An Improved Block Based Copy-Move Forgery Detection Technique

A Thesis Submitted in Partial Fulfilment of the Requirement for the Award of the Degree of

MASTER OF ENGINEERING

in

Electronics and Communication Engineering

Submitted by

PRIYANKA

801661017

Under Supervision of

Dr. Kulbir Singh

Professor, ECED

ELECTRONICS AND COMMUNICATION ENGINEERING DEPARTMENT

THAPAR INSTITUTE OF ENGINEERING AND TECHNOLOGY

(A DEEMED TO BE UNIVERSITY), PATIALA, PUNJAB-147004

JULY-2018

DECLARATION

I, Priyanka hereby declare that the work presented in this thesis entitled "**AN IMPROVED BLOCK BASED COPY-MOVE FORGERY DETECTION TECHNIQUE**" in partial fulfillment of the requirement for the award of degree of Master of Engineering (ECE) submitted at Electronics and Communication Engineering Department, Thapar Institute of Engineering & Technology (Deemed to be University), Patiala is an authentic record of work carried out under supervision of **Dr. Kulbir Singh (Professor)** Electronics and Communication Engineering Department of Thapar Institute of Engineering and Technology, (Deemed to be University), Patiala. The matter presented in this thesis has not been submitted either in part or full to any other university or institute for the award of any other degree.

Priyanka

Date: 13/7/18

Priyanka
Roll No: 801661017

It is certified that the above statement made by the student is correct to the best of my knowledge and belief.

Date: 13/7/18.

Dr. Kulbir Singh
Professor
Electronics and Communication Engineering Department
Thapar Institute of Engineering and Technology
(A Deemed To Be University), Patiala, Punjab

ACKNOWLEDGEMENT

First of all, I would like to express my gratitude to **Dr. Kulbir Singh, Professor,** Electronics and Communication Engineering Department, Thapar Institute of Engineering & Technology, Patiala (A Deemed To Be University) for his patient guidance and support throughout this report. I am truly very fortunate to have the opportunity to work with him. I found this guidance to be extremely valuable.

I am also thankful to our Head of Department, **Dr. Alpana Agarwal**. I would like to thank the entire faculty and staff of Electronics and Communication Engineering Department and then friends who devoted their valuable time and helped me in all possible ways towards successful completion of this work. I thank all those who have contributed directly or indirectly to completion of this work.

Lastly, I would like to thanks my parents for their years of unyielding love and encourage they have always wanted the best for me and I admire their determination and sacrifice.

Priyanka
ME (ECE)
801661017

ABSTRACT

With the increase in demand of identification of authenticity of the digital images, researchers are widely studying the image forgery detection techniques. Copy-move forgery is amongst the commonly used forgery, which is performed by copying a part of image and then pasting it on the same or different image. This results in concealing of image content. Most of the existing copy-move forgery detection techniques are subjected to degradation in results, under the effect of geometric transformations. In this work, a Discrete Cosine Transformation (DCT) and Singular Value Decomposition (SVD) based technique is proposed to detect the copy-move image forgery. DCT is used to transform the image from spatial domain to frequency domain and SVD is used to reduce the feature vector dimension. Combination of DCT and SVD makes the proposed scheme robust against compression, geometric transformations and noise. For classification of images as forged or authentic, Support Vector Machine (SVM) classifier is used on the feature set. Once the image is detected as forged, then for the localization of forged region, K-means clustering is used on the feature vector. According to distance threshold, similar blocks are identified and marked. The application of SVD provides the stability and invariance from geometric transformations. SVM classifier classifies the images as forged or authentic. Evaluation of the proposed scheme is done with and without post-processing operations on the images, both at pixel level and image level. Pixel level analysis shows the accuracy of proposed scheme in detecting forgery within an image. Image level analysis shows the accuracy of proposed scheme in image classification whether it is forged or original. The proposed scheme outperforms the various state-of-the-art techniques of Copy-Move Forgery Detection (CMFD) in terms of accuracy, precision, recall and F_1 parameters. Moreover, the proposed scheme also provides better results against various attacks such as rotation, scaling, noise addition and JPEG compression.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

CMFD	Copy Move Forgery Detection
JPEG	Joint Photographic Experts Group
SVD	Singular Value Decomposition
DCT	Discrete Cosine Transform
DWT	Discrete Wavelet Transform
SIFT	Scale Invariant Feature Transform
SURF	Speed Up Robust Feature
TPR	True Positive Ratio
FPR	False Positive Ratio
PCA	Principle Component Analysis
SVM	Support Vector Machine
LPT	Log-Polar Transform
LOG	Laplacian Of Gaussian
FMT	Fourier Miller Transform
VAM	Visual Attention Model
QEM	Quaternion Exponent Moment
RANSAC	Random Sample Consensus
CWT	Continuous Wavelet Transform
SATS	Same Affine Transformation Selection
MHJ	Multi-Hop Jump
QCD	Quantization Component Decomposition
PCET	Polar Complex Exponential Transform
LBP	Local Binary Pattern
SPT	Steerable Pyramid Transform
AUC	Area Under Curve
KLT	Kanade-Lucas-Tomasi
CBF	Counting Bloom Filters
LSH	Locally Sensitive Hashing

CHAPTER 1
INTRODUCTION

Revolution of digital techniques has provided ease of availability of related software. People of every age group frequently use them in their day to day life. Digital cameras and image editing software are also one of them. Purpose of using this software may vary from person to person. Someone may use this to improve or enhance the information, whereas others may use it to alter data. Nowadays, digital media has become an important part of our life, as we encounter with thousands of images in our daily life. Images play an important role in delivering useful information in various fields. So, one must be sure about the authenticity of these images. Misinterpretation of image data may lead to serious conflicts and disastrous results. Image forensics field handles the issues related to the image forgery. A Passive, popular and tampering dependent technique is copy-move forgery, which can misinterpret data. So, there is a need to explore this field [1].

1.1 IMAGE FORGERY DETECTION METHODS

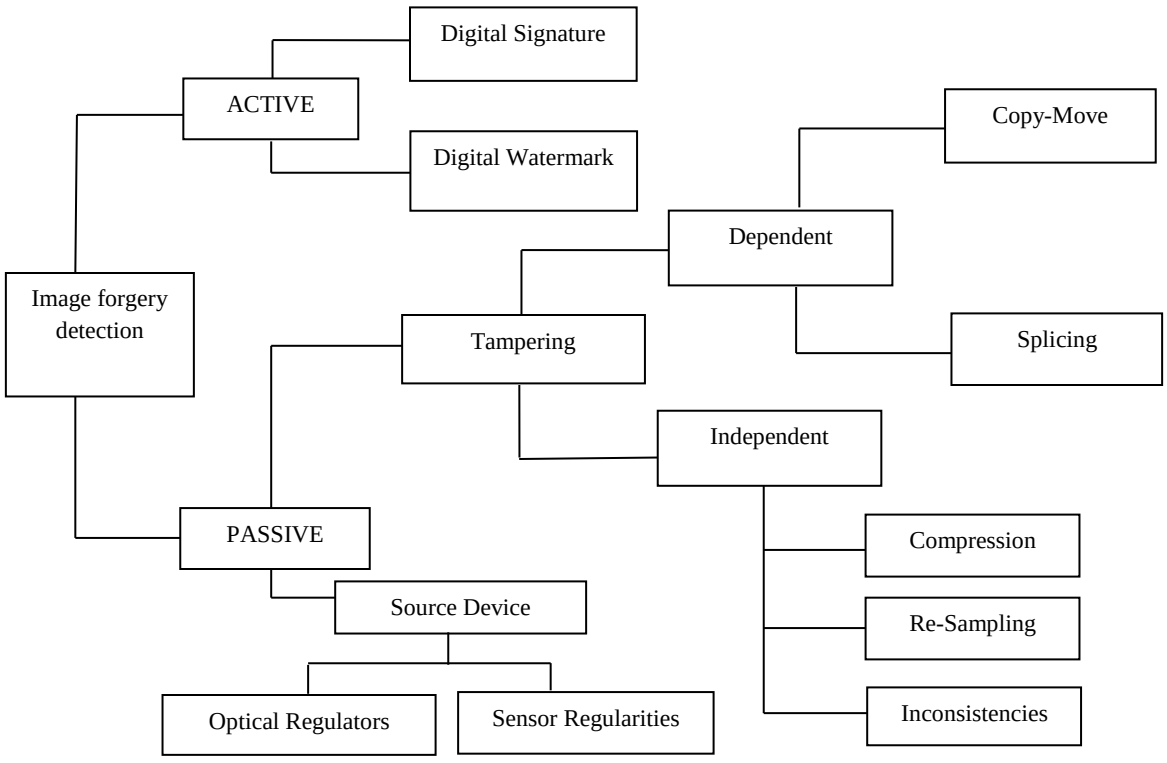


Figure 1.1: Classification of Image forgery Detection techniques [1]

1.1.1 Classification of Passive Image Forgery Detection Methods

A. Pixel Based Techniques

Pixel based techniques are based on detecting the statistical abnormalities which are introduced at the pixel level when images are being forged. Analysis of the pixel level correlation is done in order to detect forgery in the spatial or transformed domain.

B. Format Based Techniques

For efficient storage and transmission, every image undergoes compression at least once during its life cycle. Doctored images are also compressed after the tampering. This compression leaves many traces, which may reveal the tampering. JPEG is one of the widespread used compression technique.

C. Camera Based Technique

Every camera adds some artifacts in the captured image. Camera based techniques detect the traces of any modification at various stages of image processing by using color filter array, chromatic aberration, sensor noise and camera response.

D. Source Camera Identification Based Techniques

In this technique, the specifications of a camera are used for detection of tampered images by analyzing the CFA, lens aberration and sensor noise etc.

E. Physics Based Techniques

Matching the lighting conditions of the different images is a difficult task. The variations in lighting conditions of the captured image can reveal the tampering of the image.

F. Geometric Based Techniques

Every image has a principle point at its center, which shifts when these images are modified by copy move or splicing. With the help of projective geometry principles, forgery detection method can be designed [2].

As we know, forgery detection in images can be of Active and Passive class. Copy-move forgery is among the passive ones. Copy-move forgery resulted from copying-pasting of one or more than one, small or large area of images. This type of forgery is easy to perform and is very common. The main motive behind this type of activity is to hide something from an image or to manipulate the information contained by the image. In order to make this forged image undetectable by naked eyes, many conditions

are taken into consideration. Illumination and the temperature of the colors and lighting conditions should be properly matched [1]. Work process of copy-move forgery is shown in Figure 1.2.

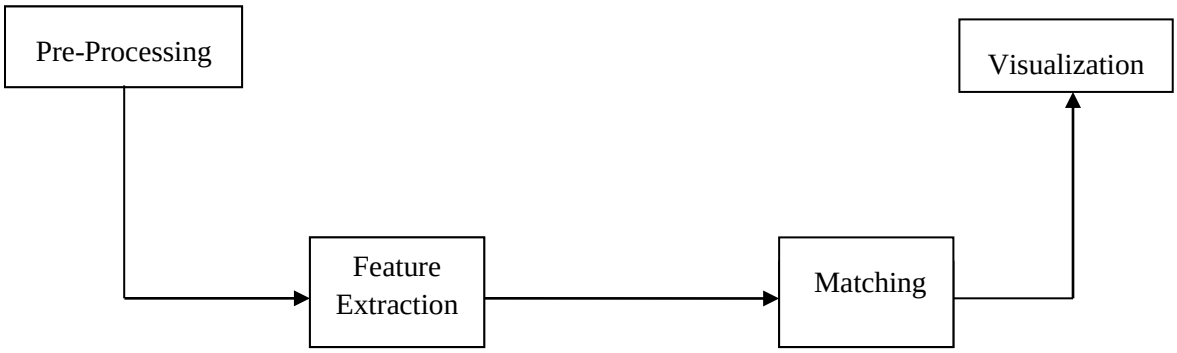


Figure 1.2: Work process of Copy-Move Forgery Detection Technique [1]

Copy-move forgery commonly consists of four stages as shown in Figure 1.2. Preprocessing is the optional one, where an improvement in the image is done by removing distortions and enhancement of important features contained by the it. Colored images are converted to gray scale image, which reduces the dimension of image data and makes the features distinct. Processing speed also increases. Most of the techniques also used block division process. Blocks help in formation of a simple algorithm for the feature matching process. Once image is transformed into blocks, features are taken out of each block, which is known as feature extraction. The features contain unique information which is also useful. Features chosen should be less affected by post-processing operations done on the tampered image. Sometimes invariant features are given more importance like as key point based, texture and intensity. After the features are extracted, then matching procedure is started on those extracted features. Similar features in the image are find out. Feature matching techniques are classified as block matching and key point matching schemes. In block matching technique, features of every block is matched with the features of another block and similar features are kept together. Whereas in key point matching technique, key points are matched in the whole image and matching points are joined by lines [1].

1.1.2 Types of Digital Image Forging

Addition of something to the image which is unwanted is called forgery. At present time we are surrounded with the digital images. Many of these images are suffered from forgery. Mainly following types of digital image forgeries are encountered in the images [3]:

Cloned: It is the easiest way of performing copy-move forgery but it is not easy to detect because of post-processing operations performed on the image.

Figure 1.3: Example of cloning used for hiding the real content of the image

Spliced: It is a procedure of making a forged image with the help of two or more than two images. Borders of the different used images are hidden in the forged image using some operations, in order to make it non-detectable.

Figure 1.4: Example of splicing used for manipulating the content of images where data from the other image is inserted into an image to make it forged.

Retouched: With the help of special dyes or digital retouching we can remove, repair and enhance features in the target image.

Figure 1.5: Example of Retouching used for changing the features of the original image

Morphed: Conversion of one image to other digitally is called as morphing. Features of one image are mixed with the features of another image slowly.

Figure 1.6: Morphed image formed by converting a human face into an alien doll

Enhanced: It is done for enhancement of the image features like contrast and edges, which is a very common practice in image processing.

Figure 1.7: Example of color, brightness, contrast and blurriness enhancement

Computer Generated: Computer software can be used to create a forged image and can perform post-processing operations too in order to hide the manipulated areas.

Figure 1.8: Example of computer generated image

1.1.3 Types of Copied Regions

Mainly there are four categories of the images that datasets usually contained namely Background, Object, Creature, Letter etc. Copied regions can also be classified as following on the basis of content they contained as shown in Figure 1.9:

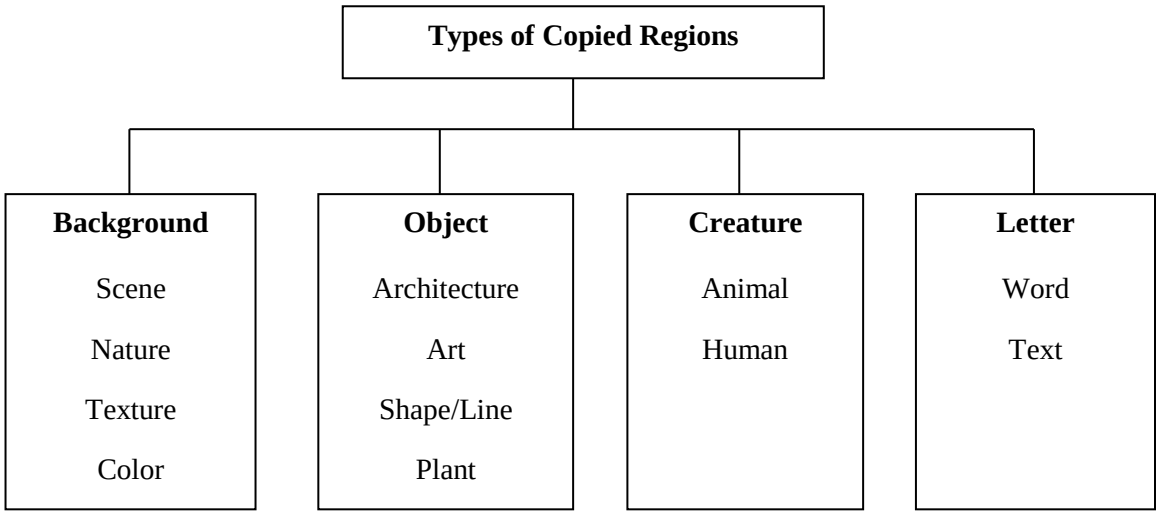


Figure 1.9: Types of copied regions [1]

1.2 BLOCK BASED COPY MOVE FORGERY DETECTION METHODS

These methods make division of image into the equal size of over-lap or nonoverlapping blocks. Features are taken out of every block and then blocks with similar features are searched. These schemes are very simple and well known. Most of the techniques also used block division process. Blocks help in the formation of a simple algorithm for feature matching process. After block division, features are extracted

from each block, which is known as feature extraction. Features contain unique information, which is also useful. Feature extraction for a block based scheme can be done in following ways [1]:

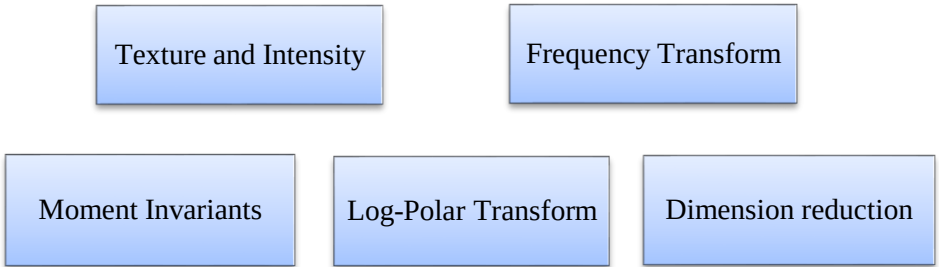


Figure 1.10: Methods of feature extraction

Copy-Move tampered image can be treated by a block basis forgery detection method by the series of steps like, preprocessing on input image and division of image into blocks for driving features out as shown in Figure 1.11.

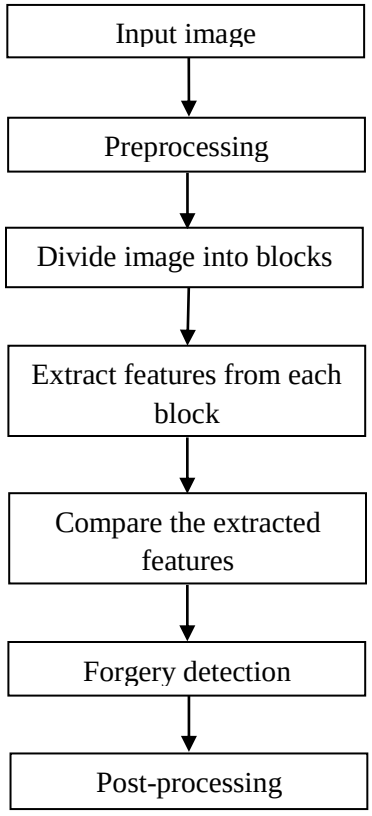


Figure 1.11: Workflow of Block based copy-move forgery detection methods [5]

Frequency Transform

It is the famous and frequently used feature extraction method. It does the best separation between the rotational component and translational one [1]. Frequency transform methods are more robust to noise and are shown as follows:

- Discrete Cosine Transform (DCT)
- Discrete Wavelet Transform (DWT)
- Dyadic Wavelet Transform (DyWT)
- Wiener Filter Wavelet

Out of above-mentioned techniques, DCT is most commonly and frequently used technique. DCT based technique is noise robust and simple to use. We can decrease the size of features in order to lower the processing time and complexity.

Table 1.1 Frequency Transform based feature Extraction method [1]

Technique	Methods	Advantages	Limitations
DCT	• Mismatch of DCT grid and block artifacts are detected. • Higher frequency coefficients of DCT are truncated. • Circular blocks are used instead of square. • SVD is applied on blocks after DCT.	Effective and robust to compression. Feature dimension is reduced. Multiple regions can be detected. Robust to blurring and compression	Complex computations. Works only for small duplicate region. Only works with post-processing. Works well with post-processing.
FWHT	Fast Walsh Hadamard Transform is used due to simplicity.	Accurate and fast.	Affected by transformations.
DWT	Used low frequency subbands.	Less complex.	Location dependent.
DyWT	Comparison between LL band and HH band is done.	Shift and jpeg compression invariant.	Tested with compression and rotation only.
Wiener Filter	Wiener filter is applied on image to extract features.	Robust to compression, scaling, rotation and noise.	Unable to self adjustments.

Moments Invariant

Some set of features are invariable to a different type of transformation like transition, rotation and scaling. Examples of some moment invariant techniques are Zernike moment and exponential moment. Zernike moments are unaffected by rotations than moment. Exponential moments are more robust to noise and smooth distortion conditions [2].

Texture and Intensity

Natural scenes including grass, clouds, trees and ground contain similarity in texture and intensity which can be used as a feature to find out the forged area. There are different types of standard texture descriptors which are named as Intensity, Pattern and Color.

Log-Polar Transform

Log-Polar consists of techniques which are rotation and scaling invariant [2]. These techniques project the points from Cartesian plane to Log polarized plane. Log-Polar Fourier Transform (LPFT) has more robustness against rotation.

Dimension reduction

These methods are used to minimize the dimension of image. Mainly used dimension reduction techniques are LLE and SVD.

1.2.1 Block Based Matching Techniques

After the extraction of features, blocks with similar features are searched. Similar blocks are kept adjacent in a matrix. Various algorithms are designed for block matching purpose. Some of them are given here, which detect the similar blocks on the basis of distance criteria or on basis of the correlation between the blocks [1].

Sorting is a method of arranging a matrix or array in a certain manner. In the matching process of block based technique, sorting helps in matching the feature vectors. Blocks having similar features come in the neighborhood of each other in a sorted matrix. Various sorting techniques are as follows [1]:

- Lexicographical: Sorting on the alphabets basis of feature values.
- Kd-Tree: It is a data structural scheme that searches for the nearest neighbor.

- Radix: For a non comparative integer, radix sorting is performed, where the individual digit is sorted out.

Out of all above three techniques, the lexicographical technique is most widely and commonly used technique. The efficiency of this technique depends upon its usage. Kd-tree is a tree-based sorting technique that divides the blocks array into two parts and neighborhood is compared with a threshold.

Hash can detect any modification in data using hash functions. Similar features have same hash value. Counting Bloom Filters (CBF) and Locally Sensitive Hashing (LSH) use hash functions to detect any sort of modifications. Nearest neighbor algorithm searches for features that have same hash values. Small hash values result in speeding up of searching operation. Hash functions are used for mapping a large size data to a fixed size. These functions are mainly used in searching operations where they find similar values in large files [1].

Correlation is finding the relation between two variables which helps in detection of duplicated region. This is done usually after sorting but can be used before also. Normally used correlation is phase correlation. Correlation peaks which are higher than the threshold value are detected [1].

