Advanced Techniques for Image Forgery Detection

Amruta Prabhakar Jagtap
M.E Computer Engineering Department
JSCOE
Pune, India.

H. A. Hingoliwala
Computer Engineering Department
JSCOE
Pune, India.

ABSTRACT
Image forgery means manipulation of digital image to conceal meaningful information of the image. The detection of forged image is driven by the need of authenticity and to maintain integrity of the image. A copy–move forgery detection theme victimization adaptive over segmentation and have purpose feature matching is proposed. The proposed scheme integrates both block-based and key point-based forgery detection methods. The proposed adaptive over-segmentation algorithm segments the host image into non-overlapping and irregular blocks adaptively. Then, the feature points are extracted from each block as block features, and the block features are matched with one another to locate the labeled feature points; this procedure can approximately indicate the suspected forgery regions. To detect the forgery regions more accurately, we propose the forgery region extraction algorithm which replaces the features point with small super pixels as feature blocks and them merges the neighboring blocks that have similar local color features into the feature block to generate the merged regions. Finally, it applies the morphological operation to merged regions to generate the detected forgery regions. In cut-paste image forgery detection, proposed digital image forensic techniques capable of detecting global and local contrast enhancement, identifying the use of histogram equalization.

Keywords
Copy-move forgery detection; Adaptive over-segmentation; Feature point matching and extraction; Cut-paste forgery detection.

1. INTRODUCTION
In this era, Digital Image Forgery has been increasingly easy to perform, so the reliability of the image is thus becoming an important issue to be focus on. It does not differ very much in nature to conventional image forgery. Instead of using photograph digital image forgery deals with the digital image. By using the tool such as Adobe Photoshop, GIMP, Coral Paint fake images can be created as some of the tools are open source. Image forgery may lead to hazards. In banking system image forgery is a big threat, this result into big frauds. Nowadays detecting these types of forgeries has become very useful to reduce these problems at present. To determine whether a digital image is original is a big challenge. To find the marks of tampering in a digital image is a challenging task. Tampering is normally done to cover objects in an image in order to either produce false proof or to make the image more pleasant for appearance. There are many cases in digital image forgery, all this cases are classified into three categories based on the process of creating fake images the group are image retouching, and image splicing, copy-move attack. Image forgery is basically a modification of image to conceal some meaningful or useful information. The common manipulations of a digital image are copy-move and cut-paste forgery.

1.1 Copy-Move Forgery Detection
Copy-move forgery, which is to paste one or several copied region of an image into other part of the same image. During the copy and move operations, some image processing methods such as rotation, scaling, blurring, compression, and noise addition are occasionally applied to make convincing forgeries. Earlier blocked based forgery detection was used to detect forged image but this algorithm faced some drawbacks such as the host image is divided into over-lapping rectangular blocks, which would be computationally expensive as the size of the image increases and it was less efficient as it take more time to be process. To avoid such drawbacks along with the blocked based forgery, we proposed an image-blocking method called Adaptive Over-Segmentation that divided the host image into non over-lapping blocks adaptively with the help of two algorithm those are Simple Linear Iterative Clustering (SLIC) to segment the host image into irregular blocks and Discrete Wavelength Transform (DWT) which is employed to analyze the frequencies of the super pixel. Further the image block formed are pass to the Block Feature Extraction method where the block feature are extracted by using Scale Invariant Feature Transform (SIFT) as it possessed constant and better performance compared with the other extraction method. Further the process of Block Feature Matching is carried out which used Simple Linear Iterative Clustering (SLIC) for calculating super pixel and Discrete Wavelength Transform for finding super pixel from one block and checking other for other blocks. When the features are extracted and matched then we get to know which regions the host image has been forged.

1.2 Cut-And Paste Image Forgery Detection
Cut-and paste image forgery consists of creating a composite image by replacing a contiguous set of pixels in one image with a set of pixels corresponding to an object from a separate image. If the two images used to create the composite image were captured under different lighting environments, an image forger may need to perform contrast enhancement on so that lighting conditions match across the composite image. Failure to do this may result in a composite image which does not appear realistic. Image forgeries created in this manner can be identified by using localized contrast enhancement detection to locate, the cut-and-pasted region.

