



Ministry of Higher
Education and Scientific
Research
University of Diyala
College of Science
Department of Computer
Science



Modified Efficient Forensic Technique for Detecting the Copy- Move Forged Digital Images

**A Thesis Submitted in Partial Fulfillment of the
Requirements for the Master Degree in Computer Science**

**Department of Computer Science/ College of
Science/University of Diyala
Iraq / Diyala**

By

Mokhles Hussein Khudhur

Supervised By

Assist. Prof. Dr. Jumana Waleed

Prof. Dr. Dhahir Abdulhade Abdullah

Acknowledgment

First of all, praise is to GOD, the lord of the whole creation, on all the blessing was the help in achieving this research to its end.

*I wish to express my thanks to my supervisors, **Assist. Prof. Dr. Jumana Waleed and Prof. Dr. Dhahir Abdulhade Abdullah** for supervising this research and for the generosity, patience and continuous guidance throughout the work. It has been my good fortune to have the advice and guidance from them. My thanks to the academic and administrative staff at the Department of the computer sciences.*

I would like to express my gratitude to my family.



Mokhles Hussein Khudhur

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

﴿ نَرْفَعُ دَرَجَاتٍ مِّنْ نَّهَارٍ وَمَن يُّفَوِّقْ كُلَّ نَبِيٍّ عَلِيمٌ ﴾

صَدَقَ اللَّهُ الْعَظِيمُ

سورة يوسف

الآية (76)

Abstract

Images represent an effective and natural communication media for humans, due to their immediacy and the easy way to understand the image content. due to the widespread availability of digital devices, various open source and commercially available image editing tools have made authenticity of image contents questionable. This will lead to increase the need of using forgery detection algorithms. Copy-move forgery (CMF) is a common technique to produce tampered images by concealing undesirable objects or replicating desirable objects in the same image. Therefore, means are required to authenticate image contents and identify the tampered areas. Many digital image copy-move detection algorithms have been developed, algorithms based on Discrete Cosine transform (DCT), algorithms using invariant image moments, algorithms using texture and intensity descriptors, algorithms using invariant key points, algorithms based on mutual information, and algorithms based on SVD to determine the existence of digital image forgery.

In this thesis, two robust techniques for CMF detection and identification in digital images are proposed. DCT based technique uses DCT coefficients to extract features and the Framing technique uses set of frames applies for each block to extract features ,these features used for exposing the forgeries in digital images and determine whether the content is authentic or modified without depending on any knowledge of prior information related to the source image. The dimension of the feature vectors is reduced by applying discrete cosine transform (DCT) in the DCT based technique and by frames applied on overlapped blocks in framing technique, to evaluate the proposed techniques, images forged by Gnu Image Manipulation Program (GIMP) common application for

experimentations has been used. The proposed forgery detection techniques can be applied to detect the tampered areas and the benefits can be obtained in image forensic applications.

MATLAB R2010 has been used to build the two techniques and GIMP application utilized to create the forgery to be used in experiments. The performance analysis showing that the DCT based technique can detect the multi-duplicated regions with 99% accuracy ratio, with 5.90075 seconds of processing time. While, the framing technique can detect the multi-duplicated regions with 99% accuracy ratio even when an image was modified by JPEG compression, rotation, and scaling conditions. Also, it reduced the processing time to 2.8708 seconds.

Table of Contents

ABSTRACT	
Table Of Contents	
List Of Tables	
List Of Figures	
List of Abbreviations	
1 INTRODUCTION	1
1.1 Introduction.....	1
1.2 Related Work	2
1.3 Problem Statment.....	5
1.4 Aim of Thesis.....	6
1.5 Thesis Layout.....	6
2 THEORITICAL BACKGROUND	7
2.1 Introduction.....	7
2.2 Image Forensics	9
2.2.1 Active Approaches.....	10
2.2.2 Passive-Blind Approaches	10
2.2.2.1 Types of Passive Image Forgery	11
2.3 Copy-Move Forensic based Algorithms	14
2.3.1 Algorithms Based on DCT.....	18
2.3.2 Algorithms Based on Invariant Image Moment.....	19
2.3.3 Algorithms Using Texture & Intensity Descriptors.....	20
2.3.4 Algorithms Based on Invariant Key Points	20
2.3.5 Algorithms Based on Reciprocal Information	21
2.3.6 Algorithms Based on SVD.....	21
2.4 Digital Image File Format.....	22
2.4.1 Binary Images	22
2.4.2 Grayscale Images	22
2.4.3 Color Images	23
2.5 K- Mean Clustering.....	24

2.6	Performance Masurment	25
3	THE PROPOSED FORENSIC TECHNIQUES	27
3.1	Introduction.....	27
3.2	The Proposed Techniques	27
3.2.1	The DCT based Proposed Technique	28
3.2.2	The Framing Proposed Technique	38
4	THE EXPERIMENT RESULTS	44
4.1	Introduction.....	44
4.2	Experiment Results for the DCT based technique	44
4.3	Experiment results for the Framing technique.....	48
5	CONCLUSION AND FUTURE WORKS	59
5.1	Introduction.....	59
5.2	Conclusion	59
5.3	Future works	60
	REFERENCES	61
	PUBICATIONS	68

List of Tables

Table 4.1: The time consumption results of detecting forgery using DCT based technique.....	46
Table 4.2 : The time consumption results of detecting forgery using Framing technique.....	47
Table 4.3 : The time consumption results of detecting forgery using Framing technique	50
Table 4.4 : Time consumption results of testing forgery detection with parameters $F=4,5$ and $K = 5,10,20$ using both techniques and comparing between them.....	52
Table 4.5: A Time comparison between different recently existing forensic techniques and the proposed techniques.....	53
Table 4.6: Time consumption to detect forgery in scaled images using the Framing technique.....	55
Table 4.7: Time consumption to detect forgery in rotated images using the Framing technique.....	56
Table 4.8: True positive and false positive for image in Figure 4.9.....	57
Table 4.9: The accuracy rate under different conditions.....	58

List of Figures

Figure 1.1: The No. of publications in the "IEEE"	2
Figure 1.2: The No. of publications in the "Science Direct"	3
Figure 2.1: Digital Image Forensic Techniques.....	9
Figure 2.2 : Copy-Move Forgery image	11
Figure 2.3: Image splicing	12
Figure 2.4: Image Resampling	13
Figure 2.5: Image retouching	13
Figure 2.6 : Image Morphing	14
Figure 2.7 : Image Created by Graphical Software	14
Figure 2.8 : The general structure of the copy-move detection.	17
Figure 2.9 : The 64 Basis Functions of an 8x8Matrix	19
Figure 2.10 : Color Image.....	23
Figure 3.1 : The structure of the DCT based proposed technique	29
Figure 3.2 : The process of feature selection	31
Figure 3.3 : An instance of radix sort	36
Figure 3.4 : The structure of the Framing proposed technique.....	39
Figure 3.5 : Block division.....	40
Figure 3.6 : An instance of block rotation	41
Figure 3.7 : The blocks clustering and sorting into classes	43
Figure 4.1 : The various potential locations of multiple duplicated regions	46
Figure 4.2 : The time consuming for detecting the forgery	47
Figure 4.3 : The time consumptions for detecting the forgery	48
Figure 4.4 : The various potential locations of multiple duplicated regions	50
Figure 4.5 : The time consumptions for detecting the forgery	51
Figure 4.6 : a comparison between both techniques	52
Figure 4.7 :Copy move detection technique Framing technique in scaling conditions	54
Figure 4.8 : copy move forgery detection technique Framing after different Rotation conditions	56
Figure 4.9 : the result of different throusholds.	57

Abbreviations table

Abbreviation	Description
SIFT	Scale Invariant Feature Transform
SIVA	Sample, Information and Value Analysis.
RGB	Red, Green and Blue
CMY	Clay, Magenta And Yellow
CMYK	Clay, Magenta, Yellow and Black
TPR	True Positive Rate
FPR	False Positive Rate
FNR	False Negative Rate
YCBCR	The Ycbcr Color Space
MSD	Most Significant Digit
RAM	Random Access Memory
GIMP	GNU Image Manipulation Program
BMP	Bitmap Image

Chapter One

Introduction

CHAPTER ONE

INTRODUCTION

1.1 Introduction

Digital images are the foremost source of information and they are the fastest means of information convey. As an evidence for any event in the court of law images can be useful. Digital images are being used in many applications like military, medical diagnosis, art piece, photography etc. The ease of use and accessibility of software tools [1] and low-cost hardware, makes it very simple to forge digital images leaving almost no trace of being subjected to any tampering. So, it becomes difficult for humans to trace these manipulations. As a result, the integrity and authenticity of digital images is lost. This modification of images can be done for hiding some important traces from an image, to change the details of an image etc. so that incorrect information is transmitted [2]. This challenge the reliability of digital images offered as medical diagnosis, as evidence in courts, as newspaper items or as legal documents because of difficulty in differentiating original and modified contents [3].

For authenticating an image, various authentication approaches have emerged. Commonly, these authentication approaches are classified into Active and Passive approaches [4]. The DCT based encompass the digital signature techniques, and data embedding techniques like watermarking [5,6] which need specific information to be included into an image through the creation, or before publication. While the Framing approach encompass of the image splicing and image copy-move forgery which work without the need to the protection techniques and without any prior information concerning the image under

analysis. the passive approaches can be considered a suitable solution for making a trustworthy decision about the authenticity of an image.

Furthermore, as the most common type in the authentication approaches, copy-move forgery detection is classified into block-based and key point-based techniques [7]. This thesis proposes a blind digital image of a block-based forensic techniques for checking the image authenticity.

1.2 Related Works

Digital image forensics techniques have developed sufficiently to resist the digital image forgery problem in different areas such as medicine, sports, legal services, and intelligence. There are a large number of works have been appeared in the areas of detecting the digital image Copy Move Forgery (CMF) , and this is obvious in the Figure (1.1) and Figure (1.2) that illustrate the papers numbers which addressed the CMF detection techniques in Science Direct and IEEE in the last ten years.

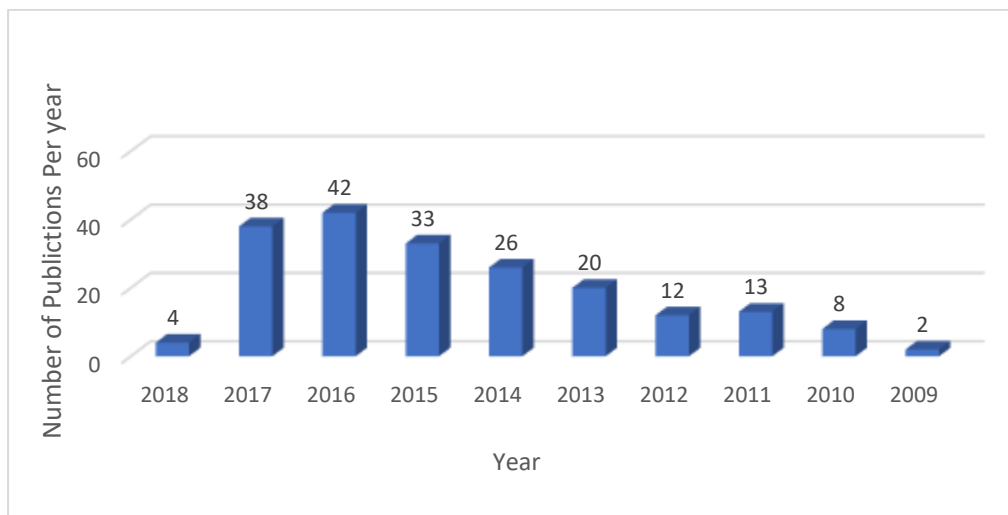


Figure 1.1: The No. of publications in the "IEEE" within the field of CMF detection techniques

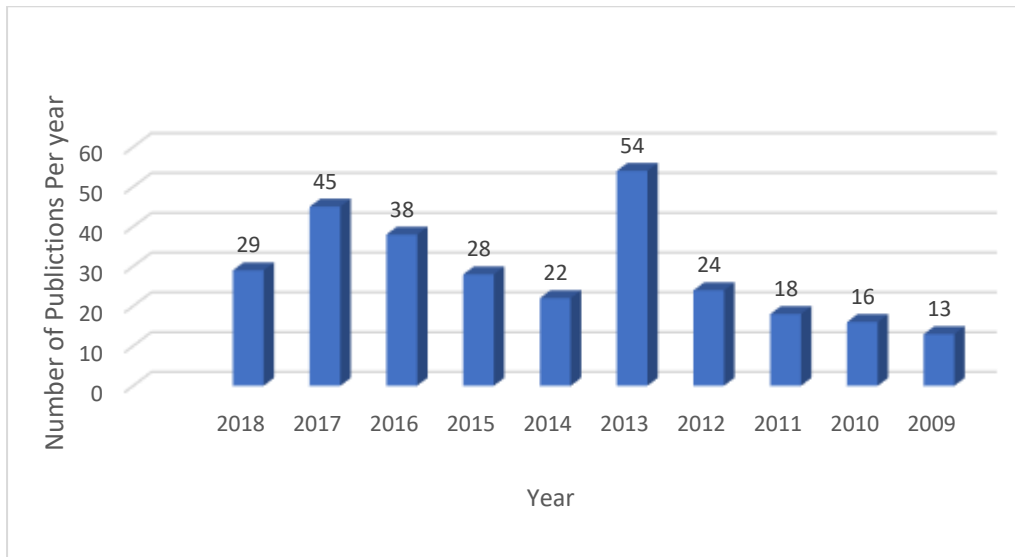


Figure 1.2: The No. of publications in the "Science Direct" within the field of CMF detection Techniques

The block-based techniques of detecting copy-move forgery work on dividing the image into overlapped blocks and utilize different methods for extracting features from these blocks. Finding a similarity between the extracted features vectors represents an evidence to exist forgery. There are lots of researchers which proposed different techniques in the topic of block-based copy-move forgery detection.

- Fridrich et al, 2003.[8] firstly, offered a technique by utilizing exhaustive search; After that suggested an effective block matching detecting technique depending on discrete cosine transform (DCT). The main idea behind using an algorithm based on DCT is to use its coefficients as a feature to be compared to find the repeated blocks. In the context of accuracy, this technique show in results multi false detections especially on flat areas such as clouds, grass, sky ..etc.
- Popescu et al, 2004 [9] suggested a technique which utilizes the principal component analysis (PCA) rather than DCT. Owing to the attributes of PCA, the features needed to represent the block was decreased to approximately half compared with the features utilized by Fridrich. Therefore, the technique which uses the PCA has a preferable

time complexity; But this technique has the low robustness to small rotations of copy-moved regions.

- Yanping Huang et al, 2011. [10] present the usage of an algorithm based on improved DCT through dividing the image into fixed-overlapping blocks and applying DCT on each block for representing its features. These features are truncated for reducing the dimension and lexicographically sorted for neighboring the duplicated blocks in the sorted list. Matching between blocks is applied to detect duplications. Here the researchers suppose that the duplicated regions are not overlapping. The process of the detection is for determining if the digital image includes duplicated regions. Because the size and shape of regions are unknown, it is computationally impossible for trying to examine each potential pair of the region with different size and shape. It is very efficient to partition the digital image into fixed sized of overlapping blocks and do the experiment if the blocks pairs are duplicated. The main advantages of this algorithm that it used fewer features to represent each block. This algorithm showed an ability to detect copy-move forgery in an image that is quite robust to JPEG compression, blurring or white Gaussian noise distortion. But the researchers did not mention to one main disadvantage that this method could not take a decision with rotated or reflected image.
- Nathalie Diane W. et al, 2013. [11] have used the same methodology to copy-move forgery detection based on DCT, except the usage of the Euclidean distance as a similarity criterion to identify the duplicated blocks, this technique showed that it could detect multi-duplicated regions but with small factor of scale, rotation and distortion.
- Davarzani et al, 2013. [12] proposed a multiresolution local binary patterns block-based technique in which lexicographical sorting and k-d

tree are used for speeding up the process of block matching, but this technique remains consuming time.

- Jen-Chun Lee,2015. [13] presented a histogram of orientated Gabor magnitude block-based technique which is capable of detecting multiple copy-move forgeries in the same image , in this technique the author developed a noise detector to reduce the probability of false matches, in practice it shows a high level of false positive result of matching ,reducing the false positive will increase time consumption.
- B. Ustubioglu et al,2016. [14] proposed a discrete cosine transform DCT block-based technique with a high degree of accuracy and low false negative. In this technique, the process of feature extracting is done by utilizing the similarity of the element by element between the feature vectors rather than utilizing the cross-correlation, Experimental results show that the method yields higher accuracy ratios ,but also show lower false negative values.
- Sondos M. Fadl et al ,2017. [15] presented a Fourier block-based technique in which the Polar representation is used for getting the representative features to the blocks. This technique also provides an efficient detection for the copy-move regions, but the execution time needs to be improved.

