics", which has a purpose to collect quantitative evidence of digital image authenticity and to detect any illegal changes made in the source image [17].

This paper provides a new digital image tamper detection system based on GLCM and Statistical parameters. In the proposed system, the important features of the test image are extracted using GLCM approach and common Statistical measures. The same procedures are applied for the reference image and the matching process are performed through the Euclidean distance to determine whether the test image is original or fake.

II. Background

The theoretical background of the techniques which used in the proposed digital image forgery detection system can be described as follows:

A) GLCM

The GLCM approach stands for Gray-Level Co-Occurrence Matrix [18], and it is used to obtain the texture information of the given image [19]. The GLCM characterizes the texture of the image based on how often pairs of pixels with specific intensity values in a specified spatial relationship occur in the image [20]. In practice, features of GLCM are useful in a number of useful applications, such as image recognition, texture recognition, image retrieval, classification of images, … etc [21]. As for the Haralick, GLCM gives a number of texture features [22], the most important of them are the following four statistical properties [23]:

- **Contrast or Variance Inertia**

This property is used to estimate the local variations in the GLCM and is computed according to the mathematical formula shown in Eq. (1).

$$\textbf{Contrast}= \sum_{x,y=0}^{N-1} P_{xy}(x-y)^2 \quad (1)$$

- **Correlation**

This property is used to measure the occurrence of specified pairs of pixels of common possibility (which are associated with their adjacent pixels over entire frames) and is computed according to the mathematical formula shown in Eq. (2).

$$\textbf{Correlation}= \sum_{x,y=0}^{N-1} P_{xy}\frac{(x-\mu)(y-\mu)}{\sigma^2} \quad (2)$$

- **Energy or Angular Second Moment or uniformity**

This property is used to measure the sum of squared elements and is computed according to the mathematical formula shown in Eq.(3).

$$\textbf{Energy}= \sum_{x,y=0}^{N-1} (P_{xy})^2 \tag{3}$$

- **Homogeneity**

This property is used to evaluate the closeness of the distribution of the elements diagonally in GLCM and is computed according to the mathematical formula shown in Eq. (4).

$$\textbf{Homogeneity}= \sum_{x,y=0}^{N-1} \frac{P_{xy}}{1+(x-y)^2} \tag{4}$$

B) Statistical Moments

Statistics is the study of the collection, organization, analysis, and interpretation of data and includes many statistical measures which can be used in extracting image features, the most important of which are the following [24, 28]:

- **Mean**

This statistical measure is the average value of the data set and is computed according to the mathematical formula shown in Eq. (5).

$$\bar{X} = \frac{x_{(1)}+x_{(2)}+\cdots+x_{(n)}}{n} = \frac{\sum_{i=1}^{n} x_i}{n} \tag{5}$$

- **Median**

This statistical measure is the middle number or the number that splits data set in half and is computed as the following steps:

Step 1: Arrange data in increasing order.
Step 2: Compute the numbers of the data set = n.
Step 3: If n is odd, the median is the middle number, otherwise, if n is even, the median is the mean of the two middle numbers.

- **Standard Deviation**

This statistical measure is the square root of the variance and is computed for the sample according to the mathematical formula shown in Eq. (6), while it is computed for the population according to the mathematical formula shown in Eq. (7).

$$S=\sqrt{S^2}=\sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}} \tag{6}$$

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n}} \tag{7}$$

- **Kurtosis**

This statistical measure is a number that expresses the kurtosis of the distribution, i.e. whether the distribution is taller or shorter than the normal distribution and is computed according to the mathematical formula shown in Eq. (8).

$$\mu^4 = \sigma^4 \sum_{i=0}^{N-1}((i-\mu)^4 . p(i)) - 3 \quad (8)$$

- **Skewness**

This statistical measure is a number that measures the symmetry of the distribution and is computed according to the mathematical formula shown in Eq. (9).

$$\mu^3 = \sigma^{-3} \sum_{i=0}^{N-1}((i-\mu)^3 . p(i)) \quad (9)$$

III. The Proposed System

A) System Overview

In scientific literature, the GLCM approach is an image texture feature extractor that first computes a co-occurrence matrix, from which it uses several statistical features. In this study, the four fundamental features were considered: contrast, entropy, energy and correlation. On other hand, the Statistical measures are a descriptive analysis approach used to summarize the characteristics of a data set which can represent the whole population or a sample of it. The major six moments: mean, median, standard division, skewness and kurtosis are the focus of interest in the current study. To determine whether an image is original or fake, the proposed system relies on several steps. The features vector of the test image is generated by combining the GLCM features and the Statistical Moments, and the same procedures are performed for the reference image. The degree of similarity between the two images was achieved by comparing the two vectors using the Euclidean distance metric.Figure.1 shows a flow diagram of the proposed system.

