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

2020 AC **1441 AH**

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

بِسۡمِ ٱللَّهِ ٱلرَّحۡمَٰنِ ٱلرَّحِيمِ

﴿ نَرۡفَعُ دَرَجَٰتٖ مَّن نَّشَآءُۗ وَفَوۡقَ كُلِّ ذِي عِلۡمٍ عَلِيمٞ ﴾

صَدَقَ ٱللَّهُ ٱلعَظِيم

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

List of Tables

List of Figures

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.

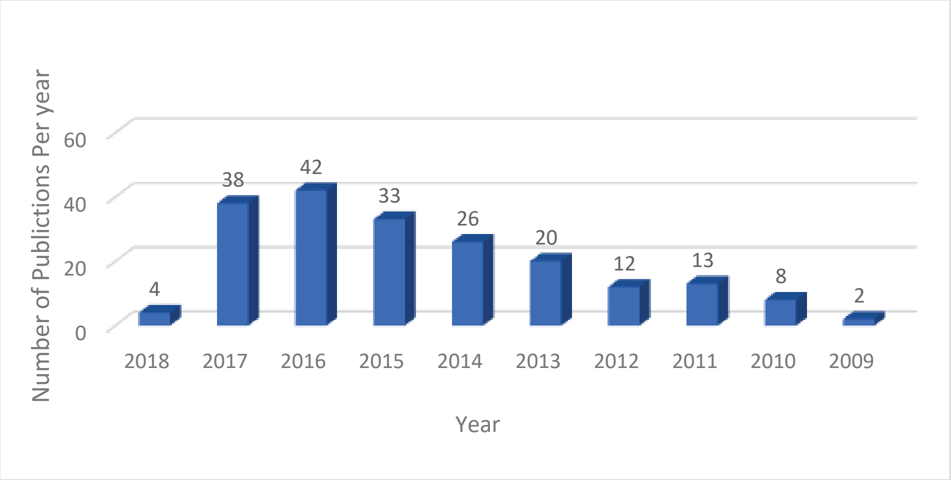

Figure 1.1: The No. of publications in the "IEEE" within the field of CMF detection techniques

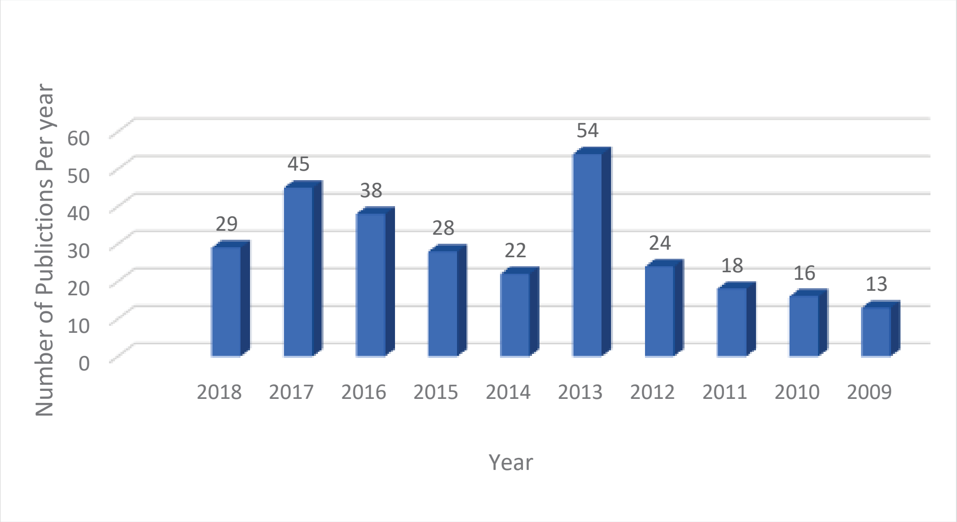

Figure 1.2: The No. of publications in the "Science Direct" within the field of CMF detection Techniques

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 demotion 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, considerately 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.