Euclidean distance is the distance calculated between the two vectors. It is also calculated after the sorting process. Similar blocks are at the neighboring position. If the distance between them is greater than threshold decided for Euclidean distance, then they are called forged region. For a two dimensional geometry, Euclidean distance between two points, c= (c_x,c_y) and d= (d_x, d_y) is defined by Equation (1.1) as [1]:

$$D(c,d) = \sqrt{(d_x - c_x)^2 + (d_y - c_y)^2} \tag{1.1}$$

1.3 KEY POINT BASED COPY MOVE FORGERY DETECTION

Key point based techniques do not divide an image into blocks. Distinct features extraction is done for the image. Some commonly used features are corners, edges, blobs etc. Descriptors are searched out around the feature region in order to increase their reliability against any sort of transformation. Together the feature and descriptors are matched and define a tampered region [7]. Some methods work on block matching techniques to decrease load of computations. They performed the block matching in the spatial

domain. Use of GPU and integral images enhanced the copy-move forgery detection techniques in a method. Sometimes images are expressed in bit plane form, then ASCII codes may assign to them for forgery detection. Sometimes images are segmented into patches and SIFT key points are used to describe the image features which are matched for forgery detection. For matching the Key points, Expectation and Maximization algorithm is used. Workflow of key point based techniques is as follows:

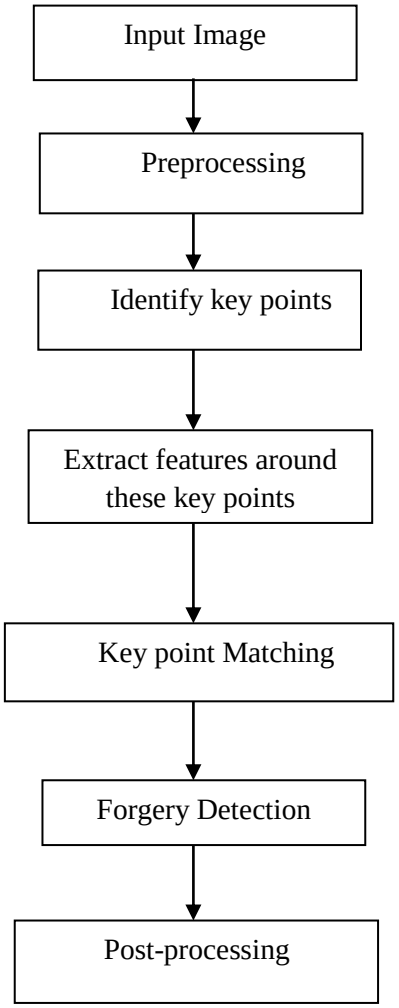


Figure 1.12: Workflow of Key point based Image Forgery [7]

Key point based feature extraction techniques: SIFT, SURF and Harris corner detector are discussed here.

SIFT: It works on local image statistical features. These features are matched in the image to detect duplicated region. These features are robust to scaling, rotation, noise, compression and any other geometric transformation. SIFT with hierarchal clustering detects multiple copied region. SIFT with Euclidean distance give better performance and is robust to scaling and rotation [7].

Harris Corner Detector: The first technique used in key point matching is Harris corner detector. It is named after Chris Harris. It calculates the differential of corner score and can easily distinguish between the edges and corners. Earlier Kanade- Lucas- Tomasi (KLT) and Harris operator were used for capturing corners. Harris corner gives better results, which are robust to illumination change and rotation. Its algorithm is as follows [7]:

- Colored image is changed to grayscale then spatial derivative is calculated.
- Structure tensor is set up.
- Harris response is calculated.
- Nonmaximum suppression is found out.

SURF: It was an improvement to SIFT in processing time and dimension reduction of features. SURF is invariant to post-processing operations and geometric transformation. The drawback of SURF is that it does not give response to copied region which is small in size. SURF descriptor and detector are made by finding key points using Fast Hessian Detector [7].

1.4 EVALUATION CRITERIA

There are various criteria on the basis of which a number of copy-move tampering detection schemes are evaluated and compared with each other. Some of them are accuracy, precision, recall, F_1 score and computational time. Accuracy is defined by True Positive Ratio (TPR) and False Positive Ratio (FPR) values. TPR, FPR, True Negative Ratio (TNR), False Negative Ratio (FNR), Precision, Recall, F_1 score can be calculated by ahead mentioned formulas [1]:

$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{1.2}$$

$$\mathrm{FPR} = \frac{\mathrm{FP}}{\mathrm{FP} + \mathrm{TN}} \tag{1.3}$$

$$\mathrm{FNR} = \frac{\mathrm{FN}}{\mathrm{FN} + \mathrm{TP}} \tag{1.4}$$

$$\mathrm{TNR} = \frac{\mathrm{TN}}{\mathrm{FP} + \mathrm{TN}} \tag{1.5}$$

$$\text{Precision ratio} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{1.6}$$

$$\text{Recall ratio} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{1.7}$$

$$\text{F}_1 = 2.\frac{\text{Precision ratio.Recall ratio}}{\text{Precision ratio} + \text{Recall ratio}} \tag{1.8}$$

where, TP is True Positive, TN is True Negative, FP is False Positive, FN is False Negative

Accuracy is determined as a factor that determines the number of correctly detected images to the total number of images. Accuracy can be calculated from the ROC curve. Area under Curve (AUC) gives the measure of detection accuracy of a classifier. In order to increase the detection accuracy of a classifier, AUC should be maximized in the ROC curve. Precision determines ratio of truly forged images to the sum of detected forged images. Recall is the ratio of truly forged images to the total number of images. Recall is also called as measure of sensitivity. F_1 score is defined as an average of Precision ratio and Recall ratio with specified weights of both [1].

1.5 CONTRIBUTION

Presently, copy-move image forgery detection methods are able to find any sort of tampering in the images with a good precision, recall and F_1 score, in the case of copy-move forgery detection without any post processing operations. But most of the existing techniques do not provide satisfactory results against post-processing operations. In the case of post-processing operations, images undergo various attacks after they are forged, like scaling, JPEG compression, rotation and noise addition. The proposed scheme takes the advantage of SVD which is scaling and rotation invariant. Hence, the proposed scheme performs better even in the presence of various attacks.

1.6 ORGANIZATION OF THESIS

Chapter 1 contains the introduction of image forgery detection techniques. Types of digital image forgeries are discussed in this chapter and main concentration is given on the copy-move forgery detection technique and their classification as Block based and key point.

Chapter 2 provides the literature survey of CMFD techniques. Observations drawn from the study of existing CMFD techniques and gaps are discussed in this chapter. Objectives selected for the proposed

work are also discussed in this chapter. The methodology of proposed scheme is also briefly discussed at the end of the chapter.

Chapter 3 contains the detailed description of the proposed scheme. Workflow of the proposed scheme and pseudo code is also described in this chapter. A detailed description of SVD, SVM and K-mean clustering is given in this chapter. Summary of the chapter is given at the end.

Chapter 4 contains the results of our proposed CMFD scheme and their discussion. Detection results are discussed for CMFD on the given datasets at pixel level and image level and their comparison is done with the existing copy-move forgery techniques with and without post-processing operations. Value of K used for clustering is varied and its effect on the detection results is discussed. The proposed technique is compared with other existing techniques both at pixel and at image level. At pixel level, geometric transformations are performed on the detection techniques and their results are compared.

Chapter 5 provides conclusion and future scope of the proposed work.

CHAPTER 2

LITERATURE REVIEW AND PROBLEM DEFINITION

This chapter contains a brief detail of work done by various researchers in the field of copy-move image forgery detection. From the literature survey, various observations and objectives have been drawn that are listed at the end of the chapter.

2.1 LITERATURE SURVEY

From the blocks of images, features are taken out and in order to reduce their dimension any of the dimension reduction algorithms like PCA, SVD can be used. Sorting of extracted features is done in order to decrease the time complexity and matching loads. On the basis of certain predefined thresholds, similar blocks are matched and forgery is detected. In frequency based methods, DCT is used as feature extraction method. DCT is applied on the images and the resulted coefficients are taken as a feature which represents that image or that particular block. Some methods [8] used quantized DCT as a feature extraction technique. Extracted features undergo sorting and blocks are compared on the basis of certain predefined threshold values, which may be its shift vector. Similar blocks are detected and this method performs well under noise and compression also but not to rotation and scaling. Another method [9] used FMT to make its feature vector which was superior to the previous method in case of rotation and scaling. This method does not use any kind of sorting and is faster than previous one.

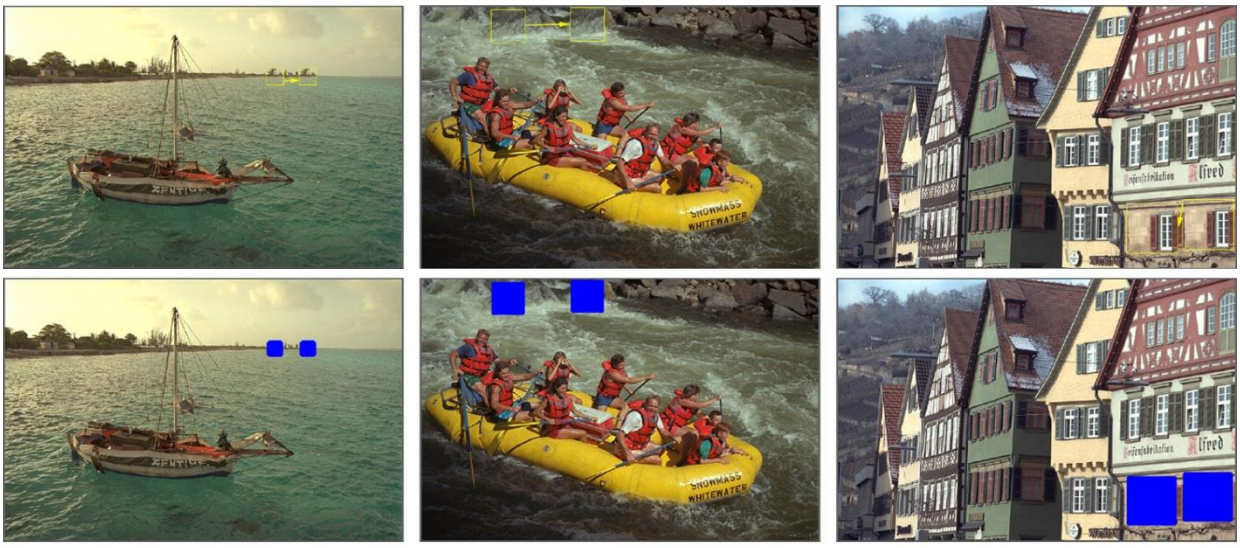


Figure 2.1: CMFD using SVD and DCT method [11]

Instead of DCT, some methods used DWT on the images to extract their features. Distance thresholds are used to detect the forged regions. DWT is affected by rotation and scaling type geometrical operations. Therefore, it is not useful in some cases so, method [10] used dyadic wavelet transform instead of DCT. This method is superior to the DWT methods. In [11] after applying DCT to the blocks, every block undergoes singular value decomposition which reduces its feature vector dimension. Combination of DCT and SVD provide robustness against noise and compression but is only applicable to post-processing operations. Some papers used above feature extraction schemes along with feature dimension reduction techniques such as [12] used DWT along with Kernel PCA. These methods give us the advantage of two combined techniques in a single method. Kernel PCA helps in improvement of technique with noise and compression presence. Sometimes SVD is also used to obtain features which are robust to geometric transformations [13]. Sorting of Blocks are done according to Chebyshev distance between them. This method performs well under blurriness, noise and compression. DWT is also used as feature dimension reduction in some methods [14]. SVD is used on DWT coefficients and sorting of the resulted vector is done.

Figure 2.2: FMT based copy-move forgery detection technique on GIRP Database [9]

Figure 2.3: Original image and tampered image detected by DWT-SVD based method [14]

Some texture based forgery detection techniques like Local binary pattern (LBP) are used in some methods [15]. In method [15] images are passed from low pass filter and then circular blocks are extracted from these images. This method performed well under the effect of noise, compression, blurring and rotation. Method [16] used multi resolution local binary pattern feature which use various types of operators. K-d tree and lexicographical methods are used for sorting of blocks which reduced the time complexity. Hessian points are unaffected for geometrical transformations so, they are combined with LBP [17].

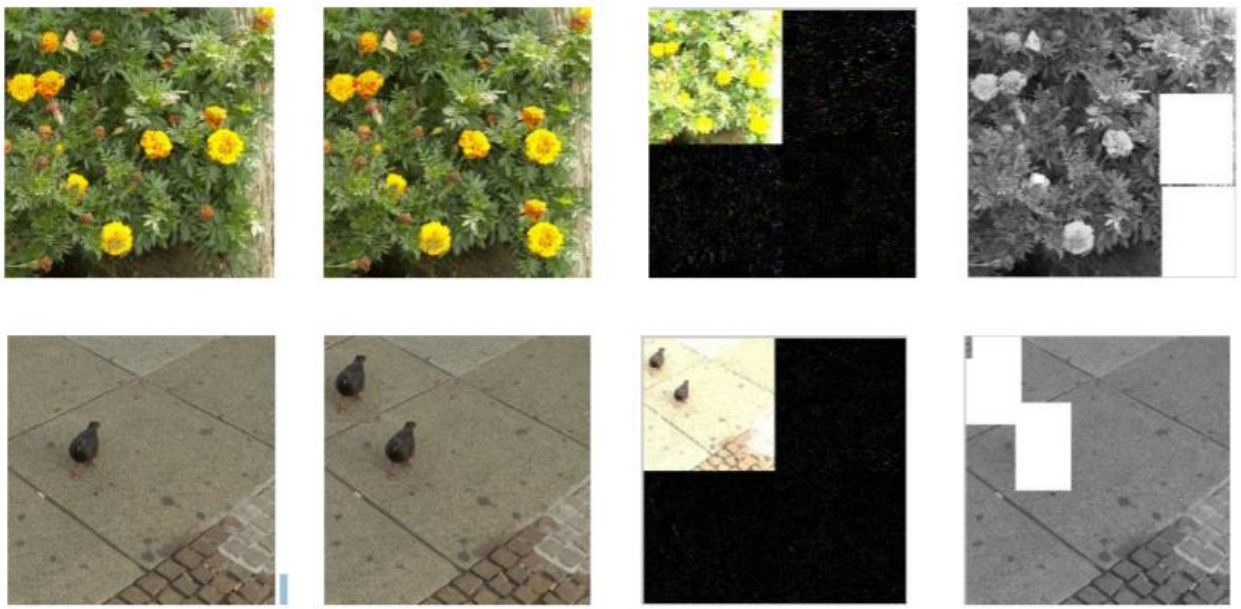


Figure 2.4: Detection of Forgery using LBP features and DWT decomposition of image [15]

Steerable Pyramid Transform (SPT) is used to extract the color content of images and then LBP can be applied to extract the features from every SPT sub-band [18]. The feature vector is given to the SVM

classifier which classifies the image as forged or unforged [19]. Some methods [20] used LBP after applying DCT on the image. LBP extract chrominance component from the image, which is later converted to frequency domain by DCT. With the help of standard deviations, the feature vector is formed from DCT coefficients and is given to SVM classifier.

There exist some texture based feature extraction techniques like in method [21], which extracts texture feature from each block that is precise and better for small forged areas. Gabor filter is also used to make a feature vector from each block using different scaling factors and rotation angles [22]. Also, a different frequency is used for filtering purpose. Method [23] used an average of gray values of images as a feature vector. Method [24] used statistical features, taken out by using HOG on every block. The Sorting of features is done and then similar blocks are searched for forgery detection. Some methods take the help of classifiers such as Extreme learning machine, which concatenate all the feature points of each block and then take it as input for classification of image as forged [25]. These methods are faster but affected by rotation, scaling and compression. Features such as moment invariant are unaffected by the geometrical transformations. Method [26] used the blur moment invariant method to create a 72-dimensional feature. Size of the feature vector is decreased and K-d tree can be used for matching of similar blocks.

Zernike moment is also invariant to geometric transformations such as rotation. The magnitude of Zernike moment is taken as a feature which represents that particular block [27]. This method used pixel level and image level analysis as its evaluation criteria. At pixel level, ability to correctly locate the tampered regions is confirmed and at image level, ability to correctly recognize whether an image has been tampered or not is confirmed. This method has better results at image level as compared to pixel level with precision ratio of 99.41%. PCT is also used as feature extraction method where similar blocks are identified by nearest neighbor search and hash functions [28]. This method detects faster than method [27]. Quaternion Exponent Moment (QEM) [29] is invariant to various transformations and scaling so, it can be used as a feature descriptor. Instead of rectangular blocks, the image is divided into circular one and this is performed on the colored image. This method applies QEM on every block and used Euclidean and hash functions for block matching. Polar Complex Exponential Transform (PCET) method is robust to rotational transformations [30]. This method suffers from time complexity. Some methods work on block matching techniques to reduce the computational complexity [31]. They performed block matching in the spatial domain. Use of GPU and integral images enhanced the CMFD techniques in a method [32]. Sometimes images are expressed in bit plane form and ASCII codes are assigned to them for forgery detection [33]. Sometimes images are segmented into patches and SIFT key points are used to describe

the image features which are matched for forgery detection [34]. For matching the Key points, Expectation and Maximization algorithm is used.

Method [8] suggested a DCT based method that converts the image into blocks mainly overlapping and convert pixel values to the frequency domain with the help of DCT. Sorting is performed on the blocks to detect manipulated regions. Use of PCA [35] on the image blocks helps in dimension reduction with the help of eigen vectors and eigen values, unique features are extracted from each block. This method performs well under the effect of noise and compression. Method [36] used DWT along with DCT technique which is also called as Quantization Coefficient Decomposition (QCD) for CMFD. In [37] used visual clues are used to find the copy-move forgery. This method used window function which distributes the image blocks into further sub-blocks and used VAM (Visual Attention Model) for feature extraction. This method searches for fixation points in image sub-blocks and makes a feature vector corresponding to each sub-block. Dimension reduction methods can be used to reduce the length of the feature vector. Method [38] used SIFT features and correlation map for the localization of copied region. Method [39] used clustering technique for localization of forged region and Random Sample Consensus (RANSAC) method for detection of any sort of geometric transformations. Zernike moments are also used as a feature in some copy-move detection techniques [39-40]. To make detection scheme rotation invariant, patch matching algorithm is used. Some techniques performed segmentation on the image for improvement in detection technique. Irregularly shaped patches are extracted from the images [41]. Hybrid features are formed from the combination of KAZE and SIFT. Filtering is also done to reduce the false detection rate [42]. This method performed well in case of rotation, scaling and noise. Method [43] used correlation coefficient of Fourier transform for matching of blocks.

Classification on the basis of pre-processing techniques is done on the color component of the given image. Gray scale converters merge the RGB components of the image into a single component. Some methods convert the RGB component to the Y, Cb, Cr color space to obtain luminance or chrominance information [44-47]. Some methods [48-51] combine the effect of the chrominance, luminance and RGB component both. Resizing of image also helps in reduction of the time required to process the image. Some other methods to reduce the image content are: low pass filtering [52-54], mean filtering, Gaussian pyramid, degraded palette, integral image [55-58].

After extracting the circular or square blocks from an image, of different sizes; small size of block results in large computational time. Large size of blocks results in false detection of forged region in the case where forgery is done in small part of image only. So, size of blocks should be chosen carefully according

to the application. Most of the methods used square blocks for division of the image whereas, some methods prefer to use the circular blocks instead of the square [50, 58]. Some approaches directly work on blocks of images. These methods suffer from computational complexity so; there is a requirement of dimensional reduction method.

Method [59] used Curvelet based Transform method for extracting features from the images. This method is used to form sub-band and from each sub-band, mean value is calculated for every level scaling. Method [60] used Fast Walsh Hadamard Transform on blocks for feature extraction. Ridgelet transform and Radon transform with predefined angles are used for feature extraction [61-62]. Dual-tree Complex Wavelet Transform applied three levels Continous Wavelet Transform (CWT) on each block and extracts energies from each sub-band. The magnitude of those energies is taken as a feature [63]. Detection of forgery on the basis of single criteria is not always sufficient so, we need a number of other criteria to reduce false matches in the cases where the images have large homogeneous area. Some examples of such methods are discussed ahead. Method [64-65] used Same Affine Transformation Selection (SATS) and method [60] used Multi-Hop Jump (MHJ) technique. Method [64] has recall ratio higher at image level than at pixel level with the recall ratio of 79.17%. Method [68] used SIFT based technique for forgery detection and transformation recovery. This method has higher value of F_1 score at image level than at pixel level with a F_1 ratio of 74.15%. Method [69] used SURF based CMFD technique and has better precision ratio of 84.50% at pixel level. Method [70-74] used Euclidean distance alone as a criterion for detecting the similar blocks, their spatial location and max norm distance should also be used. Method [71] used adaptive over segmentation and feature point matching for CMFD. In the proposed scheme, DCT and SVD are used to make a feature vector. SVM classifies images and is used to classify the images as forged or authentic. Localization of forged image is done by K-means clustering algorithm.

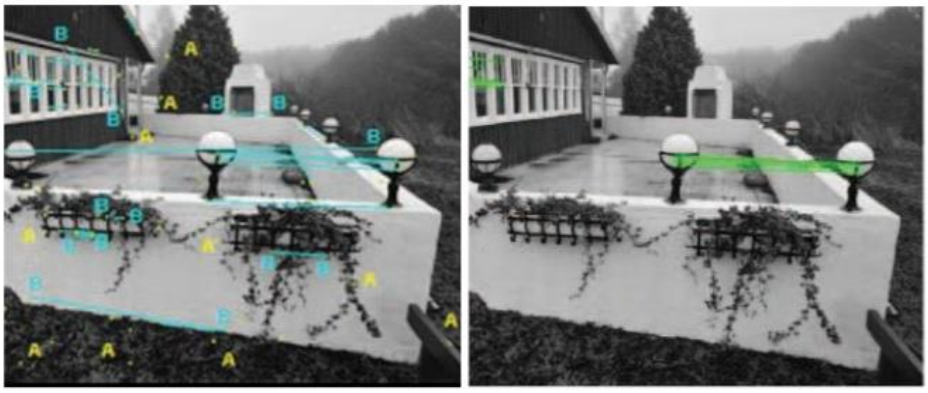


Figure 2.5: SIFT and RANSAC based Cluster matching copy-move forgery detection [39]

For CMFD, various datasets have been used by researchers. Most of the techniques used CMFD benchmark dataset [41] for evaluation of their detection scheme. This dataset contains 48 original images and their forged images. This dataset contains images with and without any post-processing. The proposed technique used the CMFD benchmark dataset [41] and Columbia image splicing dataset [66] images. Columbia image splicing dataset [66] contains fixed size images of 128×128. In the literature, methods [27], [42], [64], [68], [69], [70], [71], [73] and [74] used pixel level and image level analysis for their performance testing. Most of the techniques used common criteria for evaluation on the basis of precision, recall and F_1 parameters.

