International Journal of Hybrid Information Technology
Vol. 10, No.10 (2017), pp. 21-38
http://dx.doi.org/10.14257/ijhit.2017.10.10.03

Region Duplication Detection Based On Statistical Features of Image

Saba Mushtaq[1]* and Ajaz Hussain Mir[2]

[1, 2]Department of Electronics and Communication Engineering,
National Institute of Technology Srinagar India 190006
[1]sab.mushtaq@gmail.com, [2]ahmir@rediffmail.com

Abstract

Abrupt boom in digital world has led to an instant increase in the popularity of digital images. Easy availability of image tampering tools like Picasa, Adobe Photoshop and Gimp etc. have made image tampering widespread. As such detecting tampering in images has become an active area of research. Region duplication is most common image tampering technique because of the ease with which it can be carried out. Available techniques for region duplication detection fail to accurately locate the tampered region and lack robustness. This paper proposes duplicate region detection method based on statistical texture features using gray level co-occurrence matrix (GLCM) and gray level run length matrix (GLRLM) features. The method divides the forged image into overlapping blocks, calculate texture features based on GLCM and GLRLM of each block. Feature vectors thus obtained for each block are lexicographically sorted. Blocks with similar features are identified using feature distances. Post processing isolates the duplicate regions. Experimental results establish that the proposed method using GLRLM features can precisely locate duplicate regions in image and can effectively withstand the common post processing operation like jpeg compression, blurring, brightness and contrast change with reduced computation complexity.

Keywords*: Copy-move, Texture based features, Digital image forensics, Passive Authentication, block division, GLCM, GLRLM*

1. Introduction

With abrupt boom in the digital world, digital images have become the key source of information. Rampant developments in image tampering tools have made image forgery very easy. Thus authenticating genuineness of images has become mandatory and as such image authentication has become a widely researched area [1,2,3]. Image authentication is categorized into two techniques [4]. One is active authentication which involves insertion of security codes (watermarks and digital signatures) into the images at the time of generation [2] and the second one is passive authentication that aims at verifying the authenticity of images without any prior information about the image [1,2,5]. Passive authentication is more practical as it requires no prior information about the image at hand. Passive authentication come across various tampering practices namely region duplication, splicing, retouching, re-sampling, compression [1]. However, region duplication detection is most common due to the ease with which it is carried out [1, 2, 5, 6]. Region duplication involves copying a region of an image and pasting it somewhere else in the same image. It altogether changes the information conveyed by the image. As the copied region belongs to same image, the dynamic range, color and statistics of the forged region remains unchanged [7]. This makes detection of the tampered region even harder. An example of copy- move forgery is shown in Figure 1. It shows an original

Received (May 19, 2017), Review Result (September 28, 2017), Accepted (October 5, 2017)

ISSN: 1738-9968 IJHIT

image and its forged counterpart where a car is copied and pasted at the top right corner of the image.

Figure 1. Left is the Original Image; Right is the Tampered Image [46]

Copy-move forgery detection techniques can be broadly classified into three types [3] as shown in Figure 2. One is the block based detection; second the key-point based detection and third is the brute force detection.

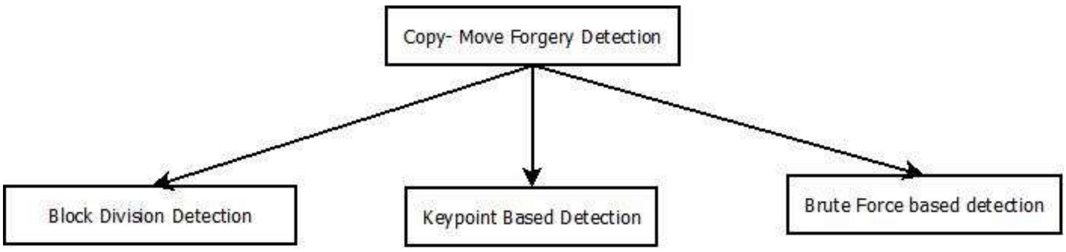

Figure 2. Copy Move Forgery Detection Techniques

The simplest solution to the problem of copy-move forgery is brute force detection which involves comparison of image to every shifted version of itself [3]. The problem with this method is its computational complexity. Autocorrelation method is an improvement over exhaustive search technique. However, it can be applied only when large image patches have been copy pasted [8].