2. PROBLEM DEFINITION
In existing blocked based forgery detection faced some drawbacks such as the host image is divided into over-lapping rectangular blocks and it was computationally expensive in terms of size, so the need to overcome this problem was necessary for accurate and efficient results. Hence, proposed an image-blocking method called Adaptive Over-Segmentation that divided the host image into non over-lapping blocks adaptively with the help of two algorithm those are Simple Linear Iterative Clustering (SLIC) to segment the host image into irregular blocks and Discrete Wavelength Transform (DWT) which is employed to analyze

the frequencies of the super pixels. Further the image block formed are pass to the Block Feature Extraction method where the block feature are extracted by using Scale Invariant Feature Transform (SIFT) as it possessed constant and better performance compared with the other extraction methods.

3. LITERATURE SURVEY

The existing block-based forgery detection methods divide the input images into overlapping and regular image blocks; then, the tampered region can be obtained by matching blocks of image pixels or transform coefficients. Fridrich [7], proposed a forgery detection method in which the input image was divided into over-lapping rectangular blocks, from which the quantized Discrete Cosine Transform (DCT) coefficients of the blocks were matched to find the tampered regions. In [8], proposed a method for detecting copy-move forgery over images tampered by copy-move. To detect such forgeries, the given image is divided into overlapping blocks of equal size, feature for each block is then extracted and represented as a vector, all the extracted feature vectors are then sorted using the radix sort.

DWT and SIFT [2] algorithms are proposed for copy-move detection. With DWT, the low frequency information or image is obtained. With SIFT robustness is introduced, here it detect forgery of the image even it is copied, rotate scale and then pasted. In [3], survey done on various image forgery detection techniques and finally conclude the comparative study with some parameters. Also tools are mentioned to detect the forged images that travel over the network or by natural way for daily forensics, image processing, and security. Salam A.Thajeel [4], discussed digital image forensics and its types, challenges and research problems and detail analysis of the existing approaches for detect image tampering. Author also discussed block based method and key point-based method and popular techniques of two methods. Moreover, most of the methods may not address the problems. Therefore, there is a need to develop techniques that is efficient to deal with these challenges.

The Speeded Up Robust Features (SURF) [6] were applied to extract features instead of SIFT. However, although these methods can locate the matched key-points, most of them cannot locate the forgery regions very well; therefore, they cannot achieve satisfactory detection results and, at the same time, a sustained high recall rate [5].

A novel copy–move forgery detection scheme using adaptive over segmentation and feature point matching [1] is proposed, that integrates both block-based and key point-based forgery detection methods.

Methods for detecting locally applied contrast enhancement as well as a method for identifying histogram equalization [9] are proposed. By observing that the intrinsic fingerprints of contrast enhancement operations add energy to the high frequency components of an image's pixel value histogram, we developed a global contrast enhancement detection technique. We extended this technique into a method for detecting locally applied contrast enhancement and demonstrated its usefulness for detecting cut and paste type forgeries.

A novel algorithm is proposed to identify the source-enhanced composite image created by enforcing contrast adjustment on either single or both source regions [10].The two source images used for creating cut-and-paste type of forged images may have different color temperature or luminance contrast. So, in order to make the forged image more real, contrast enhancement is performed on either one or both the regions.

4. MATHEMATICAL MODEL

4.1 Mapping

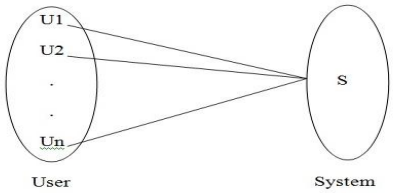

Fig 1: Many users can use this system.

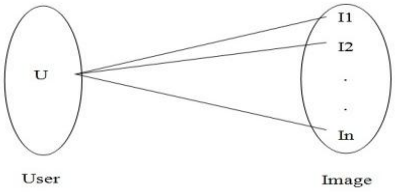

Fig 2: User can upload multiple image files.

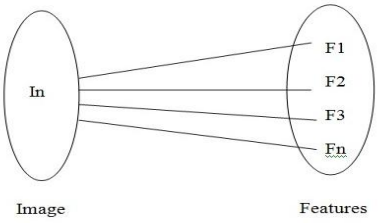

Fig 3: A single image has multiple features.

4.2 Set Theory

Proposed system can be represented as a set

$X = \{I, O, S_C, F_C, C\}$

Where,

I=set of inputs

O=set of outputs

S_C= set of outputs in success cases

F_C = set of outputs in failure cases

C = set of constraints

$I = \{I_N\}$

Where,

I_N = Set of images uploaded by User.

$O = \{C_{I,} G_S\}$

Where,

C_I = Forge Region Image.

G_S = Line Graph displaying contrast histogram.

$S_C = \{I_{Un}\}$

Where,

I_{Un} = valid set of images uploaded.