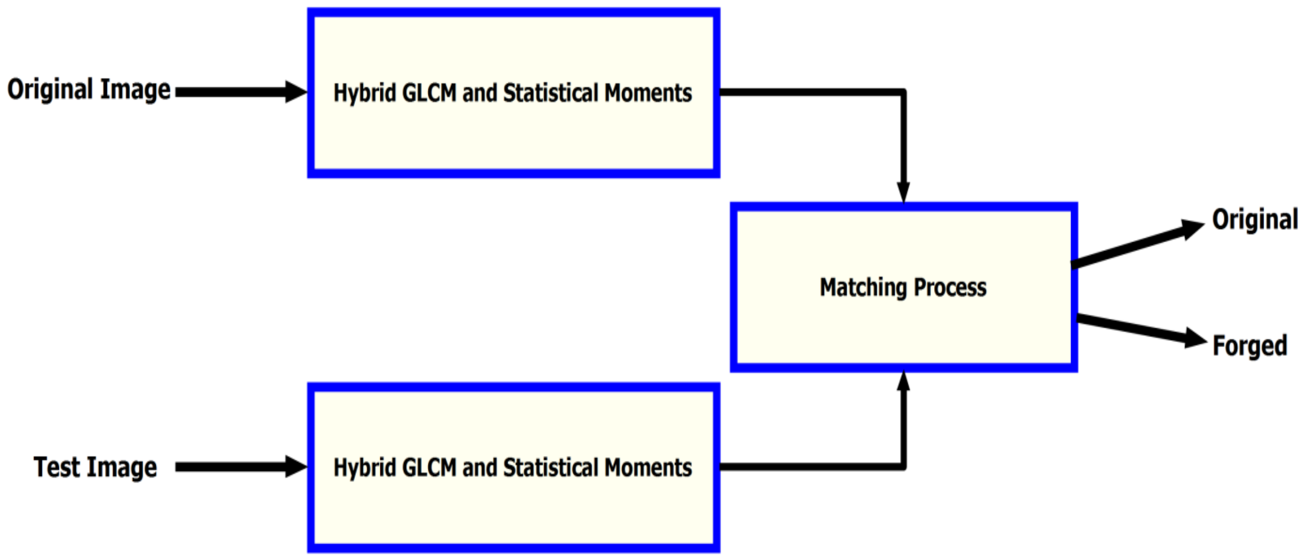

Fig.1: Flow diagram of the proposed system

From the previous Figure, the main steps can be summed up as follows:

[1]. Load the reference image.

[2]. Load the image to be tested.

[3]. Build the features vector for each of the two images as follows:

- **a)** Extract the GLCM features.
- **b)** Extract the Statistical Moments.
- **c)** Combine the extracted GLCM features and the Statistical Moments.

[4]. Match the two vectors using the Euclidean distance.

[5]. Classify the tested image as whether it is original or fake.

B) System Implementation

The proposed system runs on the MATLAB platform under Windows 7 (64-bit version). The first screen of the system is depicted in Figure.2. This screen displays crucial information to others such as (University, Faculty, Study Area, Name of Researcher, …).