1.3 Problem Statement

Digital images are most popular representation of information sharing. This popularity creates an opportunity for the researchers to ensure trustworthiness of images. The forensic detection of an image is performed using various techniques to ensure its credibility. Due to the advancement in image forgery methods, a tampered region of an image is hard to detect with bare human eyes. It has become crucial to develop methods to detect more sophisticated image forgeries in the large number of available images.

1.4 Aim of Thesis

This thesis aims to handle copy – move forgery issues in block-based techniques by proposing a modified methods in similarity matching step, considerably improves speed of the calculating process and enhance the algorithm performance.

1.5 Thesis Layout

The rest of this thesis is:

Chapter Two: Theoretical Background

This chapter gives the background and review of the basis for algorithms and techniques, especially, that are used in this thesis.

Chapter Three: The Proposed System

This chapter describes the proposed system with its design and implementation.

Chapter Four: Results and Discussion

This chapter explains the results that have been gotten from the proposed system with discussion.

Chapter Five: Conclusions, and Suggestions for Future Works

This chapter presents the conclusions about this work. Also, the suggestions for future works.

Chapter two

Theoretical Background

CHAPTER TWO

THEORETICAL BACKGROUND

In this chapter, the first section presents a brief introduction to the image forgery detection approaches and the second section summarizes these approaches and focused on passive-blind approaches. The recently used copy-move forensic based algorithms are illustrated in the third section; The digital image file format and K- Mean clustering are illustrated in the fourth and fifth sections. Finally, the performance measurements are shown in the last section.

2.1 Introduction

Nowadays, the image is the most popular manners of communication, the image can easily, correctly, and quickly be carrying any idea between the recipients. The vast domain of applications leads the image to be most affected by fraud and tamper. The swift spread of cheap and simple to use devices which qualify the visual data acquisition makes approximately everyone able to record, store, and share many digital images. The large availability of image editing software tools makes extremely simple to alter the content of the images. Therefore, there is no confidence that anything appears in a photo is a real representation of what truly occurred. The photography value should be carefully evaluated as events record. This necessity comes from a various range of applications; The most significant one is the scenario of forensics, in which the reliability of the image should be confirmed before utilizing it as a prospective evidence.

The image forensics (IF) is the science addressing the identification, validation, analyzing, and interpretation of the digital images as a prospective evidence [17]. IF aims to understand if the given image is a combination of

various shots. Generating a forgery commonly needs some steps of processing. Permanently, these steps remain some statistical traces in the signal. There are different operations which happen through forgery are; cropping, blurring, adding noise, rotation, scaling, compression, down sampling, resizing, retouching, and etcetera [18]. The revolution of digital information and matters related to the security of multimedia has created different approaches in the field of tampering detection and IF [19]. The significance of appearing different approaches of forgery detection is, when a situation between two parties, taking any decision depending on the given forged images without having the original images is extremely difficult and will lead to disastrous results. Therefore, it is tricky to detect these manipulations. As a result, the authenticity and integrity of images are lost. The alteration of digital images can be utilized in several malicious purposes such as for hiding some significant traces of an image or to transmit incorrect information. And to identify the integrity of these images, there is a necessity for detecting if received images are forged or not [20].

Recently, different approaches have been presented for tracing the digital image forgery. Generally, these approaches are categorized into active and passive-blind. Active approaches are classified into the data hiding such as watermarking [21] and digital signature approaches. In contrast to active approaches, passive-blind approaches work without using any techniques of protection and with the absence of any previous information concerning the image. For detecting the tampering traces, the blind approaches utilize the fact that the forgery can leave specified detectable modifications to the image such as statistical changes. We focused on passive-blind approaches especially CMF detection, regarded as a new trend. Different algorithms based CMF are founded to detect the digital image forgery or the image manipulation which had a potential optimization to provide the accurate decisions without depend on any previous information related to the original image as algorithms based on DCT, algorithms use the invariant image moments, algorithms using texture and

intensity descriptors, algorithms use invariant key points, algorithms based on the mutual information, and algorithms based on SVD to determine the existence of digital image forgery.

2.2 Image Forensics (IF)

The digital forensics domain is developing considerably to resist the IF problems in different fields such as sports, medical images, legal services, and intelligence [22]. These digital images can be given as proof for the court of law. In such cases, it becomes extremely significant for proving the originality of digital images. IF plays a dynamic role in these cases by examining authenticity and integrity of digital images [23]. For proving authenticity of digital images, different approaches have been presented which are generally divided into active and passive approaches; Figure (2.1) illustrated the IF techniques.

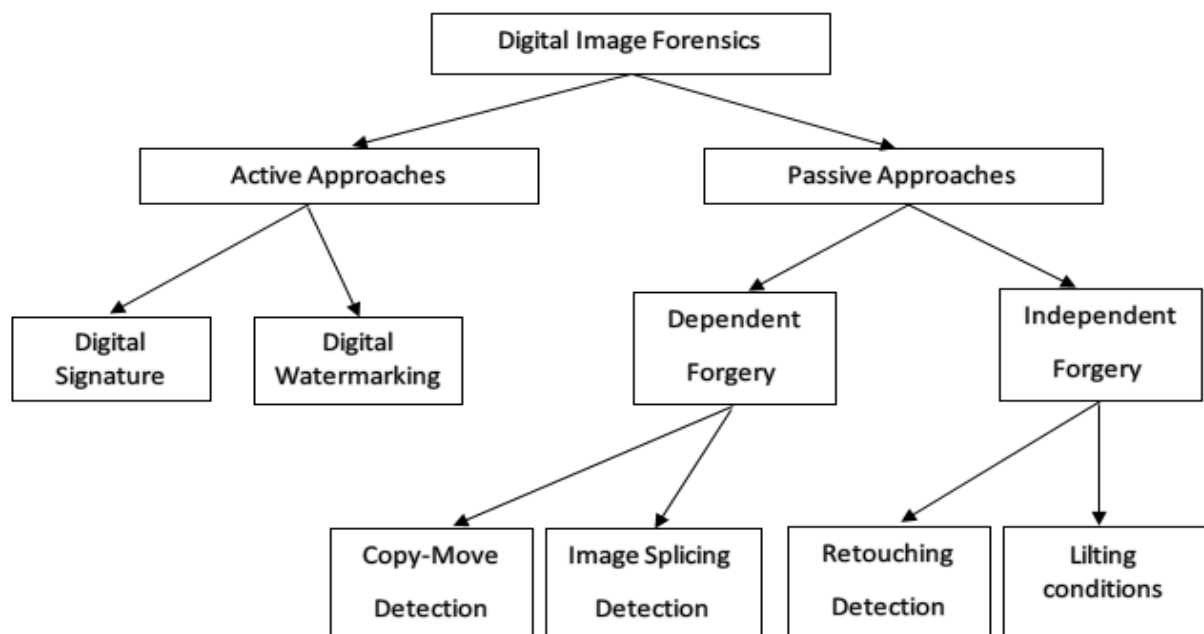


Figure 2.1: Digital Image Forensic Techniques [24].

2.2.1 Active Approaches

With the active approaches, the authentication process needs prior information concerning the image. It is related to embedding the data where at the generation time a code is hidden into the image. Proving this code will lead to authenticate the originality of the digital image. [25]. Manipulation of the image consists of many processing operations like scaling, rotating, blurring, brightness adjusting, change in contrast, etc. or any combination of these operations. Doctoring image means pasting one part of the image into another part of the image, skillfully without leaving any trace. One important tool for the authenticity of the digital image is the watermarking and digital signature [26].

2.2.2 Passive-Blind Approaches

Passive-blind approaches are proving the authenticity of the digital images without the need for prior information, only the image itself. It assumes that although the manipulation may do not leave any perceptible traces, it is probably to change the implicit statistics. These inconsistencies can be utilized for detecting the forgery. The passive-blind approaches became extremely significant to pass the difficulties of active approaches represented by the preceding knowing of images content, and the time of processing to hide a watermark or signature in a digital image; also, the wasting of processing time to examine the authenticity at the receiver side [26]. Passive-blind approaches could be categorized into dependent and independent forgery approaches [25]. The forgery-dependent detection approaches are developed for detecting specific types of forgery like splicing and copy-move that are based on the kind of forgeries accomplished on an image. Whilst the forgery-independent approaches detect forgery without reliance on the type of forgeries but depending on

tampering traces left through the operation of resampling and lighting inconsistencies [27].

2.2.2.1 Types of Passive Image Forgery

There are several types of passive image forgery, most of them are as follows:

1. Copy-Move Forgery: Copy-move forgery is the most spread forgery, especially, in forgeries that using individual image to duplicate or hide one or more objects in the same image [27]. It is performed by copying a region from the original image and pasting it into the same image to hide or duplicate specific objects in the image to produce the forged image. Copy-move forgery is simple to carry out and can be relatively effective in image manipulation, particularly when both source and destination regions are from the same image as properties of both such as color temperature, illumination conditions and noise will generally be matched between the tampered region and the source image. Therefore, it would be difficult to detect by the naked eye. In copy-move forgery, the common manipulated areas in the image are found to be grass, foliage or fabric. These areas are easy to blend in the background due to similarities in the texture and color as shown in Fig 2.2 a part of image used to hide specific object in the image [28].



Figure 2.2: Copy-Move Forgery [30].

2. Image Splicing forgery: Image splicing is the same as the copy-move forgery, but the copied regions are not collected from one image, two or more images are involved [29]. It is performed by copy one or more regions from two or more images and combine these regions into one new tampered image as shows in Figure 2.3. Using of different regions and features from different images may makes the effects of image splicing forgery cleared and difficult to detect when combine them into one new image [30].



Figure 2.3: Image splicing [30].

3. Image Resampling Forgery: Image resampling based on generating a new image with adjusting or modifying the height/width of a particular object in image or in all content of the image. Resizing the image means changing the object dimensions only to appear larger but not to improve the quality of that object. The indication step plays a key role in the resampling method and presents insignificant statistical variations. Resampling introduces certain periodic correlations obsessed by the image. These correlations can be used to detect forgery affected by resampling [31], as shown in Figure (2.4).



Figure 2.4: Image Resampling [31].

4. Image Retouching Forgery: Image retouching manipulates an image by enhancing or reducing certain features of the image without making significant changes on image content [32]. Image retouching forgery clue is to enhancing the image to show or hide a specific feature such as coloring, illumination or background altering to attract attention or to distract attention about specific object inside the image [31], as shown in Figure (2.5).



Figure 2.5: Image retouching [31].

5. Image Morphing Forgery: It is an image forgery where one object into the image is transformed into another object in the target image. An example for Morphing is shown in Figure (2.6), where left image is the source image and the right is the morphed image [33]. In Image morphing, the shape of an image is progressively changed into another shape in another image and it must be used between two images.



Figure 2.6: Image Morphing [30].

6. Image Created by Graphical Software: An image is produced by graphical software by utilizing a computer and its applications to generate a forged image not associated with realism by structure its ideas and features by computer as illustrates in Figure (2.7).



Figure 2.7: Image Created by Graphical Software [30].

2.3 Copy-Move Forensic based Algorithms

Copy-Move is the most common and popular photographs forgery approach due to the ease that can be achieved through the use of it [23]. It includes the copying of some regions on the image and moving them to other regions on the image [34]. Because the copied regions belong to the same image, so, the dynamic area and color stay consistent with the remainder of image [38].

In this part, we attempt to give a guide of the investigations that seemed in this field by focusing on the recently used copy-move forensic based algorithms; displaying and comparing the advantage besides disadvantage separately.

The general algorithm used in the CMF detection should come as follows:

1. Image conversion:

Convert the forged image with size $(M \times N)$ into a grayscale.

$$\text{Grayscale Image} = 0.22 R + 0.587 G + 0.11 B \quad (2.1)$$

Or

Convert the forged image with size $(M \times N)$ into a YCbCr

$$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} + \begin{bmatrix} 65.481 & 128.553 & 24.966 \\ -37.797 & -74.203 & 112.000 \\ 112.000 & -93.786 & -18.214 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (2.2)$$

2. Blocking stage

for an image of size $M \times N$, the image could be divided into small overlapping blocks of $b \times b$ pixels resulting in B blocks where :divide the image into overlapping blocks with size $(b \times b)$ to find $O = (M - b + 1) \times (N - b + 1)$ of overlapping blocks; And these blocks are rearranged into a row vector R_i with size $(O \times K)$; where $K = b^2$. After that, for each block, the features are extracted and stored as rows depending on used algorithm.

$$\text{Blocks Number} = (I - t + 1)(J - t + 1) \quad (2.3)$$

3. Features Extraction:

In this stage, DCT transformation is applied on each block. Assume the size of the block is $b \times b$, there are elements in the matrix [35]. As it is the nature of DCT that not all the elements are equally important. To

facilitate length reduction step, DCT coefficients have been reshaped to a row vector in zigzag order Fig. 3. Coefficients of indices greater than l (e.g. $l=9$) have been truncated to reduce the dimension of the features and reduce processing time. The final feature vector contains l DCT coefficients (assume $l=9$) as well as 2 indices for block position.

4. The Block Clustering

K-means is a clustering algorithm that groups similar objects based on features into number of group. is positive integer number. The grouping is done by minimizing the sum of squares of distances between data and the corresponding cluster centroid. The accelerated (Fast K-Means) algorithm avoids unnecessary distance calculations by applying the triangle inequality and keeping track of lower and upper bounds for distances between points and centers.[36]

5. Block sorting stage:

Perform the lexicographical sorting on the rows of the extracted feature matrix. This leads to getting identical blocks.

6. Apply the matching process. In the conventional matching, the pairs of the feature vector are searched among the nearest neighbors utilizing a threshold. If R_{ij} is the matched pair including the features R_i and R_j , where $i \neq j$ represents indices of the feature; After that, the shift vector SV between two matching blocks is computed as:

$$SV_{ij}(d_x, d_y) = (X_i - X_j, Y_i - Y_j) \quad (2.4)$$

The counter of shift vectors CO (SV) is taken out and incremented by one for each matching pair of blocks with the same shift [37]. as in the following equation:

$$CO(d_x, d_y) = CO(d_x, d_y) + 1 \quad (2.5)$$

Matching pairs of blocks with the same shift are grouped. Ignore the groups of blocks with a count of shift vector below θ . The threshold θ controls the size of the little copy-move detectable segment [24].

Figure (2.8) illustrates the general structure of the copy-move forensic based algorithms.

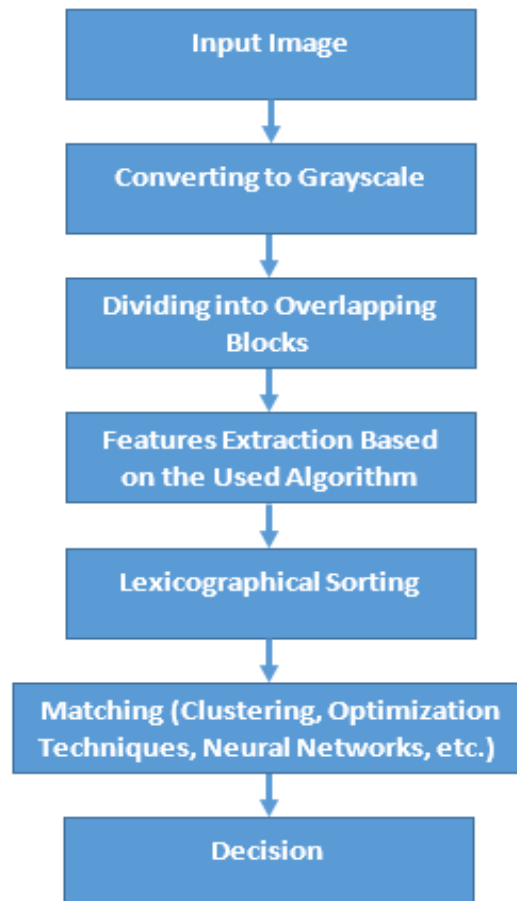


Figure 2.8: The general structure of the copy-move detection [24]

2.3.1 Algorithms Based on DCT

The discrete cosine transform (DCT) represents an image as a sum of sinusoids of varying magnitudes and frequencies. The DCT has the property that, for a typical image, most of the visually significant information about the image

is concentrated in just a few coefficients of the DCT. For this reason, the DCT is often used in image compression applications. For example, the DCT is at the heart of the international standard lossy image compression algorithm known as JPEG.

The definition of the two-dimensional DCT for an input image A and output image B is:

$$B_{pq} = \alpha_p \alpha_q \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} A_{mn} \cos \frac{\pi(2m+1)p}{2M} \cos \frac{\pi(2n+1)q}{2N}, \quad \begin{matrix} 0 \leq p \leq M-1 \\ 0 \leq q \leq N-1 \end{matrix} \quad (2.6)$$

$$\alpha_p = \begin{cases} \frac{1}{\sqrt{M}}, & p = 0 \\ \sqrt{2/M}, & 1 \leq p \leq M-1 \end{cases} \quad \alpha_q = \begin{cases} \frac{1}{\sqrt{N}}, & q = 0 \\ \sqrt{2/N}, & 1 \leq q \leq N-1 \end{cases} \quad (2.7)$$

Where M and N are the row and column of A , respectively. The values B_{pq} are called the DCT *coefficients* of A .