2.2 OBSERVATIONS

From the above section, few observations have been drawn that are discussed here:

- Block based CMFD techniques are most popular with high matching performance.
- Frequency domain based feature extraction methods do not perform well with post-processing operations.
- Use of PCA, DWT, and SVD in block based CMFD methods lead to reduction of feature vector dimension, thus reducing the computational time.
- Log-polar and Moments based CMFD techniques perform well under geometric transformations.
- Key point based methods suffer degradation in performance in case of uniform or smooth image area.
- Most of the existing CMFD techniques used pixel level and image level analysis for their performance testing.

2.3 GAPS AND PROBLEM FORMULATION

From the above observation, following gaps have been drawn:

- Existing Block based CMFD not perform well in presence of geometric transformations, noise and compression.
- Block based CMFD techniques suffers from the problem of high computational time.

2.4 OBJECTIVES

After considering the observations and gaps, following objectives have been drawn:

- To study and analyze the various existing copy-move forgery detection techniques.
- To propose an improved block based copy-move forgery detection technique based on DCT and SVD.
- To confirm the capability of proposed scheme against various attacks such as rotation, scaling, noise and JPEG compression.

2.5 RESEARCH METHODOLOGY

The basic idea used in this work is to make a DCT and SVD based CMFD technique that detects the copy-move forgery efficiently and provide robustness against rotation, scaling, noise and JPEG compression. DCT is used to make the features and SVD is applied to reduce the feature vector dimension. SVD provides stability and invariance from geometric transformations. SVM used to classify the image into forged or authentic. Once the images are detected as forged, K-means clustering is applied on them for the detection of the forged area. Forged areas are marked in the resulted image. All the experiments were performed with MATLAB R2016a. For evaluation of proposed detection scheme; Precision, Recall and F1 score parameters are compared with the existing CMFD techniques both at pixel, image level on the CMFD benchmark dataset [41] and Columbia image splicing dataset [66]. Performance of the proposed scheme is also analyzed under the post-processing operations such as rotation, scaling, noise addition and JPEG compression.

CHAPTER 3

PROPOSED COPY-MOVE FORGERY DETECTION SCHEME

In block based CMFD schemes, images are firstly divided into overlapping or non overlapping blocks, then feature extraction is performed on each block. In order to reduce the dimension, size reduction algorithm like PCA, SVD are used. On the basis of certain predefined thresholds, similar blocks are matched and forgery is detected. In frequency based methods, DCT is used as a feature extraction method. DCT is applied on the images and the resulted coefficients are taken as a feature which represents that image or that particular block. Whole detection process of the proposed scheme is discussed here. After dividing the gray scaled version of image into overlapping blocks of fixed size of 8×8, DCT is performed on each block. DCT is a method of conversion of an image pixel to the frequency domain. DCT has an ability to contain most of the energy of the pixels, which is very useful in presence of noise. Two-dimensional DCT of $P \times Q$ block matrix can be calculated as in following [11]:

$$D_{\mathrm{rs}} = \alpha_{\mathrm{r}}\alpha_{\mathrm{s}} \sum_{p=0}^{P-1} \sum_{q=0}^{Q-1} B_{pq} \cos\frac{\pi(2p+1)r}{2P} \cos\frac{\pi(2q+1)s}{2Q} \tag{3.1}$$

where, $0 \le r \le P-1$ and $0 \le s \le Q-1$

$$\alpha_r = \begin{cases} \frac{1}{\sqrt{P}}, & r = 0 \\ \frac{\sqrt{2}}{P}, & 1 \le r \le P-1 \end{cases} \qquad \alpha_s = \begin{cases} \frac{1}{\sqrt{Q}}, & s = 0 \\ \frac{\sqrt{2}}{Q}, & 1 \le s \le Q-1 \end{cases} \tag{3.2}$$

After applying DCT on blocks, quantization is performed on each coefficient by dividing every coefficient with quantization factor and then round off operation is done as follows [11]:

$$D = \text{round}\left(\frac{D_{mn}}{Q_{mn}}\right) \qquad \text{where, } m, n \in \{0,1,2 \ldots,7\} \tag{3.3}$$

After the DCT operation, images undergo SVD dimension reduction method. SVD is unique value for a matrix and used to reduce the dimension of a feature vector [5]. Given image matrix A can be decomposed using SVD by the following method [11]:

$$A = U \sum V^T \tag{3.4}$$

For A to be a $P \times Q$ matrix, U is a $P \times P$ matrix containing eigenvectors of AA^T as its column values and V is $Q \times Q$ matrix of eigenvectors of A^TA. Matrix V is a rotation of the corresponding column of a matrix

to Cartesian coordinates. Then $\sum$ does scaling on them and contains singular values which are the square root of eigen values of AA^T as shown in Equation (3.5) [11].

$$\Sigma = \begin{bmatrix} \sum k & 0 \\ 0 & 0 \end{bmatrix} \tag{3.5}$$

$\sum k$ is a square diagonal matrix with positive values. U transforms the scaled vectors to the original coordinates. SVD represents a linear transformation. It shows that every linear transformation contains a series of operations which are rotation, scaling and again rotation. Geometrical representation of SVD decomposition of a matrix is shown in Figure 3.1 as follows:

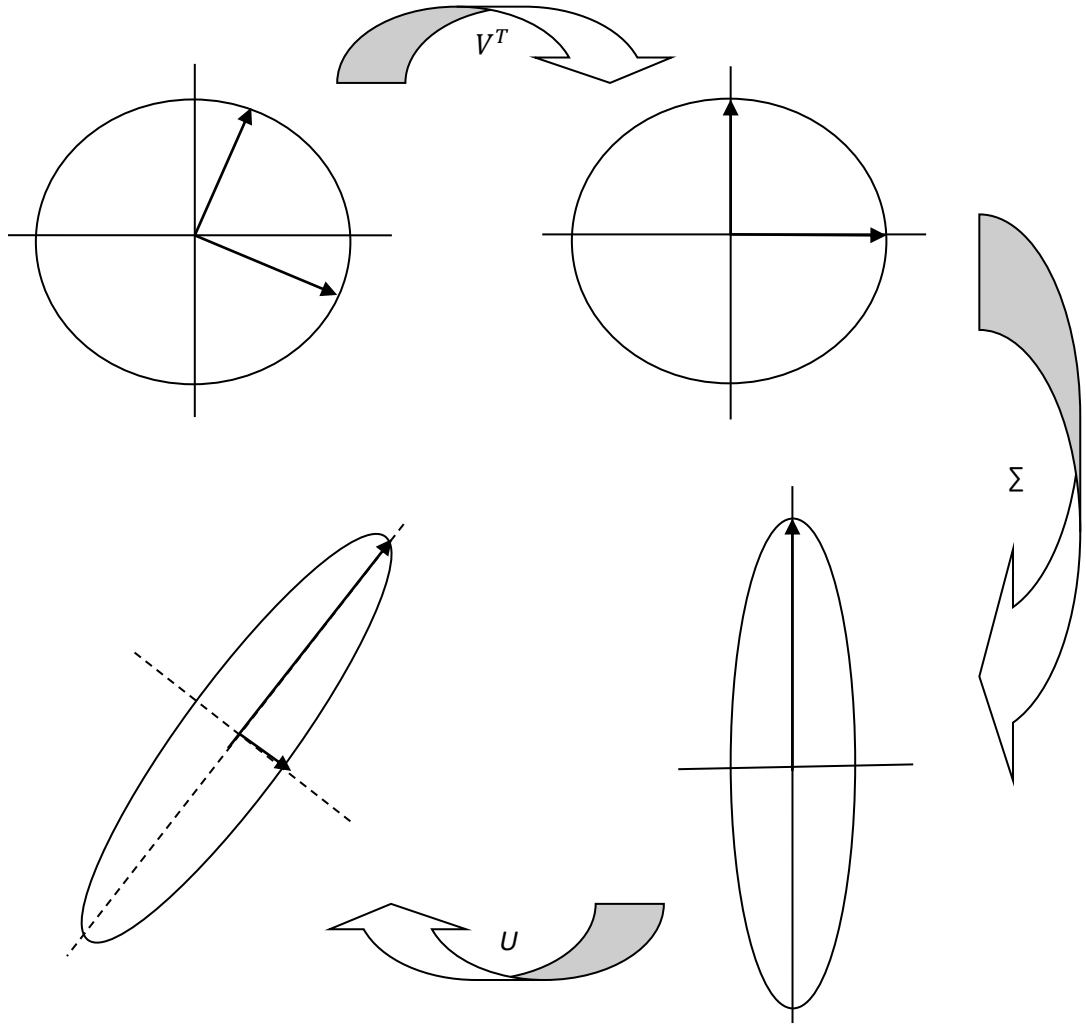


Figure 3.1: Transformation of matrix A by the composition of rotation, scaling and rotation.

In our proposed scheme we apply SVD on each block by creating sub-blocks of size 2×2 and then select the max SVD value corresponding to each sub-block. Thus as a total, we get 16 SVD values from one block of 8×8. A feature vector of dimension 1×64 is reduced to 1×16 by the use of SVD [11]. From the total no. of blocks of images, we construct one feature vector of dimension 1×16 for a single image as shown in Figure 3.2.

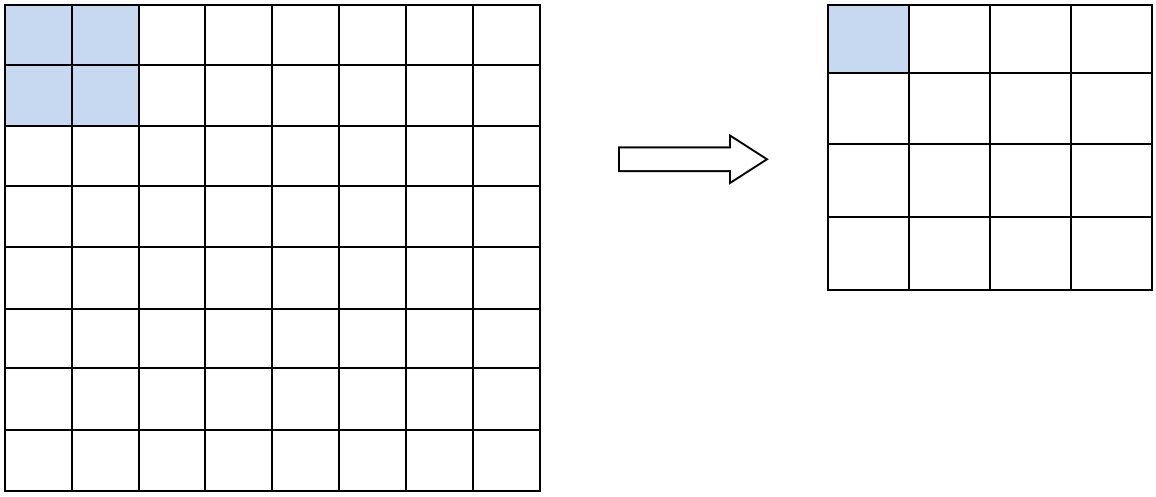


Figure 3.2: SVD decomposition of a matrix [11]

We construct a feature vector matrix for n no. of images of training data, where every row represents an image and every column represents a feature vector corresponding to that particular image. This features matrix is given as an input to SVM classifier for training on given set of images. Make another set of completely different images for testing purpose and extract the features matrix from them. SVM classifier separates the two-class datasets by making a decision boundary that separates them as far as possible. SVM classifier will predict that whether the image is forged or authentic with higher accuracy. If classifier classifies the input forged image as forged then it is called as True Negative (TN) event and if it classifies the input original image as original then it is called as True Positive (TP). If classifier classifies the input forged image as original then it is called as False Positive (FP) event and if it classifies the input original image as forged then it is called as False Negative (FN) event [1]. We can draw confusion matrix from these values and then ROC curve can be plotted. Other parameters that can be obtained from these values are sensitivity, recall, hit rate and precision. Figure 3.3 shows the workflow of proposed CMFD technique.

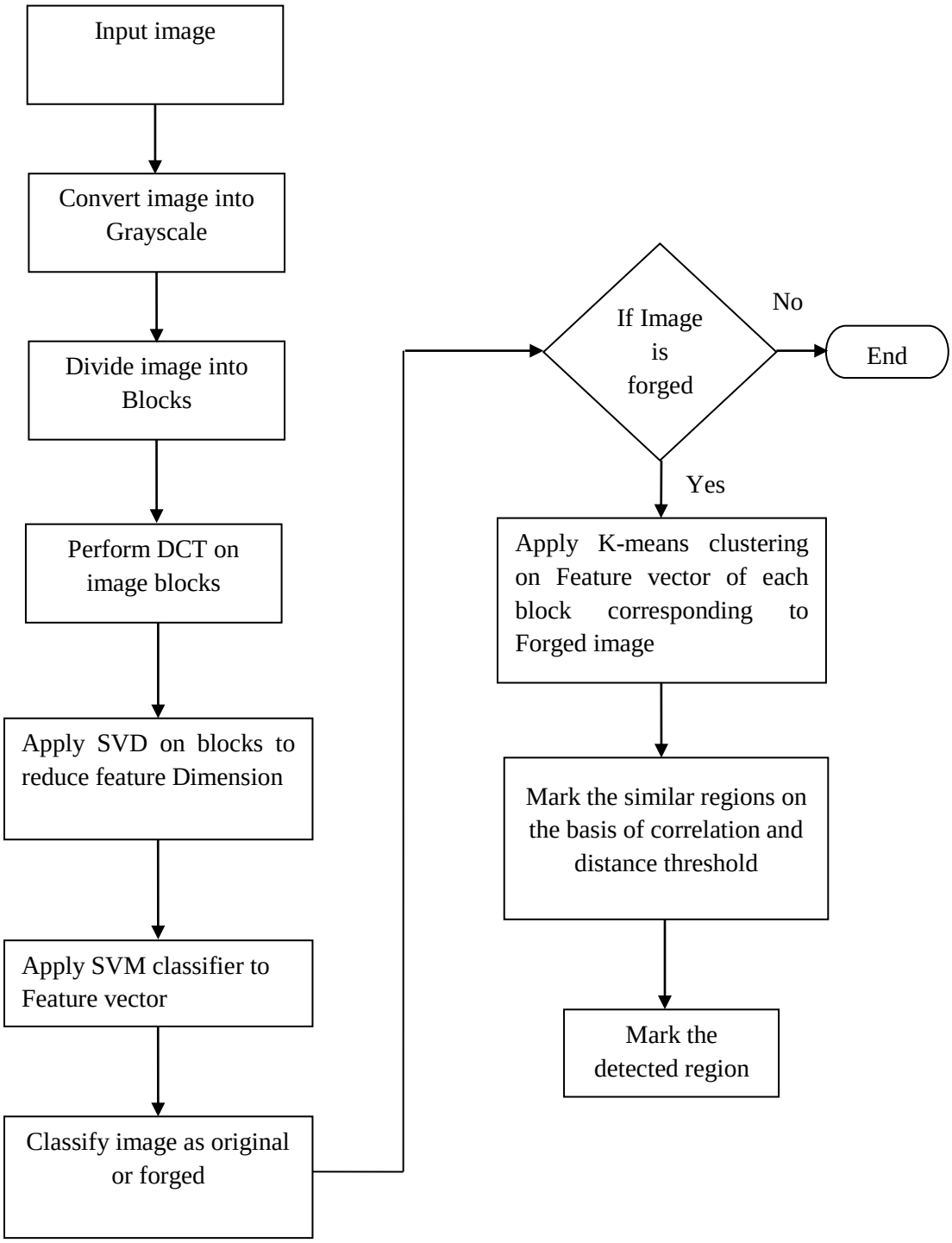


Figure 3.3: Workflow of the proposed detection scheme

3.1 PSEUDO CODE OF THE PROPOSED SCHEME

Input: Forged image; **I**

Output: Detected forged images, Forged region

```
for v= 1 to number of images
    [r c n] = size (I)
        if   n >1
            im = RGB to gray conversion of I
        else
            im = I
        end
    Block size b×b
    Block = Block divisions (im, Block size)          % total no. of blocks are (P-b+1)×(Q-b+1)
        for a= 1 to number of blocks
            block' = DCT (blocks)
            block matrix = Quantization (blocks')
            for sz=1 to size of block matrix
                s = SVD (block matrix)                %SVD is used to dimension reduction
            end
        end
    features set = s of image
end
SVM classifier = feature of training images           % input the feature set of training images to SVM
Detection result = Prediction of SVM on test images   % Detection result contains no. of images detected
                                                        as forged as well as original.
```

%% For Localization of Forged region in an image

```
for n= 1 to number of forged images
    Apply K-means clustering on Feature matrix corresponding to each block of image.
    Calculate the Correlation and Distance Threshold.
        if threshold criteria is satisfied
            Mark the detected regions.
        end
end
Resulted image = copied and pasted parts are marked in forged image.
```

For detection of forged region in an image, detection result of SVM classifier is taken under consideration. SVM classifier separates two-class datasets by making a decision boundary that separates them as far as possible as shown in Figure 3.4. SVM classifier will predict that whether the image is forged or authentic with higher accuracy.

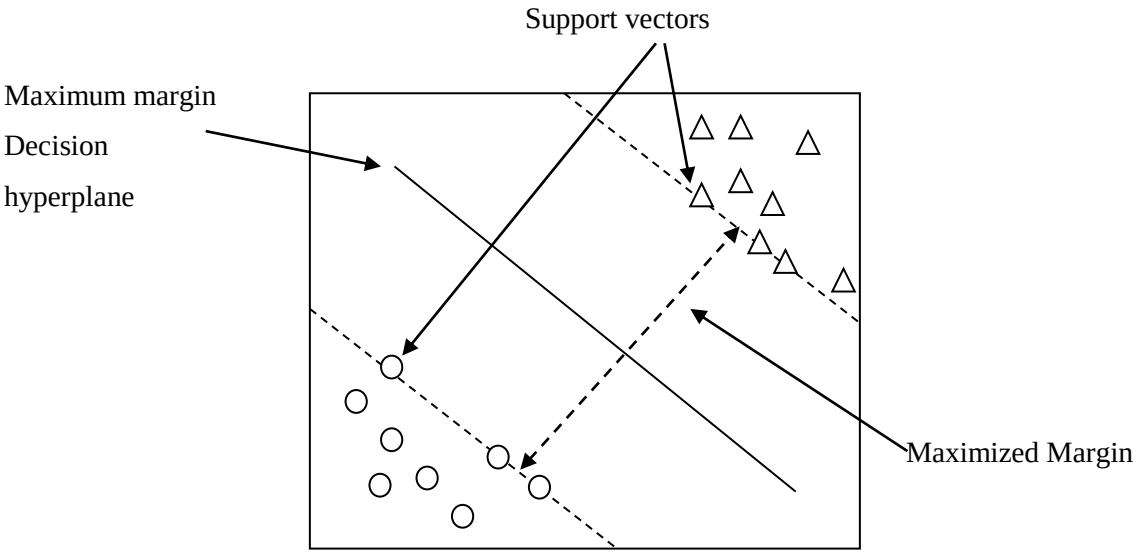


Figure 3.4: Classification of data-points using SVM

3.2 K-MEANS CLUSTERING ALGORITHM

Once the images are declared as forged by the classifier, K-means clustering is performed on the feature set related to every block in an image. K-means clustering is an unsupervised learning method which classifies a given data into K number of clusters. First of all centers for each cluster is find out as far as possible from each other. Data is assigned to each cluster on the basis of minimum distance between the centroid and data point. On the consideration of assigned points to a cluster, its center is recalculated and data is reassigned to each cluster. This loop ends where center points and data points assigned becomes almost invariant. K-means clustering method tries to minimize the squared error function as [23]:

$$J(y) = \sum_{i=1}^{K} \sum_{j=1}^{K_i} \left\| X_i - Y_j \right\|^2 \tag{3.6}$$

$\left\| X_i - Y_j \right\|$ is Euclidean distance between X_i and Y_j, K is number of clusters, K_i is number of data points in K_{th} cluster.

On the basis of feature set blocks are clustered into K number of classes where K is no. of clusters, which are predefined by the user. Grouping is done on the basis of minimum sum of square distances between the cluster centroid and data point. Figure 3.5 shows the flowchart of K-means clustering algorithm.