The key point based method depends on extraction of important key points like corners, edges, blobs in the image. Scale invariant feature transform (SIFT)[9,10,11,12] and speed up robust features (SURF) [13-15] have been widely used to extract key points in image . SIFT techniques are highly robust against post processing and intermediate operations [16]. However, they are computationally complex and incapable to determine forgeries in areas which are flat due to lack of reliable key points [9]. SURF features where proposed to improve the performance of SIFT [13,14]. SURF reduces the false acceptance rate considerably that too for high resolution images but lacks in detection if the copy pasted area is very small [13]. Recently Harris corner detectors have also been employed for key point extraction to improve performance of SIFT features [16-18]. The problem with key point based approach is that they do not give the exact shape and location of the forged region [6].

The block based approach relies on dividing image into blocks either overlapping or non-overlapping [8, 19-25]. Features are extracted from each block and compared against each other to locate matching blocks to identify copy- pasted regions. The advantage with the block based method over key point based method is that these give the exact extent and shapes of the copied areas [8, 19,20,21]. Moreover, if the forged area exhibits certain structure it may be entirely missed by key point based method [19]. However, the main problem with the block based methods is the computation time required in the block matching step [26]. The computation time is based on the number of blocks in which the image is divided, the number of features that represent each block and the sorting algorithm used [26].

This paper focuses on block based method for duplicate region detection. We have proposed use of texture features of image blocks to isolate the forgery. We have tried to reduce the computation time by using feature vector having fewer numbers of features in comparison to the available techniques. The next section gives the related work in the field of block based copy-move detection techniques.

2. Related Work

A number of methods have been developed for block based copy-move forgery detection. The preliminary attempt was carried out by Fridrich *et al*. They gave a method that carries out exhaustive search followed by block matching technique based on DCT [8]. This technique gives no account of robustness against post processing techniques like jpeg compression and the feature vector dimension is 64. Popescu and Farid proposed a technique that proposes use of principal component analysis (PCA) as block matching technique [19]. This method is considered to be efficient as features used are half (32) of that used by Fridrich [8]. Haung *et al.*, proposed a scheme based on DCT features and reduced the feature vector by truncating high frequency of coefficients to only 16 features for each block [20]. This results in better performance other than for small copy pasted areas. Cao *et al.*, [21] proposed a technique which is also based on DCT using 4 features per block. However, this technique is based on DCT of circular blocks instead of square blocks. Mahdian *et al.*, proposed use of blur invariant moments which demonstrated their usefulness against preprocessing operations like blur degradation, additive noise and random changes in contrast[22].The number of features generated is 24 for gray scale image blocks and 72 for RGB image blocks. Zhang et al presented a method that uses low frequency sub bands from DWT exhibiting low computational complexity but with a drawback of dependability of speed on location of copy-pasted region[23]. Muhammad *et al.*, proposed a technique that decomposes the image into approximate LL sub band and detail HH sub band from the DyWT technique and make comparison of these sub bands [24]. Though this method is tested for rotation and jpeg compression and performs better than DWT in shift invariance but it is not tested for any post processing operation.

All the aforementioned methods for block based copy- move detection generate very high number of feature vectors resulting in high time complexity [27]. A slight decrease in the number of features in the feature vector results in a considerable improvement in speed and reduction in computational complexity [25]. This limitation has been a motivation to explore a new feature extraction technique that reduces the number of features representing each block and is robust against post processing operations. A step in this direction, we have made an attempt to explore the texture based technique GLCM and GLRLM. This paper proposes a block division method based on texture features using GLCM and GLRLM. A number of evaluations conducted on realistic image show the effectiveness of our method. The proposed method locates the copy-pasted regions appropriately and exhibit robustness against post processing operations like compression, blurring, brightness change and contrast changes. The next section introduces the texture methods GLCM and GLRLM and features extracted from them.