The inverse DCT equation can be interpreted as meaning that any M -by- N Matrix A can be written as:

$$\alpha_p \alpha_q \cos \frac{\pi(2m+1)p}{2M} \cos \frac{\pi(2n+1)q}{2N}, \quad \begin{matrix} 0 \leq p \leq M-1 \\ 0 \leq q \leq N-1 \end{matrix} \quad (2.8)$$

These functions are called the *basis functions* of the DCT [38]. The DCT coefficients B_{pq} , then, can be regarded as the *weights* applied to each basis function. For 8-by-8 matrix, the 64 basis functions are illustrated by this image. The DCT of an image are shown in Figure (2.9), where, the horizontal frequencies increase from left to right, and vertical frequencies increase from top to bottom [39]. The constant-valued basis function at the upper left is often called the *DC basis function*, and the corresponding DCT coefficient B_{00} is often called the *DC coefficient*.

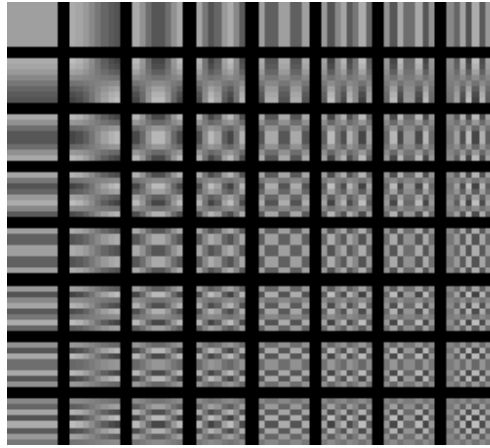


Figure 2.9: The 64 Basis Functions of an 8×8 Matrix [41].

In forensics, the provided image is split into blocks and DCT factors are extracted and caches in an array. The array is lexicographically sorted and quantized values of DCT coefficients for every single block are compared [40]. Additionally, seems at the reciprocal positions of each identical block pair and outputs a specific block pair only if there are many other matching pairs in the similar reciprocal position. DCT based detection technique robust to multiple copy move forgery, noise contamination, Gaussian obliterating, rotation up and about to five degrees, shifting and scaling [41].

2.3.2 Algorithms Based on Invariant Image Moment

The concept of image moments refers to a certain particular weighted average for the intensities of the image pixel. These moments are selected for giving a uniform image interpretation which assists in the analysis of the shape. It is beneficial for describing the objects of an image when separation or describing the entire image central point, intensity, as well as info of objects adjustment for detecting scaling, transformation, and rotation. The image instants connect districts in a twofold shape for translating scaling, and rotation in a suitable classification shape and part recognition [42]. The translation invariance is the main characteristic of the centralized moments. Even scaling or rotating the image, the centralized moments are still not changed. It is extremely

significant for different applications to construct these moments with scaling, rotation, and translation invariance [43]. CMF can be traced by computing Hu, Blur invariant, and Zernike moments.

2.3.3 Algorithms Using Texture & Intensity Descriptors

Texture and density occur in pure parts such as grass, tree, cloud, and earth, the characteristics of the image like coarseness, smoothness, and regularity characterize the texture subjects. So, the texture and density can be utilized as features to detect the similarity in the tampered image. In CMF detection, texture and density have been computed and distinguished by pattern and density [44].

A block-based technique utilizes the texture and intensity as features for detecting the CMF. The blocks of the image are partitioning into several directional sub blocks; After that, these sub-blocks are used to obtain the energy features. This technique works with grayscale and color images [45,46].

2.3.4 Algorithms Based on Invariant Key Points

The algorithm based on the key point extracts the feature point from specific regions of an image without any partition of an image [47]. It is reducing in the computation time and memory consumption, also it has robustness against rotation, scaling, Gaussian noise, JPEG compression, and illumination. The size of the extracted feature is extremely smaller in size. the key point-based algorithm is being the perfect selection for CMF detection in the images of large size. It extracts the feature point by utilizing several approaches such as Scale-Invariant Feature Transform (SIFT), Speeded-Up Robust Feature Extraction (SURF), etcetera. with no need to partition the image. The Feature points are matched with each other by utilizing several approaches such as Euclidean distance and clustering [48].

2.3.5 Algorithms Based on Reciprocal Information

The idea of reciprocal information was firstly suggested by Soleimani and Khosravifard [49] in which the template matching of an image refers to the dependency between two arbitrary variables. Reciprocated information is in its extreme value when two regions or arbitrary variables remain dependent. In this briefcase, the common probability matrix is diagonal, and in the situation of autonomous areas or variables, the reciprocated information equal to zero. Chakraborty [50] proposed a new technique for detecting the CMF depending on reciprocal information searches for repeated regions with no need to extract any features, depends on a mathematical way. The major advantages of this technique that it is easy and high speed and it is perfectly robust against illumination alterations.

2.3.6 Algorithms Based on Singular Value Feature

Singular value feature vectors (SVD) is provided to generate algebraic and geometric invariant and feature vectors. The algorithm divides up the image into overlapping blocks. Then to every single block utilizes the SVD and find reduced ranking dimension and obtain singular value feature vectors and storing it in a matrix. Lexicographically sort and successive rows indicate the same blocks. When the resemblance between blocks is greater than a fixed value, the CMF is detected. This technique locates the copy move tampering, has lower computational complication and greater noise immunity [51]. The SVD based technique supplies the spatial position of the image part that has been forged by comparing the recovered image and the forged image that the recipient receives [52].

2.4 Digital Image File Format

Digital image is generated, gathered, and put in storage in a large range of proprietary and standard formats. Image folder format continue to develop, becoming more complicated as amended software versions include new features or functionality. Professionals, practitioners, students, and beginner users are overcome with the diverse image file format [44].

Digital images can broadly have classified under several categories: Binary Images, Grayscale Images and Color Images.

2.4.1 Binary Images

Binary images are imaging whose pixels have only two possible intensity values. Numerically, the two values are often 0 for black, and either 1 or 255 for white. The main reason binary images are particularly useful in the field of Image Processing is because they allow easy separation of an object from the background. The process of segmentation allows to label each pixel as 'background' or 'object' and assigns corresponding black and white colours. [53].

2.4.2 Grayscale Images

Grayscale image is one in which the value of each pixel is a single sample representing only an amount of light, that is, it carries only intensity information. a kind of black-and-white or gray monochrome, are composed exclusively of shades of gray. Starting from black at the lowest level of intensity to the white at the highest level of intensity. The grayscale images differ from one bit black and white images, while it is only two-color images in the computer imaging context [54]. it is not like the two black and white images, the grayscale images have many levels or shades of gray, grayscale images named monochromatic in some references. Grayscale images are often the result of measuring the intensity of

light at each pixel in a separate group of electromagnetic spectrum and in these cases, they are monochromatic suitable when only a given frequency is captured. But also, they can be produced from a full color image. In grayscale images each pixel has 8 bits. Hence there are 0 to 255 intensity level are in pixel. [55].

2.4.3 Color Images

In the actual life we usually get colored images. Some systems use the RGB color model. There are 2^{24} probable levels for each key color. Once the image is changed as $R = G = B$ then the image is well-known as 16-bit grayscale. In 8-bit grayscale image the lightness of the gray is immediately proportional to the number representing the brightness levels of the major colors. A 16-bit digital grayscale image uses far more memory or storage space than the same image with the similar physical aspects in 8-bit digital grayscale [32]. The colored image is shown in Figure (2.10).



Figure 2.10: Color Image [32].

2.5 K- Mean Clustering

K-Means is perhaps the most well-known clustering algorithm. It's taught in a lot of introductory data science and machine learning classes. It clusters similar entities based on Common features into number of groups under a positive integer number. The clustering is performed by reducing the sum of squares of spaces between data and the corresponding cluster centroid.

Step 1. Initialization. A set of objects to be partitioned, the number of groups and a centroid for each group are defined.

Step 2. Classification. For each object, its distance to each of the centroids is calculated, the closest centroid is determined, and the object is incorporated to the group related to this centroid.

Step 3. Centroid calculation. For each group generated in the previous step, its centroid is recalculated.

Step 4. Convergence condition. Several convergence conditions have been used from which the most utilized are the following:

stopping when reaching a given number of iterations, stopping when there is no exchange of objects among groups, or stopping when the difference among centroids at two consecutive iterations is smaller than a given threshold.

If the convergence condition is not satisfied, steps two, three and four of the algorithm are repeated.

K-Means has the advantage that it is fast, it is handling the distances between points and group centers; very few computations, it thus has a linear complexity $O(n)$.

On the other hand, K-Means has a some of drawbacks, it must initially select number of groups/classes should be used for the clustering. This is not always trivial and ideally with a clustering algorithm it to figure those out because the point of it is to gain some insight from the data. K-means also starts with a random choice of cluster centers and therefore it may yield different clustering results on different runs of the algorithm. Thus, the results may not be repeatable and lack consistency.

In other words, the K-means algorithm identifies k number of centroids, and then allocates every data point to the nearest cluster, while keeping the centroids as small as possible [56].

2.6 Performance Measurements

To measure the performance rate of the proposed techniques [57], equation (2.6) is used as follows.

$$Accuracy = \frac{(Forged\ Blocks \cap Detected\ Blocks)}{Detected\ Blocks} \quad (2.9)$$

Also, the True and False Positive rates and False negative rate are utilized to evaluate the performance of the proposed techniques. True positive rate (TPR) or the proportion of actual detection measures the proportion of actual positives that correctly identified, in copy move forgery detection it refer to percentage of blocks that are detected and they are identified as forged blocks.

$$TPR = \frac{TP}{TP+FN} \quad (2.10)$$

False positive rate (FPR) refer to the percentage of blocks that are not forged but it is detected and identified in wrongly as forged.

$$FPR = \frac{FP}{FP+TN} \quad (2.11)$$

False negative rate (FNR) refers to percentage of blocks that forged but it identified as original blocks

$$FNR = \frac{FN}{TP+FN} \quad (2.12)$$

TP (True positive) , FN (False Negative) ,FP (False positive) TN (True Negative)

Chapter three

The proposed
forensic techniques

CHAPTER THREE

THE PROPOSED FORENSIC TECHNIQUES

This chapter describes the design and implementation of two proposed digital image copy-move forensic techniques.

3.1 Introduction

Digital images consider as active and natural media for carrying worth information over the world. With the release of high-resolution digital cameras, modern PCs, and developed photo editing applications, the digital images forgery becomes very popular. Therefore, the need for using forensic techniques is increased, and the most active research subfield is copy-move forgery detection.

Many block-based techniques showed good results in detecting forgery, but it still need to increase robustness and reducing processing time. especially when working against scaling and rotation conditions. To deal with these problems, an efficient and fast techniques have been designed for checking the image authenticity.

3.2 The Proposed Techniques

In this thesis, two digital image copy-move forensic techniques are proposed. The DCT-based technique represents a blind digital image of a block-based forensic technique to detect and identify the digital image copy-move forgery without any prior information concerning the image under analysis. the Framing technique can handle rotation and scale preprocesses with small factors. While, the Framing technique can detect the copy-move rotation in the spatial

domain and accelerating the process of matching the standard block utilizing auto-clustering for the blocks and comparing blocks parallel.

3.2.1 The DCT based proposed Technique

To perform the process of colour image copy-move forgery detection, the suspicious image is transformed from its colour space to its equivalent YCbCr. Because the duplicated regions are taken from the image itself, when the process is finished, similar areas in the same image are gotten. The size of these duplicated regions is unknown; therefore, it is not easy to compute the operation of regions pairs comparison of different sizes. So, for providing more efficiency, in this proposed technique, the image is sub-divided into overlapping-blocks of fixed size. Then, a DCT is applied to these blocks for extracting features out of the DCT coefficients based on the Zigzag method to reduce or compress the information (energy) of each block and concentrate it into few coefficients. The fast K-mean is applied since it is the best clustering algorithm for reducing the time cost which is utilized to cluster the blocks vectors to cope K group of features, after that, these vectors are sorted lexicographically using Radix sorting algorithm according the MOST Significant Digit(MSD) . The correlation between each pair of similar group vectors is computed for reducing false copy-move forgery detection, Correlation compared with the correlation standard threshold $T = \{ 0.1 \dots 0.9 \}$,if greater of equal to T then the distance (D) should be calculated to approve it is duplicated values .after calculating the distance between pair of similar groups, S refer to the distance threshold , if D grater or equal than S this mean Copy move forgery exist but if it is lower than S ,this mean that it comparing the same pixel with itself or nearest pixel from background , Finally, the decision of forgery existence is taken, and the duplicated blocks are identified. Figure (3.1) shows the overall structure of the proposed technique.

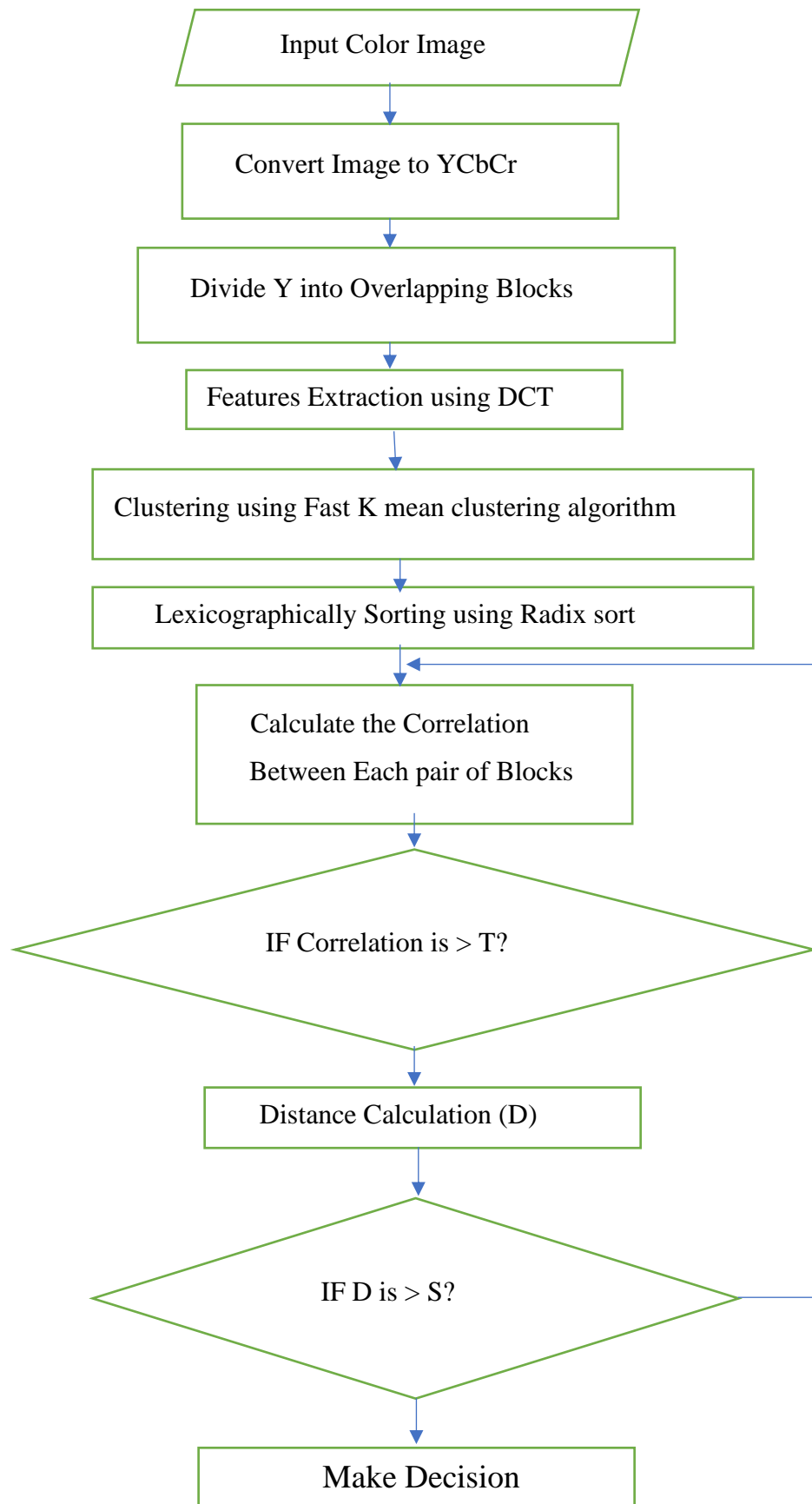


Figure 3.1: The Structure of the DCT based Proposed Technique.

3.2.1.1 Color Space Conversion

The digital image is transformed from RGB Colour space to YCbCr. The first component "Y" denotes the luminance, whilst the second and third components "Cb and Cr" denote the chrominance. The conversion process from RGB to YCbCr is presented in the equation (2.2).