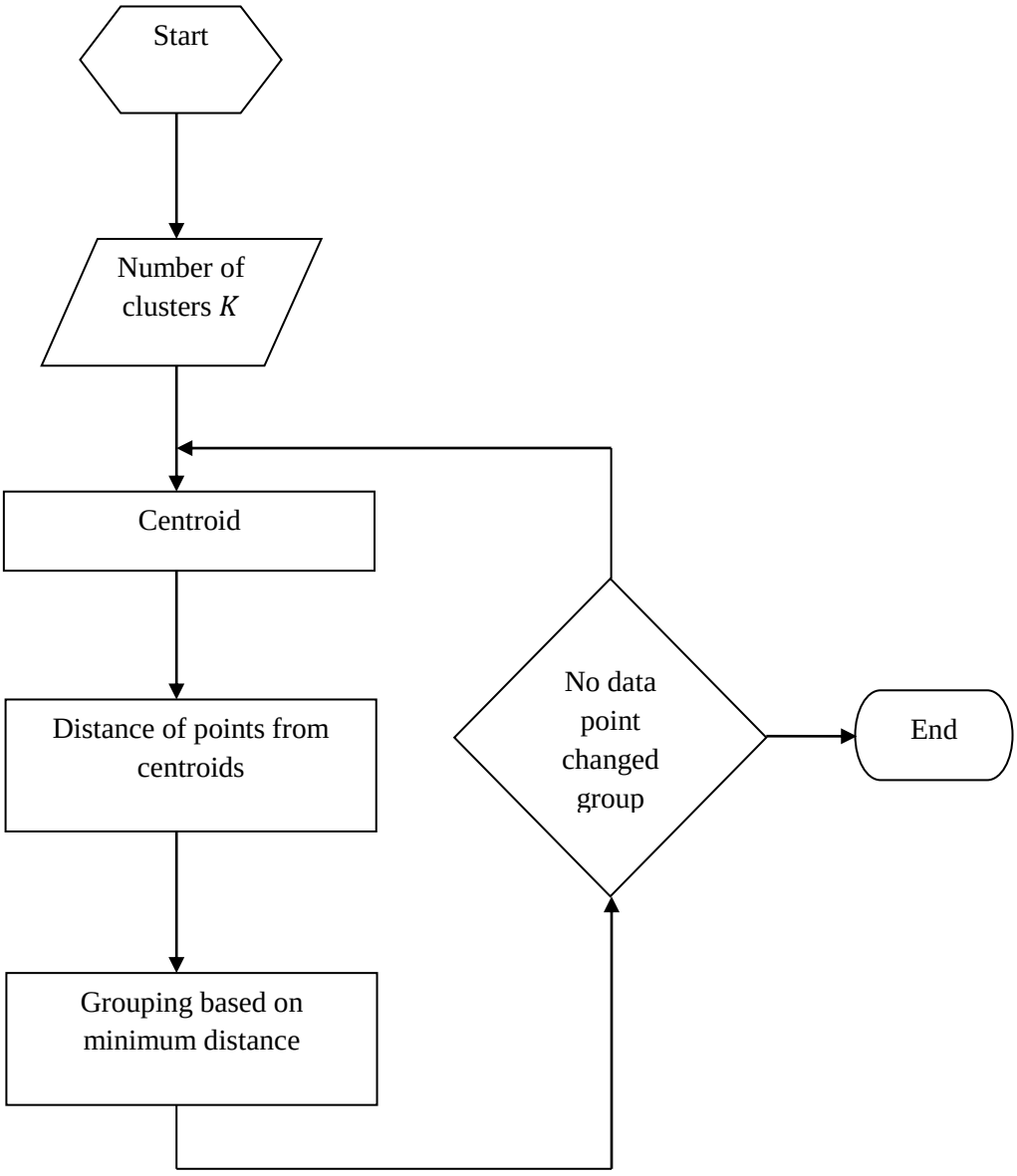


Figure 3.5: Flowchart of K-means clustering [23]

Explanation of flowchart of K-means clustering is as follows:

1. Select the centers C of the clusters randomly.

2. Calculate the distance between the cluster center and each point from the dataset.
3. Assign the points to the centers of clusters accordingly to minimum distance criteria.
4. Calculate the new centers of the clusters from the assigned points using the Equation (3.7) as [23]:

$$C_i = \left(\frac{1}{K_i}\right)\sum_{y=1}^{K_i} X_i \tag{3.7}$$

where, K_i is number of points in K_{th} cluster.

5. Recalculate the distance between new cluster's centers and data points.
6. If no new point is assigned then stop, else go to step 3.

For each class, Block features are radix sorted. Let us assume that sorted matrix is stored in Gs, then each row of Gs_i is compared with Gs_{i+1}. After the blocks are sorted, Correlation can be calculated by the Equation (3.8) as [23]:

$$corr = \frac{\sum_{i=1}^{n}(q_x - \overline{q_x}).(q_y - \overline{q_y})}{\sqrt{\sum_{i=1}^{n}(q_x - \overline{q_x})^2 \cdot \sum_{i=1}^{n}\left(q_y - \overline{q_y}\right)^2}} \tag{3.8}$$

where, q_x, q_y are the block DCT coefficients of i and $(i+1)^{th}$ blocks respectively, $\overline{q_x}$, $\overline{q_y}$ are their mean values, n is the total number of coefficients of a block. If the value of $corr$ found from above Equation (3.8) is greater than our decided threshold Th, then the two blocks are declared to be similar. To reduce false matches, distance threshold is also calculated for similar bocks by the Equation (3.9) as [23]:

$$D = \sqrt{((a_1 - a_2)^2 + (b_1 - b_2)^2)} \tag{3.9}$$

where, (a_1, b_1), (a_2, b_2) are coordinates of similar blocks. If D is greater than the predefined threshold then corresponding blocks in the image are marked as forged.

3.3 SUMMARY

Proposed scheme is a DCT and SVD based CMFD scheme, that detects the copy-move forgery in the images efficiently and provide robustness against transformations. DCT is used to drive the features and then SVD is applied to reduce the feature vector dimension. SVM classifies the image as forged or original. Once the images are detected as forged, K-means clustering is applied on them for the detection of forged area. K-means clustering is an unsupervised learning method which classifies a given data into K number of clusters.

CHAPTER 4
RESULTS AND DISCUSSIONS

For evaluation of the proposed technique, we have performed the experiments on CMFD benchmark dataset [41] and Columbia image splicing dataset [66] at both image level and pixel level. CMFD benchmark dataset has been used in most of the copy-move forgery discovering techniques. This dataset contains 48 original images and their forged versions. Ground truth images are also provided in this dataset, which shows copied and pasted regions in the forged image. Columbia image splicing dataset contains 1845 images of fixed size of 128×128 and is easily available. For experimentation, images are divided into block size of 8×8. Number of clusters used are 20, correlation threshold is chosen in between 0 to 1 and distance threshold is taken near about 16. All the experimentation was done on a desktop computer with processor of Intel (R) Core™ i5-4210U CPU, 4 GB RAM, Windows 8.1 and 64-bit OS with MATLAB R2016a.

4.1 DATASET AND SETTINGS

For image forgery detection, various datasets have been used by researchers. Some of them are CMFD benchmark dataset [41], Columbia image splicing dataset [66], CASIA dataset [75], MICC F220 and MICC F2000 [76], GRIP dataset [77], Coverage dataset [78] and COMOFOD dataset [79]. Out of these datasets, CMFD benchmark dataset [41] is used by most of the copy-move forgery detection techniques. Columbia image splicing datasets [66] contains fixed size images and is easily available. So, for the evaluation of proposed technique, we have used two datasets: CMFD benchmark dataset [41] and Columbia image splicing dataset [66]. Both datasets contain original as well as forged images. CMFD benchmark dataset [41] contains 48 high resolution PNG images with their ground truth images as well as forged images. These images are uncompressed with an average size of 1500×1500. This dataset also contain the images with post-processing operations applied on them. Images of this dataset are categorized on the basis of objects they contained and on the basis of their texture. Nature, man-made and living things are the different categories on the object basis. On the basis of texture, images can be classified as rough, smooth and structured images. We have presented the results on the complete dataset in this work. Columbia image splicing dataset [66] contains a total of 1845 images with same number of original and tampered images. Size of images is 128×128 and all are Bitmap images. For the experiments, we have created two sets from each dataset. One set contains images for training purpose, whereas other one is for testing. Both the sets should not contain the similar images because training and testing should be done on completely different images. Two sets are created from the dataset, which contains both original as well as forged images. Feature matrix is extracted from the both testing and training images.

All the images are classified into two types of labels. If they are forged, their label value will be one and if they are original their label value will be zero. Classifier learns about the forged and original images from this given training set. Similarly feature matrix is extracted from the testing images and is given to the SVM classifier for prediction of their labels. Evaluation of the proposed CMFD is done for copy-move tampering in the absence of any post-processing attacks such as geometric transformations. After that detection results of the proposed scheme with the varying number of clusters is shown. Comparative analysis of proposed scheme is discussed at pixel level and image level, which is followed by detection results of proposed scheme and various existing CMFD techniques under different attacks. Figure 4.1 shows the flowchart of organization of work discussed in this chapter.

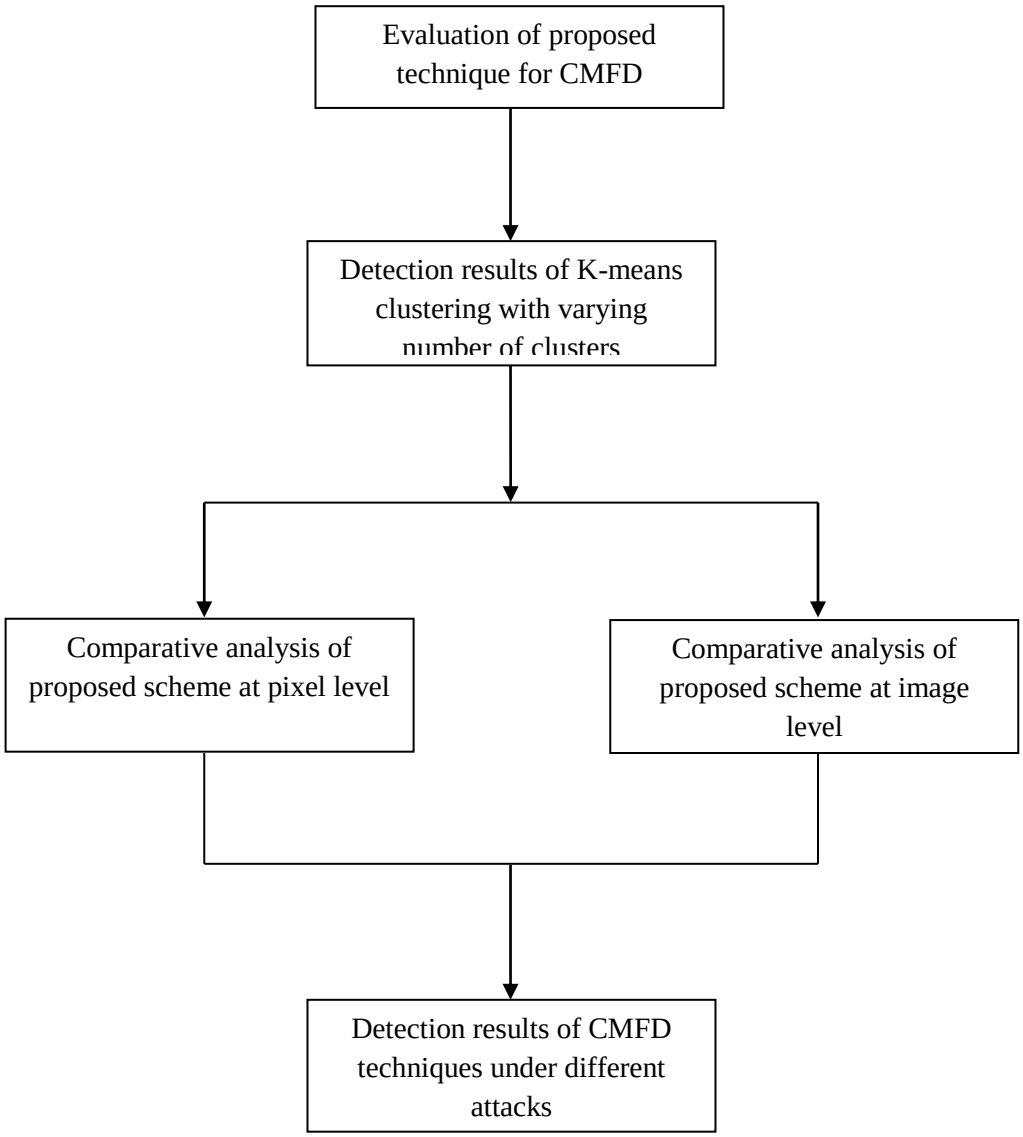
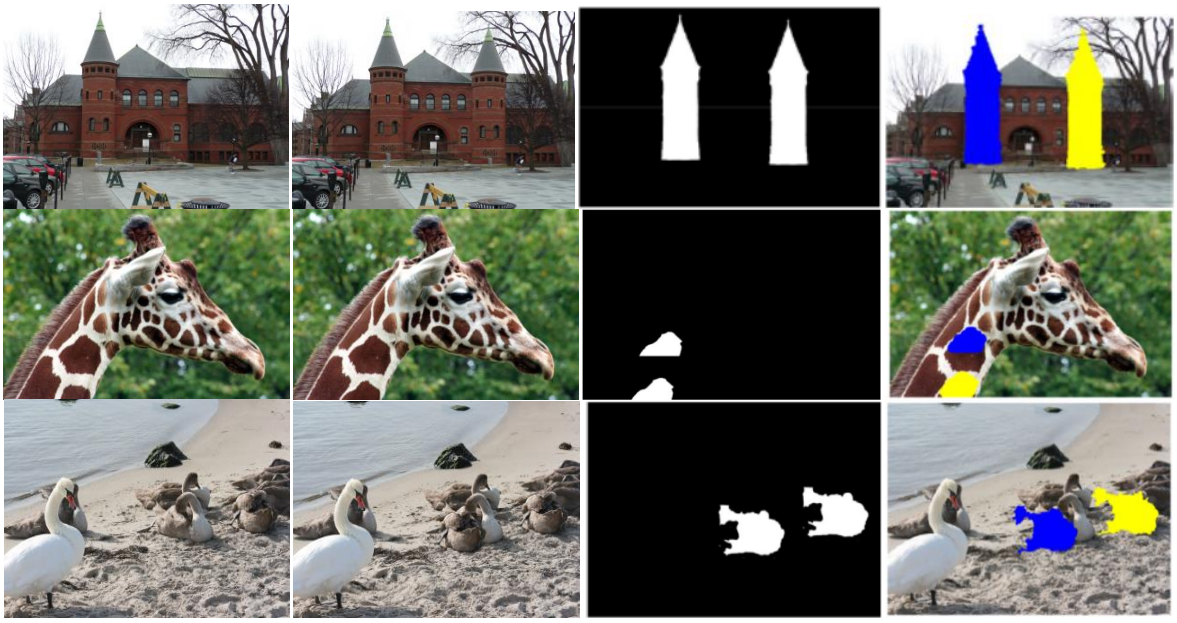


Figure 4.1: Organization of work

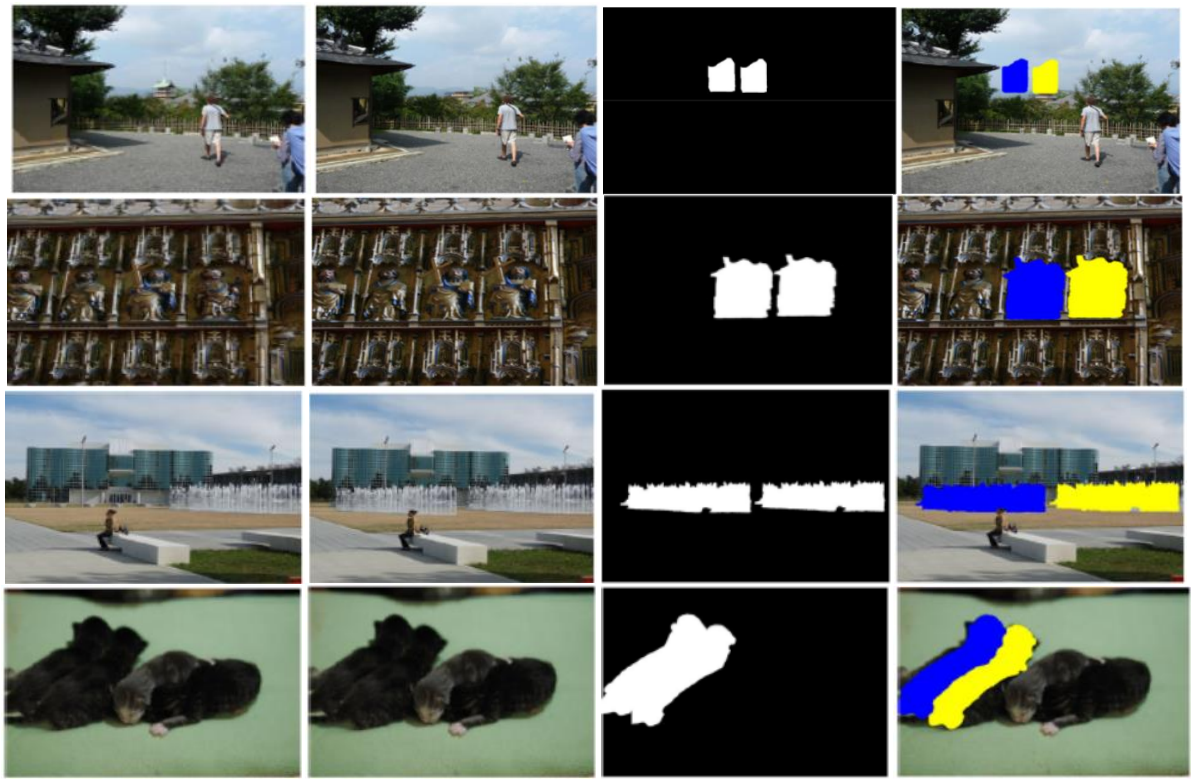
4.2 EVALUATION OF THE PROPOSED TECHNIQUE FOR CMFD

In this section, we are considering the copy-move forgery which does not suffers from post-processing attacks. For forgery creation, an arbitrary portion of any image is cropped and is pasted over the same image in an unpredictable manner. In our proposed method first of all, SVM classifier detects that whether an image is forged or not, if it is then K-means clustering method detects the forged region and marked it. Figure 4.2 contains the detection results of our proposed scheme on CMFD benchmark dataset [41]. CMFD benchmark dataset [41] contains 48 high resolution PNG images with their ground truth images as well as forged images. These images are uncompressed with an average size of 1500×1500. First column of Figure 4.2 is of original images; second column is of forged images and the third column is ground truth of forgery region in images, which tells about the exact forged region, Fourth column is of detection result of our proposed scheme, which shows the original and detected regions marked with yellow and blue color. Figure 4.3 contains the detection results of our proposed scheme on Columbia image splicing dataset images. First column of Figure 4.3 is of original images, second column is of forged images, third column is of detected forged images, in which copied parts and pasted parts are detected and marked with blue and yellow colors respectively.



(i) (ii) (iii) (iv)

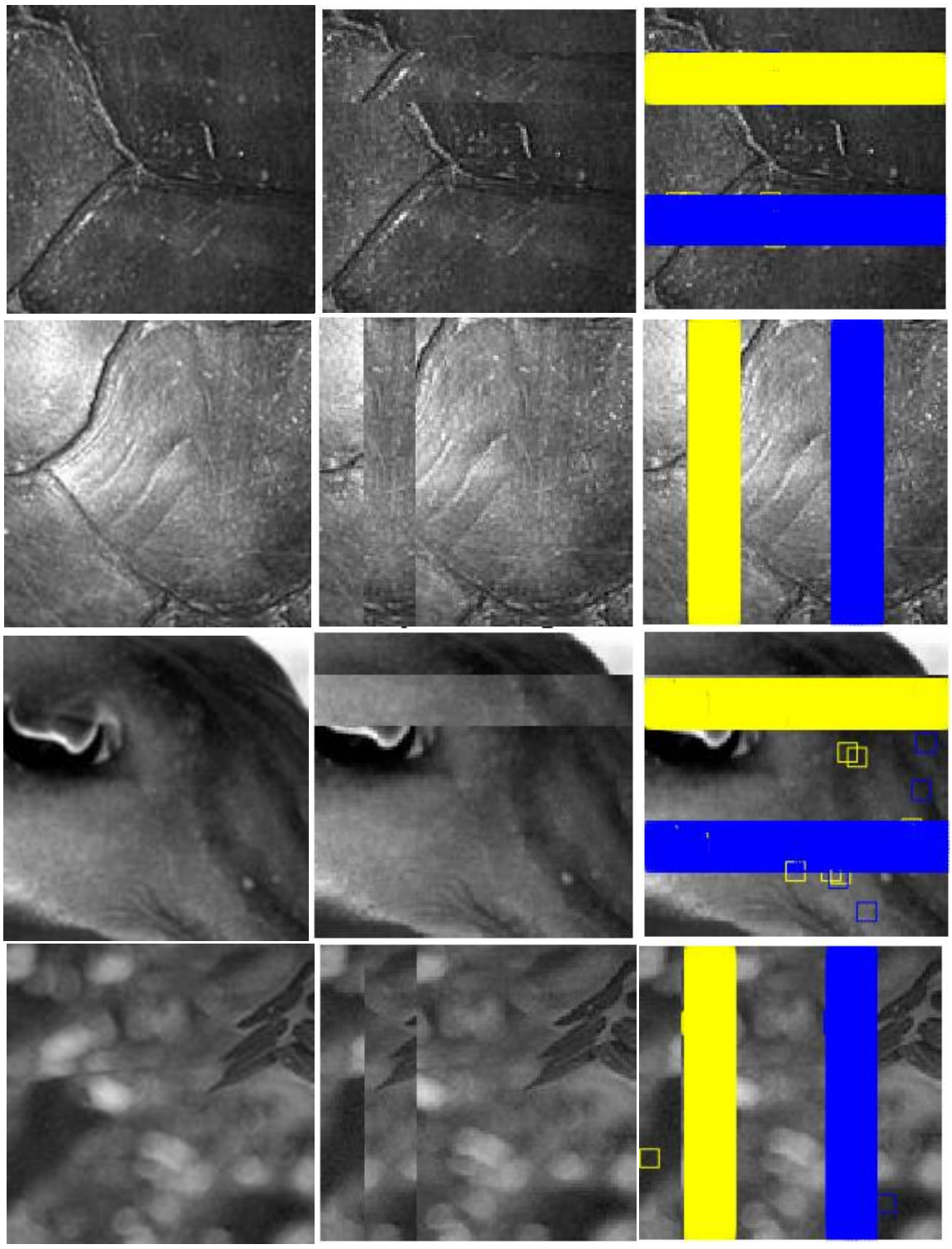
Figure 4.2: Forgery detection results on CMFD benchmark dataset: (i) original images (ii) forged images (iii) ground truth of forgery images (iv) detected copied and pasted regions (contd.)