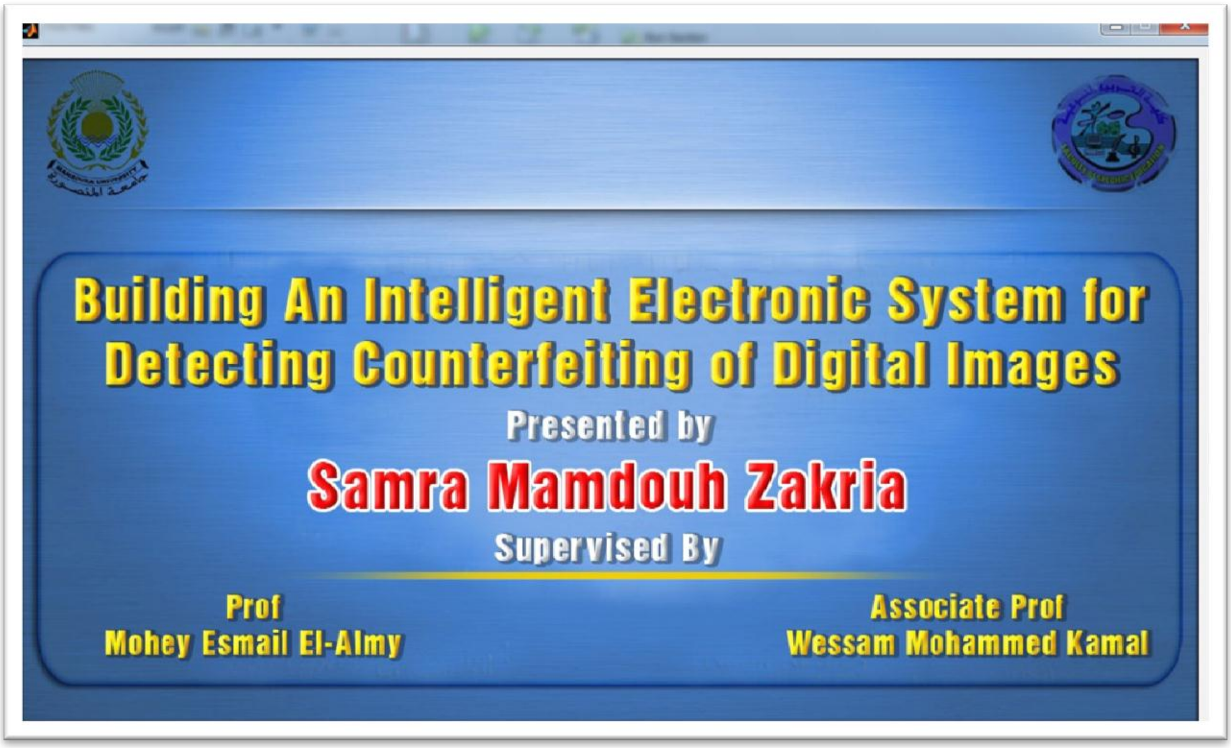

Fig.2: The start screen of the proposed system

After a few seconds, the screen automatically disappears and the Login screen of the proposed system shown in Figure.3 will be displayed. This screen enables the end users to use the proposed system by inserting the right login data.

Fig.3: The Login screen of the proposed system

If the authorized users enter the correct password, the Main screen of the proposed system shown in Figure.4 will be displayed.