3.2.1.2 The Overlapping Blocks

The Y component image is sub-divided into overlapping pixels blocks. For an image of size ($I \times J$) and a block of size (8×8), the overlapped blocks number as in equation (2.3), Pseudo- code (3.1) show the blocking stage on Y component of the image

Pseudo-code (3.1): Finding the Overlapped Blocks
Input: Y component image
Output: Overlapped blocks
<pre>for j=1:overlapp:(c-blocksize) + 1 for i=1:overlapp:(r-blocksize) + 1 Matrix(a). block=im(i:i+blocksize-1,j:j+blocksize-1); Matrix(a). position=[i j]; Matrix(a). index=a; a=a+1; end end</pre>

3.2.1.3 DCT based Feature Extracting

The Discrete Cosine Transform (DCT) is closely related to the discrete Fourier transform, it is a separable liner transformation, that is the two dimensional transform is equivalent to a one dimension DCT performed along a single dimension followed by a one dimensional DCT in the other dimension.

The two-dimensional DCT algorithm is applied to each image blocks to yield three sub-bands: low-frequency, middle-frequency, and high-frequency.

After applying this conversion, The DCT coefficients for each block are reshaped into a zigzag order row vector, see figure (3.2).

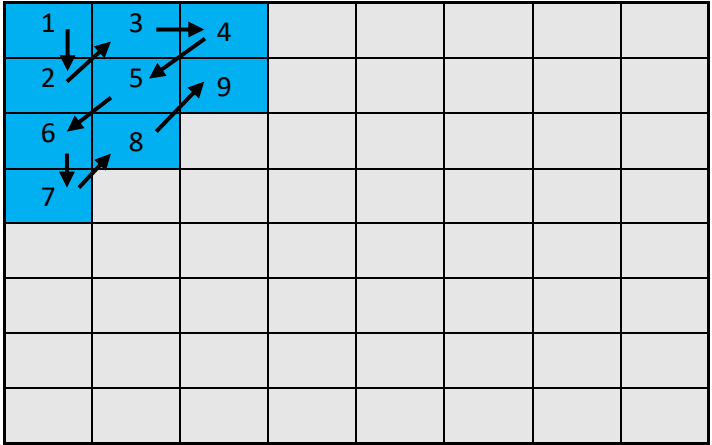


Figure 3.2: The process of feature selection.

The low-frequency sub-band include the most important coefficients in the frequency image which carry most of the information about the original image. In order to compress the information of each block into few coefficients and decrease the time of processing, the features F (coefficients) are extracted from the low-frequency sub-band as shown in figure (3.2). Finally, all blocks with selected features are stored in a matrix M .

DCT coefficients are used as input to the sorting algorithm ,zig-zag algorithm scans values diagonally to extract features starting from the most left upper corner of block array.it extract l of features l=9,5 as the code shown in Pseudo code (3.2).

Pseudo-code (3.2): Extracting the features using Zigzag
Input: Each image block DCT coefficients
Output: The first nine Zigzag order row vector
<pre> function [M] = tozig(x) [row col]=size(x); y=zeros(row*col,1); count=1; for s=1:row </pre>

```

if mod(s,2)==0
for m=s:-1:1
y(count)=x(m,s+1-m);
count=count+1;
end;
else
for m=1:s
y(count)=x(m,s+1-m);
count=count+1;
end; end; end;
if mod(row,2)==0
flip=1;
else
flip=0;
end;
for s=row+1:2*row-1
if mod(flip,2)==0
for m=row:-1:s+1-row
y(count)=x(m,s+1-m);
count =count +1;
end;
else
for m=row:-1:s+1-row
y(count)=x(s+1-m,m);
count=count+1;
end end;
flip=flip+1;
M=y.';
end;

```

3.2.1.4 The Block Clustering

The algorithm of Fast k-mean clustering on M is suggested to cluster the similar extracted feature vectors into k group. The core attribute of the Fast k-mean clustering is to avoid the calculations of not needed distance through apply the triangle inequality and keep track of the minimum and maximum bounds to the distances between points and centers.

In Pseudo – code (3.3) the features extracted from framing process will be used as inputs to the clustering algorithm ,initially K mean will select random values for the number of groups that features extracted clustered

,suppose that $K = 5, 10, 20$,according to selected centroids ,each feature block assigned to one group of K depending on the closest value of each centroid ,then select new centroids by recalculate centroids values ,and keep iteration process between clustering and recalculating centroids till reach one of stopping conditions ,algorithm will stop iterations when reaching a given number of iterations, stopping when there is no exchange of objects among groups, or stopping when the difference among centroids at two consecutive iterations is smaller than a given threshold.

Pseudo-code (3.3): K – Mean clustering
Input: Features vector
Output: clustered groups of blocks
<pre> function [centers,mincenter,mindist,q2,quality] = if nargin < 3 method = 2; end [n,dim] = size(data); if max(size(initcenters)) == 1 k = initcenters; [centers, mincenter, mindist, lower, computed] = anchors(mean(data),k,data); total = computed; skipstep = 1; else centers = initcenters; mincenter = zeros(n,1); total = 0; skipstep = 0; [k,dim2] = size(centers); if dim ~= dim2 error('dim(data) ~= dim(centers)'); end; end nchanged = n; iteration = 0; oldmincenter = zeros(n,1); while nchanged > 0 % do one E step, then one M step computed = 0; if method == 0 & ~skipstep for i = 1:n for j = 1:k distmat(i,j) = calcdist(data(i,:),centers(j,:)); end end end </pre>

```

[mindist,mincenter] = min(distmat,[],2);
computed = k*n;

elseif (method == 1 | (method == 2 & iteration == 0)) & ~skipestep
mindist = Inf*ones(n,1);
lower = zeros(n,k);
for j = 1:k
    jdist = calcdist(data,centers(j,:));
    lower(:,j) = jdist;
    track = find(jdist < mindist);
    mindist(track) = jdist(track);
    mincenter(track) = j;
end
computed = k*n;

elseif method == 2 & ~skipestep
    computed = 0;

    nndist = min(centdist,[],2);
    mobile = find(mindist > nndist(mincenter));

    mdm = mindist(mobile);
    mcm = mincenter(mobile);

    for j = 1:k
        track = find(mdm > centdist(mcm,j));
        if isempty(track) continue; end
        alt = find(mdm(track) > lower(mobile(track),j));
        if isempty(alt) continue; end
        track1 = mobile(track(alt));

        redo = find(~recalculated(track1));
        redo = track1(redo);
        c = mincenter(redo);
        computed = computed + size(redo,1);
        for jj = unique(c)'
            rp = redo(find(c == jj));
            udist = calcdist(data(rp,:),centers(jj,:));
            lower(rp,jj) = udist;
            mindist(rp) = udist;
        end
        recalculated(redo) = 1;

        track2 = find(mindist(track1) >
centdist(mincenter(track1),j));
        track1 = track1(track2);
        if isempty(track1) continue; end

        % calculate exact distances to center j
        track4 = find(lower(track1,j) < mindist(track1));
        if isempty(track4) continue; end
        track5 = track1(track4);
        jdist = calcdist(data(track5,:),centers(j,:));
        computed = computed + size(track5,1);
        lower(track5,j) = jdist;

        % find which points really are assigned to center j

```

```

        track2 = find(jdist < mindist(track5));
        track3 = track5(track2);
        mindist(track3) = jdist(track2);
        mincenter(track3) = j;
    end % for j=1:k
end % if method
oldcenters = centers;
diff = find(mincenter ~= oldmincenter);
diffj = unique([mincenter(diff);oldmincenter(diff)]);
diffj = diffj(find(diffj > 0));

if size(diff,1) < n/3 & iteration > 0
    for j = diffj
        plus = find(mincenter(diff) == j);
        minus = find(oldmincenter(diff) == j);
        oldpop = pop(j);
        pop(j) = pop(j) + size(plus,1) - size(minus,1);
        if pop(j) == 0 continue; end
        centers(j,:) = (centers(j,:)*oldpop +
sum(data(diff(plus,:),:),1) - sum(data(diff(minus,:),:),1))/pop(j);
    end
else
    for j = diffj
        track = find(mincenter == j);
        pop(j) = size(track,1);
        if pop(j) == 0 continue; end

        centers(j,:) = mean(data(track,:),1);
    end
end

if method == 2
    for j = diffj
        offset = calcdist(centers(j,:),oldcenters(j,:));
        computed = computed + 1;
        if offset == 0 continue; end
        track = find(mincenter == j);
        mindist(track) = mindist(track) + offset;
        lower(:,j) = max(lower(:,j) - offset,0);
    end

    % compute distance between each pair of centers
    % modify centdist to make "find" using it faster
    recalculated = zeros(n,1);
    realdist = alldist(centers);
    centdist = 0.5*realdist + diag(Inf*ones(k,1));
    computed = computed + k + k*(k-1)/2;
end

nchanged = size(diff,1) + skipstep;
iteration = iteration+1;
skipstep = 0;
oldmincenter = mincenter;

total = total + computed;
end % while nchanged > 0

udist = calcdist(data,centers(mincenter,:));
quality = mean(udist);

```



```
q2 = mean(udist.^2);
```

3.2.1.5 Sorting the Clustered Blocks

MSD radix sorts are most suitable for sorting strings or fixed-length integer representations. A sequence like [b, c, e, d, f, g, ba] would be sorted as [b, ba, c, d, e, f, g]. If lexicographic ordering is used to sort variable-length integer in base 10, then numbers from 1 to 10 would be output as [1, 10, 2, 3, 4, 5, 6, 7, 8, 9], as if the shorter keys were left-justified and padded on the right with blank characters to make the shorter keys as long as the longest key. MSD sorts are not necessarily stable if the original ordering of duplicate keys must always be maintained.

The Most Significant Digit (MSD) Radix sorting algorithm is used to sort ascendingly the clustered blocks vectors in M. Pseudo code (3.4) show the critical code for Sorting the Clustered Blocks using Radix sort (MSD) ,The similar values are sorted lexicographically from left to right to acquire new series of vectors that allow comparing between each neighbor vectors. Figure (3.3) provides an instance of radix sort.

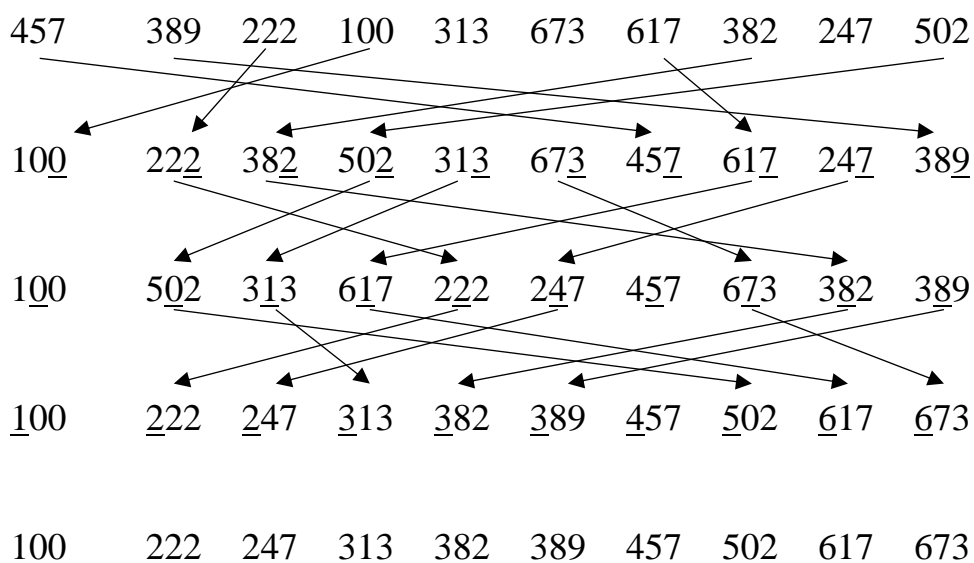


Figure 3.3: An instance of radix sort.

Pseudo-code (3.4): Sorting the Clustered Blocks
Input: The Clustered Blocks
Output: The Sorted Clustered Blocks
<pre> i=digit; Asorted=A; while i>=1 [x,inx]=sort(Asorted(:,i)); Asorted(:,i)=Asorted(inx(:,i),:); i=i-1; end end </pre>

3.2.1.6 Calculating the Correlation

The Correlation is calculated between every two similar blocks, as in the following equation.

$$Correlation = \frac{\sum_{i=1}^c (x_i - x') \cdot (y_i - y')}{\sqrt{\sum_{i=1}^c (x_i - x')^2 \cdot \sum_{i=1}^c (y_i - y')^2}} \quad (3.1)$$

Where x , and y are the DCT block coefficients, and x' , y' are the mean values respectively, c is the number of block coefficients. When the correlation is bigger than the threshold value T , the two blocks are assumed to be identical and the distance should be found between similar blocks for eliminating the false positives, otherwise, skip this block to the next one.

The distance between two similar blocks could be found by:

$$Distance = \sqrt{(M^x_i - M^x_{i+1}) + (M^y_i - M^y_{i+1})} \quad (3.2)$$

Where (M^x_i, M^y_i) is the location of block (i) and (M^x_{i+1}, M^y_{i+1}) is the location of block $(i+1)$.

3.2.1.7 Comparing the Distance

Comparing the distance between similar blocks if $Distance > S$, Where distance calculated by equation (3.2) and S could be selected by training $S=16$ and $D>16$ that mean there is duplication between two blocks and the distance between these blocks is more than 16 pixel , then take a decision about the existence of forgery and then pointing out the duplicated blocks.

3.2.2 The Framing Proposed Technique

The main objective of the forensic techniques is to detect if an image includes duplicated regions with or without modifications like blurring, reflection, and rotation. Because the size and the shape of the regions are not known, computationally, it is not possible to attempt for examining each potential pair of regions with various sizes and shapes. Therefore, it is very efficient to partition an image into overlapped blocks of fixed size.

In the framing proposed technique as illustrated in Figure (3.4) the image is separated into overlapping blocks. For every block, the features are extracted via separating it into nested frames and calculating the average to every frame. The extracted features are utilized in the step of clustering to group identical blocks into several classes. The feature vectors in every class are lexicographically sorted via radix sort. A comparison is calculated between every close pair of blocks, when it is less than a specified threshold, two blocks are regarded as identical. And, for reducing false detection, a spatial distance between these blocks is computed. The steps of the framing proposed technique are explained in detail in the next subsections.

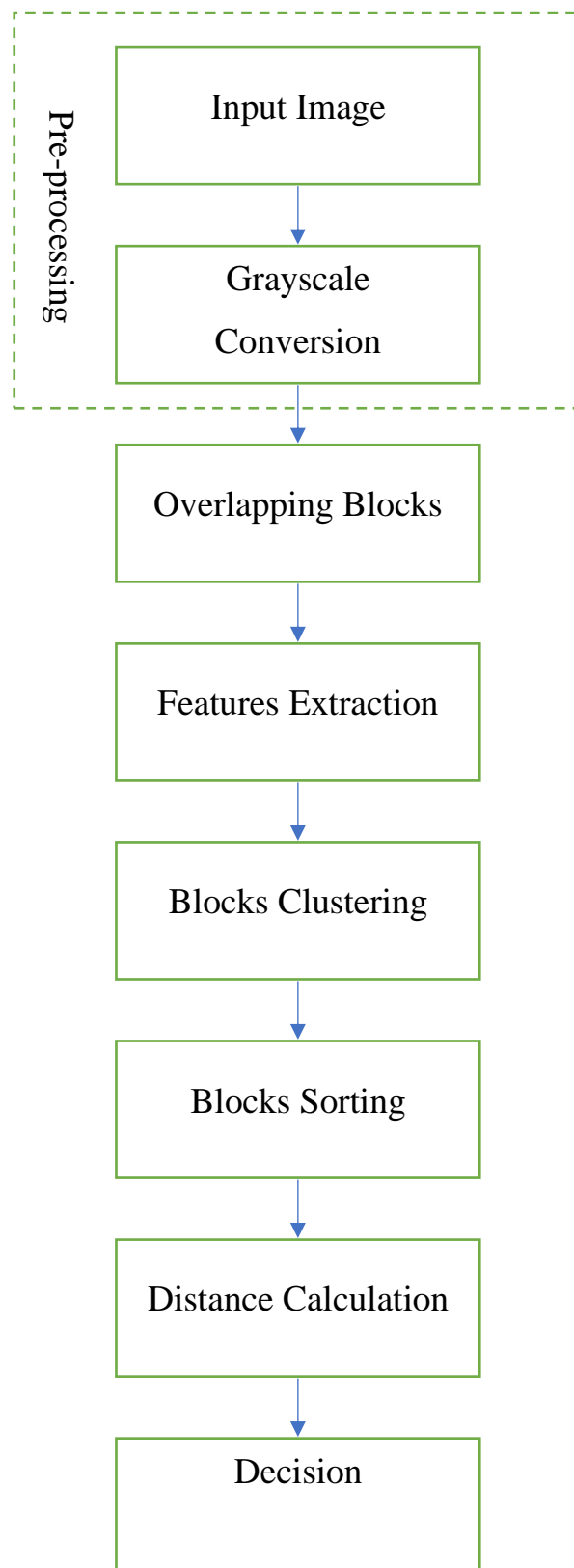


Figure 3.4: The structure of the framing proposed technique.