(i) (ii) (iii) (iv)

Figure 4.2: Forgery detection results on CMFD benchmark dataset: (i) original images (ii) forged images (iii) ground truth of forgery images (iv) detected copied and pasted regions

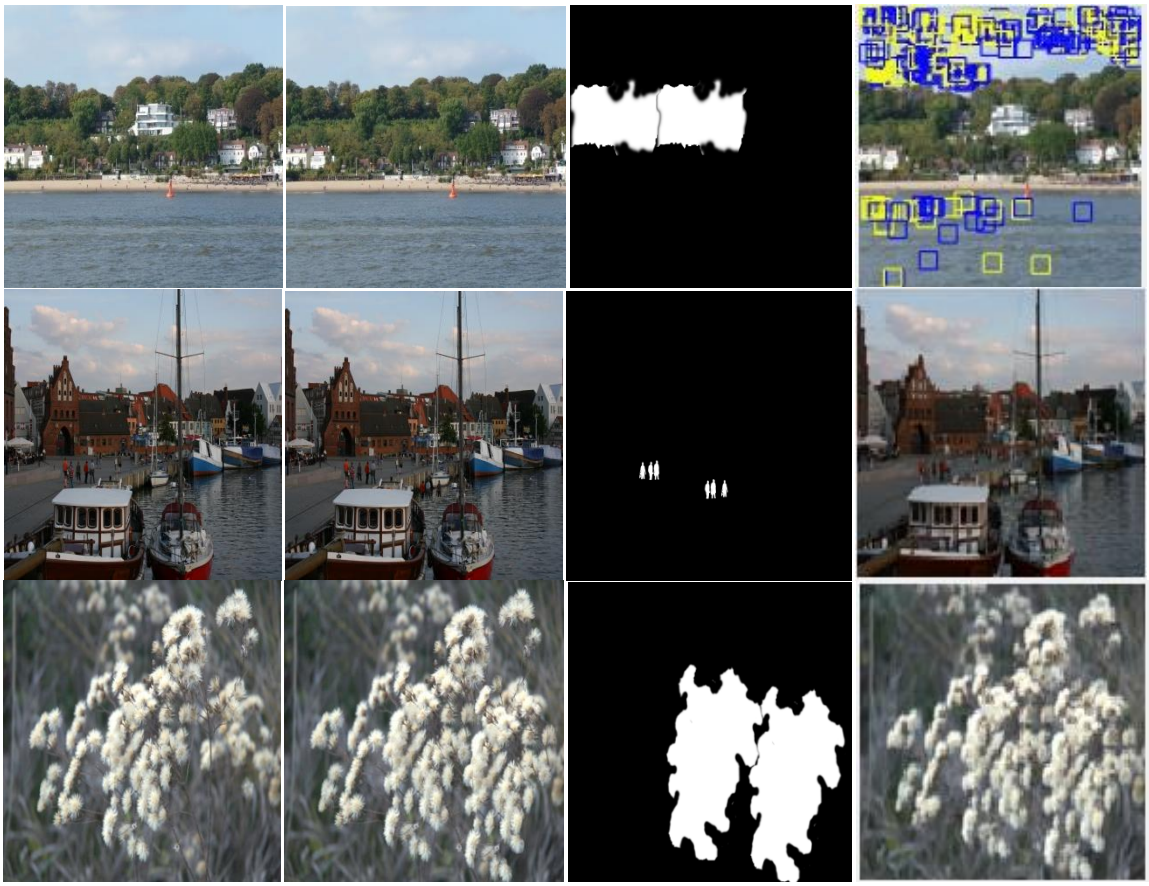
Figure 4.3 contains detection results of our proposed copy-move forgery detection scheme on Columbia image splicing dataset [66]. Columbia image splicing dataset contains a total of 1845 images with same number of original and tampered images. Size of images is 128×128 and all are Bitmap images. First column of Figure 4.3 contains original images; our second column is of forged images; and our third column of the figure represents detected copy-move forgery in the images. Copied part is marked in the detected image with blue color and where it is pasted is marked with the yellow color. In order to create forgery, an arbitrary portion of an original image is cropped and is paste down over the same or maybe different image in an unpredictable manner. In our proposed method first of all, SVM classifier detects that whether an image is forged or not, if it is then K-means clustering method detects the forged region and marked it.



(i) (ii) (iii)

Figure 4.3: Forgery detection results on Columbia image splicing dataset: (i) original images (ii) forged images (iii) detected copied and pasted regions

Some false detections done by our classifier are also shown here in Figure 4.4 for CMFD benchmark dataset, where original image given as input is classified as forged and vice versa.



(i) (ii) (iii) (iv)

Figure 4.4: Forgery detection results on CMFD benchmark dataset: (i) original images (ii) forged images (iii) ground truth of forgery images (iv) false detection done by the classifier

For the experiments, we have created two sets from each dataset. One set contains images for training purpose, whereas other one is for testing. Both the sets should not contain the similar images because training and testing should be done on completely different images. Two sets are created from the dataset, which contains both original as well as forged images. Feature matrix is extracted from the both testing and training images. All the images are classified into two types of labels as shown in Table 4.1 for CMFD benchmark dataset [41]. If they are forged, their label value will be one and if they are original their label value will be zero. Feature matrix of the training set images along with their labels is fed to the classifier. Classifier learns about the forged and original images from this given training set. Classifier predicts label for every image that whether the given image is original or forged.

Table 4.1 Training images given to the classifier

Training Images	Dimension	Format	Size	Classification of images
Barrier	1944×1296	PNG	3.02 MB	Original
Beach wood	1632×1224	PNG	3.07 MB	Original
Berries	1520×1007	PNG	2.14 MB	Original
Bricks	1944×12961	PNG	3.87 MB	Original
Cattle	640×427	PNG	505 KB	Original
Central-park	1632×1224	PNG	3.53 MB	Original
Christmas hedge	1520×1007	PNG	2.12 MB	Original
Clean walls	512×377	PNG	339 KB	Original
Dark and bright	1944×1296	PNG	3.85 MB	Original
Disconnected	1296×1944	PNG	3.71 MB	Original
Mask	1007×1520	PNG	2.35 MB	Forged
Motorcycle	1632×1224	PNG	3.75 MB	Forged
Mykene	1936×1296	PNG	4.47 MB	Forged
No beach	1944×1296	PNG	3.61 MB	Forged
Noise pattern	1944×1296	PNG	3.93 MB	Forged
Port	1632×1224	PNG	3.83 MB	Forged
Red tower	1632×1224	PNG	3.25 MB	Forged
Sailing	1306×1950	PNG	4.29 MB	Forged
Sails	1944×1296	PNG	4.02 MB	Forged
Scotland	1152×1536	PNG	2.08 MB	Forged
Giraffe	400×266	PNG	215 KB	Original
Hedge	1632×1224	PNG	2.29 MB	Original
Horses	1520×1007	PNG	2.43 MB	Original
Japan tower	1632×1224	PNG	3.67 MB	Original
Jellyfish chaos	1944×1296	PNG	2.70 MB	Original
Stone ghost	400×266	PNG	214 KB	Original
Supermarket	1632×1224	PNG	3.52 MB	Original
Swan	1944×1296	PNG	4.10 MB	Original
Sweets	1520×1007	PNG	3.10 MB	Original
Tree	512×512	PNG	282 KB	Original
Egyptian	1520×1007	PNG	2.61 MB	Forged
Extension	1944×1296	PNG	4.33 MB	Forged
Fisherman	1632×1224	PNG	3.85 MB	Forged
Fountain	1632×1224	PNG	2.99 MB	Forged
Four babies	1504×1000	PNG	2.16 MB	Forged
Knight moves	720×720	PNG	850 KB	Forged
Kore	1936×1296	PNG	4.01 MB	Forged
Lone cat	1520×1007	PNG	2.89 MB	Forged
Malawi	1536×1152	PNG	3.04 MB	Forged
Statue	1330×1052	PNG	2.11 MB	Forged
Ship number	1632×1224	PNG	3.36 MB	Original
Tapestry	1632×1224	PNG	3.09 MB	Original
Three hundred	1181×790	PNG	1.86 MB	Original
Wading	1944×1296	PNG	3.41 MB	Original
White	1520×1007	PNG	2.74 MB	Original
Window	1520×1007	PNG	2.58 MB	Original
Wood carvings	1296×1944	PNG	4.60 MB	Original

Table 4.2 Detection Result of classifier on testing images

Test Images	Dimension	Format	Size	Classification	Detection results
Egyptian	1520×1007	PNG	2.61 MB	Original	Original
Extension	1944×1296	PNG	4.33 MB	Original	Original
Fisherman	1632×1224	PNG	3.85 MB	Original	Original
Fountain	1632×1224	PNG	2.99 MB	Original	Original
Four babies	1504×1000	PNG	2.16 MB	Original	Original
Giraffe	400×266	PNG	215 KB	Forged	Forged
Hedge	1632×1224	PNG	2.29 MB	Forged	Forged
Horses	1520×1007	PNG	2.43 MB	Forged	Forged
Japan tower	1632×1224	PNG	3.67 MB	Forged	Original
Jellyfish chaos	1944×1296	PNG	2.70 MB	Forged	Forged
Knight moves	720×720	PNG	850 KB	Original	Original
Kore	1936×1296	PNG	4.01 MB	Original	Original
Lone cat	1520×1007	PNG	2.89 MB	Original	Original
Malawi	1536×1152	PNG	3.04 MB	Original	Original
Statue	1330×1052	PNG	2.11 MB	Original	Forged
Stone ghost	400×266	PNG	214 KB	Forged	Forged
Supermarket	1632×1224	PNG	3.52 MB	Forged	Forged
Swan	1944×1296	PNG	4.10 MB	Forged	Forged
Sweets	1520×1007	PNG	3.10 MB	Forged	Forged
Tree	512×512	PNG	282 KB	Forged	Forged
Mask	1007×1520	PNG	2.35 MB	Original	Original
Motorcycle	1632×1224	PNG	3.75 MB	Original	Original
Mykene	1936×1296	PNG	4.47 MB	Original	Original
No beach	1944×1296	PNG	3.61 MB	Original	Original
Noise pattern	1944×1296	PNG	3.93 MB	Original	Original
Barrier	1944×1296	PNG	3.02 MB	Forged	Original
Beach wood	1632×1224	PNG	3.07 MB	Forged	Forged
Berries	1520×1007	PNG	2.14 MB	Forged	Forged
Bricks	1944×12961	PNG	3.87 MB	Forged	Forged
Cattle	640×427	PNG	505 KB	Forged	Forged
Port	1632×1224	PNG	3.83 MB	Forged	Forged
Red tower	1632×1224	PNG	3.25 MB	Forged	Forged
Sailing	1306×1950	PNG	4.29 MB	Forged	Forged
Sails	1944×1296	PNG	4.02 MB	Forged	Forged
Scotland	1152×1536	PNG	2.08 MB	Forged	Forged
Central-park	1632×1224	PNG	3.53 MB	Forged	Forged
Christmas hedge	1520×1007	PNG	2.12 MB	Forged	Forged
Clean walls	512×377	PNG	339 KB	Forged	Forged
Dark and bright	1944×1296	PNG	3.85 MB	Forged	Forged
Disconnected	1296×1944	PNG	3.71 MB	Forged	Forged
Ship number	1632×1224	PNG	3.36 MB	Original	Original
Tapestry	1632×1224	PNG	3.09 MB	Original	Original
Three hundred	1181×790	PNG	1.86 MB	Original	Original
Wading	1944×1296	PNG	3.41 MB	Original	Original
White	1520×1007	PNG	2.74 MB	Original	Original
Window	1520×1007	PNG	2.58 MB	Forged	Forged

Table 4.2 shows the results of classifier on testing image set of CMFD benchmark dataset [41]. It shows that out of 48 images given for testing; only three images (underlined) are falsely detected. Detection result for these three images is shown in Figure 4.4, where it is shown that when original image is given as an input to the detection algorithm, it is classified as forged and vice versa. Thus the accuracy for the given sets of images is about 93%. For the calculation of parameters: Accuracy, Recall, Precision and F_1, we have given the complete dataset to the classifier by dividing half of the number of images into training set and another half into testing set. Results obtained for both the datasets are discussed here in Table 4.3 and Table 4.4. Table 4.3 shows the parameters obtained for CMFD benchmark dataset [41] and Table 4.4 shows the parameters obtained for Columbia image splicing dataset [66].

Table 4.3 Parameters obtained for CMFD benchmark dataset [41]

Parameters	**Values (%)**
AUC	93.00
Precision	91.42
Recall	83.30
F_1	86.81

Table 4.4 Parameters obtained for Columbia image splicing dataset [66]

Parameters	**Values (%)**
AUC	90.03
Precision	93.08
Recall	85.24
F_1	87.58

From the accuracy point of view, our proposed copy-move detection method has about 93% accuracy on CMFD benchmark dataset [41] and 90% Accuracy on Columbia image splicing dataset [66] as shown in Figure 4.5 and Figure 4.6 for copy-move forgery detection. Accuracy has been calculated from ROC curve on the basis of FPR and TPR values. Where TPR value is correctly detected forged images from the total input forged images to the classifier and FPR is incorrectly detected forged images when input images were the original one. AUC determines the accuracy of a classifier on a particular dataset. So, AUC should be as maximum as possible.

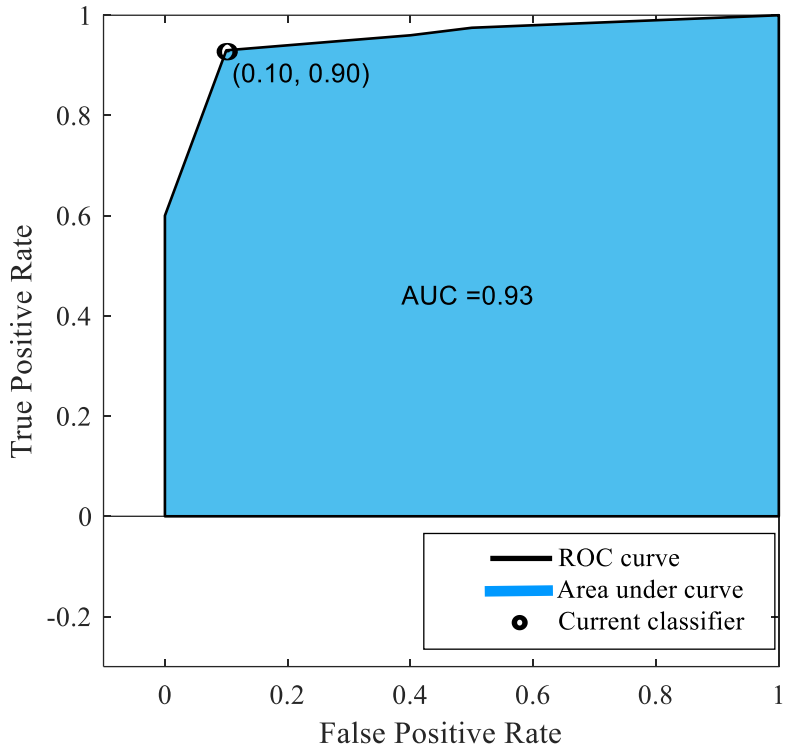


Figure 4.5: 93 % accuracy achieved with the proposed method on the CMFD benchmark Dataset [41]

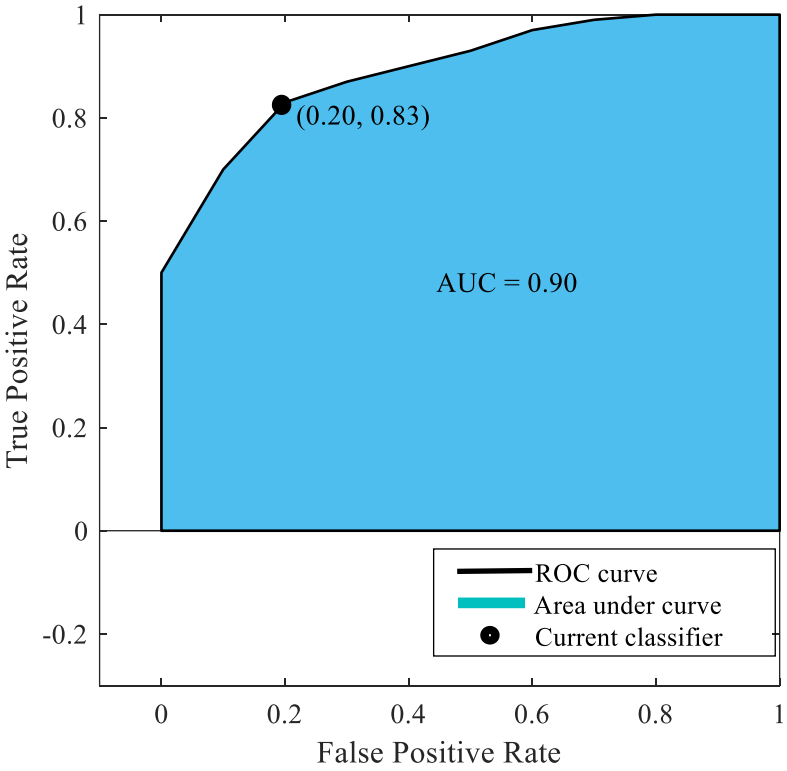
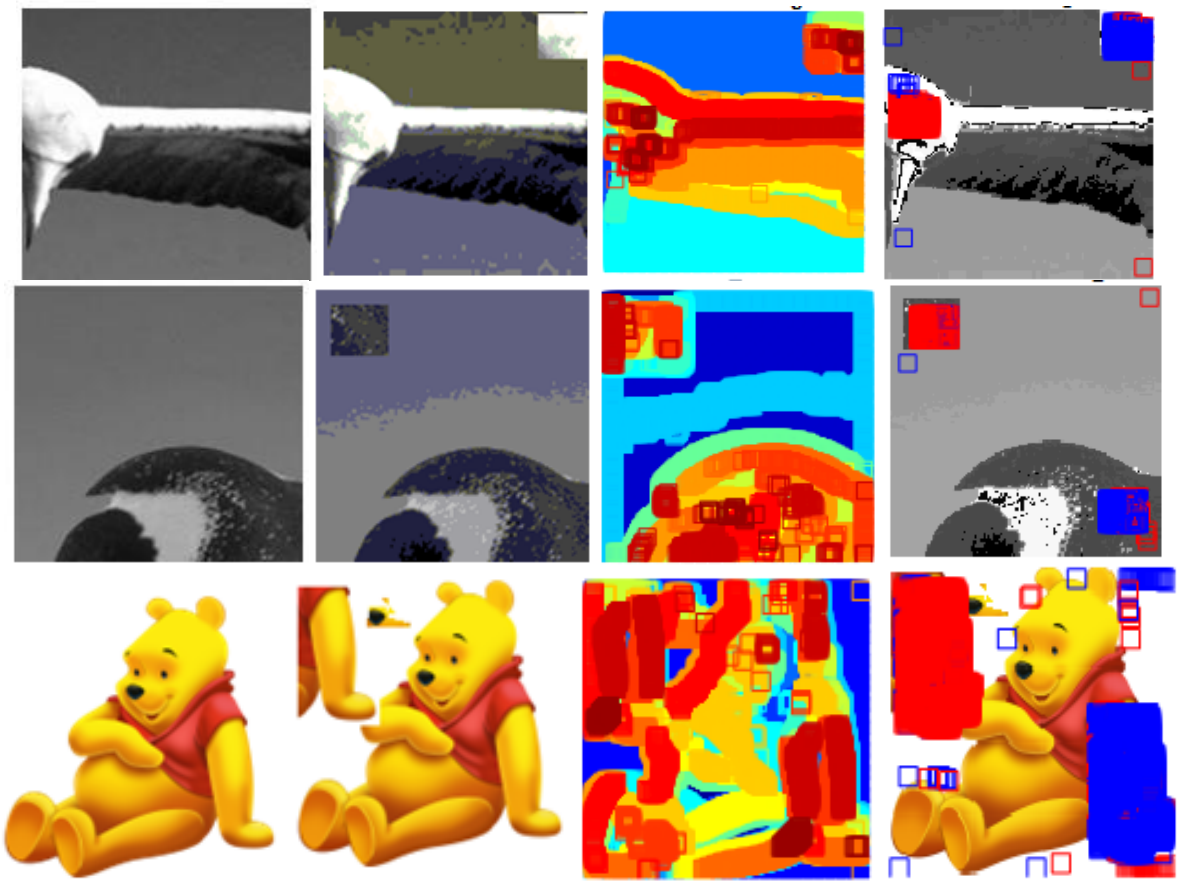


Figure 4.6: 90 % accuracy achieved by proposed method on Columbia image splicing Dataset [66]

4.3 DETECTION RESULTS OF K-MEANS CLUSTERING ALGORITHM WITH VARYING NUMBER OF CLUSTERS

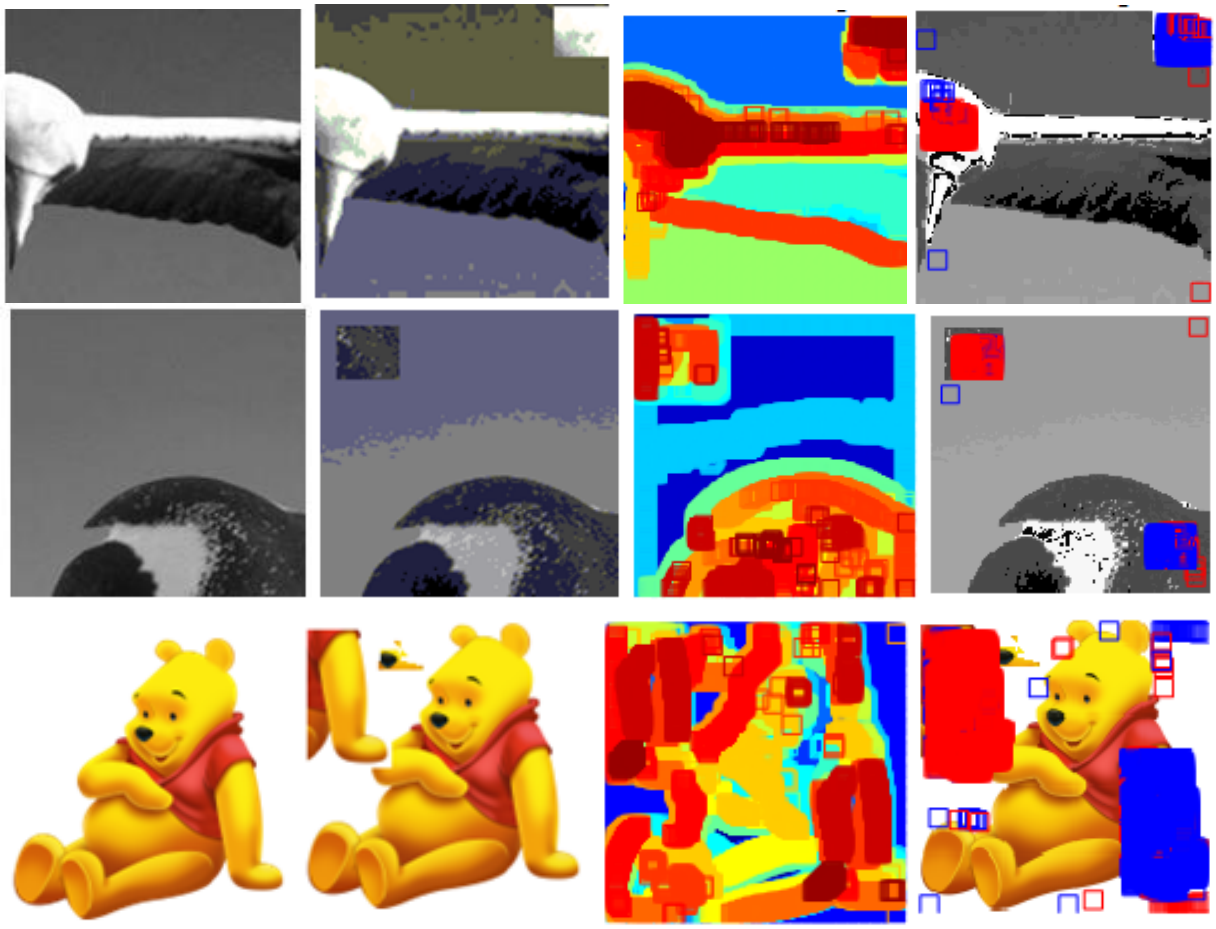
In our proposed scheme we use clustering for detection of forged region in an image. *K* is the number of clusters taken by the method. The correct choice of *K* depends upon number, shape and distribution of data points [72]. As we increase the number of clusters, false detection reduces. K-means clustering is a method of clustering the data points into given number of clusters. On the basis of distance between the cluster centroid and data points, points are assigned to each cluster. In our experiment, we have varied the number of clusters as *K*= 3, 4, 10 and 20, and their results are shown in Figure [4.7-4.10]. Figure 4.7 shows the detection results of our proposed technique for number of clusters, *K*=3. Column (i) of Figure 4.7 is of original images, column (ii) is of forged images, column (iii) is of clustered images, obtained after applying K-means clustering, column (iv) is of detected forged image.