$F_C = \{I_{Fn}, NULL\}$

Where,

I_{Fn} = invalid set of images (failed to upload)

$C = \{C_1, C_2\}$

Where,

C_1 = "System only accepts images of filetypes

such as bmp, jpeg, png"

I_U, I_{Uo}, I_{Un} are in the form

$I = \{I_1, I_2, \ldots, I_n\}$

where,

$I_1, I_2, \ldots, I_n$ are images.

5. PROOPSED SYSTEM OVERVIEW

5.1 Copy-Move Forgery Detection

5.1.1 Adaptive Over-Segmentation

In this block based forgery method the image can be divided into blocks. Earlier in block based forgery detection schemes size of blocks divided into over-lapping regular blocks with predefined block size thus forgery region detected by matching the blocks in turn size of the host image is increased simultaneously by increasing computation of over-lapping blocks so we used adaptive over-segmentation method which can segment host image into non over-lapping region of irregular shape and image blocks as the host image into non-overlapping region of irregular shape and in addition to this super pixels can be obtained by over-segmentation.

For this purpose we employed SLIC algorithm to segment the host image into meaningful irregular super pixels for each block SLIC algorithm make use of k-means clustering approach to generate super pixels by using this we get rid of over-lapping block and hence decreased the computational expenses.

DWT- Discrete Wavelength transform is employed to analyze the frequency distribution of host image here low frequency energy is discarded and high frequency energy of the host image is considered as smooth image.

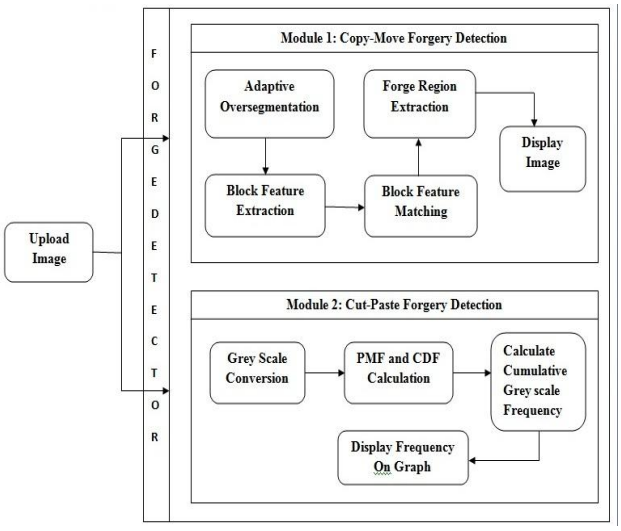

Fig 4: System Architecture

5.1.2 Block Feature Extraction

In this section, we extract block features from the image blocks .The feature points extraction methods SIFT and SURF have been widely used in the field of computer vision. The SIFT possessed more constant and better performance compared with the other image feature extraction method. Therefore in our proposed algorithm, we chose SIFT as the feature point extraction method.

SIFT Algorithm:

SIFT isn't just scale invariant. You can change the scale, rotation, illumination and viewpoint, and still get good results.

Steps involved in SIFT algorithm;

Step 1: Constructing a scale space

This is the initial preparation. Create internal representations of the original image to ensure scale invariance. This is done by generating a "scale space". Generate several octaves of the original image. Each octave's image size is half the previous one. Within an octave, images are progressively blurred using the Gaussian Blur operator. The creator of SIFT suggests that 4 octaves and 5 blur levels are ideal for the algorithm.

$$L(x,y,\sigma)=G(x,y,\sigma)*I(x,y) \qquad (1)$$

The symbols:

L- is a blurred image

G -is the Gaussian Blur operator

I -is an image

x,y -are the location coordinates

σ- is the "scale" parameter. Think of it as the amount of blur. Greater the value, greater the blur.

The * is the convolution operation in x and y. It "applies" gaussian blur G onto the image I.

$$G(x,y,\sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2} \qquad (2)$$

This is the actual Gaussian Blur operator.

Step 2: LoG Approximation

LoG is a little costly, so SIFT algorithm uses Difference of Gaussians (DoG) which is an approximation of LoG. Two consecutive images in an octave are picked and one is subtracted from the other. Then the next consecutive pair is taken, and the process repeats. This is done for all octaves. The resulting images are an approximation of scale invariant laplacian of Gaussian (which is good for detecting key points).