3.2.2.1 The Pre-processing Step

In the framing proposed technique, when the input images are grayscale, it will be used directly, else, when the input images are RGB color images, it will be converted to grayscale images, this conversion is illustrated in equation (2.1)

3.2.2.2 Overlapping blocks Step

The grayscale images with the size of $(M \times N)$ can be separated into overlapping blocks of $(t \times t)$ pixels, where t is an odd number, to result in T of blocks, as explained in the equation (2.3)

3.2.2.3 Features Extraction Step

In this step, each block is separated into four frames as shown in figure (3.5), supposing t is equal to eight. For all blocks, the features are extracted by computing the averages of these frames. Feature vector includes $(Number=t/2)$ coefficients, in addition to two indices to the block location, as explained in the following equation:

$$FV_i = Average(F_i), 1 \leq i \leq Number \quad (3.3)$$

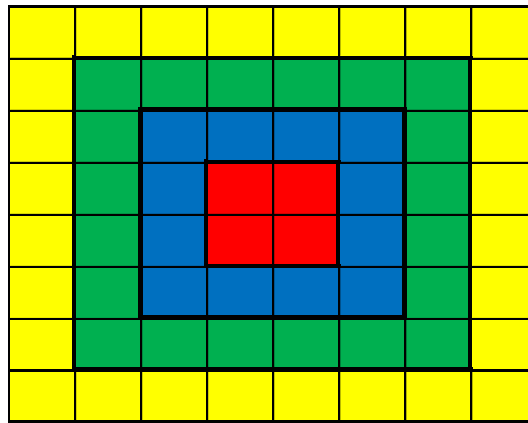


Figure 3.5: Block division; Frame1: Yellow, Frame2: Green, Frame3: Blue, Frame 4: Red.

An instance of rotated block has been given in figure (3.6) in which the block has been rotated via the fundamental angles (90, 180 and 270). It is noticeable that, the block values are unchanged along the block frame. Lastly, the whole blocks are stored in an "A" array with size " $T \times (Number+2)$ ". To be obvious, suppose that t is equal to 8, consequently, the feature vector length is the frames averages "4+2" for block center leads to six features, the Pseudo code (3.5) show the features extraction process , N and n is counters ,block refers to the blocks extracted from overlapping process , tot refer to the buffer that calculate frames summation ,feature is an array refer to the average of each frame.

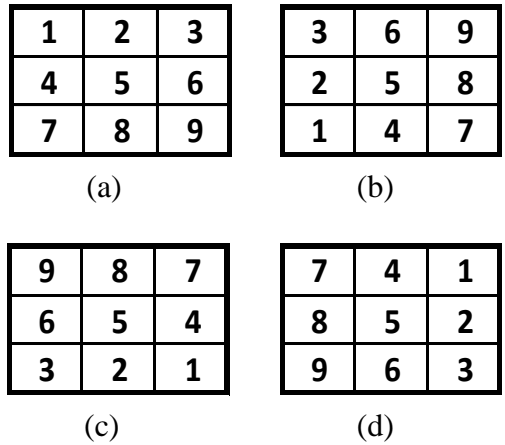


Figure 3.6: An instance of block rotation: (a) The original block, (b) Rotation with angle 90, (c) Rotation with angle 180 and (d) Rotation with angle 270.

Pseudo-code (3.5): Feature Extraction
Input: The Overlapped Block
Output: Reduced features
<pre> block; tot=[]; n=4;N=5; while (n>0) </pre>

```

t=0;
for i=n:N
    for j=n:N
        t=double(t)+double (block(i,j));
        block(i,j)=0;
    end
end
tot(n)=t;
n=n-1;
N=N+1;
end
feature=[tot(1)/28 tot(2)/20 tot(3)/12 tot(4)/4];
vector=0;
end

```

3.2.2.4 Blocks Clustering Step

The cluster technique has been used for blocks clustering to considerable classes for parallel comparison. K-means algorithm represents a fast clustering which works on grouping similar objects depending on features into the "K" number of groups, where "K" refers to an integer "positive" number. The process of grouping is accomplished via reducing the sum of squares of distances between data and the corresponding cluster centroid. For speeding up this algorithm, a Fast K-Means algorithm could be used for avoiding needless distance computations through implementing the triangle inequality and keeping track of upper and lower bounds of distances between centers and points.

In the Framing proposed technique, the Fast K-Means algorithm is used for the purpose of clustering. The values of features in every vector are stored in the "C" matrix to implement the Fast K-Means algorithm. The blocks feature vectors in each class are sorted lexicographically via radix sort. This sorting algorithm is a fast, stable and non-comparison-based sorting algorithm.

3.2.2.5 Blocks Sorting Step

The most significant digit (MSD) Radix sorting algorithm has been used in the Framing proposed technique to sort ascendingly the clustered blocks

vectors in M . The similar values are sorted lexicographically from left to right to acquire new series of vectors that allow comparing between each neighbor vectors. Figure (3.7) explains the blocks clustering and sorting into classes.

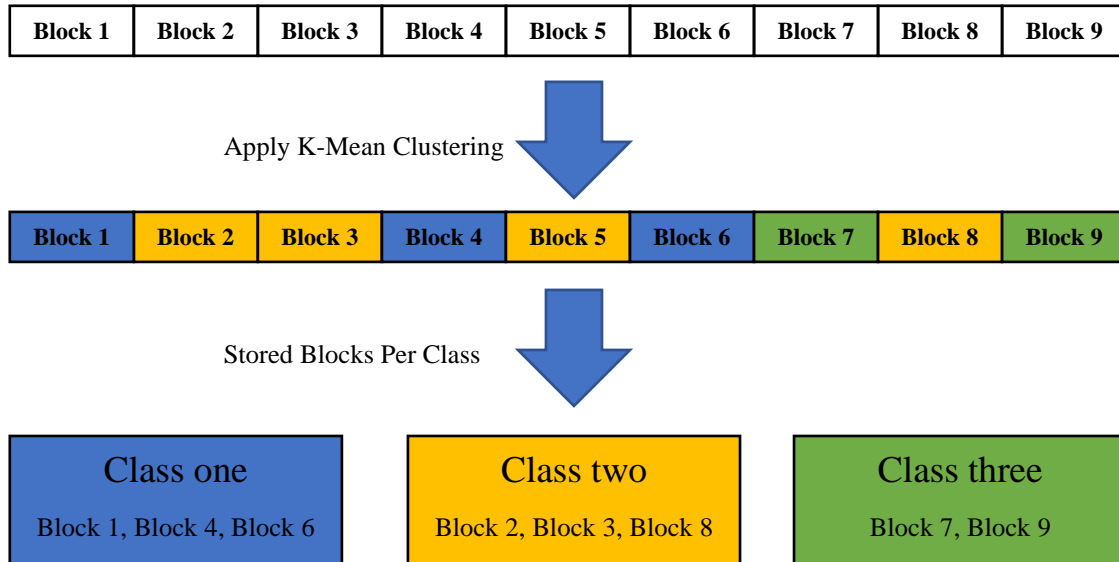


Figure 3.7: The blocks clustering and sorting into classes.

3.2.2.6 Distance Calculation Step

The Correlation is calculated between every two similar blocks, as in equation (3.1). When the correlation is bigger than the threshold value, the two blocks are assumed to be identical and the distance should be found between similar blocks for eliminating the false positives; Otherwise, skip this block to the next one. And the distance between two similar blocks could be found by using the equation (3.2).

3.2.2.7 Decision Making Step

Comparing the distance between similar blocks as referred in the DCT-based technique.

Chapter four

The Experiment results

CHAPTER FOUR

THE EXPERIMENT RESULTS

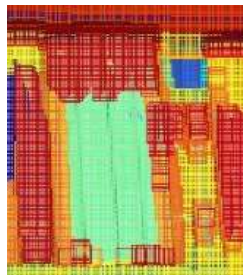
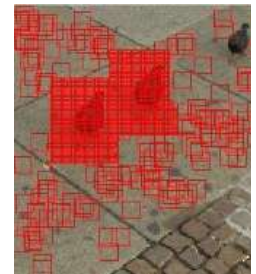
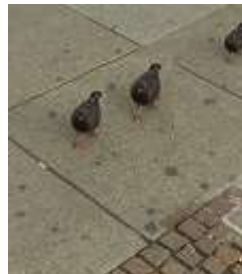
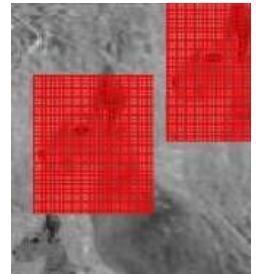
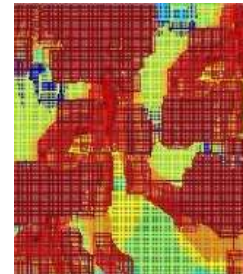
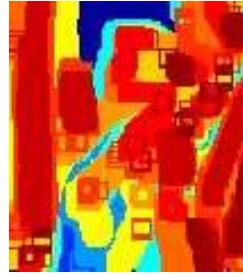
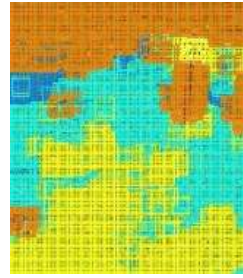
This chapter describes the experimental results of implementation of the two proposed digital image copy-move forensic techniques.

4.1 Introduction

The implementation of the proposed techniques is evaluated based on tampered images produced via the program of image manipulation GIMP 2.6.12. The experiments are implemented on MATLAB R2010b, RAM 12 GB and processor 2.30 GHz. All the used images are 128×128 pixels, and the colour image is saved in BMP file format. The parameters in the experiment are set to; block size $(t \times t) = 8 \times 8$, $S=16$, $F=4,5,9$ and $K=5,10, 20$.

4.2 Experiment Results for the DCT based technique

The results of copy-move detection marked on the forged images are shown in Figure 4.1. Each row in this figure is encompassed of four images: from left to right, the original, forged, clustered, and obtained images. The obtained or detected images refer to the ability of the proposed technique to detect multiple duplicated regions efficiently under different distortions such as blurring, scaling and rotation with small factors. Table 4.1 and Figure 4.2 illustrate the time consuming for detecting the forgery, when $F=5$. And, Table 4.2 and Figure 4.3 illustrate the time consuming for detecting the forgery, when $F=9$.



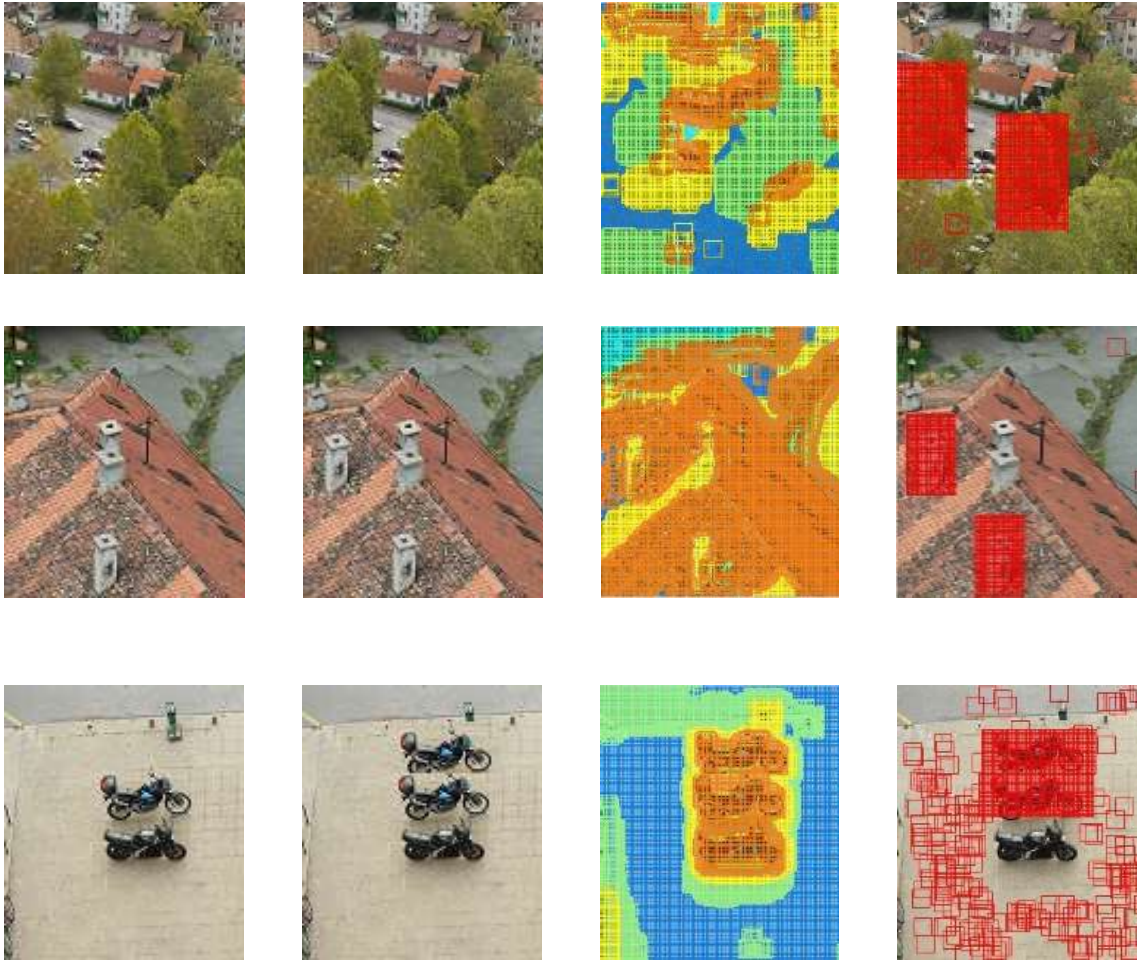


Figure 4.1: The various potential locations of multiple duplicated regions in image1, image2, ..., image8 using DCT based technique, $K=20$.

Table 4.1: The time consumption results of detecting forgery using DCT based technique when $F=5$ in seconds unit.

#	$K=5$	$K=10$	$K=20$
Image 1	3.9856	3.9487	3.9702
Image 2	4.2570	4.0686	4.3070
Image 3	3.7979	3.9974	4.4419
Image 4	3.9560	4.0961	5.3332
Image 5	3.5688	3.7086	3.9367
Image 6	3.8013	3.9883	5.5069
Image 7	3.8013	3.9794	4.2850
Image 8	4.0381	4.0509	4.8062

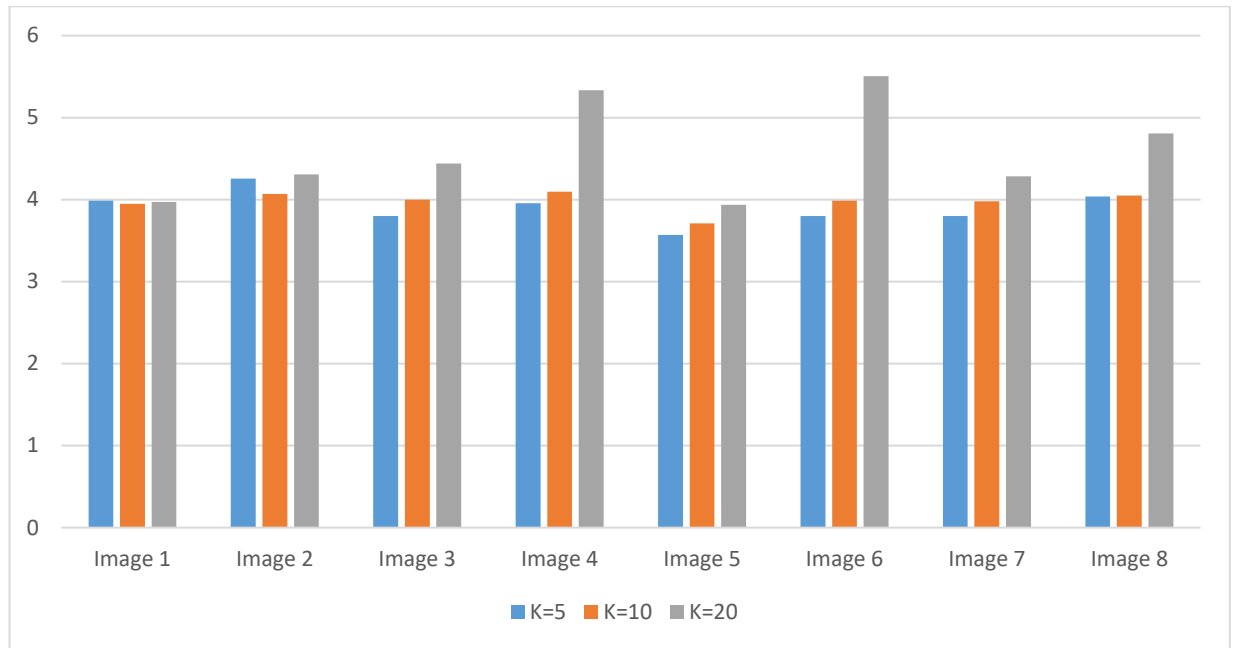


Figure 4.2: The time consumption for detecting the forgery, when F=5 using the DCT based Technique.