3. Texture Analysis

Texture is a measure of surface roughness, coarseness, and regularity [28]. Texture analysis is achieved by performing numerical manipulation of digitized images to get quantitative measurements. Texture analysis techniques have been widely explored in medical imaging [28,29], signature verification[40,] steganalysis[30] and to reasonable extent in the field of forgery detection[31]. In all these applications, texture analysis has revealed the information which otherwise is not perceptible to the naked eye. This very concept is applicable to image forgery detection as well, since forged part is concealed from naked eye. As such texture analysis is one potential technique that can be applied for forgery detection. There are a number of texture analysis techniques available like model based, structural, transform based and statistical [32,33]. However, usefulness of each technique depends on the area of application. Structural methods may not be useful in forgery detection since the forged image region will not have a regular shape. In the proposed method we have established the effectiveness of statistical texture features since, statistical features have proven superior than Fourier transform based methods [34,35] for the purpose of image classification. Specifically, gray level co-occurrence method (GLCM) and gray level run length method (GLRLM) are used to establish use of texture for copy-move detection. Brief description of the texture analysis techniques is given below.

3.1. Gray Level Co-occurrence Matrix

GLCM is a second order statistical texture feature extraction technique which was proposed by Harlick et al [36]. GLCM defines the frequency of one gray tone appearing in a specified spatial linear relationship with another gray tone, within the region of interest [37]. GLCM has been widely used in applications like CT image analysis [28,29] ink type analysis [38], face recognition [39], signature verification [40], brain tumor classification [41] and steganalysis [42]. GLCM for a given image I(x,y) of size M*N having G_t as total distinct gray levels illustrates the number of times a pixel I at position (x,y) occur in accordance with pixel j at position (x+ Δx, y+ Δy). This frequency of occurrence is denoted by A(i,j,d,θ) and is mathematically expressed as

$$A(i,j,d,\theta) = \sum_{x=1}^{M}\sum_{y=1}^{N}\begin{Bmatrix} 1, if\ I(x,y) = i\ and\ I(x+\Delta x, y+\Delta y) = j \\ 0\ otherwise \end{Bmatrix} \tag{1}$$

Where'd' specify the offset distance Δx,Δy between the pixel and its neighbor and θ represents the direction. GLCM is often taken in four directions (0̊, 45̊, 90̊, 135̊). In the proposed technique GLCM in other directions (180̊, 225̊, 270̊,315̊) is not calculated as it adds no significant texture information about the image. Harlick[36] proposed fourteen GLCM features to extract valuable information from image out of which nine have been used in our technique. These are correlation, homogeneity, entropy, energy, contrast, angular second moment, sum Entropy, difference entropy and variance. These features are defined as under:

$$\text{Correlation} = \sum_{i=1}^{G_t}\sum_{j=1}^{G_t}\frac{(1-\mu_i)}{\sigma_x\sigma_y} \tag{2}$$

$$\text{Homogeneity} = \sum_{i=1}^{G_t}\sum_{j=1}^{G_t}\frac{A(i,j)}{1+|i-j|} \tag{3}$$

$$\text{Entropy} = -\sum_{i=1}^{Gt}\sum_{j=1}^{Gt}A(i,j)Log(A(i,j)) \tag{4}$$

$$\text{Energy} = \sum_{i=1}^{Gt}\sum_{j=1}^{Gt}A^2(i,j) \tag{5}$$

$$\text{Contrast} = \sum_{n=0}^{G_t-1} n^2 \sum_{l=1}^{Gt} \sum_{j=1}^{Gt} A(i,j)\ ,\ \ |i-j| = n \tag{6}$$

$$\text{Angular Second Moment (ASM)} = \sum_{i=0}^{Gt-1} \sum_{j=0}^{Gt-1}, [A(i,j)]^2 \tag{7}$$

$$\text{Sum Entropy} = -\sum_{i=0}^{2G_t-2} A_{(x+y)}(i) Log(A_{(x+y)}(i)) \tag{8}$$

$$\text{Difference Entropy} = -\sum_{i=0}^{G_t-1} A_{(x+y)}(i) Log(A_{(x+y)}(i) \tag{9}$$

$$\text{Variance} = \sum_{i=0}^{G_t-1} \sum_{j=0}^{G_t-1} (i-\mu)^2 (A(i,j)) \tag{10}$$