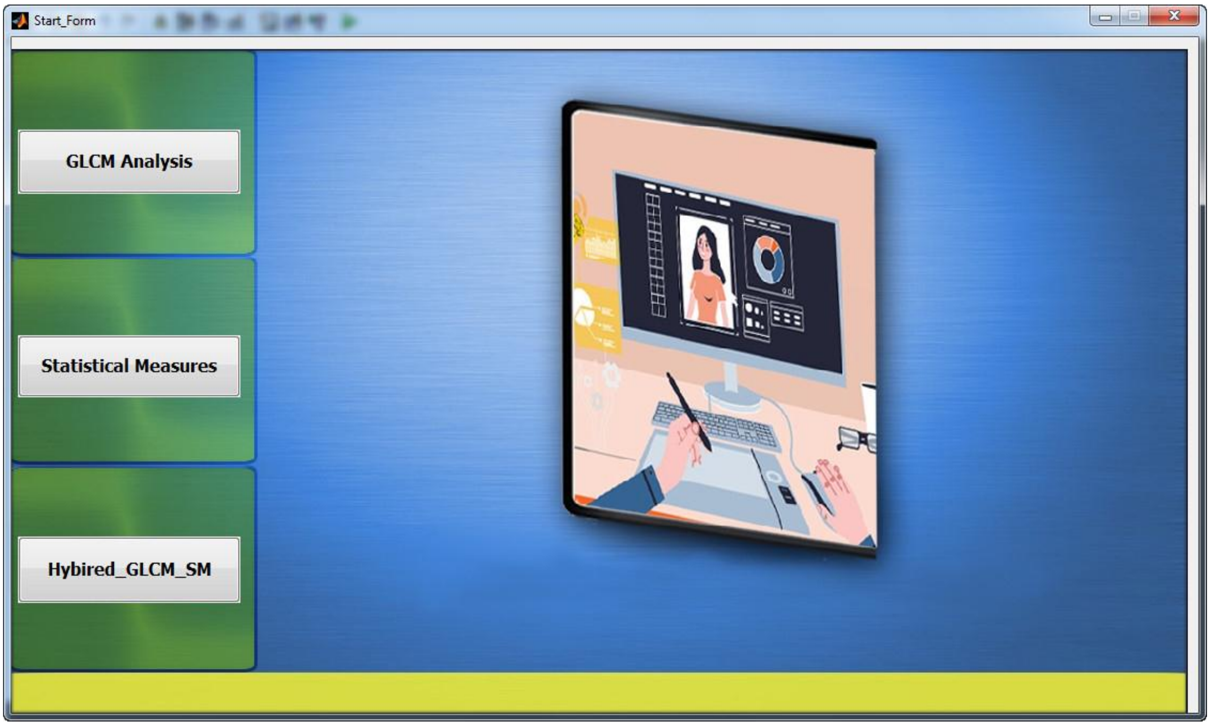

Fig.4: The main screen of the proposed system

Once the user clicks the "Hybired_GLCM_SM" button, the Matching screen of the proposed system shown in Figure.5 will be displayed. From this screen, after the user clicks the above "Browse" button, a pop-out box will be presented to enable the user to select the reference color image. After that, the source image features information will be shown in the small boxes and the selected reference image will be shown in the box which locates at the top of the screen. On other hand, the same operations are repeated at the bottom part of the screen once the bottom "Browse" button is pressed. Finally, the matching result appears by pressing the "Result" button.

Fig.5: The matching result screen of the proposed system

IV. Experimental Results

This section describes the experimental and assessment results on several digital images. Initially, the image datasets are identified, followed by the performance evaluation and findings. Finally, in a comparison process, the suggested methodology are compared to some major image analysis methods. To analyze the efficiency of the proposed system, it was liable for a serval of experimental tests, which was applied at two branches as follows:

A) Validity Test

The major goal of the validity test is to ensure the viability of the proposed system for use both inside and outside several universities. The test covered four major areas: content quality, design quality, organization quality, and user interface quality. Content quality is referring to the depth of information and insight contained within content. Design quality is referring to ultimate validation of design is to build the program and have its users reported their satisfaction. Organization quality is used to ensure that screen elements are organized properly. User interface quality is referring to ensure that interface has elements that are easy to access, understand and use.

When the final model of the proposed system was completed, it was sent to a group of computer science experts from Mansoura University for validation. To establish which components of the proposed system were accepted and which were not, each member of the assessment team was requested to complete a printed copy of the performance appraisal form, which included a number of Likert scale questions spanning the above-mentioned assessment scope.

The questionnaire forms were collected, filtered, and analyzed using quantitative statistical methods to analyze the raters' responses, the overall frequencies and the percentages of approved and disapproved for pre-specified evaluation aspects were calculated, as detailed in Table.1 and summarized in Figure.6.

Table 1: Validity Test Results

Evaluation Scope	Approved		Disapproved	
	Frequency	Percentage	Frequency	Percentage
Content Quality	46	92	4	8
Design Quality	47	94	3	6
Organization Quality	47	94	3	6

User Interface Quality	48	96	2	4
Total	**188**	**---**	**12**	**---**
Overall Average	**---**	**94**	**---**	**6**