Table 4.2: The time consumption results of detecting forgery using DCT Based technique when F=9.

#	K=5	K=10	K=20
Image 1	6.1669	6.5601	9.1743
Image 2	5.9111	6.8175	8.3281
Image 3	5.9276	5.8426	6.0431
Image 4	6.184	6.8591	6.9561
Image 5	5.6944	6.2432	7.0537
Image 6	6.178	6.5608	9.7578
Image 7	6.1801	8.1465	7.4169
Image 8	6.512	6.7791	9.066

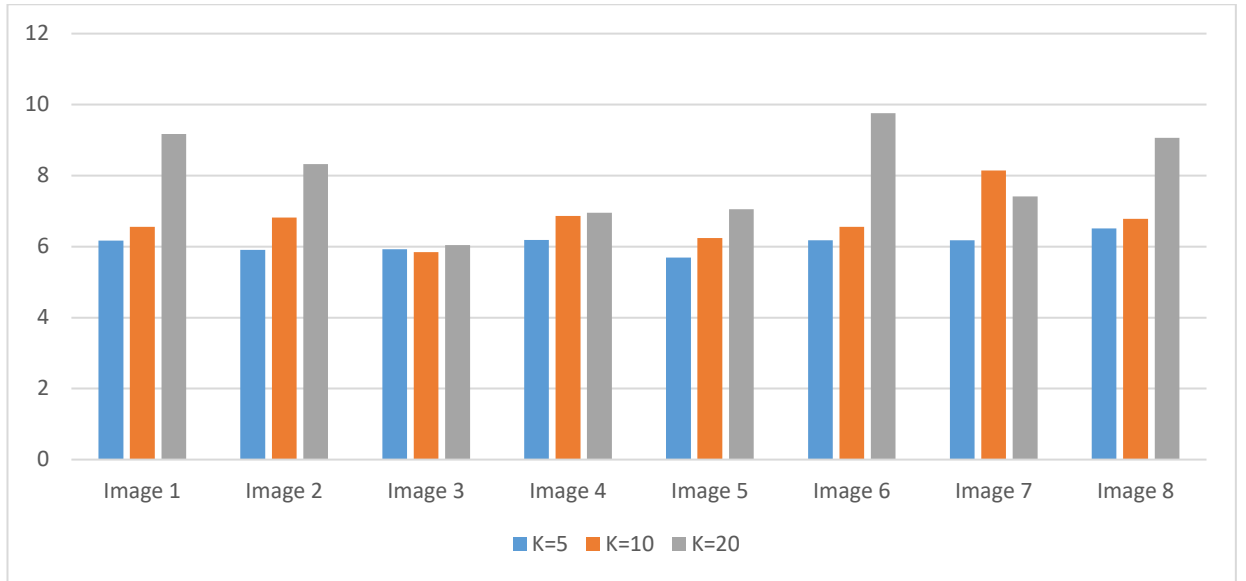
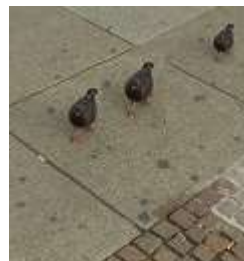
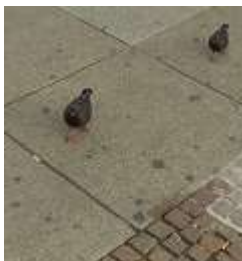
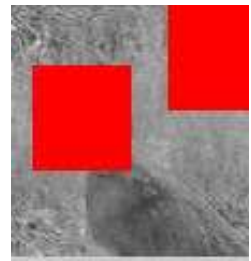
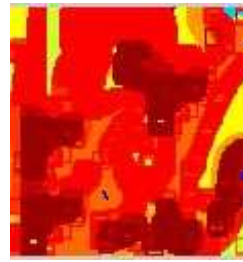
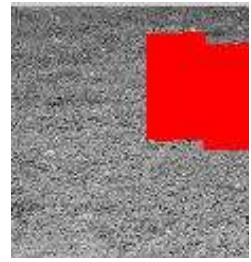


Figure: 4.3 The time consumptions, when $F=9$ using DCT based technique and different K .

After the experience, the selection of minimum features ($F=5$) with a minimum number of grouping K will provide best results; minimum memory and time consumptions.

4.3 Experiment results for the Framing technique

The results of copy-move detection marked on the forged images are shown in Figure 4.4. Each row in this figure is encompassed of four images: from left to right, the original, forged, clustered, and obtained images. The obtained or detected images refer to the ability of the proposed technique to detect multiple duplicated regions efficiently under different distortions such as blurring, scaling and rotation with small factors. Table 4.3 and Figure 4.5 illustrate the time consuming for detecting the forgery, and the selection of minimum number of grouping K will provide best results; minimum memory and time consumptions.



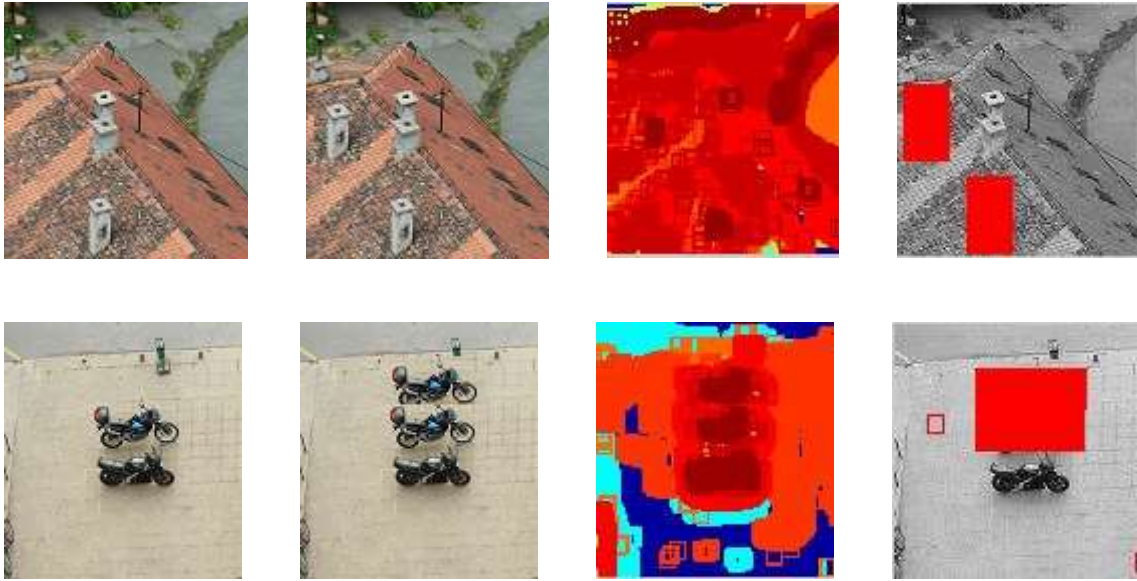


Figure 4.4: The various potential locations of multiple duplicated regions in image1, image2, ..., image8 using Framing technique with $k=20$.

Table 4.3: The time consumption results of detecting forgery using Framing technique

#	K=5	K=10	K=20
Image 1	2.521	2.7269	4.4671
Image 2	2.3135	2.5287	3.4524
Image 3	2.4298	2.6037	2.7765
Image 4	2.4563	2.9627	4.3568
Image 5	2.1432	2.1025	2.7263
Image 6	2.338	3.2772	3.7089
Image 7	2.2478	2.5805	4.426
Image 8	2.2855	2.5005	2.9676

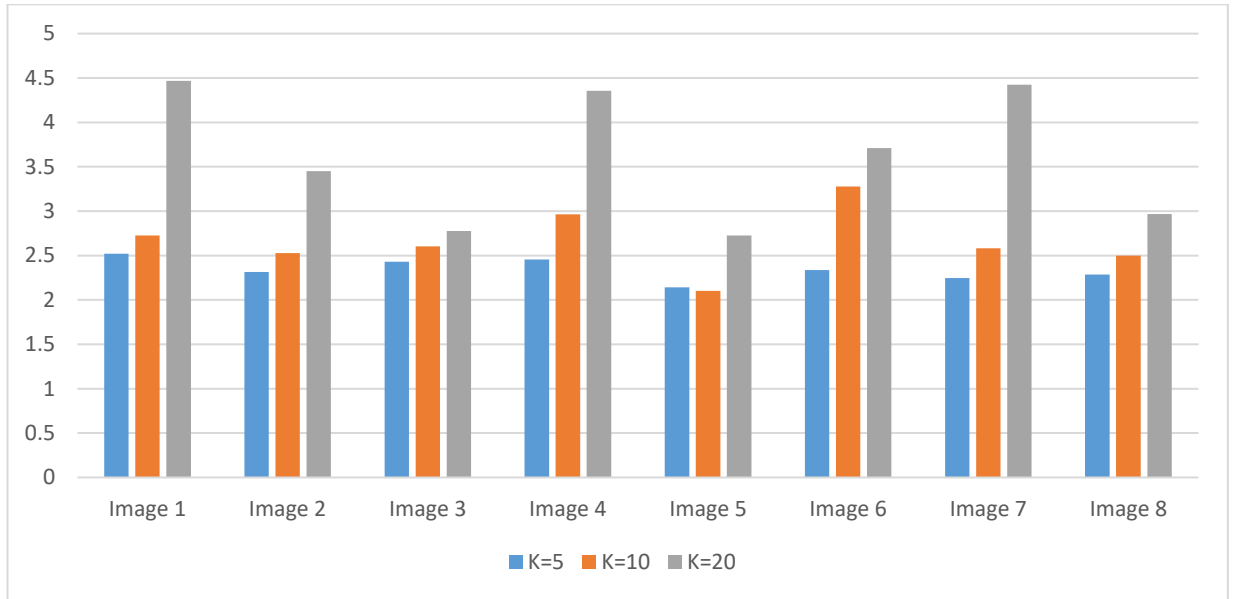


Figure 4.5: The time consumptions for detecting the forgery, when $F=4$ using Framing Technique.

Figure 4.6 and Table 4.4 represent a comparison between the DCT based and Framing technique in time consumption to detect images forgery with various values parameters for number of classes and extracted features. Where, in the DCT based technique, the block size ($t \times t$) is 8×8 , $S=16$, $F=5$ and $K=5, 10, 20$. And, in the Framing technique; the block size ($t \times t$) is 8×8 , $S=16$, $F=4$ and $K=5, 10, 20$.

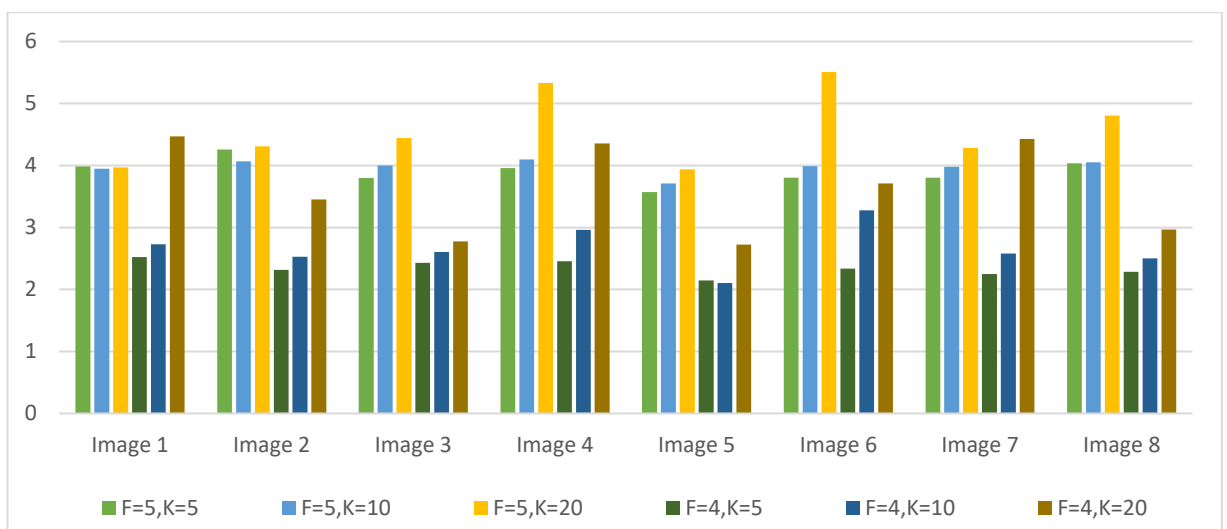


Figure 4.6: A comparison between DCT Based and Framing techniques when applied on the same series of images.

Table 4.4 Time consumption results of testing forgery detection with parameters F=4,5 and K = 5,10,20 using both techniques and comparing between them.

#	DCT based Proposed Technique F=5			Framing Proposed Technique F= 4		
	K=5	K=10	K=20	K=5	K=10	K=20
Image 1	4.0	3.9	4.0	2.5	2.7	4.5
Image 2	4.3	4.1	4.3	2.3	2.5	3.5
Image 3	3.8	4.0	4.4	2.4	2.6	2.8
Image 4	4.0	4.1	5.3	2.5	3.0	4.4
Image 5	3.6	3.7	3.9	2.1	2.1	2.7
Image 6	3.8	4.0	5.5	2.3	3.3	3.7
Image 7	3.8	4.0	4.3	2.2	2.6	4.4
Image 8	4.0	4.1	4.8	2.3	2.5	3.0

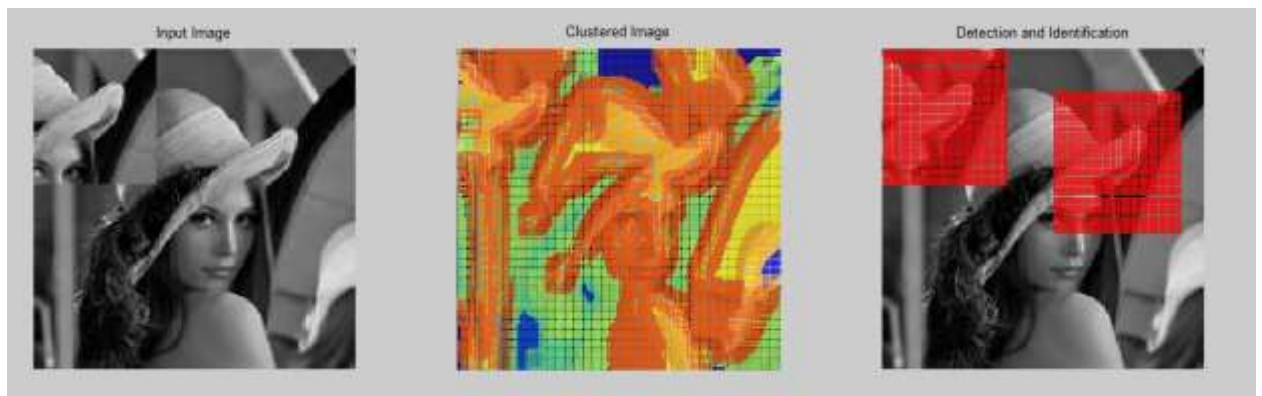
Table 4.5 shows a comparison between different recently existing forensic techniques and the proposed techniques based on the used feature extraction algorithm, matching, and the time-consuming measure.

Table 4.5: A Time comparison between different recently existing forensic techniques and the proposed techniques.

Author(s), Year	Ref. No.	Algorithm	Matching	Average Time (Seconds)
R. Davarzani et al., 2013	[6]	Patterns of Multiresolution Local Binary	Lexicographical Sort, and K-D Tree	223.6
Jen-Chun Lee, 2015	[7]	Gabor Magnitude	Radix Sort	18.34
Sondos M. Fadl et al., 2017	[9]	DCT	Radix Sort	6.8

The DCT based Technique	-	DCT	Clustering - Lexicographical Sort	5.90075
The Framing proposed forensic Technique	-	Framing	Clustering - Lexicographical Sort	2.8708

From the review of the previous table, we can notice that Framing technique is faster in comparison with the DCT based technique. Also, the Framing technique has more robustness against several scaling conditions, where the scaling of the copied segment could be used, hence the overlapping methods are used to find the similarities between multiple segments with the source image. The Framing technique could easily detect more scaling increased or decreased by 20% of the whole size of forged images. Figure 4.7 demonstrates the copy move forgery detection after different scaling conditions; scaling segment size +10%, multiple copy move forgery detection with scaling, decreasing segment size by -50%, and increase segment size +20% consecutively. Table 4.6 demonstrates the time consumption to detect forgery in scaled images.