Step 3: Finding key points

Detect the maxima and minima in the DoG images generated in the previous step. This is done by comparing neighboring pixels in the current scale, the scale "above" and the scale "below". Using the available pixel data, sub-pixel values are generated. This is done by the Taylor expansion of the image around the approximate key point.

Mathematically, it's like this:

$$D(x) = D + \frac{\partial D^T}{\partial x} x + \frac{1}{2} x^T \frac{\partial^2 D}{\partial x^2} x \qquad (3)$$

Step 4: Get rid of bad key points

Once potential key points locations are found, they have to be refined to get more accurate results. Edges and low contrast regions are bad key points. Eliminating these makes the algorithm efficient and robust. A technique similar to the Harris Corner Detector is used here. So, it eliminates any low-

contrast key points and edge key points and what remain is strong interest points.

Step 5: Assigning an orientation to the key points

An orientation is calculated for each key point. Any further calculations are done relative to this orientation. This effectively cancels out the effect of orientation, making it rotation invariant.

Step 6: Generate SIFT features

Take a 16x16 window of "in-between" pixels around the key point. Split that window into sixteen 4x4 windows. From each 4x4 window generate a histogram of 8 bins. Each bin corresponding to 0-44 degrees, 45-89 degrees, etc. Gradient orientations from the 4x4 are put into these bins. This is done for all 4x4 blocks. Finally, normalize the 128 values you get.

5.1.3 Adaptive Block Feature Matching

In block feature matching algorithm unlike that of Adaptive Segmentation where we use to calculate the super pixel by using simple linear iterative clustering (SLIC) we used to partition the image into individual blocks or set of blocks and by using Discrete Wavelet Transform (DWT) we find the super pixel from one block and checks for the other blocks pixel frequency , if we get the higher frequency of the pixel then that block is added to the block of super pixel and irregular blocks are formed in this way. while in the block feature matching in this instead of the super pixel, sub pixels are taken into consideration and again the block is formed of regular size unlike irregular.

Algorithm:

Input**:** Block Features (BF);

Output**:** Labeled Feature Points (LFP).

STEP-1**:** Load the Block Features *BF* = {*BF*1, *BF*2, . . *BFN* },

Where, *N* means the number of image blocks; and calculates the correlation coefficients *C* of the image blocks.

STEP-2: Calculate the block matching threshold *TRB* according to the distribution of correlation coefficients.

STEP-3: Locate the matched blocks *MB* according to the block matching threshold *T RB*.

STEP-4: Label the matched feature points in the matched blocks *MB* to indicate the suspected forgery regions.

5.1.4 Forgery Region Extraction

Although we have extracted the labeled feature points (LFP), which are only the locations of the forgery regions, we must still locate the forgery regions. Considering that the super-pixels can segment the host image very well, we proposed a method by replacing the LFP with small super-pixels to obtain the suspected regions (SR), which are combinations of labeled small super-pixels. Furthermore, to improve the *precision* and *recall* results, we measure the local color feature of the super-pixels that are neighbors to the suspected regions (SR); if their color feature is similar to that of the suspected regions, then we merge the neighbor super pixels into the corresponding suspected regions, which generates the merged regions (MR). Finally, a close morphological operation is applied to the merged regions to generate the detected copy-move forgery regions.
Algorithm:

Input**:** Labeled Feature Points (LFP)

Output**:** Detected Forgery Regions.

STEP-1**:** Load the Labeled Feature Points (LFP), apply the SLIC algorithm with the initial size *S* to the host image to segment it into small superpixels as feature blocks, and replace each labeled feature point with its corresponding feature block, thus generating the Suspected Regions (SR).

STEP-2**:** Measure the local color feature of the superpixels neighbor to the SR, called neighbor blocks; when their color feature is similar to that of the suspected regions, we merge the neighbor blocks into the corresponding SR, therefore creating the merged regions (MR).

STEP-3**:** Apply the morphological close operation into MR to finally generate the detected forgery regions.

5.2 Cut-Paste Forgery Detection

5.2.1 Grey Scale Conversion

The aim of pre-processing is the improvement of image data that suppresses unwanted distortions or enhances some image features important for further detection. The given image is converted into grey-scale (color conversion), when applicable.

Take average of the contribution from each channel;

$$\frac{R+G+B}{3} \qquad (4)$$

Calculate weighted average:

Grey = 0.2126R + 0.7152G + 0.0722B. (5)

5.2.2 PMF and CDF Calculation

In the histogram equalization, the first and the second step are PMF (probability mass function) and CDF (cumulative distributive function). In histogram equalization, we have to equalize all the pixel values of an image. So PMF helps us calculating the probability of each pixel value in an image. And CDF gives us the cumulative sum of these values.