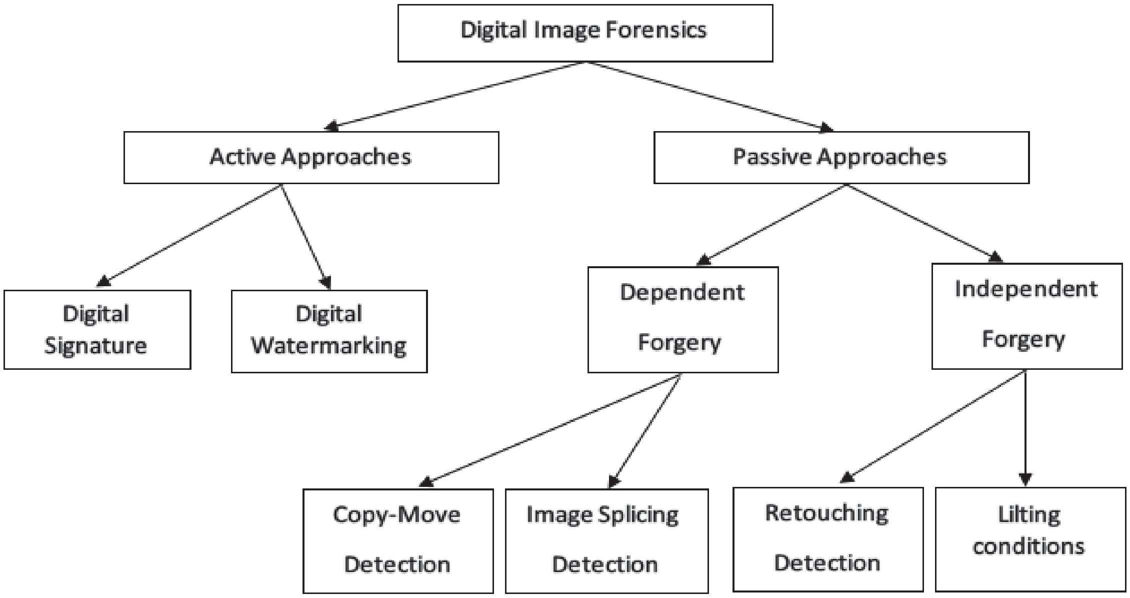

Figure 2.1: Digital Image Forensic Techniques [24].

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].

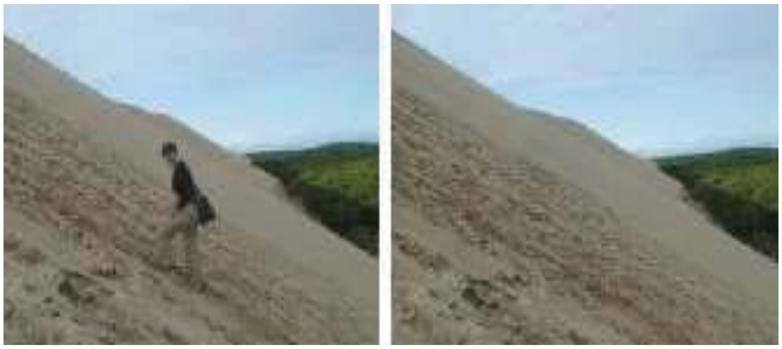

Figure 2.2: Copy-Move Forgery [30].

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].

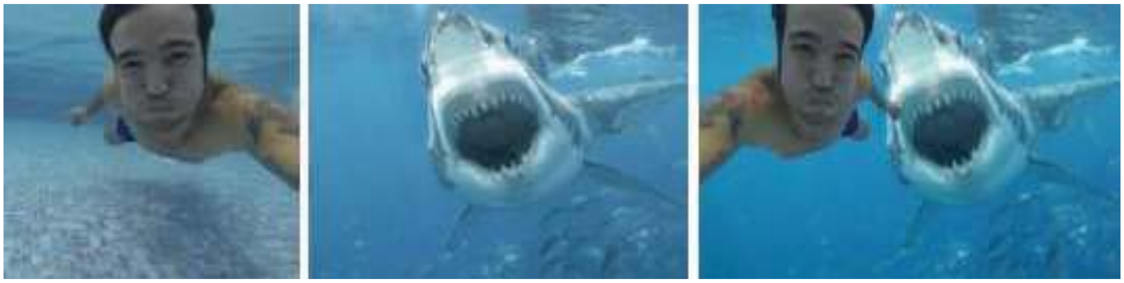

Figure 2.3: Image splicing [30].

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).

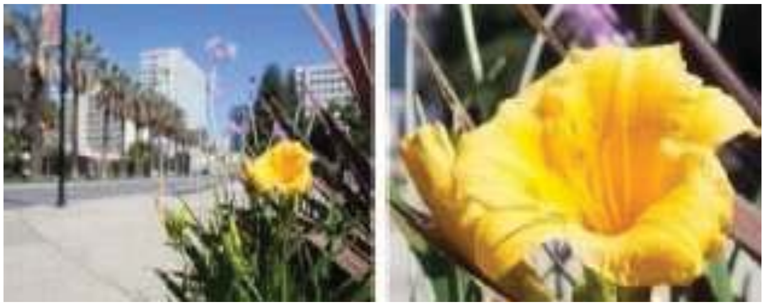

Figure 2.4: Image Resampling [31].

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).

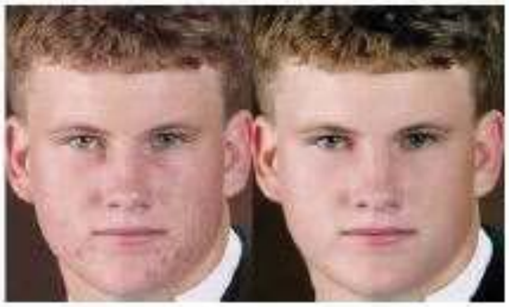

Figure 2.5: Image retouching [31].

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.