μ= mean

3.2. Gray Level Run Length Matrix

Gray-Level Run-Length Matrix Texture is a pattern of grey intensity pixel in a particular direction from a reference pixel in an image [43]. This method examines the image in a given direction for pixels having same gray level intensities. Run length is the number of adjacent pixels that have the same grey intensity in a particular direction. GLRLM is defined by intensity of runs (i) , length of runs (l) and run direction (θ) from a reference pixel.

$$\text{GLRLM}(\theta) = (\ S(i,l)\ |\ \theta\ ) \tag{11}$$

S (i, l) is the number of times there is a run of length l having gray level i in direction θ. There are four Run Length Matrix that can be computed for 4 directions of run (θ = 0°, 45°, 90°, 135°). In this paper we have considered GLRLM at 0^0 only since experimentation revealed that considering GLRLM at other angles does not improve the efficacy of the algorithm significantly.
From this GLRLM 7 features are calculated out of which SRE,LRE,GLN,RLN,RP are given by Gallow[44] and HGRE , LGRE by chu et al[45]as:

$$\text{Short Run Emphasis} = \frac{1}{N} \sum_{i=0}^{Gt-1} \sum_{l=1}^{Rt} \frac{S(i,l)|\theta}{l^2} \tag{12}$$

$$\text{Long run emphasis} = \frac{1}{N} \sum_{i=0}^{Gt-1} \sum_{l=1}^{Rt} l^{\,2}\ S(i,l)|\theta \tag{13}$$

$$\text{Gray Level Distribution} = \frac{1}{N} \sum_{i=0}^{Gt-1} [\ \sum_{l=1}^{Gt-1} S(i,l|\theta)]^2 \tag{14}$$

$$\text{Run Length Distribution} = \frac{1}{N} \sum_{i=0}^{Rt} [\ \sum_{l=1}^{Gt-1} S(i,l|\theta)]^2 \tag{15}$$

$$\text{Run Percentage} = \frac{1}{Nt} \sum_{i=0}^{Gt-1} \sum_{l=1}^{Rt} S(i,l)|\ \theta \tag{16}$$

$$\text{Low Gray Level Run Emphasis} = \frac{1}{N} \sum_{i=0}^{Gt-1} \sum_{l=1}^{Rt} \frac{S(i,l)|\theta}{i^2} \tag{17}$$

$$\text{High Gray Level Run Emphasis} = \frac{1}{N} \sum_{i=0}^{Gt-1} \sum_{l=1}^{Rt} i^2\ S(i,l)|\theta \tag{18}$$

Where
Gt is the number of grey levels
Rt is the number of run lengths in the matrix
Nt is the number of points in image.

$$N = \sum_{i=0}^{Gt-1} \sum_{l=1}^{Rt} S(i,l)|\theta$$

4. Proposed Scheme

This paper proposes a block based method. Test image is divided into overlapping blocks Texture features are extracted from each overlapping block. These features are compared against each other to find matching block pairs. The proposed scheme is diagrammatically shown in Figure3 and the steps are subsequently explained.