(i) (ii) (iii) (iv)

Figure 4.7: **For *K*=3**: (i) Original images (ii) Forged images (iii) Clustered images (iv) Detected forged images

Above shown results are for K=3 which means number of clusters used are 3. From the above results, we can say that detection accuracy is very poor as there are many false matches in the resulted image and detection scheme is not able to detect the forged region properly. Figure 4.8 shows the detection results for K=4. Column (i) of Figure 4.8 is of original images, column (ii) is of forged images, column (iii) is of clustered images, obtained after applying K-means clustering, column (iv) is of detected forged image.

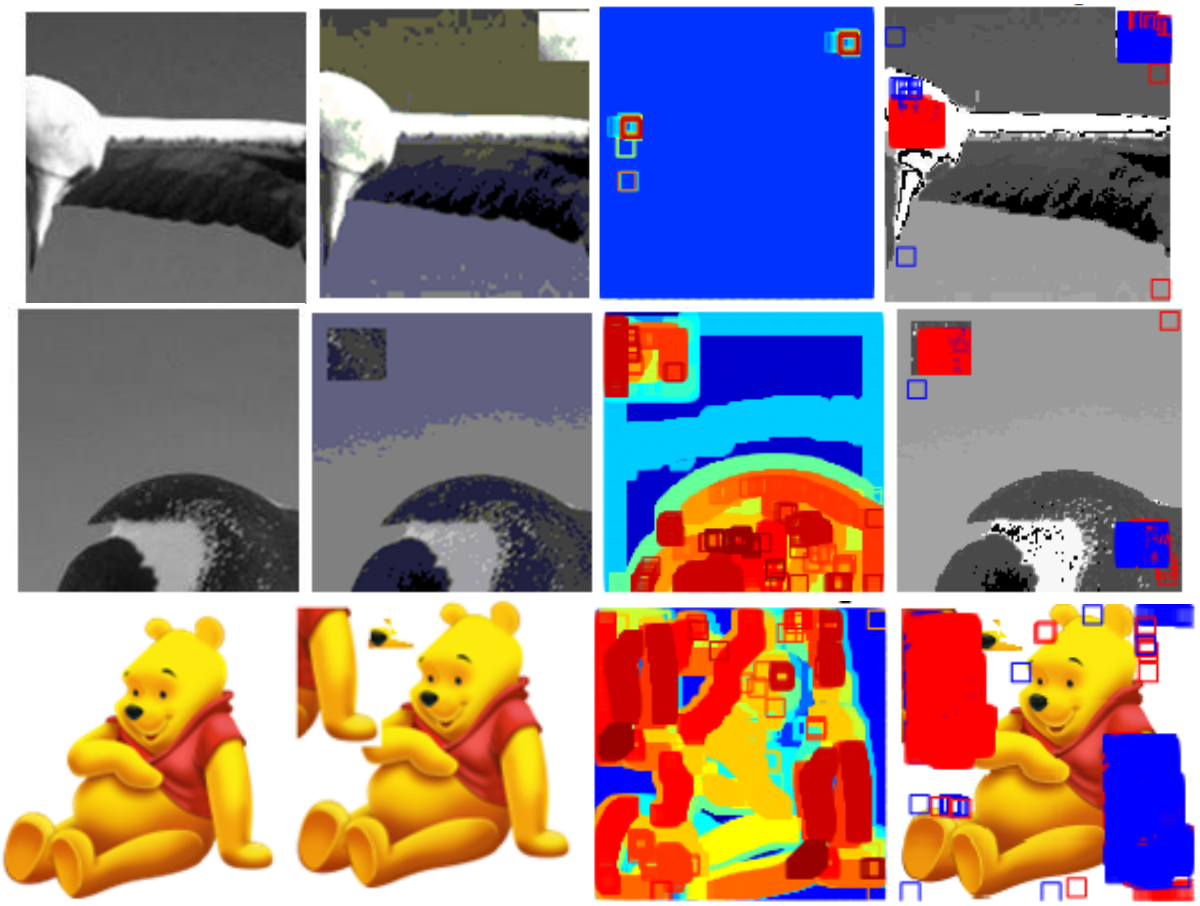


(i) (ii) (iii) (iv)

Figure 4.8: For ***K*=4**: (i) Original images (ii) Forged images (iii) Clustered images (iv) Detected forged images

Above shown results are for K=4 which means number of clusters used are 4. From the above results, we can say that detection accuracy is poor but there is some improvement than K=3, as there are less false matches in the resulted image. Figure 4.9 shows the detection results for K=10. Column (i) of Figure 4.9

is of original images, column (ii) is of forged images, column (iii) is of clustered images, column (iv) is of detected forged image.

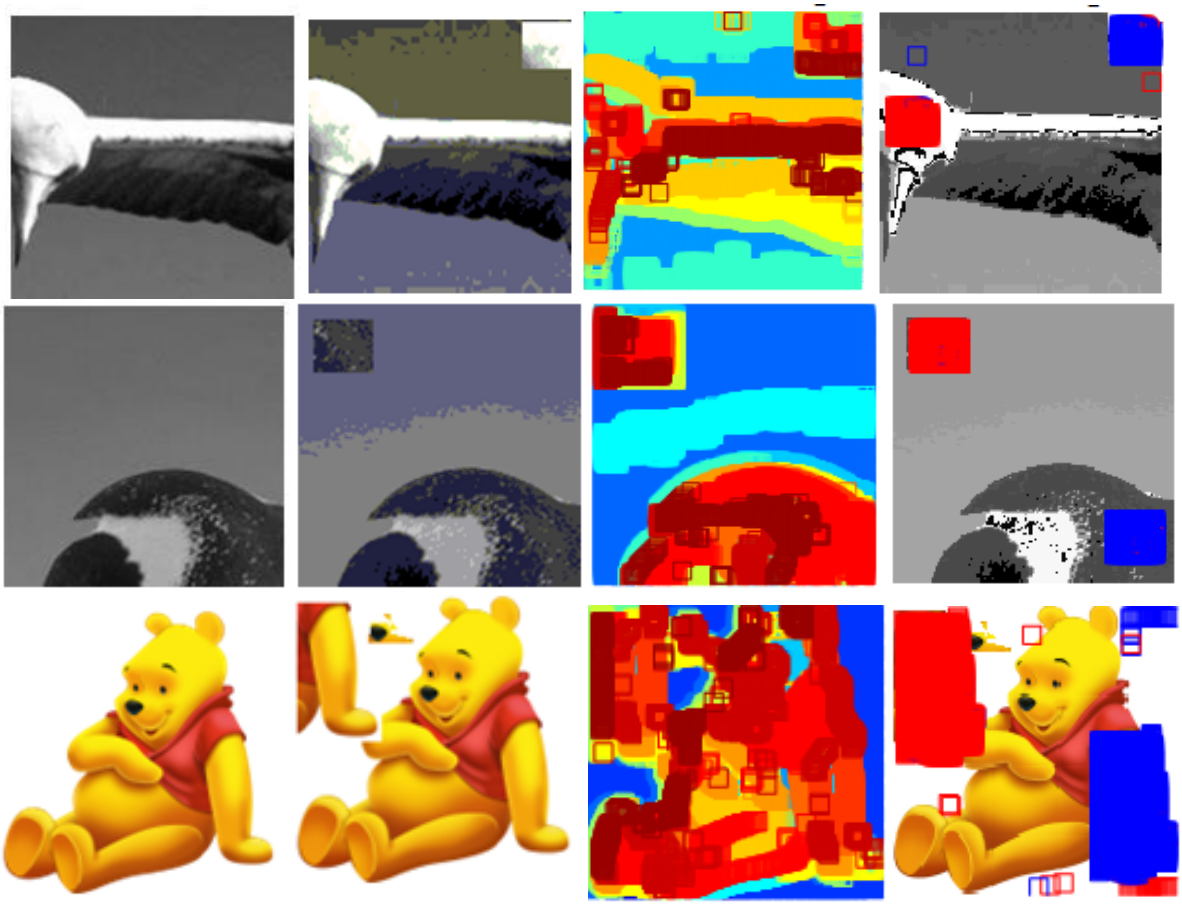


(i) (ii) (iii) (iv)

Figure 4.9: **For *K*=10**: (i) Original images (ii) Forged images (iii) Clustered images (iv) Detected forged images

Above shown results are for K=10 which means number of clusters used are 10. From the above results, we can say that detection accuracy is better than K=3 and K=4, as there are few false matches in the resulted image. Detected copied and pasted regions are shown by rectangular boxes of color blue and red. For false matches, they dislocate from their actual position and split in other parts of the figure, where forgery has not been performed. We can see that number of dislocated rectangular boxes shown in the resultant figure of K=10 is less than in the case of K=3 and K=4. Figure 4.10 shows the detection results for K=20. Column (i) of Figure 4.10 is of original images, column (ii) is of forged images, column (iii) is of clustered images, obtained after applying K-means clustering, column (iv) is of detected forged image.

From the Figure 4.10, we can say that as the number of clusters increase, number of dislocated square boxes decreases. So, accuracy of forgery detection scheme increases with the increase in value of K.



(i) (ii) (iii) (iv)

Figure 4.10: **For *K*=20**: (i) Original images (ii) Forged images (iii) Clustered images (iv) Detected forged images

From the experiments conducted for various values of number of clusters in clustering algorithm, we concluded that for K= 20, we got better results and minimum error in forged area detection as shown above. Numbers of false matches are less than K=3, 4 and 10. Also at K= 20, the proposed technique is able to detect and localize the forged region properly in the detected image. Thus, with the increase in number of clusters used, detection accuracy improves.

4.4 COMPARISON OF PROPOSED DETECTION TECHNIQUE WITH THE EXISTING TECHNIQUES

For testing of any forgery detection technique two levels are considered: Pixel level and Image level. In case of image level, we check whether the method is able to detect the image as forged or tampered correctly or not. In case of pixel level, we check whether a method can correctly detect the forged region inside a forged image or not. Parameters like precision, recall and F_1 score can be calculated on the basis of both pixel level and image level. In this work, we have calculated the above mentioned parameters on both pixel level and image level and the results of proposed technique are compared with the various other existing techniques. Precision determines truly forged images ratio to the sum of detected forged images. Recall is ratio of truly forged images to the total number of images. Recall is also called as measure of sensitivity. F_1 score is defined as average with specified weights of both precision ratio and recall ratio [1]. Detection results of various techniques for copy-move detection on CMFD benchmark dataset are compared at pixel level and are shown below in Table 4.5, which shows that proposed scheme has better Precision ratio than [68], [69], [70], [71] and [42]. Recall ratio and F_1 score ratio of proposed scheme is better than all other existing CMFD schemes [64], [68], [69], [27], [70], [71], [42].

Table 4.5 Comparison of copy-move forgery detection techniques at pixel level on CMFD benchmark dataset [41]

Method	**Precision (%)**	**Recall (%)**	**F_1 score (%)**
Bravo et al. (2011) [64]	**99.82**	32.58	49.12
(SIFT) Amerin et al. (2011) [68]	82.17	65.66	72.99
(SURF) Baboo et al. (2011) [69]	84.50	44.83	58.58
(Zernike) Ryu et al. (2013) [27]	99.41	68.76	81.29
Cozzolino et al. (2014) [70]	82.17	56.78	85.40
Pun et al. (2015) [71]	84.08	74.83	79.73
Yang et al. (2017) [42]	90.27	78.61	84.04
Proposed	91.42	**83.30**	**86.81**

For this case, considered copy-move forgery detection techniques are free from any kind of post-processing attacks, which include rotation, scaling, noise addition and JPEG compression. Table 4.5 shows that method proposed by Bravo [64] has highest precision ratio of 99.82%, after that precison ratio

of Zernike based method of Ryu [27] is highest with ratio 99.41%. Recall ratio of SURF based method [69] is lowest. F_1 score of method proposed by Bravo is 49.12% which is lowest. F_1 score of methods [42] and [70] is better after our proposed scheme. Table 4.6 shows the comparison of proposed detection technique with the other existing CMFD techniques at image level on CMFD benchmark dataset.

Table 4.6 Comparison of copy-move forgery detection techniques at image level on CMFD benchmark dataset [41]

Method	**Precision (%)**	**Recall (%)**	**F_1 score (%)**
Bravo et al. (2011) [64]	**92.68**	79.17	85.39
(SIFT) Amerin et al. (2011) [68]	80.49	68.75	74.15
(SURF) Baboo et al. (2011) [69]	77.27	70.83	73.90
(Zernike) Ryu et al. (2013) [27]	88.46	95.83	91.99
Cozzolino et al. (2014) [70]	81.32	71.64	83.21
Pun et al. (2015) [71]	89.28	79.61	89.73
Yang et al. (2017) [42]	78.33	96.92	87.04
Proposed	90.03	**97.12**	**93.00**

Table 4.6 shows the comparison of CMFD techniques at image level without any post-processing. At image level, techniques are tested on the basis of their ability of detecting a forged image. At image level, whether the detection scheme is able to detect the forgery or not is checked. Since, our detection scheme is able to classify the images with a high accuracy, so our parameters: Precision, Recall and F_1 are having high value for image level than pixel level. So, from Table 4.6 we can infer that precision ratio of techniques [64] is 92.68% which is better than our proposed technique whereas technique proposed by method [69] has the lowest precision ratio with ratio 77.27%. Recall ratio and F_1 score values of our proposed detection method are 97.12% and 93% respectively, which are better than all existing techniques. Method [68] has lowest recall ratio which is 68.75% and method [69] has a lowest F_1 score with ratio 73.90%. Thus, from Table 4.5 and 4.6, we can conclude that both at image level and pixel level, our detection technique has highest recall and precision ratio than the techniques [64], [68], [27], [70], [71] and [42]. Precision ratio of our detection scheme is also higher than the most of existing techniques but is lower than a few of them. Table 4.7 shows the comparison of proposed technique with the existing CMFD techniques at pixel level without any post-processing on Columbia image splicing dataset [66]. Precision ratio of our proposed schme at pixel level for second dataset is lower than the method [74] but

higher than the method [73]. Recall ratio of our method is high. F_1 score ratio of proposed scheme is higher than method [73] but lower than method [74].

Table 4.7 Comparison of copy-move detection techniques at pixel level on Columbia image splicing dataset [66]

Method	**Precision (%)**	**Recall (%)**	**F_1 score (%)**
R.C. (2011) [73]	80	75	73
Lynch G. (2013) [74]	**97**	95	**94**
Proposed	90.70	**95**	92

Table 4.8 shows the comparison of CMFD techniques at image level without any post-processing. At image level techniques are tested on the basis of their detection accuracy of detecting an image as a forged or original.

Table 4.8 Comparison of copy-move detection techniques at image level on Columbia image splicing dataset [66]

Method	**Precision (%)**	**Recall (%)**	**F_1 score (%)**
R.C. (2011) [73]	79	72	74
Lynch G. (2013) [74]	**94**	92	93
Proposed	91	**95**	**94**

From the above tables it is clear that precision ratio of our proposed scheme is not highest among the existing CMFD techniques but the other two parameters are having better values.

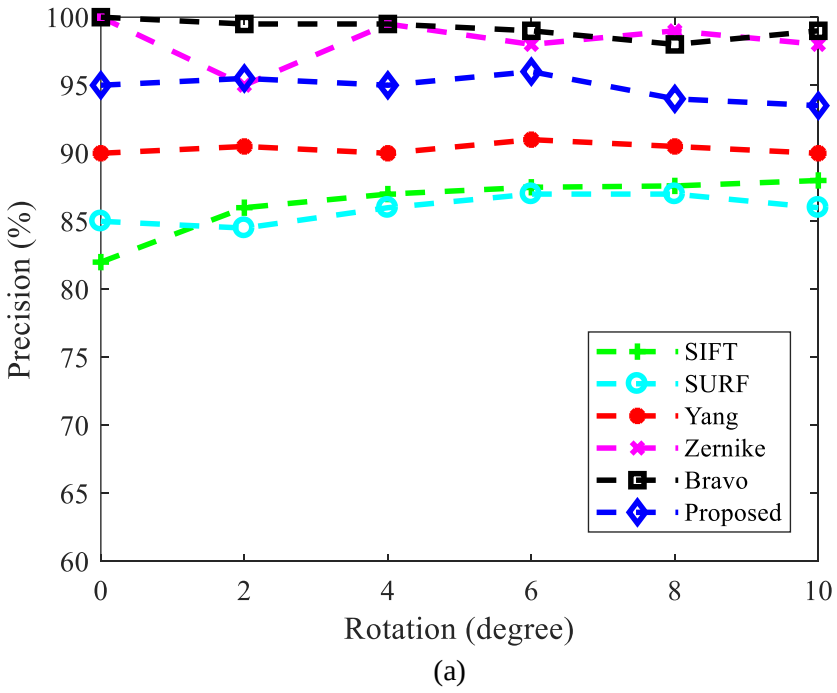
4.5 DETECTION RESULTS OF CMFD TECHNIQUES WITH VARIOUS TRANSFORMATIONS AT PIXEL LEVEL

The proposed scheme is also tested under the effect of various geometric transformations and compression etc. Comparison of the proposed technique is also done with the existing forgery detection techniques, under the effect of geometric transformations at pixel level. For this purpose, forged images undergoes scaling, rotation, compression and noise addition. To check the performance of detection techniques in the presence of all these parameters, we have taken the images from the CMFD benchmark dataset [41] that have undergone post-processing opertions. For evaluation of proposed scheme under

these attacks, values of Precision, Recall and F_1 score are analysed. Figure 4.11 to 4.14 are the results of detection schemes with rotation, scaling, noise addition and JPEG compression.

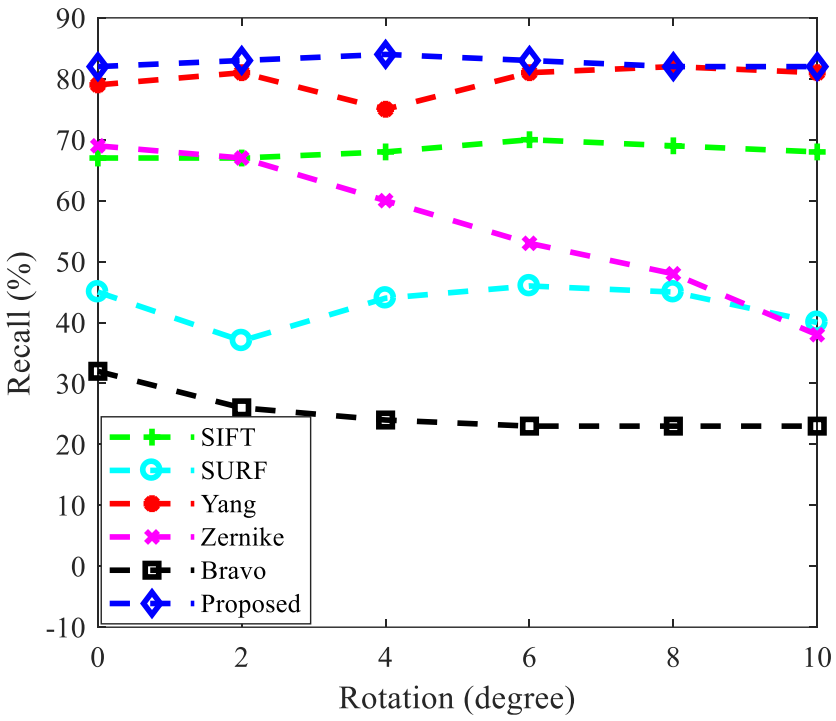
4.5.1 Performance of CMFD techniques under the effect of Rotation

In case of rotation, forged part in the image is rotated at some angle in order to make the forgery unpredictable. Rotation affects the forgery detection scheme performance. Features selected form the images for forgery detection should be less affected by rotation. In our proposed technique, we have taken DCT and SVD based features. SVD is rotation invariant. So, our detection scheme is less affected by rotation and perform well under the effect of rotation. For testing of the performance of techniques under rotation effect, we have rotated the forged region with number of rotation angles. Figure 4.11 shows the effect of change of rotation angles on the precision ratio, recall ratio and F_1 score ratio with the change in rotation angles from 0 degree to 10 degrees at the step length of 2 degrees.

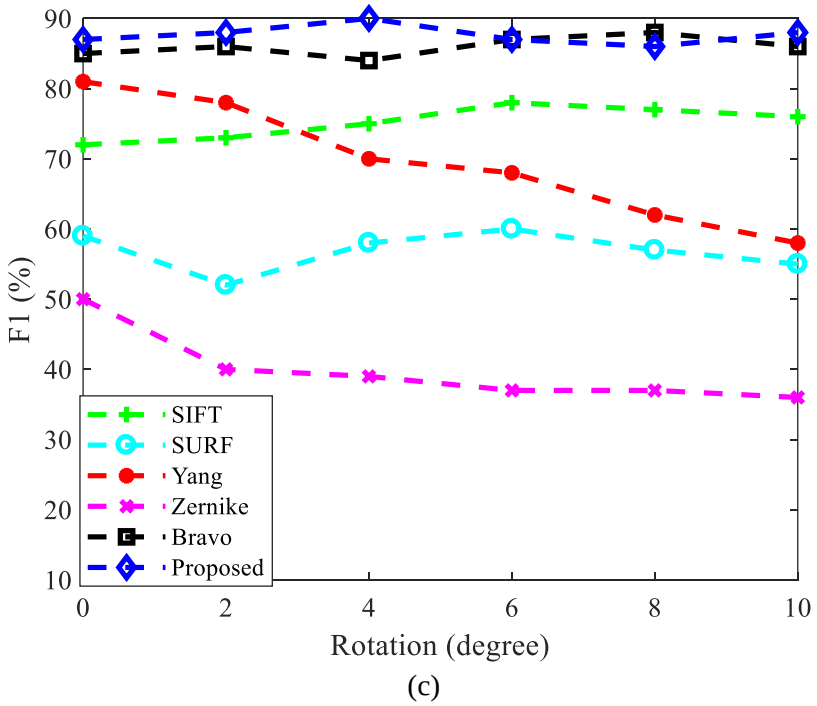


(a)

Figure 4.11: (a) Precision ratio with change in rotation angels (b) Recall ratio with change in rotation angles (c) F_1 score with change in rotation angles (contd.)