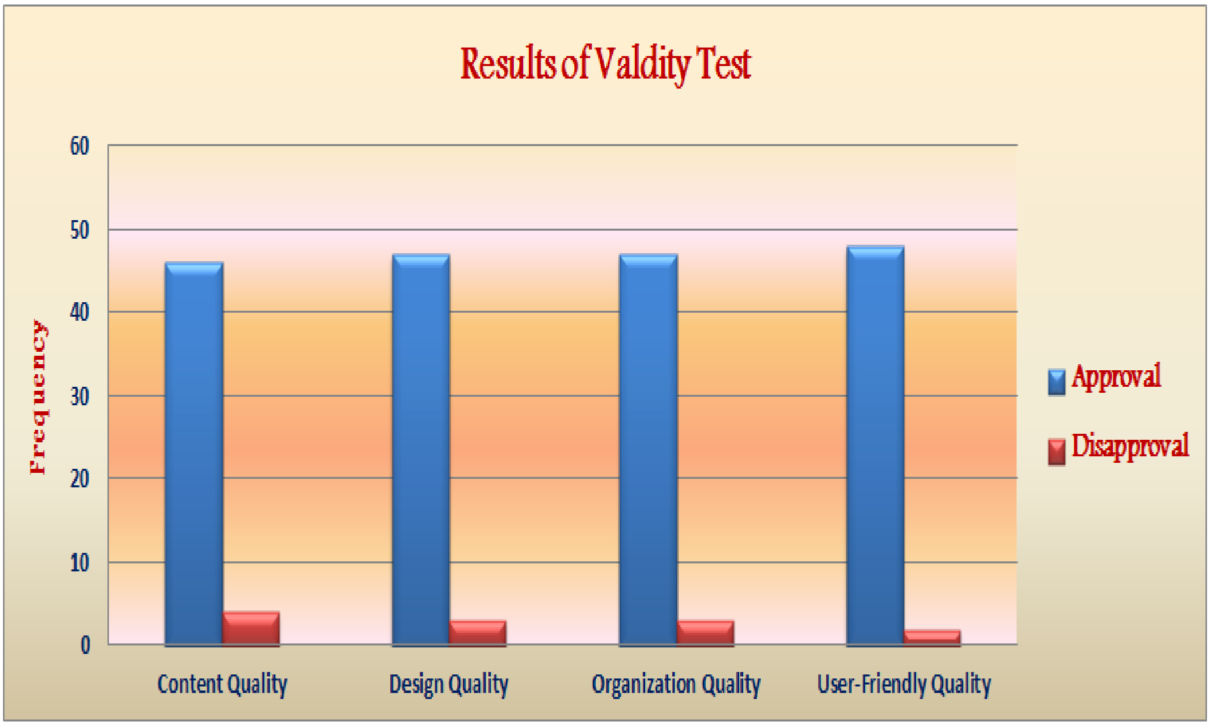

Fig.6: The validity test outcomes

The aforementioned findings reveal that, the overall percent of approved replies 94% for recognized evaluation features was higher than the overall percent of disapproved responses 6%.The data also demonstrates that the overall approval replies vary depending on the aspect, with the total approved of the content quality being 46 frequencies with 92%. The total application approved of design quality is 47 frequencies with 94%, total acceptance of organization quality is 47 frequencies with 94%, and total approved oval of user interface quality is 48 frequencies with 96%.

According to the numerical analysis of the acquired results, the overall average percentage of application approved replies for the predefined evaluation scope is 94%, showing a high degree of agreement among the evaluators on the appropriateness of the proposed system for general use.

B) Efficiency Test

The results of this test were based on the numerical values of the features extracted from the three methods of interest: GLCM, Statistical Moments and Proposed Hybrid Approach. The matching process was performed using the Euclidian Distance formula. The results of the comparison between the three previously mentioned methods are presented in the Table.2 and summarized in Figure.7.

Table 2: Efficiency Test Results

Test Image	GLCM		Statistical Moments		Hybrid Approach	
	Detection Type	**Euclidean Distance**	**Detection Type**	**Euclidean Distance**	**Detection Type**	**Euclidean Distance**
Image 1- Original	Original	0	Original	0	Original	0
Image 2 - Original	Fake	0.923	Fake	0.867	Original	0
Image 3 - Fake	Fake	0	Fake	0	Fake	0
Image 4 - Original	Original	0	Fake	0.902	Original	0

…..	…..	…..	…..	…..	…..	…..
…..	…..	…..	…..	…..	…..	…..
…..	…..	…..	…..	…..	…..	…..
Image 419 - Fake	Original	0.422	Fake	0	Original	0.306
Image 420 - Fake	Fake	0	Original	0.657	Fake	0
No. of Correctly Detected Images	**345**		**345**		**365**	
No. of Incorrectly Detected Images	**75**		**75**		**55**	
Overall Average	**True: 82.14** **False: 17.86**		**True: 82.14** **False: 17.8**		**True: 86.90** **False: 13.10**	