5.2.2.1 PMF

PMF stands for probability mass function. As it name suggest, it gives the probability of each number in the data set or you can say that it basically gives the count or frequency of each element.

To calculate the PMF from a matrix;

For Example,

Consider matrix;

$$\begin{pmatrix} 1 & 2 & 7 & 5 & 6 \\ 7 & 2 & 3 & 4 & 5 \\ 0 & 1 & 5 & 7 & 3 \\ 1 & 2 & 5 & 6 & 7 \\ 6 & 1 & 0 & 3 & 4 \end{pmatrix}$$

Fig 5: Matrix

0	2	2/25
1	4	4/25
2	3	3/25
3	3	3/25

Fig 6: PMF from Matrix

At first, we will take the first value in the matrix, and then we will count, how much time this value appears in the whole matrix. After count they can be represented in a histogram.

In fig 7, histogram shows frequency of gray level values for per pixel image. Now if we have to calculate its PMF, we will

simple look at the count of each bar from vertical axis and then divide it by total count as shown in figure 8.

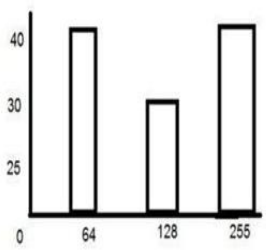

Fig 7: Histogram

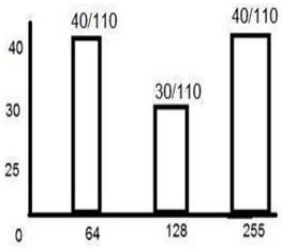

Fig 8: PMF from Histogram

5.2.2.2 CDF

CDF stands for cumulative distributive function. It is a function that calculates the cumulative sum of all the values that are calculated by PMF. It basically sums the previous one.

We will calculate CDF using a histogram. Consider the histogram which shows PMF. We will simply keep the first value as it is, and then in the 2nd value, we will add the first one and so on.

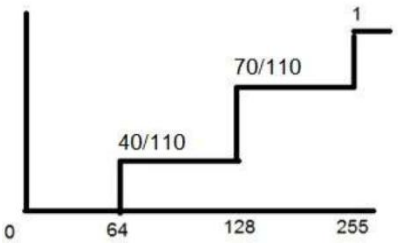

Fig 9: Histogram for CDF

5.2.3 Calculate CDF according to grey levels

In this step we will multiply the CDF value with (Gray levels (minus) 1).

Gray Level Value CDF = CDF * (Levels-1) (6)

5.2.4 Display Frequency on Graph

If the graph is smooth then forgery is detected, otherwise forgery is not detected.

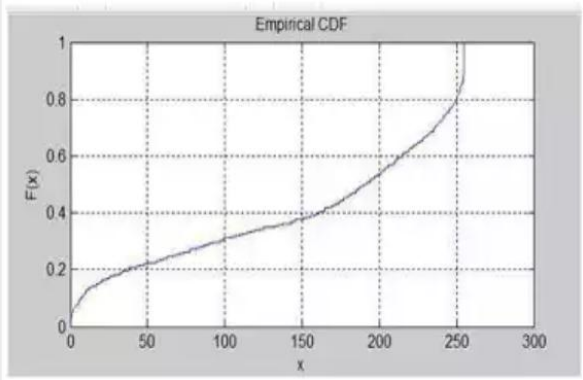

Fig 10: CDF graph

6. RESULT ANALYSIS

The two characteristics *precision* and *recall* are used to evaluate the performance of the proposed forgery detection scheme. *Precision* is the probability that the detected regions are relevant, and it is defined as the ratio of the number of correctly detected forged pixels to the number of totally detected forged pixels. *Recall* is the probability that the relevant regions are detected, and it is defined as the ratio of the number of correctly detected forged pixels to the number of forged pixels in the ground-truth forged image.

$$\textbf{\textit{Precision}} = \frac{|\alpha \cap \alpha'|}{|\alpha|} \qquad (7)$$

$$\textbf{\textit{Recall}} = \frac{|\alpha \cap \alpha'|}{|\alpha'|} \qquad (8)$$

Where, α means the detected forgery regions with the proposed scheme from the dataset, and α' means the ground-truth forgery regions of the dataset.