(b)



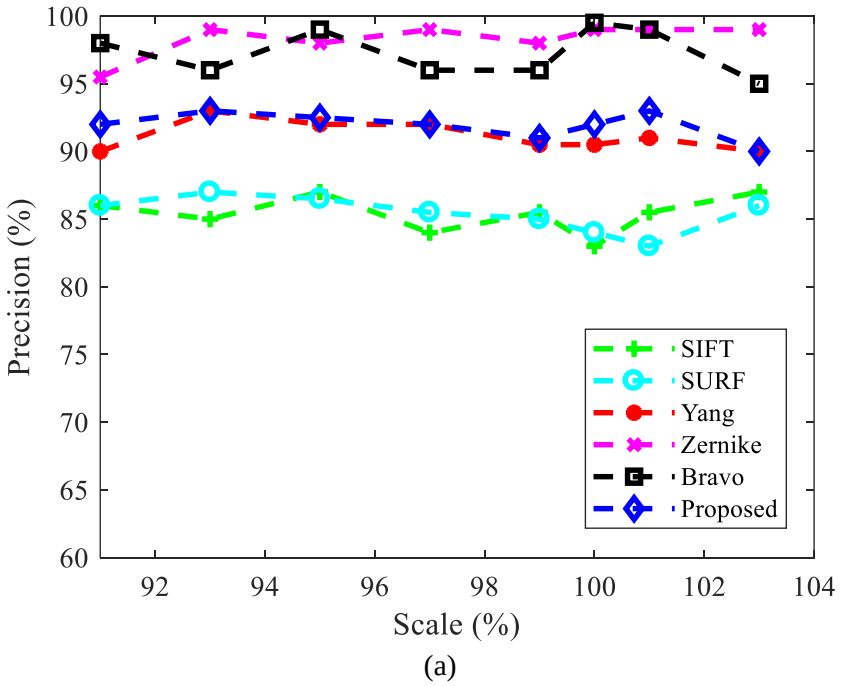
(c)

Figure 4.11: (a) Precision ratio with change in rotation angels (b) Recall ratio with change in rotation angles (c) F_1 score with change in rotation angles

Figure 4.11 (a) shows the precision ratio of the forgery detection techniques with the change in rotation angles. Precision ratio of proposed technique is 95% and is higher than SIFT [68], SURF [69] and Yang [42] methods. Precison ratio of Zernike [27] and Bravo [64] methods is higher than our proposed scheme because their extracted features provide better rotation invariance. Figure 4.11 (b) shows the effect of rotation on the recall ratio of forgery detection techniques from which it is clear that recall ratio of proposed method is between 80 to 90% and is better than all existing techniques even under the effect of rotation on the forged images. Recall ratio of Bravo [64] is worst (30% or lower) with the change in rotation angles. Figure 4.11 (c) shows the effect of rotation on the F_1 score of forgery detection techniques. We observed that with the variation in rotation angles, F_1 score of our proposed scheme is always greater than 85% which shows that, it is better than methods [64], [68], [69], [42] and [27]. F_1 score of Zernike [27] is worst (50% or lower) with the change of rotation angles.

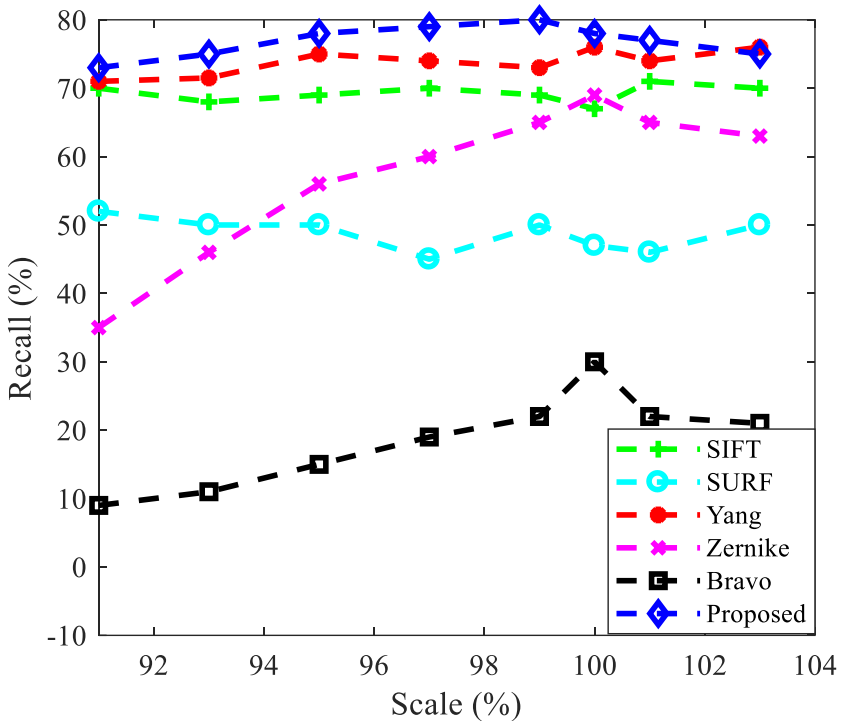
4.5.2 Performance of CMFD techniques under the effect of Scaling

For testing the performance of forgery detection techniques under scaling effect, we have rescaled the forged region with scale factors from 91% to 103% with the step length of 2%. Figure 4.12 shows the effect of change of scaling factors on the precision ratio, recall ratio and F_1 score ratio.

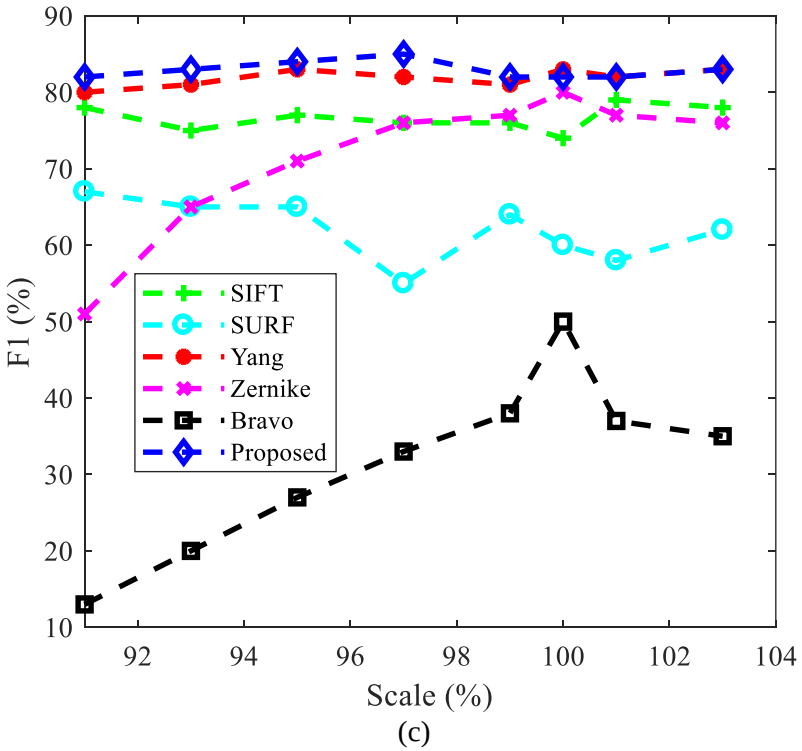


(a)

Figure 4.12: (a) Precision with variation in scaling factor (b) Recall with variation in scaling factor (c) F_1 score with variation in scaling factor (contd.)



(b)



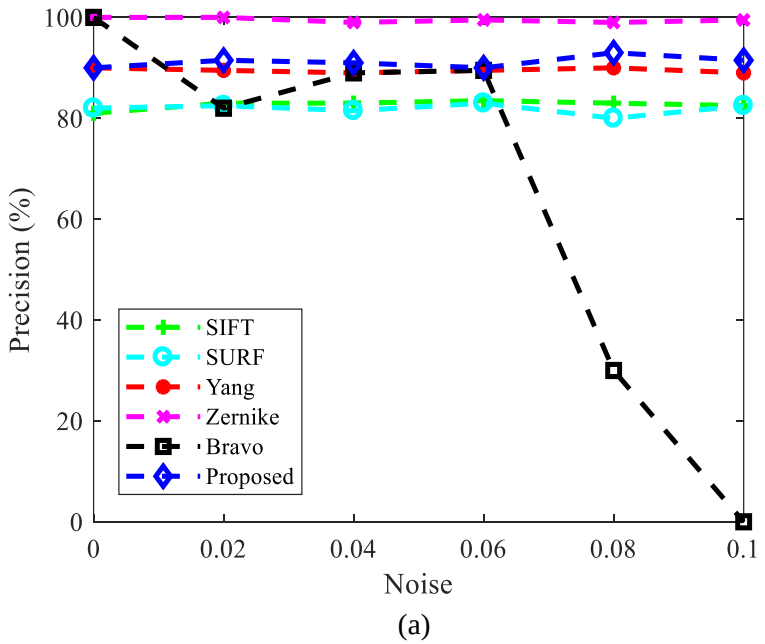
(c)

Figure 4.12: (a) Precision with variation in scaling factor (b) Recall with variation in scaling factor (c) F_1 score with variation in scaling factor

Figure 4.12 (a) shows the precision ratio of the forgery detection techniques with the change in scaling factors. Precision ratio of proposed technique is 93% and is higher than SIFT [68], SURF [69] and Yang [42] methods. Precison ratio of Zernike [27] and Bravo [64] methods is higher than our proposed scheme because their extracted features provide better scaling invariance. Figure 4.12 (b) shows the effect of scaling on the recall ratio of forgery detection techniques from which it is clear that recall ratio of proposed method is in between 70 to 80% and is better than all existing techniques even under the effect of scaling on the forged images. Recall ratio of Bravo [64] is worst (10 to 30%) with the change in scaling factors. Figure 4.12 (c) shows the effect of scaling on the F_1 score of forgery detection techniques. We observed that with the variation in scaling factor, F_1 score of our proposed scheme is always greater than 80%, which shows that it is better than methods [64], [68], [69], [42] and [27]. F_1 score of Bravo [64] is worst (10 to 50 %) under the scaling factors.

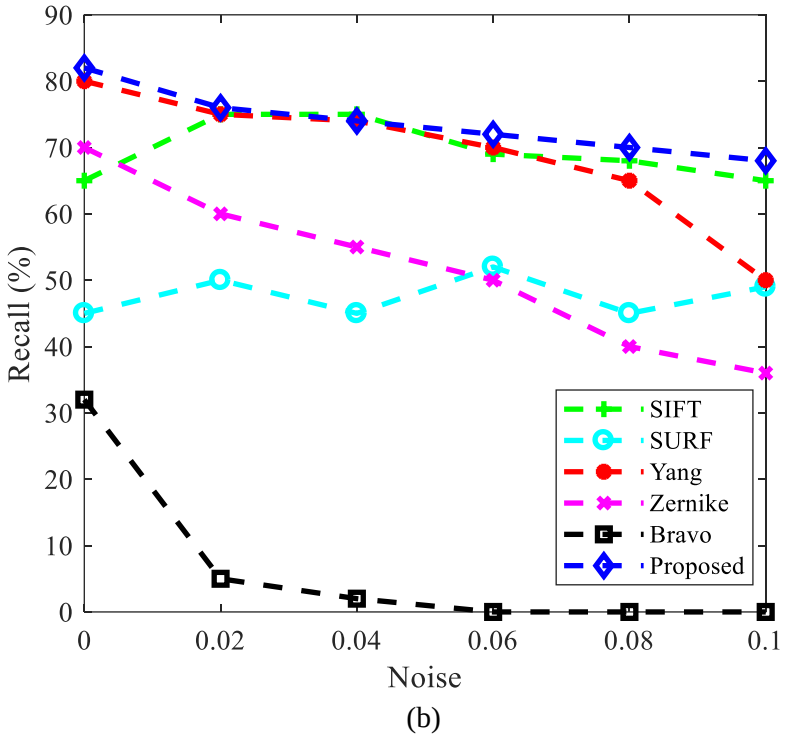
4.5.3 Performance of CMFD techniques under the effect of added Noise

For testing the performance of forgery detection techniques under noise effect, we have added the noise to the forged region with varying sd. Addition of noise to the forged image makes the forgery undetectable and hence affect the forgery detection techniques. Figure 4.13 shows the effect of change of standard deviation of added noise on the precision ratio, recall ratio and F_1 score ratio.

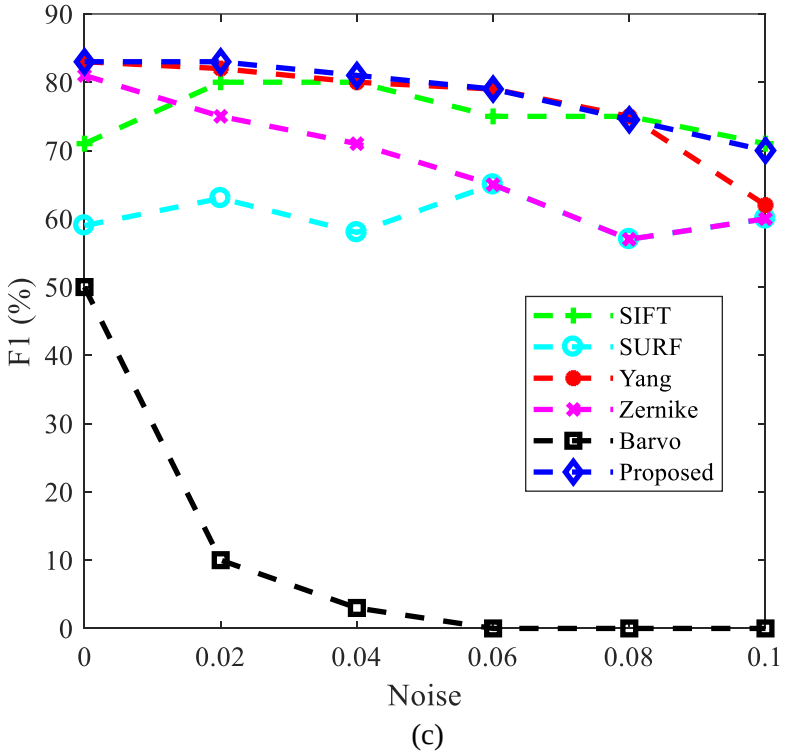


(a)

Figure 4.13: (a) Precision with variation in noise (b) Recall with variation in noise (c) F_1 score with variation in noise (contd.)



(b)



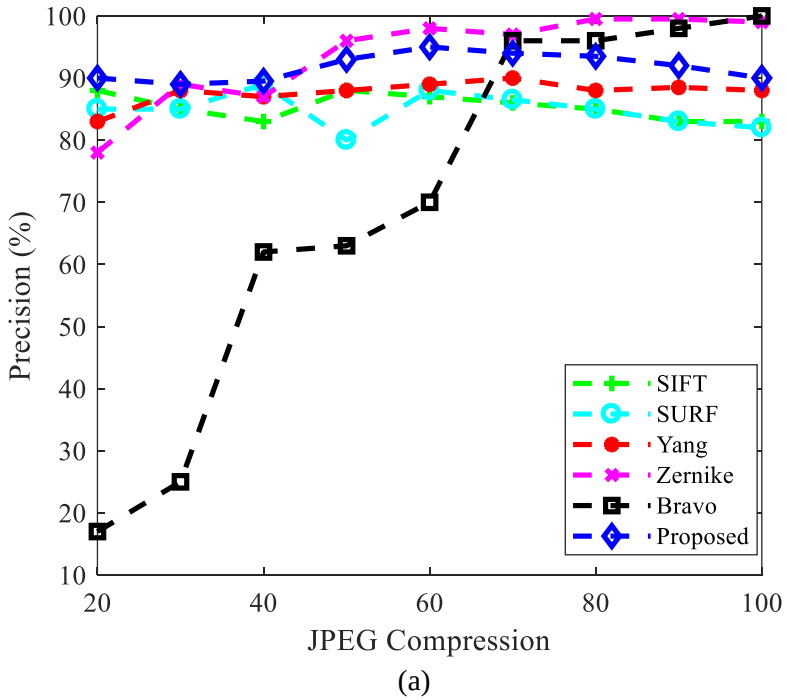
(c)

Figure 4.13: (a) Precision with variation in noise (b) Recall with variation in noise (c) F_1 score with variation in noise

Figure 4.13 (a) shows the precision ratio of the forgery detection techniques with the change in noise added. Precision ratio of proposed technique is 85% and is higher than SIFT [68], SURF [69], Bravo [64] and Yang [42] methods. Precison ratio of Zernike [27] method is higher than our proposed scheme. Figure 4.13 (b) shows the effect of noise on the recall ratio of forgery detection techniques from which it is clear that recall ratio of proposed method is between 70 to 80% and is better than all existing techniques even under the effect of noise on the forged images. Recall ratio of Bravo [64] is worst (30% or lower) with the change in added noise amount. Figure 4.13 (c) shows the effect of noise on the F_1 score of forgery detection techniques. We have observed that with the variation in noise, F_1 score of our proposed scheme is 75 to 80% which shows that it better than methods [64], [68], [69], [42] and [27]. F_1 score of Bravo [64] is worst less than 50 % under the effect of added noise.

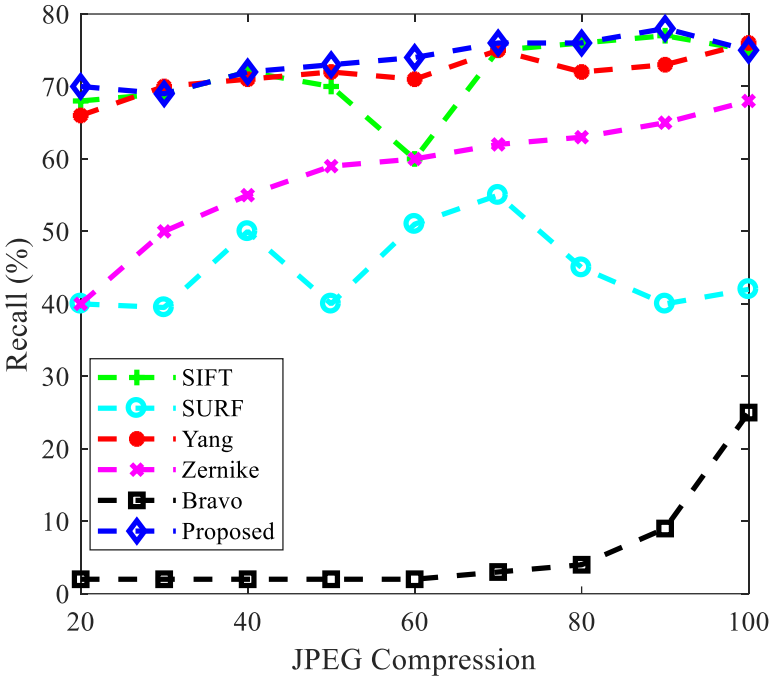
4.5.4 Performance of CMFD techniques under the effect of JPEG Compression

For testing the performance of forgery detection techniques under JPEG compression effect, we have compressed the forged image with various quality factors varying from 20 to 100% at the step size of 10 %. Figure 4.14 shows the effect of change of compression factor on the precision ratio, recall ratio and F_1 score ratio.

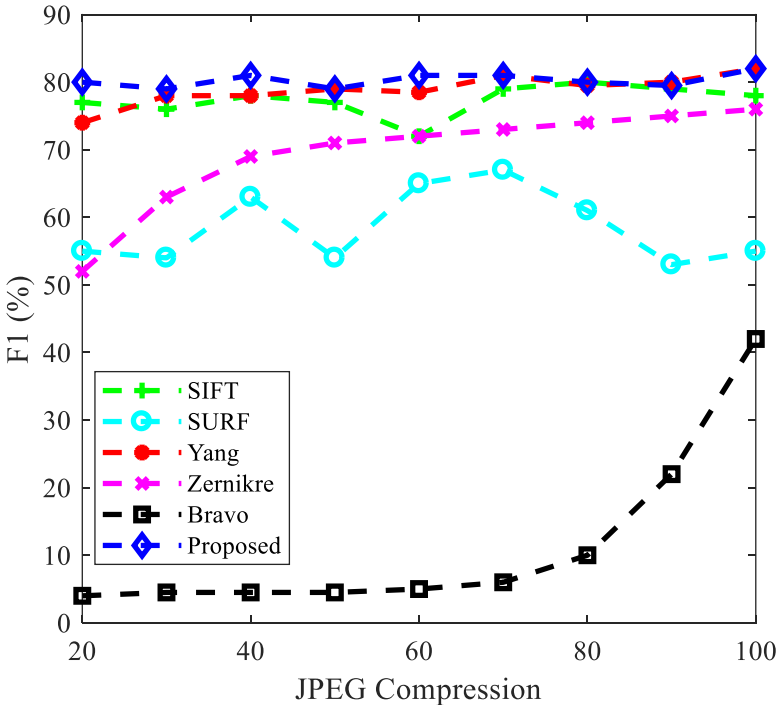


(a)

Figure 4.14: (a) Precision ratio with variation in JPEG compression (b) Recall ratio with variation in JPEG compression (c) F_1 score with variations in JPEG compression (contd.)



(b)



(c)

Figure 4.14: (a) Precision ratio with variation in JPEG compression (b) Recall ratio with variation in JPEG compression (c) F_1 score with variations in JPEG compression

JPEG compression of forged image makes the forgery undetectable as with low quality compression, image data becomes unclear and hence affect the forgery detection techniques. JPEG compression also introduce artifacts to the forged images, which can act as clue for forgery detection. Figure 4.14 (a) shows the precision ratio of the forgery detection techniques with the change in quality factor of JPEG compression. Precision ratio of the proposed technique is 90% and is higher than SIFT [68], SURF [69] and Yang [42] methods. Precison ratio of Zernike [27], Bravo [64] methods are higher than our proposed scheme. Figure 4.14 (b) shows the effect of JPEG compression on the recall ratio of forgery detection techniques, from which it is clear that recall ratio of the proposed method is about 70% and is better than all existing techniques even under the effect of JPEG compression of the forged images. Recall ratio of Bravo [64] is worst (2 to 25%) with the change in quality factor of JPEG compression. Figure 4.14 (c) shows the effect of JPEG compression on the F_1 score of forgery detection techniques. F_1 score of our proposed scheme is 80% or higher with the change in compression factor. F_1 score of Bravo is worst (5 to 40 %) under the effect of JPEG compression.