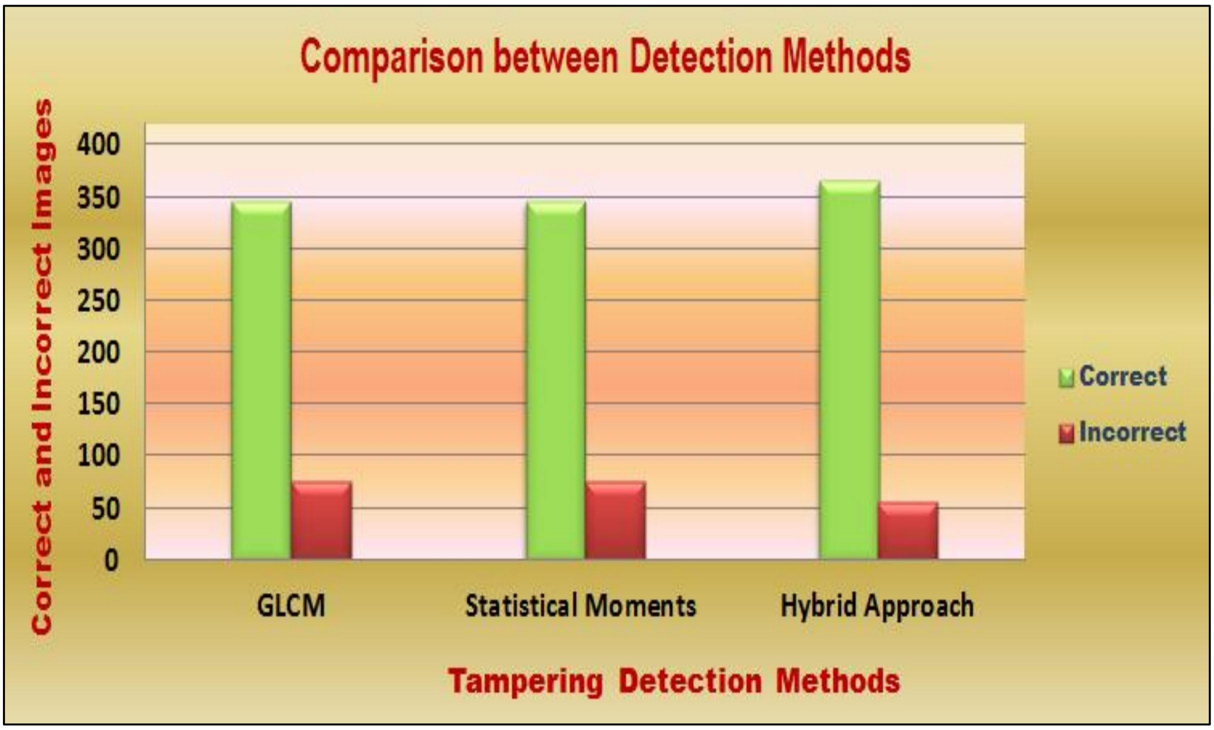

Fig.7: The number of correct and incorrect detection images

The aforementioned outcomes show that, the number of incorrect images using the proposed Hybrid approach was 55 images which it is less in comparison when both the GLCM and Statistical are used separately. In addition, the obtained outcomes indicates the average of correct detected images using the GLCM approach is 82.14 % and the average of incorrect detected images is 17.86 %. While, the average of correct detected images using the Statistical Moments approach is 82.14 % and average of incorrect detected images is 17.86 %. Finally, the average of correct detected images using the proposed Hybrid approach is 86.90 % and the average of incorrect detected images is 13.10 %.

V. Conclusion & Future Work

Current advances in digital multimedia processing techniques have made the field of image and video frames manipulation detection an important scientific research area. In the present research paper, the efficiency of the GLCM approach and the Statistical Moments approach were compared and then combined together to provide a new tool for detecting the authenticity of digital images and video frames. The proposed system is based on the hybrid method, where the features vector of the image under test is built by combining the GLCM features and the Statistical Moments features, and the same case is for the reference image, then the two vectors are compared using the Euclidean distance formula. Experiment results showed that, the proposed system does not consume much time, and is very effective in detecting tampering with digital images and videos.

Future plans for developing the proposed system emphasize providing the proposed system with many services to help visually impaired groups detect tampering with various digital multimedia files.