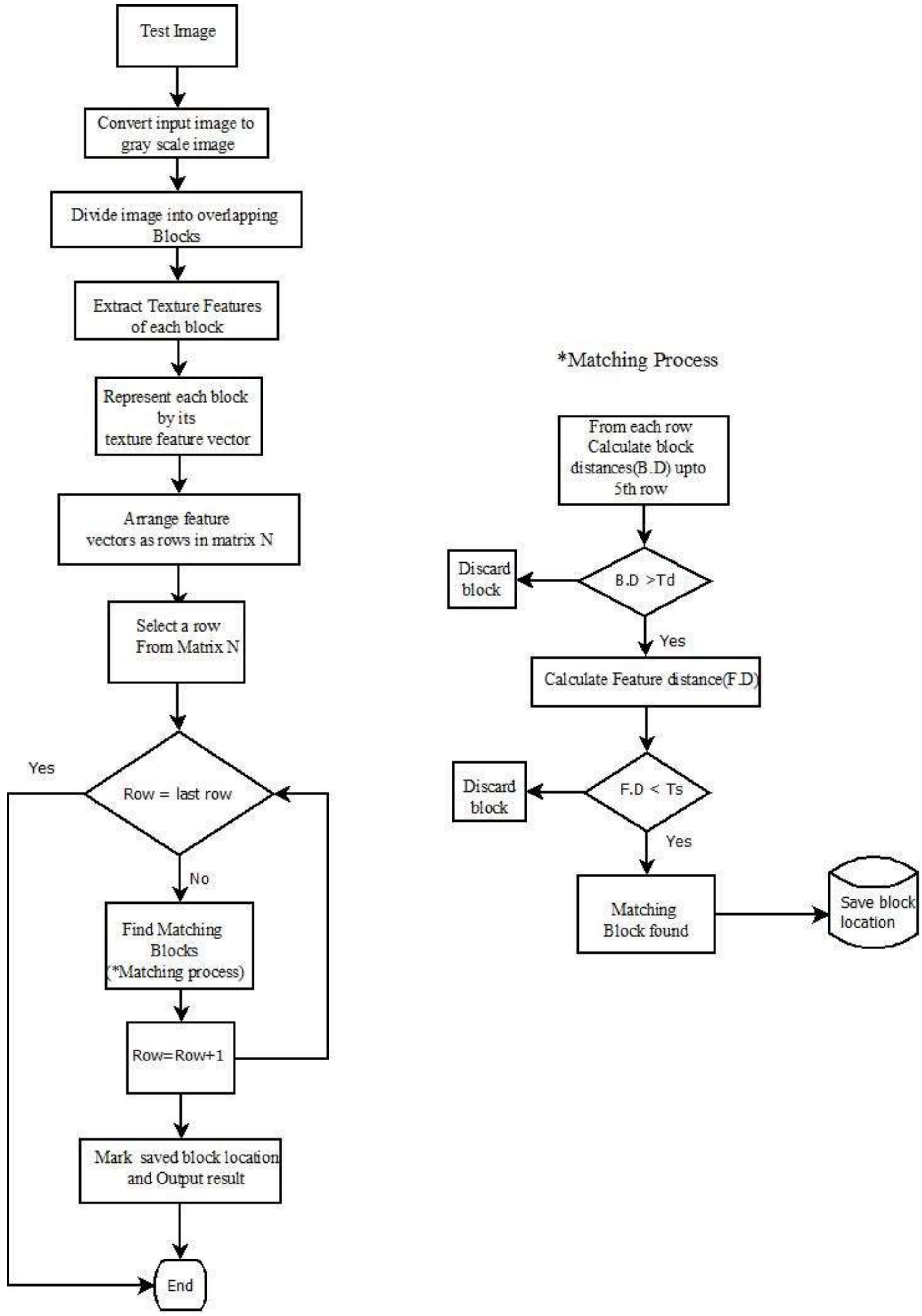

Figure 3. Proposed Algorithm for Copy-move Detection

4.1. Preprocessing the Input Image

The input color image is transformed into gray scale image using the standard formula

$$Im = 0.299R + 0.587G + 0.114B \tag{19}$$

Where R,G,B represent channels of input color image and Im is the luminance component. This formula reveals the information that human eye is more sensitive to certain color wavelengths than other colors which in turn causes a change in the perceived brightness of a given color. Moreover, in YCbCr space luminance component contains maximum spatial information than any other space.

4.2. Dividing Image into Overlapping Blocks

Image Im of size M×N is divided into fixed size square over lapping blocks of size A×A such that two consecutive blocks differ in a row or a column only. Thus image is divided into (M-A+1)(N-A+1) blocks.

4.3 Extracting Texture Features of the Blocks to Represent Each Block by a Feature Vector

4.3.1 GLRLM features: Of each overlapping block GLRLM is calculated. GLRLM features are extracted for each block. GLRLM feature vector obtained for each block represents the block. A block X is represented by 1×7 GLRLM feature vector as $X_1 = (a_1, a_2, a_3, a_4, a_5, a_6, a_7)$. Thus if the image is divided into (M-A+1)(N-A+1) blocks there will be as many feature vectors.