Precision determines how much precise our detection algorithm is, in locating forged region in an image. Whereas, recall determines how much correct our detection scheme is, in classification of image as forged or original. In our proposed detection scheme, SVM classifier classifies the image as forged or original. So, recall ratio depends upon SVM classification. K-means clustering algorithm locates the forged area in an image. So, precision ratio depends upon K-means clustering algorithm. Inappropriate clustering or non optimal threshold values can be the reasons for lower precision ratio of the proposed scheme.

4.6 SUMMARY

In this chapter, we have evaluated our proposed image forgery detection scheme on CMFD benchmark dataset [41] and Columbia image splicing dataset [66] and their results are discussed for copy-move forgery without any post processing operations performed on the forged image. For comparison of detection techniques of copy-move forgery, results are compared at pixel, image level. At the pixel level, our proposed technique has highest recall ratio and F_1 score but precision ratio is lower than the Bravo [64] and Zernike [27]. At image level, our proposed scheme has highest recall ratio and F_1 score. Precision ratio at image level is higher than many of existing techniques but lower than Bravo [64].

Then at pixel level, proposed scheme is evaluated with various post-processing operations along with other existing CMFD techniques with the same operations applied on them. Considering all the transformation parameters; the precision ratio of our proposed scheme is higher than SIFT [68], SURF

[69] and Yang [42] but lower than Zernike [27] and Bravo [64]. Recall ratio and F_1 ratio of our proposed technique is higher than all of the existing techniques. Bravo [64] and Zernike [27] has lowest Recall ratio and F_1 score. Hence, overall our proposed scheme is better than the existing image copy-move forgery detection techniques.

CHAPTER 5
CONCLUSION AND FUTURE SCOPE OF RESEARCH

5.1 CONCLUSION

In the proposed block based scheme, a DCT and SVD based technique is used to detect the copy-move image forgery. Feature set is generated by dividing an image into blocks and then applying DCT on it. Afterwards, SVD is used to reduce the feature vector dimension. The proposed technique is efficient in CMFD and provides robustness against transformations. The datasets used for the evaluation of the proposed scheme are CMFD benchmark dataset [41] and Columbia image splicing dataset [66]. The proposed scheme used DCT for feature extraction and SVD for feature dimension reduction. SVM classifier is used for classification of images and for localization of forged region, K-means clustering is used. Evaluation and comparison of the proposed scheme is done at pixel level and image level. Comparison of the proposed scheme and existing CMFD techniques is done for copy move forgery with post-processing operations applied on images and without it.

The proposed technique has highest recall ratio and F_1 score but precision ratio is lower than the Bravo [64] and Zernike [27] at pixel level. At image level, our proposed scheme has highest recall ratio and F_1 score. Precision ratio at image level is higher than many of existing techniques but lower than Bravo [64]. Then at pixel level, the proposed scheme is evaluated with various post-processing operations along with other existing CMFD techniques with the same operations applied on them. Considering all the transformation parameters: rotation, scaling, noise and JPEG compression; the precision ratio of our proposed scheme is higher than SIFT [68], SURF [69] and Yang [42] but lower than Zernike [27] and Bravo [64]. Recall ratio and F_1 score of our proposed technique is higher than all of the existing techniques. Bravo [64] and Zernike [27] has lowest Recall ratio and F_1 score. The proposed detection scheme has precision ratio lower than one or two of the discussed techniques. Possible reason for the lower ratio of precision can be the inappropriate clustering done by K-means clustering algorithm or non optimal threshold values. Largest ratio of our recall parameter shows that our proposed scheme can correctly classify an image into one of the two classes. F_1 score is average ratio of both precision and recall. Hence, overall our proposed scheme is better than the existing image copy-move forgery detection techniques.

5.2 FUTURE SCOPE OF RESEARCH

Even though our proposed scheme performed better than many existing techniques and following issues can be considered for future work:

- The proposed scheme can be made more effective for the detection in case of multiple forgery attacks.
- The proposed work is only restricted to the forgery detection in the images only and it can be done for audios and videos also.

REFERENCES

[1] Warif NBA *et al*. (2016). Copy-move forgery detection: Survey, challenges and future directions, *Journal of Network and Computer Applications*, 75, 259-278.

[2] Ansari MD, Ghrera SP and Tyagi V (2014). Pixel-based image forgery detection: A review, *IETE Journal of Education*, 55(1), 40-46.

[3] Kashyap A *et al*. (2017). An Evaluation of Digital Image Forgery Detection Approaches, *arXiv:1703.09968*.

[4] Soni B, Das PK and Thounaojam DM (2017). CMFD: A detailed review of block based and key feature based techniques in image copy-move forgery detection, *IET Image Processing*, 12(2), 167-178.

[5] Oommen RS, Jayamohan M and Sruthy S (2015). A Survey of Copy-Move Forgery Detection Techniques for Digital Images, *International Journal of Innovations in Engineering and Technology*, 5(2), 419-426.

[6] Basseville, M (1989). Distance measures for signal processing and pattern recognition, *Signal Processing*, 18(4), 349-369.

[7] Warbhe AD, Dharaskar RV and Thakare VM (2016). A survey on key-point based copy-paste forgery detection techniques, *Proceedings of International Conference on Information Security and Privacy*, [Nagpur, India: December 2015], pp. 61-67.

[8] Fridrich AJ, Soukal BD and Lukas AJ (2003). Detection of copy-move forgery in digital images, *Proceedings of Digital Forensic Research Workshop*, [Cleveland, Ohio: August 2003], pp. 1-10.

[9] Bayram S, Sencar H T and Memon N (2009). An efficient and robust method for detecting copy-move forgery, *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, [Taipei, Taiwan: April 2009], pp. 1053-1056.

[10] Fattah SA *et al.* (2014). A scheme for copy-move forgery detection in digital images based on 2D-DWT, *Proceedings of IEEE 57th International Midwest Symposium on Circuits and Systems*, [57th: TX, USA: August 2014], pp. 801-804.

[11] Zhao J and Guo J (2013). Passive forensics for copy-move image forgery using a method based on DCT and SVD, *Forensic Science International*, 233(1-3), 158-166.

[12] Bashar M *et al*. (2010). Exploring duplicated regions in natural images, *IEEE Transactions on Image Processing*, 99, 1-40.

[13] Kang X and Wei S (2008). Identifying tampered regions using singular ratio decomposition in digital image forensics, Proceedings of IEEE *International Conference on Computer Science and Software Engineering*, [Hubei, China: December 2008], pp. 926-930.

[14] Li G *et al.* (2007). A sorted neighborhood approach for detecting duplicated regions in image forgeries based on DWT and SVD, *Proceedings of* IEEE *International Conference on Multimedia and Expo*, [Beijing, China: July 2007], pp. 1750-1753.

[15] Li L *et al*. (2013). An efficient scheme for detecting copy-move forged images by local binary patterns, *Journal of Information Hiding and Multimedia Signal Processing*, 4(1), 46-56.

[16] Davarzani R *et al*. (2013). Copy-move forgery detection using multiresolution local binary patterns, *Forensic Science International*, 231(1-3), 61-72.

[17] Uliyan DM, Jalab HA and Wahab AWA (2015). Copy move image forgery detection using Hessian and center symmetric local binary pattern, *Proceedings of IEEE Conference on Open Systems*, [Bandar Melaka, Malaysia: August 2015], pp. 7-11.

[18] Muhammad G *et al.* (2013). Copy move image forgery detection method using steerable pyramid transform and texture descriptor, *Proceedings of IEEE EUROCON 2013*, [Zagreb, Croatia: July 2013], pp. 1586-1592.

[19] Muhammad G *et al*. (2014). Image forgery detection using steerable pyramid transform and local binary pattern, *Machine Vision and Applications*, 25(4), 985-995.

[20] Alahmadi A *et al*. (2017). Passive detection of image forgery using DCT and local binary pattern, *Signal, Image and Video Processing*, 11(1), 81-88.

[21] Ardizzone E, Bruno A and Mazzola G (2010). Copy-move forgery detection via texture description, *Proceedings of the 2nd ACM Workshop on Multimedia in Forensics, Security and Intelligence*, [2nd: Firenze, Italy: October 2010], pp. 59-64.

[22] Hsu HC and Wang MS (2012). Detection of copy-move forgery image using Gabor descriptor, *Proceedings of International Conference on Anti-Counterfeiting, Security and Identification (ASID)*, [Taipei, Taiwan: August 2012], pp. 1-4.

[23] Fadl SM and Semary NA (2014). A proposed accelerated image copy-move forgery detection, *Proceedings of IEEE Visual Communications and Image Processing Conference*, [Valletta, Malta: December 2014], pp. 253-257.

[24] Lee JC, Chang CP and Chen WK (2015). Detection of copy–move image forgery using histogram of orientated gradients, *Information Sciences*, 321(13), 250-262.

[25] Alhussein M (2016). Image tampering detection based on local texture descriptor and extreme learning machine, *Proceedings of 18th International Conference on Computer Modelling and Simulation (UKSim)*, [*18th*: Cambridge, UK: April 2016], pp. 196-199.

[26] Mahdian B and Saic S (2007). Detection of copy–move forgery using a method based on blur moment invariants, *Forensic Science International*, 171(2-3), 180-189.

[27] Ryu SJ, Lee MJ and Lee HK (2010). Detection of copy-rotate-move forgery using Zernike moments, Bohme R, Fong PW and Safavi-Naini R (eds.), Information Hiding. Berlin, Heidelberg, 2010, pp. 51-65.

[28] Li Y (2013). Image copy-move forgery detection based on polar cosine transform and approximate

nearest neighbor searching, *Forensic Science International*, 224(1-3), 59-67.

[29] Wang XY *et al*. (2018). Robust copy–move forgery detection using quaternion exponent moments, *Pattern Analysis and Applications*, 21(2), 451-467.

[30] Wo YK *et al*. (2016). Copy–move forgery detection based on multi-radius PCET, *IET Image Processing*, 11(2), 99-108.

[31] Shin YD (2016). Fast detection of copy-move forgery image using two step search algorithm, *International Journal of Security Applications*, 10(5), 203-214.

[32] Singh J and Raman B (2012). A high performance copy-move image forgery detection scheme on GPU, *Proceedings of International Conference on Soft Computing for Problem Solving (SocProS)*, [Roorkee, India: December 2011], pp. 239-246.

[33] Ardizzone E and Mazzola G (2009). Detection of duplicated regions in tampered digital images by bit-plane analysis, *Proceedings of International Conference on Image Analysis and Processing*, [Vietri Sul Mare, Italy: September 2009], pp. 893-901.

[34] Li J *et al*. (2015). Segmentation-based image copy-move forgery detection scheme, *IEEE Transactions on Information Forensics and Security*, 10(3), 507-518.

[35] Popescu AC and Farid H (2005). Exposing digital forgeries by detecting traces of resampling, *IEEE Transactions on Signal Processing*, 53(2), 758-767.

[36] Ghorbani M, Firouzmand M and Faraahi A (2011). DWT-DCT (QCD) based copy-move image forgery detection, *Proceedings of 18th IEEE International Conference on Systems, Signals and Image Processing (IWSSIP)*, [18th: Sarajevo, Bosnia-Herzegovina: June 2011], pp. 1-4.

[37] Qu Z, Qiu G and Huang J (2009). Detect digital image splicing with visual cues, *Proceedings of International workshop on information hiding*, [Vietri Sul Mare, Italy: June 2009], pp. 247-261.

[38] Chihaoui T, Bourouis S, and Hamrouni K (2014). Copy-move image forgery detection based on SIFT descriptors and SVD-matching, *Proceedings of 1st IEEE International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)*, [1st: Sousse, Tunisia: March 2014], pp. 125-129.

[39] Mohamadian Z and Pouyan AA (2013). Detection of duplication forgery in digital images in uniform and non-uniform regions, *Proceedings of 15th IEEE International Conference on Computer Modelling and Simulation (UKSim)*, [15th: Cambridge, UK: April 2013], pp. 455-460.

[40] Zheng J *et al*. (2016). Fusion of block and keypoints based approaches for effective copy-move image forgery detection, *Multidimensional Systems and Signal Processing*, 27(4), 989-1005.

[41] Image Manipulation Dataset, Department of computer science, Friedrich Alexander University. Available at https://www5.cs.fau.de/research/data/image-manipulation (Accessed on 16th October 2017).

[42] Hashmi MF, Hambarde AR and Keskar AG (2013). Copy move forgery detection using DWT and SIFT features, *Proceedings of 13th IEEE International Conference on Intelligent Systems Design and Applications (ISDA)*, [13th: Bangi, Malaysia: December 2013], pp. 188-193.

[43] Sadeghi S, Jalab HA and Dadkhah S (2012). Efficient copy-move forgery detection for digital images, *World Academy of Science, Engineering and Technology*, 6(11), 1339-1342.

[44] Yao H *et al.* (2011). Detecting copy-move forgery using non-negative matrix factorization, *Proceedings of 3rd IEEE International Conference on Multimedia Information Networking and Security (MINES)*, [3rd: Shanghai, China: November 2011], pp. 591-594.

[45] Wu Q, Wang S, and Zhang X (2010). Detection of image region-duplication with rotation and scaling tolerance, *Proceedings of 3rd IEEE International Conference on Computational Collective Intelligence*, [3rd: Kaohsiung, Taiwan: November 2010], pp. 100-108.

[46] Muhammad G *et al.* (2013). Copy move image forgery detection method using steerable pyramid transform and texture descriptor, *Proceedings of IEEE EUROCON*, [Zagreb, Croatia: July 2013],

pp. 1586-1592.

[47] Lin SD and Wu T (2011). An integrated technique for splicing and copy-move forgery image detection, *Proceedings of 4th IEEE International Conference on Image and Signal Processing (CISP)*, [4th: Shanghai, China: October 2011], pp. 1086-1090.

[48] Wandji ND, Xingming S and Kue MF (2013). Detection of copy-move forgery in digital images based on DCT, *arXiv:1308.5661*.

[49] Wandji ND and Xingming S (2013, January). Robust detection of copy-move forgery in color images, *Proceedings of the International Conference on Image Processing, Computer Vision, and Pattern Recognition (IPCV)*, [Las Vegas, USA: January 2013], pp. 1-4.

[50] Bravo-Solorio S and Nandi AK (2011, May). Exposing duplicated regions affected by reflection, rotation and scaling, *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, [Prague, Czech Republic: May 2011], pp. 1880-1883.

[51] Luo W, Huang J and Qiu G (2006). Robust detection of region-duplication forgery in digital image, *Proceedings of IEEE 18th International Conference on Pattern Recognition*, [18th: Hong kong, China: August 2006], pp. 746-749.

[52] Li L *et al*. (2013). An efficient scheme for detecting copy-move forged images by local binary patterns, *Journal of Information Hiding and Multimedia Signal Processing*, 4(1), 46-56.

[53] Li L *et al*. (2014). Detecting copy-move forgery under affine transforms for image forensics, *Computers and Electrical Engineering*, 40(6), 1951-1962.

[54] Dybala B, Jennings B and Letscher D (2007). Detecting filtered cloning in digital images, *Proceedings of the 9th workshop on Multimedia & security*, [9th: Texas, USA: September 2007], pp. 43-50.

[55] Yong L, Meishan H and Bogang L (2012). Robust evidence detection of copy-rotate-move forgery in image based on singular value decomposition, *Proceedings of International Conference on Information and Communications Security*, [Hong Kong, China: October 2012], pp. 357-364.

[56] Liu G *et al*. (2011). A passive image authentication scheme for detecting region-duplication forgery with rotation, *Journal of Network and Computer Applications*, 34(5), 1557-1565.

[57] Zhong L and Xu W (2013). A robust image copy-move forgery detection based on mixed Moments, *Proceedings of 4th IEEE International Conference on Software Engineering and Service Science (ICSESS)*, [4th: Beijing, China: May 2013], pp. 381-384.

[58] Wang J *et al*. (2009). Detection of image region duplication forgery using model with circle block, *Proceedings of IEEE International Conference on Multimedia Information Networking and Security*, [Hubei, China: November 2009], pp. 25-29.

[59] Qiao M *et al*. (2011). A novel approach for detection of copy-move forgery, *Proceedings of 5th International Conference on Advanced Engineering Computing and Applications in Sciences (ADVCOMP)*, [5th: Lisbon, Portugal: November 2011], pp. 44-47.

[60] Yang B *et al*. (2013). An Efficient Forensic Method for Copy--move Forgery Detection Based on DWT-FWHT, *Radioengineering*, 22(4), 1098-1105.

[61] Sheng G *et al*. (2012). Robust algorithm for detection of copy-move forgery in digital images based on ridgelet transform, *Proceedings of International Conference on Artificial Intelligence and Computational Intelligence*, [Chengdu, China: October 2012], pp. 317-323.

[62] Nguyen HC and Katzenbeisser S (2012). Detection of copy-move forgery in digital images using radon transformation and phase correlation, *Proceedings of 8th IEEE International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP)*, [8th: Piraeus, Greece: July 2012], pp. 134-137.

[63] Wu Y *et al*. (2014). Dual tree complex wavelet transform approach to copy-rotate-move forgery detection, *Science China Information Sciences*, 57(1), 1-12.

[64] Bravo-Solorio S and Nandi AK (2011). Automated detection and localisation of duplicated regions affected by reflection, rotation and scaling in image forensics, *Signal Processing*, 91(8), 1759-1770.

[65] Christlein V, Riess C and Angelopoulou E (2010). On rotation invariance in copy-move forgery detection, *Proceedings of IEEE International Information Forensics and Security (WIFS)*, [Seattle, USA: December 2010], pp.1-6.

[66] Columbia image splicing detection evaluation dataset, DVMM Laboratory of Columbia University. Available at https://www.ee.columbia.edu/ln/dvmm/AuthSplicedDataSet/photographers.htm (Accessed on 6th August 2017).

[67] Pan X and Lyu S (2010). Region duplication detection using image feature matching, *IEEE Transactions on Information Forensics and Security*, 5(4), 857-867.

[68] Amerini I *et al*. (2011). A sift-based forensic method for copy–move attack detection and transformation recovery, *IEEE Transactions on Information Forensics and Security*, 6(3), 1099-1110.

[69] Shivakumar BL and Baboo LDSS (2011). Detection of region duplication forgery in digital images using SURF, *International Journal of Computer Science Issues*, 8(4), 199-205.

[70] Cozzolino D, Poggi G and Verdoliva L (2015). Efficient dense-field copy–move forgery detection, *IEEE Transactions on Information Forensics and Security*, 10(11), 2284-2297.

[71] Pun CM, Yuan XC and Bi XL (2015). Image forgery detection using adaptive over segmentation and feature point matching, *IEEE Transactions on Information Forensics and Security*, 10(8), 1705-1716.

[72] Data clustering algorithms. Available at https://sites.google.com/site/dataclusteringalgorithms/k-means-clustering-algorithm (Accessed on 6th August 2017).

[73] Singh VK and Tripathi RC (2011). Fast and efficient region duplication detection in digital images using sub-blocking method, *International Journal of Advanced Science and Technology*, 35, 93-102.

[74] Lynch G, Shih FY and Liao HYM (2013). An efficient expanding block algorithm for image copy-move forgery detection, *Information Sciences*, 239, 253-265.

[75] CASIA Image Tampering Detection Evaluation Database, National Laboratory of Pattern Recognition. Available at http://forensics.idealtest.org (Accessed on 17th January 2018).

[76] MICC F220 and MICC F2000, Image Communication Laboratory. Available at lci.micc.unifi.it/labd (Accessed on 17th January 2018).

[77] GRIP Dataset, Global patterns of current and future road infrastructure. Available at www.globio.info (Accessed on 5th December 2017).

[78] COVERAGE Dataset. Available at https://pan.baidu (Accessed on 5th December 2017).

[79] COMOFOD- Image Database for copy-move forgery detection, Visual Communication Laboratory. Available at www.vcl.fer.hr/comofod/ (Accessed on 5th December 2017).

LIST OF PUBLICATIONS

1. Priyanka Thakur and Kulbir Singh, "An Improved Block Based Copy-Move Forgery Detection Technique," Communicated to Forensic Science International, SCI-Indexed (Impact factor: 1.974).

Turnitin Originality Report

An improved block based copy move forgery detection technique by Priyanka Thakur

From An improved block based copy move forgery detection technique (kss)

Similarity Index	Similarity by Source	
8%	Internet Sources:	4%
	Publications:	7%
	Student Papers:	1%

Processed on 11-Jul-2018 11:46 +0530
ID: 981775440
Word Count: 15732

sources:

1 < 1% match (Internet from 20-Nov-2017)
https://link.springer.com/content/pdf/10.1007/978-81-322-0997-3.pdf

2 < 1% match (publications)
Badal Soni, Pradip K. Das, Dalton Meitei Thounaojam. "CMFD: a detailed review of block based and key feature based techniques in image copy-move forgery detection", IET Image Processing, 2017

3 < 1% match (publications)
"Smart and Innovative Trends in Next Generation Computing Technologies", Springer Nature, 2018

4 < 1% match (publications)
Lecture Notes in Computer Science, 2016.

5 < 1% match (Internet from 01-Feb-2018)
http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.94.6804&rep=rep1&type=pdf

6 < 1% match (publications)
Isaac, Meera Mary, and M. Wilscy. "A key point based copy-move forgery detection using HOG features", 2016 International Conference on Circuit Power and Computing Technologies (ICCPCT), 2016.

7 < 1% match (Internet from 14-Apr-2015)
http://www.sersc.org/journals/IJAST/vol73/2.pdf

8 < 1% match (publications)
XiuLi Bi, Chi-Man Pun, Xiao-Chen Yuan. "Multi-scale feature extraction and adaptive matching for copy-move forgery detection", Multimedia Tools and Applications, 2016

9 < 1% match (publications)
Savita Walia, Krishan Kumar. "Digital image forgery detection: a systematic scrutiny", Australian Journal of Forensic Sciences, 2018

10 < 1% match (publications)
Xiuli Bi, Chi-Man Pun, Xiao-Chen Yuan. "Multi-Level Dense Descriptor and Hierarchical Feature Matching for Copy–Move Forgery Detection", Information Sciences, 2016

11 < 1% match (Internet from 19-Oct-2015)
http://www.archive.org/stream/IjaetVolume2Issue1/Volume2Issue1_djvu.txt

12 < 1% match (publications)
Lecture Notes in Computer Science, 2013.

13 < 1% match (publications)
Beste Ustubioglu, Guzin Ulutas, Mustafa Ulutas, Vasif V. Nabiyev. "A new copy move forgery detection technique with automatic threshold determination", AEU - International Journal of