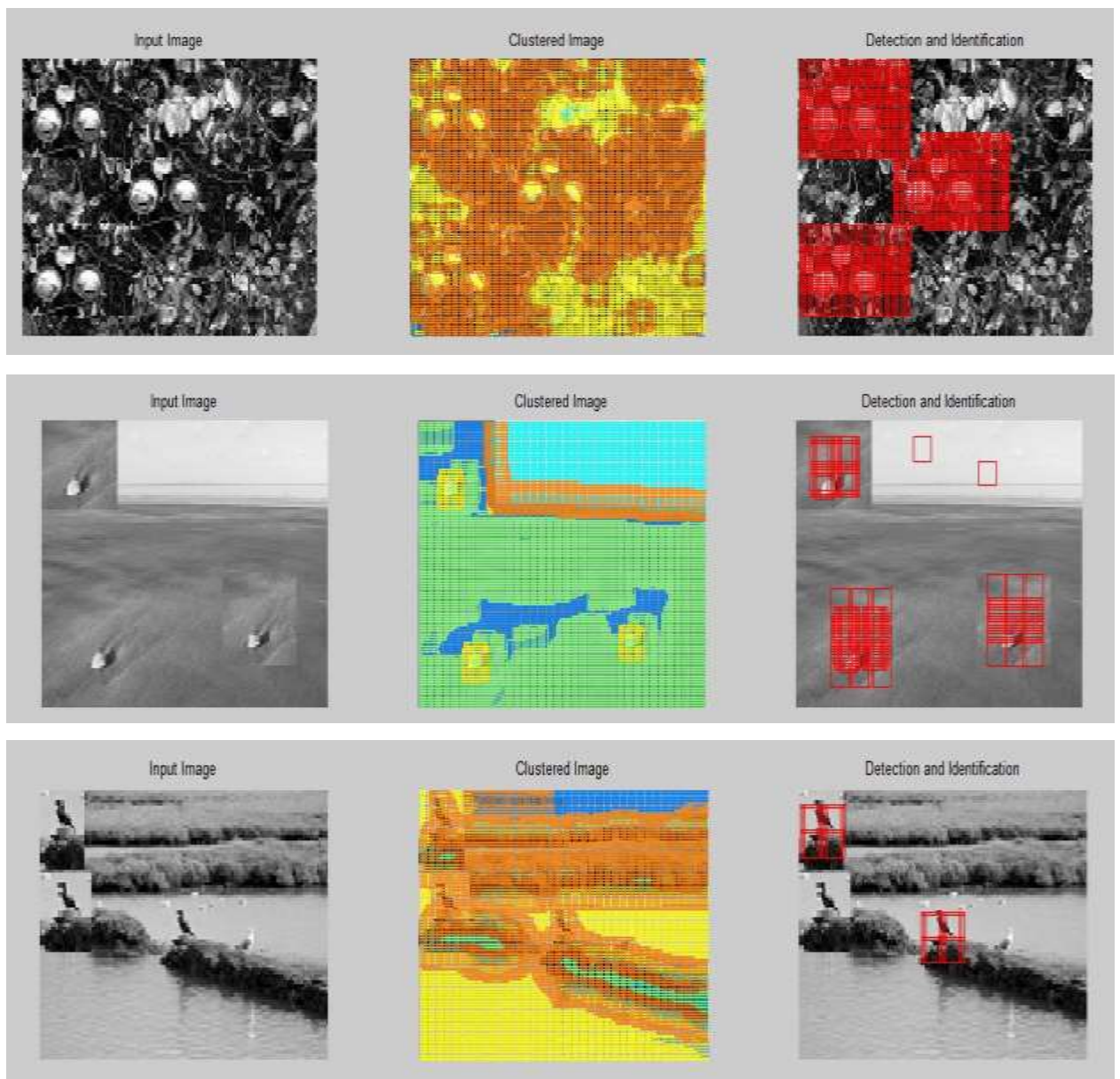
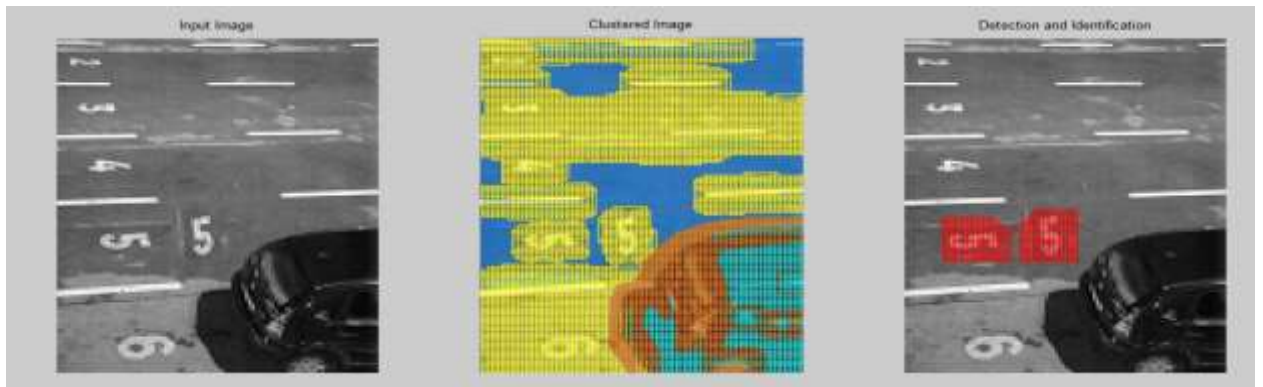


Figure 4.7: Framing technique results after different scaling conditions.

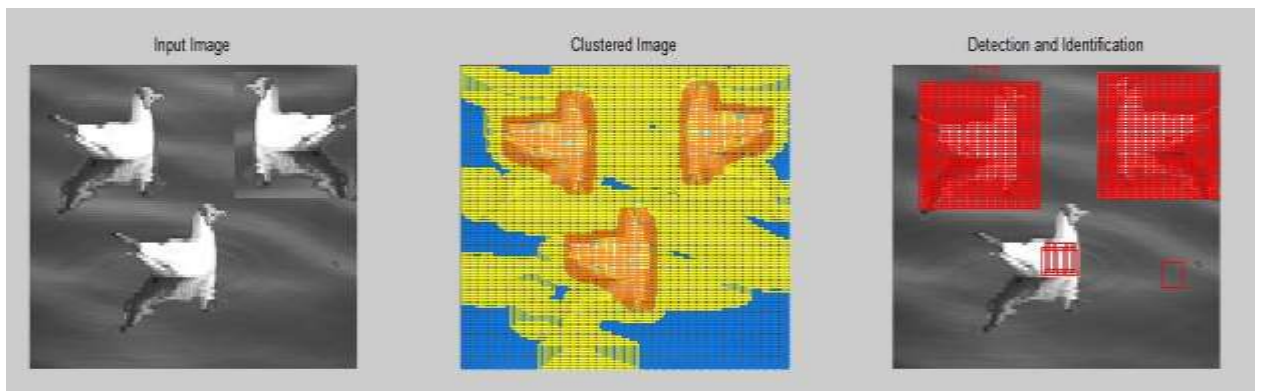
Table 4.6: Time consumption to detect forgery in scaled images using the Framing technique.

Image No.	Time
1	2.3373
2	1.81
3	1.7424
4	1.4871

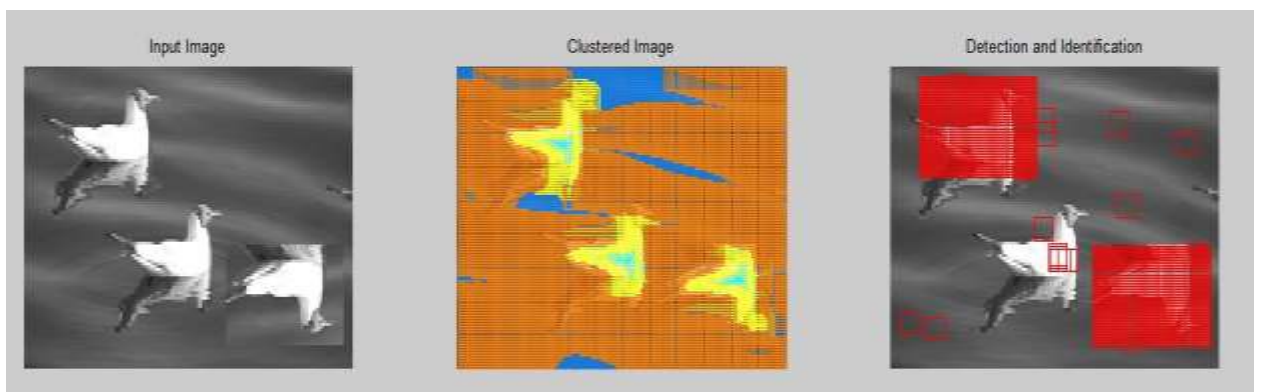
Also, the rotation could be used and applied to the copied area/segment to add more consistency to the forgery. In the Framing proposed technique, it could be used against different types of rotations 90° , 180° , and 270° degrees (clockwise and anti-clockwise), flip /reflection vertical and horizontal, as shown in Figure 4.8; (a) segment rotates by 90 degree anticlockwise, (b) horizontal reflection, (c) vertical reflection and (d) segment rotates by 90° degree anticlockwise. Table 4.7 demonstrates the time consumption to detect forgery in rotated images.



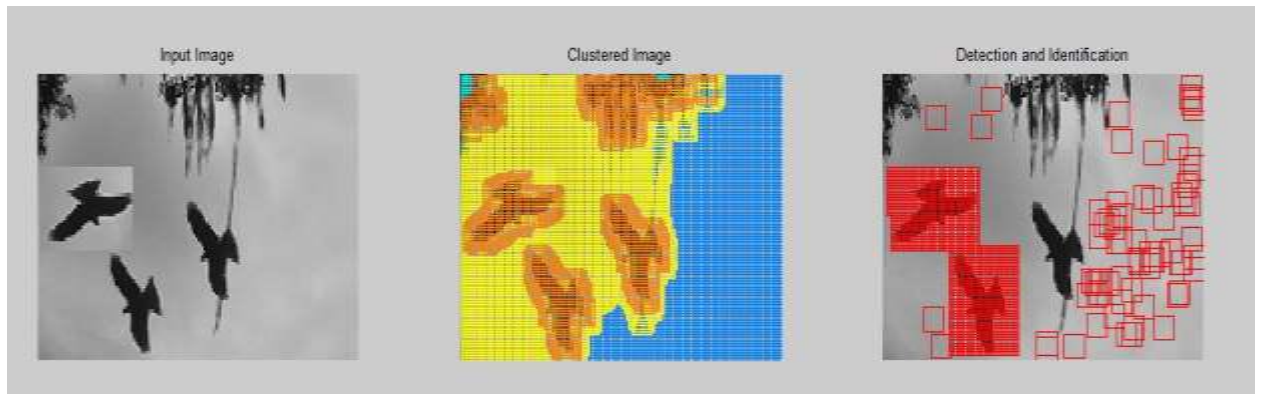
(a)



(b)



(c)



(d)

Figure 4.8: Framing technique results after different rotation conditions.

Table 4.7: Time consumption to detect forgery in rotated images using the Framing technique.

Image Name	Time
1	1.5386
2	1.6936
3	1.9691
4	1.6193

Figure 4.9 shows the result for different Thresholds using Framing technique: up row shows original image and the Copy-Move image from left to right respectively and down row shows detection result with $T=0.1$, $T=0.2$, $T=0.3$ and $T=0.4$, and $T=0.5$. Table 4.8 shows the True Positive and False Positive for detecting the Copy-Move in Lena image with $T=0.1$, $T=0.2$, $T=0.3$ and $T=0.4$, and $T=0.5$.

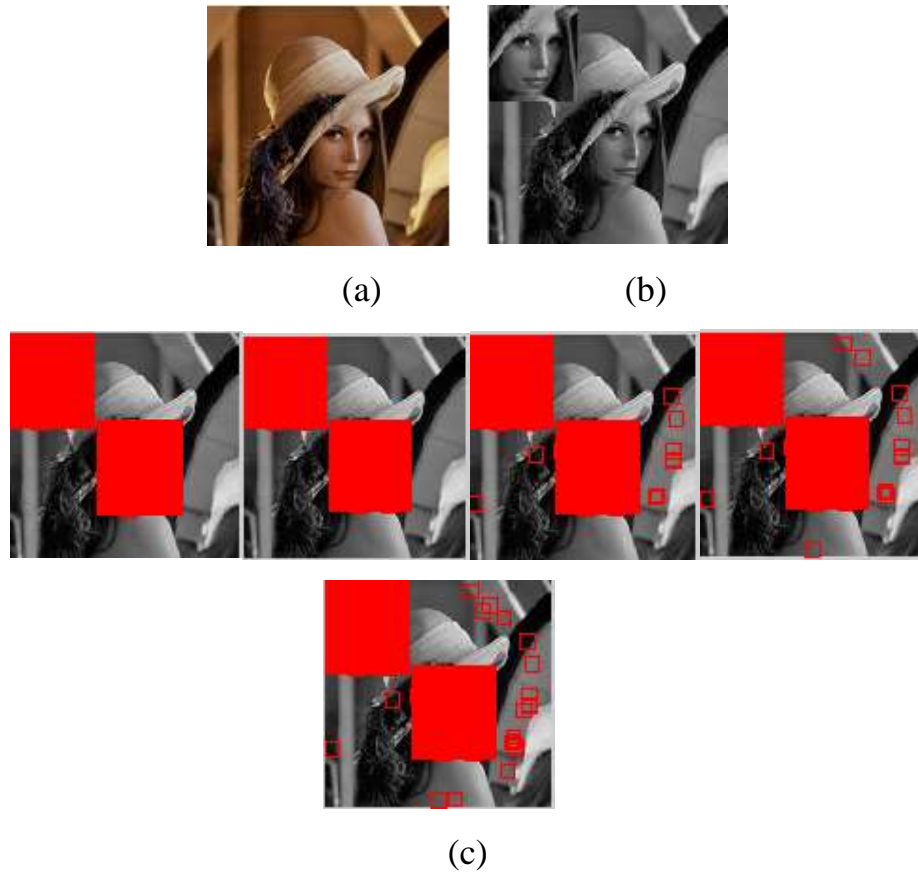


Figure 4.9: The results for different Thresholds: (a) Original Lena image; (b) Forged image; and (c) The result with $T=0.1$, $T=0.2$, $T=0.3$, $T=0.4$, and $T=0.5$.

Table 4.8: True positive and false positive for image in Figure 4.9.

Threshold	Number of Blocks		
	Detection	TP	FP
0.1	1802	1802(100%)	0 (0%)
0.2	1083	1802(99.9%)	1(0.1%)
0.3	1827	1802(98.63%)	25(1.37%)
0.4	1832	1802(97.99%)	37(2.01%)
0.5	1852	1802(97.3%)	50(2.7%)

Table 4.9 demonstrates the performance comparison among the proposed techniques and the other related works.

Table 4.9: The accuracy rate under different conditions.

	Rotation	Scaling	Gaussian blur	Multiple detection	Without modification
G. Lynch [57]	0%	0%	30%	0%	97%
Y. Huang [10]	99.9%	0%	90%	0%	99.9%
DCT based technique	100% (when less than 2 degrees)	100% (when less than 10%)	0%	99%	99.9%
Framing technique	90%	99.9%	90%	99%	99.9%

Chapter five

Conclusion and Future Works

CHAPTER FIVE

CONCLUSION AND FUTURE WORKS

5.1 Introduction

This chapter finishes this thesis by touching some important conclusions in section (5.2) and some suggestions for future work in section (5.3).

5.2 Conclusion

In this thesis, two techniques have been proposed by using the strengths of fast k-mean to detect the copy-move forgery. These techniques fall under passive authentication, and they don't need any prior information related to the original image.

The main important conclusions are presented as follows:

- 1- The experimental results present that the techniques can detect copy-move, and multiple copy-move forgeries in the same image.
- 2- they are robust to some common processing as blurring, scaling, and rotation.
- 3- Moreover, these techniques can detect the duplicated regions efficiently, with a minimal processing time.
- 4- The obtained accuracy is high and the time to find tampering is low.
- 5- The experiment results show that these proposed techniques have the ability to detect copy-move, and multiple copy-move forgeries in an image faster than other recently existing techniques.

5.3 Future works

The following improvements can be investigated in future works to improve the performance of the proposed techniques:

1. Use Algorithms Based on invariant key point to develop a modified technique faster and more accurate using its characteristics.
2. Working on developing fan mechanism starting in image center and identifies suspicious parts in the image before starting the detection process to focuses on small parts and reduce the cost of efforts and time.
3. work on detecting modified videos such as deep fake videos.