4.3.2 GLCM features: Similarly GLCM is calculated for each image block. GLCM features are extracted for each block forming a feature vector of size 1×9. This feature vector in turn represents the block as $Y_1= (b_1, b_2, b_3, b_4, b_5, b_6, b_7, b_8, b_9)$

4.4 Matching of Texture Features to Locate Similar Blocks

Same matching procedure is adopted for both feature extraction techniques. After feature extraction, feature vectors representing each block are arranged in a matrix. This matrix represents the image blocks arranged as rows. GLRLM features are stored in a matrix P of order (M-A+1)(N-A+1)×7 as given in equation (20).

$$P= \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_{(M-A+1)(N-A+1)} \end{bmatrix} \tag{20}$$

Likewise GLCM features are arranged in matrix Q of order (M-A+1)(N-A+1)×9 as given in equation (21)

$$Q= \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_{(M-A+1)(N-A+1)} \end{bmatrix} \tag{21}$$

From equation (20) and (21) the size of the matrices P and Q depend on the size of image(M×N) and the size of the block i.e. A×A. Smaller the fixed size block, better will be the matching accuracy but higher will be the computational complexity. For a large number of blocks, finding similar blocks will be cumbersome if brute force search is used. Thus lexicographical sorting of feature vectors is proposed by many authors [8, 19, 20, 22, 25]. Lexicographical sorting places similar feature vectors in adjacent rows. Sorting reduces the comparison time for matching process. The above matrices are lexicographically sorted into P_L and Q_L respectively. After lexicographical sorting, each row is tested with its neighboring rows for detection of copy-move forgery. In the matrix P_L starting with first row calculate its feature distance D with neighboring rows up to i^{th} row using Euclidean distance formula as :

$$D = \sqrt{(a_{a1} - a_{b1})^2 + (a_{a2} - a_{b2})^2 + \cdots + (a_{a7} - a_{b7})^2} \tag{22}$$

In our proposed method we calculated distance up to 5^{th} row from a given row. Since the blocks are overlapping and the copy pasted regions are non-overlapping we have to calculate distance D between two rows only when their corresponding block distance in image is more than the distance threshold T_d so as to check if the blocks are non-overlapping. Distance threshold (T_d) is set according to length of block 'A'. If two blocks satisfy this condition then they are candidates for copy-move forgery. For this second condition is to be checked i.e. $D < T_s$ Where T_s is similarity threshold. T_s is predefined and set according to image features. If two blocks are found to be matching their positions are saved. Same procedure is repeated for all rows of P_L, matching block positions are saved in matrix α. Similarly matching blocks are found using GLCM features and saved in matrix β.

The matching process thus checks two conditions

i. Feature distance is calculated between blocks only if they are non-overlapping in the original image. This condition is set by $T_{d.}$

ii. Similar blocks are matching if feature distance $D < T_{s.}$

4.5 Post Processing of Results

After matching process, the block locations saved are those of the copied and the pasted block. Thus marking all these locations in the image will isolate the copy-pasted blocks.

5. Results & Comparison

All experimentation is performed on Matlab R2009b installed on a computer with i3 2.3 GHz processor and 4 Gb memory. Images used were from dataset [46]. The dataset has 200 images in small category of size 512×512 pixels and stored in png format. Post processing operations like jpeg compression, blurring, contrast adjustments and brightness changes are also applied to original as well as forged images. The area of copy-move regions varies with images in the database varying from approximately 28× 14 to about 180×180 pixels i.e about 0.15% to 12% of the image size. Performance evaluation is carried out using two measures. True detection accuracy rate 'Ṫ' and false detection rate 'Ḟ' defined as:

$$\dot{T} = \frac{|c_o \cap c_r| + |P_o \cap P_r|}{|c_o| + |P_o|} \tag{23}$$

$$\dot{F} = \frac{|c_r - c_o| + |P_r - P_o|}{|c_r| + |P_r|} \tag{24}$$

Where c_o , P_o are pixels in the copied and pasted regions in the original image to be tested for authenticity and c_r , P_r are the pixels in copied and pasted regions in the result image after detection. || and ∩ represent area of the region and intersection of the two regions respectively. While – means difference of the two regions. Ṫ indicates the performance of the proposed method to correctly detect pixels of copy pasted regions in the image and $\dot{F}$ indicate the pixels that were not copied but were indicated to be copied by the proposed method. Closer the value of $\dot{T}$ to 1 and $\dot{F}$ to 0 higher is the precision of the system to detect copy move forgery. In the following experimentation we selected more than 200 images from CoMoFoD dataset to evaluate the proposed method. The block size was varied between 20×20, 32×32 and 50×50.