VI. References

[1]. Lee, J.H., Handling digital images for publication. Sci Ed, 2014. 1(2): p. 58-61.
[2]. Salim Al Jumaili, M., Digital image processing and applications. 2022.
[3]. Odabas Yildirim, E., et al., Color image splicing localization based on block classification using transition probability matrix. Wireless personal communications, 2023. 129(3): p. 1893-1919.
[4]. Tyagi, V., Understanding digital image processing. 2018: CRC Press.
[5]. Bhat, M., Digital image processing. International journal of scientific & technology research, 2014. 3(1): p. 272-276.
[6]. Luo, C., Y. Hao, and Z. Tong, Research on digital image processing technology and its application. 2018 8th International conference on management, education and information (MEICI 2018), 2018: p. 587-592.
[7]. Vijayaraghavan, S., Digital image processing. SCSVMV Deemed University, Kanchipuram.
[8]. Medjahed, S.A., A comparative study of feature extraction methods in images classification. International journal of image, graphics and signal processing, 2015. 7(3): p. 16.
[9]. Kumar, G. and P.K. Bhatia, A detailed review of feature extraction in image processing systems. 2014 Fourth international conference on advanced computing & communication technologies: p. 5-12.
[10]. Mohanaiah, P., P. Sathyanarayana, and L. GuruKumar, Image texture feature extraction using GLCM approach. International journal of scientific and research publications, 2013. 3(5): p. 1-5.
[11]. Goel, R., et al., A review of feature extraction techniques for image analysis. International journal of advanced research in computer and communication engineering, 2017. 6(2): p. 153-155.
[12]. Kiaee, N., E. Hashemizadeh, and N. Zarrinpanjeh, Using GLCM features in Haar Wavelet transformed space for moving object classification. IET Intelligent Transport Systems, 2019. 13(7): p. 1148-1153.
[13]. Santoni, M.M., et al., Cattle race classification using gray level co-occurrence matrix convolutional neural networks. Procedia computer science, 2015. 59: p. 493-502.
[14]. Alazawi, S.A., N.M. Shati, and A.H. Abbas, Texture features extraction based on GLCM for face retrieval system. Periodicals of engineering and natural sciences, 2019. 7(3): p. 1459-1467.
[15]. Öztürk, Ş. and B. Akdemir, Application of feature extraction and classification methods for histopathological image using GLCM, LBP, LBGLCM, GLRLM and SFTA. Procedia computer science, 2018. 132: p. 40-46.
[16]. Babu, S.B.G.T. and C.S. Rao, An optimized technique for copy–move forgery localization using statistical features. ICT express, 2022. 8(2): p. 244-249.
[17]. Shinde, A., G. Patil, and S. Kumar, Document image forgery and detection methods using image processing techniques–a review. International research journal of modernization in engineering technology and science. 4(8): p. 1077-1101.
[18]. Dhikhi, T., et al., Credit card transaction based on face recognition technology. Journal of Physics: Conference Series, 2019. 1362: p. 012109.
[19]. Gornale, S.S., G. Pattil, and R. Benne, Document image forgery detection using RGB color channel. Transactions on machine learning and artificial intelligence, 2022. 10(5): p. 1-14.
[20]. Ram, M.K., M.V. Rao, and C. Sujana, An overview on image retrieval using image processing techniques. International journal of computer science trends and technology (IJCST), 2017. 5(3): p. 83-88.
[21]. Goel, S.K. and M. Meduri, Ear biometric system using GLCM algorithm. I.J. Information technology and computer science, 2017. 10: p. 68-76.
[22]. Partio, M., et al., Rock texture retrieval using gray level co-occurrence matrix. Proc. of 5th Nordic signal processing symposium, 2002. 75: p. 1-6.
[23]. Lohithashva, B.H., V.N. Aradhya, and D.S. Guru, Violent video event detection based on integrated lbp and glcm texture features. Revue d'Intelligence Artificielle, 2020. 34(2): p. 179-187.
[24]. Kumar, V. and P. Gupta, Importance of statistical measures in digital image processing. International Journal of Emerging Technology and Advanced Engineering, 2012. 2(8): p. 56-62.
[25]. https://www.csueastbay.edu/scaa/files/docs/student-handouts/marija-stanojcicmean-median-mode-variance-standard-deviation.pdf
[26]. Groeneveld, R.A. and G. Meeden, Measuring skewness and kurtosis. Journal of the royal statistical society series D: The statistician, 1984. 33(4): p. 391-399.
[27]. Malliga, L., A novel statistical based methodology for the feature extraction of both MRI and CT images. International journal of engineering and advanced technology, 2019. 8(6S3): p. 1788-1794.
[28]. Kardian, A.R., S.A. Sudiro, and S. Madenda, Efficient implementation of mean, variance and skewness statistic formula for image processing using FPGA device. Bulletin of electrical engineering and informatics, 2018. 7(3): p. 386-392.