Table IForgery Detection Results With/Without the Proposed Adaptive Over-Segmentation Algorithm

Host Image	Fixed-size S=150	Fixed-size S=250	Adaptive-size S(11)=199 S(12)=159 S(13)=224
I1 Precision (%)	91.44	91.91	**93.85**
Recall (%)	69.99	69.74	**99.12**
I2 Precision (%)	93.07	93.26	**96.60**
Recall (%)	**90.75**	77.43	78.90
I3 Precision (%)	**96.90**	95.59	95.28
Recall (%)	81.49	89.46	**95.19**

Table I shows the comparison results for the forgery detection with and without the proposed Adaptive Over-Segmentation algorithm. It can be easily observed that for host image I1, the proposed Adaptive Over-segmentation method can produce more accurate forgery detection results with a higher *Precision=93.85%* and, at the same time, gain a much better *Recall=99.12%*; for host image I2, the proposed Adaptive Over-segmentation method can produce more accurate forgery detection results with higher *Precision=96.60%*; and for host image I3, the proposed Adaptive Over-segmentation method can produce more accurate forgery detection results with higher *Recall=95.19%* and, at the same time, maintain good *Precision=95.28%*. The comparison results indicate that the proposed Adaptive Over-Segmentation algorithm can achieve much better forgery detection results than the other forgery detection methods with fixed-size blocks.

Table 2 Cut-Paste Forgery Detection Result

Histogram Equalization graph	Forgery Detection Result
If CDC is smooth	Image is Forged
If CDC is spiky	Image is not forged

7. CONCLUSION AND FUTURE SCOPE

Proposed system, Advanced Techniques for Image Forgery Detection uses the Adaptive Over-Segmentation algorithm to segment the host image into non-overlapping and irregular blocks adaptively according to the given host images; using this approach, for each image, we can determine an appropriate block initial size to enhance the accuracy of the copy-move forgery detection results and, at the same time, reduce the computational expenses. In Cut-paste image forgery detection, proposed digital image forensic techniques capable of detecting global and local contrast enhancement, identifying the use of histogram equalization. Characteristic features of histogram equalization's intrinsic fingerprint were identified and used to propose a scheme for cut and paste forgery detection.

In Future, I would like to implement the same concept on other types of forgery, such as splicing, multiple extension support or other types of media for example, gif images, videos.

8. ACKNOWLEDGMENTS

I take this opportunity to express my profound gratitude to my guide Prof. H. A. Hingoliwala, for his inspiration, guidance, monitoring and constant encouragement which helped me in successful completion of this task in various stages. The blessings, help and guidance given by him time to time shall carry me to long way in the journey of life which I am about to embark.

9. REFERENCES

[1] Soumen Chakrabarti, Martin van den Berg 2, Byron Domc, "Image Forgery Detection Using Adaptive Over-segmentation and Feature Point Matching", IEEE Transactions On Information Forensics And Security, Vol. 10, No. 8, August 2015.

[2] Aditya R Hambarde, Avinash G Keskar,"Copy-move Forgery Detection Using DWT and SIFT Features", proceeding Department of Electronics Engineering Visvesvaraya National Institute of Technology, Nagpur, India 78699.

[3] Mr.Arun Anup M,"Image forgery And Its Detection: A survey (2015)", Department of computer engg and science, MES college of Engineering.

[4] Salam A.Thajeel, Ghazali Sulong,"A Survey Of Copy-Move Forgery Detection Techniques", Journal of Theoretical and Applied Information Technology, 10th December 2014.

[5] Vincent Christlein, " An Evaluation Of Popular Copy-Move Forgery Detection Approaches", Student member IEEE,vol.07.no 6 December 2012.

[6] X.bo.w.Junwen,"Image Copy-Move forgery detection based on SURF", in proc. Int. conf., multimedia inf. Netw.(MINES). Nov.2010.

[7] Jessica Fridich, David Soukal,"Detection of Copy Move Forgery in Digital Image", Department of computer Science, NY 13902-6000.

[8] Hwei-Jen Lin, Chun-Wei Wang, Yang-Ta Kao, ''Fast Copy-Move Forgery Detection", WSEAS Transactions On Signal Processing, Issue 5, Volume 5, May 2009.

[9] Matthew C. Stamm,, "Forensic Detection of Image Manipulation Using Statistical Intrinsic Fingerprints", IEEE Transactions On Information Forensics And Security, Vol. 5, No. 3, September 2010.

[10] Rani Mariya Joseph, Chithra A.S.,"Literature Survey on Image Manipulation Detection" ,International Research Journal of Engineering and Technology (IRJET) ,Volume: 02 Issue: 04,July-2015.