Images shown in Figure 4 below give the detection results as obtained with the proposed technique.

a) b) c)

Figure 4. a) Original Images b) Images Forged with Region Duplication c) Detection Output

The proposed technique detect duplicate regions precisely as shown in Figure 4 with all window sizes as long as the window was less than the size of the forged area. Non regular duplicate regions are also detected as shown in Figure 4(c). In order to make convincing forgeries the forged images are often subjected to various procedures like jpeg compression, brightness changes, contrast changes and blurring. Thus it is mandatory that the forgery detection technique be robust against these operations. As such robustness tests are carried out on the proposed technique and the results obtained are discussed as below:

5.1 Jpeg Compression : Most tampared images are stored in JPEG format[47]. Thus its very important that the proposed algorithm is able to handle the distortion caused by jpeg compression. Experimnetation revealed that the method performs well on images with jpeg compression of varying quality factors. The dataset has images saved with jpeg compression of quality factor ranging from Q= 20,30,40,50,60,70,80,90. But irrespective of the quality factor the method works well in detecting the copy-pasted regions. Figure 5 shows the detection results obtained for an image forged and compressed using Jpeg with Q= 60, the proposed algorithm exposes the copy-pasted regions. Table 1 and Table 2 gives the results as obtained using jpeg compressed images using GLRLM and GLCM respectively.

Figure 5. Showing Original Image, Image Forged and Compressed using Jpeg and Detection Result Output

Table 1. Result Obtained for Image with Varying Quality Factor for jpeg Compression using GLRLM

Quality Factor (Q)	Q=30	Q=40	Q=50	Q=60	Q=70	Q=80	Q=90
Ṫ	.62	.78	.90	.92	.94	.95	.96
Ḟ	.36	.20	.17	.15	.12	.08	.03

Table 2. Result Obtained for Image with varying Quality Factor for jpeg Compression using GLCM

Quality Factor (Q)	Q=30	Q=40	Q=50	Q=60	Q=70	Q=80	Q=90
Ṫ	.10	.10	.20	.35	.43	.43	.44
Ḟ	.85	.85	.80	.80	.68	.68	.68

The results indicated by Table 1 and Table 2 show that the values of Ṫ remains close to 0.9 even for images with jpeg quality factor close to 50 when using GLRLM features however the performance of GLCM features is not good.

5.2 Blurring

Images are manipulated with post processing operations like Gaussian blurring to further conceal the forgery. Our method invariably performs well with images which have been subjected to these post processing operations. The dataset [46] has 120 images which are blurred by convolving the forged image by averaging filter of mask size 3×3, 5×5 and 7×7. Images are blurred to a considerable limit due to application of these averaging masks. Figure 6 below shows original image, image forged and blurred with averaging mask of size 5×5 and the detection result. Clearly the result image isolates the forged regions.

Figure 6. Showing Original Image, Forged Image which has been Blurred by 5×5 Mask and Detection Results

Table 3 and table 4 below presents the performance of the proposed techniques on blurred images. The algorithm works well on image blurred with averaging mask of size 3×3. However, performance reduces with increased mask size for GLRLM features. Results obtained using GLCM deteriorates with post processing operations.

Table 3. Results Obtained for Images blurred with varying Mask and Block Sizes using GLRLM

Block size	20×20			34×34			50×50		
Mask size	3×3	5×5	7×7	3×3	5×5	7×7	3×3	5×5	7×7
Ṫ	.97	.95	.93	.96	.94	.91	.89	.88	.84
Ḟ	.025	.054	.079	.028	.059	.081	.031	.060	.089

Table 4. Results Obtained for Images blurred with Varying Mask and Block Sizes using GLCM

Block size	20×20			34×34			50×50		
Mask size	3×3	5×5	7×7	3×3	5×5	7×7	3×3	5×5	7×7
Ṫ	.20	.20	.18	.18	.18	.18	.18	.18	.18
Ḟ	.80	.82	.82	.84	.84	.84	.90	.90	.90

The detection results obtained for blurred image shown in Table3 and Table 4 establish the fact that the performance degrades for larger block sizes but is still reliable for GLRLM features.