References

REFERENCES

- [1] G. Liu, J. Wang, S. Lian and Z. Wang, "A passive image authentication scheme for detecting region-duplication forgery with rotation", *Journal of Network and Computer Applications*, vol. 34, no. 5, (2010), pp. 1557–1565.
- [2] A. Kaur, and R. Sharma, "Copy-move forgery detection using DCT and SIFT," *International Journal of Computer Applications* vol. 70, no. 7 pp:30-34 (2013).
- [3] B. Mahdian and S. Saic, "Blind methods for detecting image fakery", *IEEE Aerosp. Electron. Syst. Mag.*, vol. 25, (2010), pp. 18–24.
- [4] J. Waleed, D. A. Abdullah and M. H. Khudhur, "Comprehensive Display of Digital Image Copy-Move Forensics Techniques," 2018 International Conference on Engineering Technology and their Applications (IICETA), Al-Najaf, pp. 155-160, 2018.
- [5] J. Waleed, Huang D. J. and Saad H., "An optimized digital image watermarking technique based on cuckoo search (CS)", *ICIC Express Letters, Part B: Applications*, Vol. 6, No. 10, pp. 2629-2634, 2015.
- [6] J. Waleed, Huang Dong Jun, Sarah Saadoon, Saad H., Hiyam H., "An Immune Secret QR-Code Sharing based on a Twofold Zero Watermarking Scheme," *International Journal of Multimedia and Ubiquitous Engineering* Vol.10, No.4, pp.399-412, 2015.
- [7] A. D. Warbhe, R. V. Dharaskar, V. M. Thakare, "Computationally Efficient Digital Image Forensic Method for Image Authentication", *Procedia Computer Science*, Vol. 78, pp. 464 – 470, 2016.
- [8] J. Fridrich, D. Soukalm, J. Luka , "Detection of copy-move forgery in digital images," *Digital Forensic Research Workshop*, Cleveland, OH, pp. 19–23, 2003.
- [9] A. C. Popescu and H. Farid, "Exposing digital forgeries by detecting duplicated image regions," *Dept. Computer. Science, Dartmouth College*, Tech. Rep. TR2004-515, 2004.

[10] Yanping Huang, Wei Lu, Wei Sun and Dongyang Long, "Improved DCT-based Detection of Copy-Move Forgery in Images", Forensic Science International, vol. 206, pp. 178-184, 2011.

[11] Nathalie Diane Wandji, Sun Xingming, and Moise Fah Kue, " Detection of copy-move forgery in digital images based on DCT", IJCSI International Journal of Computer Science Issues, Vol. 10, No 1, March Y. Huang 2013.

[12] Davarzani.R, Yaghmaie.K, Mozaffari.S, Tapak.M "Copy-move forgery detection using multiresolution local binary patterns", Forensic science international 231(1-3):61-72 · September 2013.

[13] Jen-Chun Lee, "Copy-move image forgery detection based on Gabor magnitude", Journal of Visual Communication and Image Representation, Vol. 31, pp. 320-334, 2015.

[14] B. Ustubioglu, G. Ulutas, M. Ulutas, V. V. Nabiyev, "A new copy move forgery detection technique with automatic threshold determination", AEU - International Journal of Electronics and Communications, Vol. 70, No. 8, pp. 1076-1087, 2016.

[15] Sondos M. Fadl, Noura A. Semary, "Robust Copy–Move forgery revealing in digital images using polar coordinate system", Neurocomputing, Vol. 265, No. 22, pp. 57-65, 2017.

[16] G. Lynch, F. Y. Shih and H. Y. M. Liao, "An efficient expanding block algorithm for image copy-move forgery detection." Elsevier Information Sciences, vol. 239, pp. 253-265, 2013.

[17] Jumana Waleed, Taha Mohammed Hasan and Thekra Abbas, "Comprehensive Expansion in Anti-Forensics Techniques (AFTs) Based Compressed Image", Annual Conference on New Trends in Information & Communications Technology Applications-(NTICT'2017), pp. 156-161, March 2017.

[18] Charmil Nitin Bharti, Purvi Tandel, "A Survey of Image Forgery Detection Techniques", IEEE WiSPNET 2016 conference, pp. 877-881, 2016.

- [19] Marco Fontani, Tiziano Bianchi, Alessia De Rosa "A Forensic Tool for Investigating Image Forgeries", International Journal of Digital Crime and Forensics, Vol. 5, No. 4, pp. 15-13, October 2013.
- [20] H. Ling, H. Cheng, Q. Ma, F. Zou, and W. Yan, "Efficient image copy detection using multiscale fingerprints," IEEE Magazine of Multimedia, Vol. 19, No. 1, pp. 60-69, 2012.
- [21] Jumana Waleed, Huang Dong Jun and Saad Hameed, "An optimized digital image watermarking technique based on cuckoo search (CS)", ICIC Express Letters, Part B: Applications, Vol. 6, No. 10, pp. 2629-2634, October 2015.
- [22] E. Ardizzone, A. Bruno and G. Mazzola, "Copy-move forgery detection via texture description", MiFor'10 – Proceedings of the 2010 ACM Workshop on Multimedia in Forensics, Security and Intelligence, Co-located with ACM Multimedia, pp. 59–64, 2010.
- [23] Anil Dada Warbhe, Rajiv V. Dharaskar, Vilas M. Thakare, "Digital image forensics: An affine transform robust copy-paste tampering detection," 2016 10th International Conference on Intelligent Systems and Control (ISCO), Coimbatore, pp. 1-5, 2016.
- [24] Mokhles, Jumana , Dhahir "Comprehensive Display of Digital Image Copy-Move Forensics Techniques" 2018 2nd International Conference for Engineering, Technology and Sciences of Al-Kitab (ICETS) ,pp.155-160,4-6 Dec 2018.
- [25] G. K. S. Gaharwar, Prof. V. V. Nath, R. D. Gaharwar, "Comprehensive Study of Different Types Image Forgeries", International Conference on Recent Advances in Engineering Science and Management, August 2015.
- [26] Nikhilkumar P. Joglekar1 "A Compressive Survey on Active and Passive Methods for Image Forgery Detection "IJECS Volume 4 Issue 1 January, pp.10187-10190, 2015.
- [27] J. A. Redi, W. Taktak and J. L. Dugelay, "Digital image forensics: a booklet for beginners", Multimedia Tools Appl., vol. 51, no. 1, pp. 133–162, 2011.
- [28] Nor Bakiah A., Ainuddin Wahid A. W., Mohd Yamani I. I., Roziana Ramli, Rosli Salleh, Shahaboddin Shamshirband, Kim-Kwang R. C., "Copy-Move Forgery Detection: Survey, Challenges and Future Directions", Journal

of Network and Computer Applications, Vol. 75, pp. 259-278, November 2016.

[29] Bayram, S., Sencar, H.T., Memon, N.: ‘A survey of copy-move forgery detection techniques. IEEE Western New York Image Processing Workshop, 2008.

[30] Elaskily, M. A., Aslan, H. K., Elshakankiry, O. A., Faragallah, O. S., El-Samie, F. E. A., & Dessouky, M. M. “Comparative study of copy-move forgery detection techniques”. 2017.

[31] Ali Qureshi, M., & Deriche, M. “A review on copy move image forgery detection techniques”. *IEEE 11th International Multi-Conference on Systems, Signals & Devices (SSD14) -2014*

[32] Granty Regina Elwin J.; Aditya T.S.; Madhu Shankar . “Survey on passive methods of image tampering detection”. *International Conference on Communication and Computational Intelligence (INCOCCI),2010*.

[33] Rani Susan Oommen, Jayamohan M., Sruthy S., “A Survey of Copy-Move Forgery Detection Techniques for Digital Images”, International Journal of Innovations in Engineering and technology. April 2015.

[34] J. Fridrich, D. Soukalm, J. Lukáš, “Detection of copy-move forgery in digital images,” Digital Forensic Research Workshop, Cleveland, OH, pp. 19–23, 2003.

[35] Fadl, S. M., & Semary, N. A. (2014). A proposed accelerated image copy-move forgery detection. 2014 IEEE Visual Communications and Image Processing Conference.

[36] Fahim, A. M., Salem, A. M., Torkey, F. A., & Ramadan, M. A. (2006). An efficient enhanced k-means clustering algorithm. Journal of Zhejiang University-SCIENCE A, 7(10), 1626–1633.

[37] C. Elkan, "Using the triangle inequality to accelerate k-means." ICML. Pp. 147-153, 2003.

[38] Parveen, A., Khan, Z. H., & Ahmad, S. N. (2019). Block-based copy–move image forgery detection using DCT. Iran Journal of Computer Science.

- [39] H. Huang, W. Guo, and Y. Zhang, "Detection of Copy-Move Forgery in Digital Images Using SIFT Algorithm," in Proceedings of IEEE Pacific-Asia Workshop on Computational Intelligence and Industrial Application, Vol. 2, pp. 272-276, 2008.
- [40] S. Bravo-Solorio and A. K. Nandi, "Automated detection and localization of duplicated regions affected by reflection, rotation and scaling in image forensics", Signal Proc., vol. 91, no. 8, pp. 1759–1770, 2011.
- [41] R.S. Oommen, M. Jayamohan, S. Sruthy, "A survey of copy-move forgery detection techniques for digital images", Int J Innov Eng Technol, 5 (2) (2015), pp. 429-436.
- [42] M. A. Elaskily, H. K. Aslan, O. A. Elshakankiry, O. S. Faragallah, F. E. A. El-Samie and M. M. Dessouky, "Comparative study of copy-move forgery detection techniques," 2017 Intl Conf on Advanced Control Circuits Systems (ACCS) Systems & 2017 Intl Conf on New Paradigms in Electronics & Information Technology (PEIT), Alexandria, pp. 193-203, 2017.
- [43] Guangjie Liu, Junwen Wang, Shiguo Lian, Zhiquan Wang, "A passive image authentication scheme for detecting region-duplication forgery with rotation", Journal of Network and Computer Applications, Vol. 34, No. 5, pp. 1557-1565, September 2011.
- [44] Nor Bakiah A., Ainuddin Wahid A. W., Mohd Yamani I. I., Roziana Ramli, Rosli Salleh, Shahaboddin Shamshirband, Kim-Kwang R. C., "Copy-Move Forgery Detection: Survey, Challenges and Future Directions", Journal of Network and Computer Applications, Vol. 75, pp. 259-278, November 2016.
- [45] Hwei-Jen Lin, Chun-Wei Wang and Yang-Ta Kao, "Fast copy-move forgery detection", WSEAS Transactions on Signal Processing, vol. 5, no. 5, pp. 188-197, 2009.
- [46] Vivek Kumar Singh and R.C. Tripathi, "Fast and efficient region duplication detection in digital images using sub-blocking method", International Journal of Advanced Science and Technology, Vol. 35, pp. 93-102, October 2011.
- [47] Devanshi Chauhan, Dipali Kasat, Sanjeev Jain, Vilas Thakare, "Survey On Keypoint Based Copy-move Forgery Detection Methods on Image", Procedia Computer Science, Vol. 85, pp. 206-212, 2016.

[48] Anil Dada Warbhe, R. V. Dharaskar, V. M. Thakare, "A Survey on Keypoint Based Copy-paste Forgery Detection Techniques", *Procedia Computer Science*, Vol. 78, pp. 61-67, 2016.

[49] H. Soleimani and M. Khosravifard, "Mutual Information-Based Image Template Matching with Small Template Size," 2011 7th Iranian Conference on Machine Vision and Image Processing, Tehran, pp. 1-5, 2011.

[50] S. Chakraborty, "Copy move image forgery detection using mutual information," 2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT), Tiruchengode, pp. 1-4, 2013.

[51] X. Kang and S. Wei, "Identifying Tampered Regions Using Singular Value Decomposition in Digital Image Forensics," 2008 International Conference on Computer Science and Software Engineering, Wuhan, Hubei, pp. 926-930, 2008.

[52] S. Bhosale, G. Thube, P. Jangam and R. Borse, "Employing SVD and Wavelets for Digital Image Forensics and Tampering Detection," 2012 International Conference on Advances in Mobile Network, Communication and Its Applications, Bangalore, pp. 135-138, 2012.

[53] T.Y.Kong, "Critical kernels, minimal non-simple sets, and hereditarily simple sets in binary images on n-dimensional polytopal complexes", *Journal - Science direct . Theory, Methods and Applications*, 2017, Pages 211-256.

[54] Mizuho Nakajima and Yasushi Yamaguchi, "Extended Visual Cryptography for Natural Images", The University of Tokyo 3-8-1 Komaba, Meguro-ku, Tokyo 153-8902, Japan.

[55] Young-Chang Hou, "Visual cryptography for color images", *Pattern Recognition*, Vol. 36, No. 7, pp. 1619-1629, 2003.

[56] Xinyang, Henan, "A Clustering Method Based on K-Means Algorithm" *Trans Tech Publications*, Switzerland. Pp. 1697-1700, 2013.

[57] S.Fadl , N.Semary ,M.Hadhoud, "Copy-Rotate-Move Forgery Detection Based on Spatial Domain", pp.136-141, 2014.

Publications

[1] M. H. Khudhur et al., "An Efficient and Fast Digital Image Copy-Move Forensic Technique," *2018 2nd International Conference for Engineering, Technology and Sciences of Al-Kitab (ICETS)*, Karkuk, Iraq, 2018, pp. 78-82.

[2] J. Waleed, D. A. Abdullah and M. H. Khudhur, "Comprehensive Display of Digital Image Copy-Move Forensics Techniques," *2018 International Conference on Engineering Technology and their Applications (IICETA)*, Al-Najaf, 2018, pp. 155-160.

الملخص

تمثل الصور وسائط اتصال فعالة وطبيعية للبشر ، نظرًا لسهولة فهم محتوى الصورة. ونظرًا لتوفر الأجهزة الرقمية على نطاق واسع ، فإن أدوات تحرير الصور مفتوحة المصدر والمتاحة تجاريًا جعلت من موثوقية محتويات الصور محل تساؤل. وادى ذلك إلى زيادة الحاجة إلى استخدام خوارزميات الكشف عن التزوير. ادوات نسخ وتحريك لتزوير الصور الرقمية (CMF) وهو أسلوب شائع لإنتاج صور يتم العبث بها عن طريق إخفاء الأشياء غير المرغوب فيها أو تكرار الكائنات المرغوبة في نفس الصورة. لذلك هناك حاجة إلى وسائل لاثبات مصداقية محتويات الصورة وتحديد المناطق التي تم العبث بها. لقد تم تطوير العديد من خوارزميات الكشف عن نسخ الصور الرقمية ، هذه الخوارزميات المستندة إلى (DCT)، وخوارزميات تستخدم invariant image moments ، وخوارزميات تستخدم texture and intensity descriptors ، والخوارزميات التي تستخدم invariant key points والخوارزميات التي تستخدم mutual information ، والخوارزميات التي تستخدم SVD للحكم على الصورة الرقمية اذا ما كانت تحتوي على تزوير.

في هذه الاطروحة ، تم اقتراح تقنيات قوية للكشف عن التزوير بالنسخ والتحريك CMF وتحديد ما في الصور الرقمية. تستخرج هذه التقنيات الخصائص للكشف عن عمليات التزوير في الصور الرقمية وتحديد ما إذا كان المحتوى أصليًا أم معدلًا دون الاعتماد على أي معرفة بالمعلومات السابقة المتعلقة بالصورة المصدر. يتم تقليل الخصائص المستخرجة من الصورة باستخدام (DCT) أو باستخدام طريقة التآطير (التقسيم)، لتقييم التقنيات المقترحة تم استخدام الصور المزورة بواسطة تطبيق IGMP الشائع لغرض التجربة. يمكن تطبيق تقنيات الكشف عن التزوير المقترحة للكشف عن المناطق المزورة ويمكن الحصول على فوائد في تطبيقات الكشف عن التزوير في الصور الرقمية.

يوضح تحليل الأداء أن التقنية المقترحة الأولى يمكنها اكتشاف المناطق متعددة التكرار بكفاءة ، مع الحد الأدنى من وقت المعالجة. بينما ، يمكن أن تكتشف التقنية الثانية المناطق

متعددة التكرار بكفاءة حتى عندما يتم تعديل الصورة بالضغط JPEG ، اوتدوير المنطقة المستنسخة ، او تغيير حجمها ، اوتتعيمها أيضا ، وبالرغم من ذلك فهي تقلل من وقت المعالجة.



جمهورية العراق
وزارة التعليم العالي والبحث العلمي
جامعة ديالى - كلية العلوم - قسم
علوم الحاسبات



تقنية حديثة و فعالة للكشف عن التزوير في الصور الرقمية

رسالة مقدمة الى قسم علوم الحاسوب في كلية العلوم/ جامعة ديالى وهي جزء من
متطلبات نيل درجة الماجستير في علوم الحاسوب

من قبل

مخلص حسين خضر

بإشراف

أ.د. ظاهر عبد الهادي عبد الله

أ.م.د. جمانة وليد