5.3 Contrast and Brightness Changes

Images are often subjected to contrast adjustments and brightness changes to further conceal the forgery. In the dataset [46] 120 images are post processed by contrast change by mapping the intensity values to interval [0 1] into three ranges with lower bound and upper bound values as [(0.01, 0.95), (0.01, 0.9), (0.01, 0.8)]. The changes in contrast by mapping to range [0.01 .8] cause a significant change to the image however the other two range values bring an insignificant change. The proposed approach works well on images which undergo significant contrast change.

From the same dataset 120 images with brightness altered are tested for forgery. The brightness values are also mapped to three ranges as [(0.01, 0.95), (0.01, 0.9), (0.01, 0.8)]. Noticeable change due to range (.01 .08) are incurred which upon being tested by the proposed method yield good results. Figure 7 below give the detection result for forged images subjected to contrast adjustments and brightness changes.

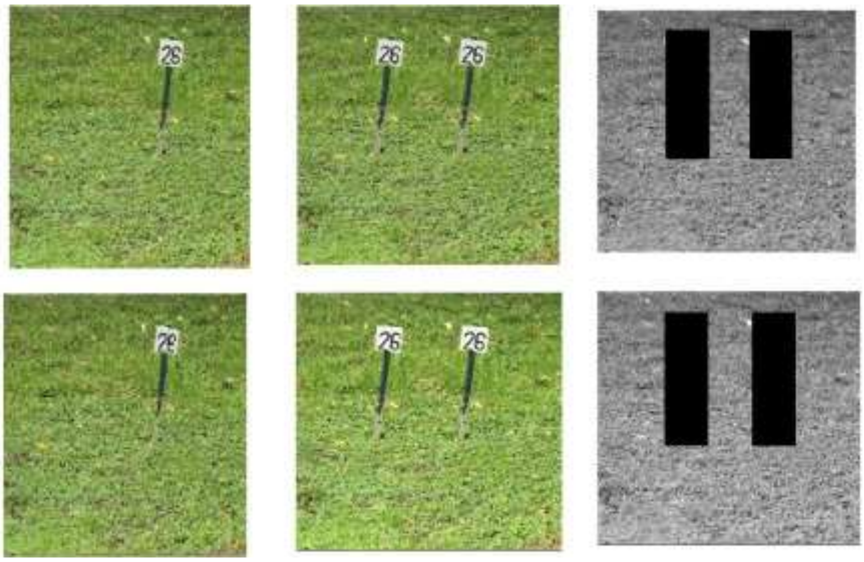

Figure 7. Shows Original Image, Images Forged and Subjected to Contrast Adjustment and Brightness Changes Respectively and the Detection Results

Table 5 and Table 6 present results achieved by the proposed methods on images subjected to contrast adjustment and brightness changes. Values of Ṫ remain close to 0.9 signifying efficacy of the method.

Table 5. Results for Image Subjected to Contrast Change and Brightness Changes using GLRLM

Forgery	Range	Ṫ	Ḟ
Contrast Adjustments	0.01-.95	.98	.02
	0.01- .80	.95	.05
	0.01-.90	.97	.027
Brightness Change	0.01-.95	.99	.010
	0.01- .80	.98	.029
	0.01-.90	.99	.018

Table 6. Results for Image Subjected to Contrast Change and Brightness Changes using GLCM

Forgery	Range	Ṫ	Ḟ
Contrast Adjustments	0.01-.95	.15	.85
	0.01- .80	.15	.85
	0.01-.90	.15	.85
Brightness Change	0.01-.95	.20	.88
	0.01- .80	.20	.88
	0.01-.90	.20	.88

6. Comparison

The results obtained with the proposed method using GLRLM and GLCM are compared to the techniques based on DCT [8], PCA [19] and Gabor magnitude [25]. The comparison is made on the basis of True detection accuracy (Ṫ) and false detection rate (Ḟ). Figure 8, Figure 9 and Figure 10 gives performance comparison for the proposed techniques with the state of art techniques [8,19,25].

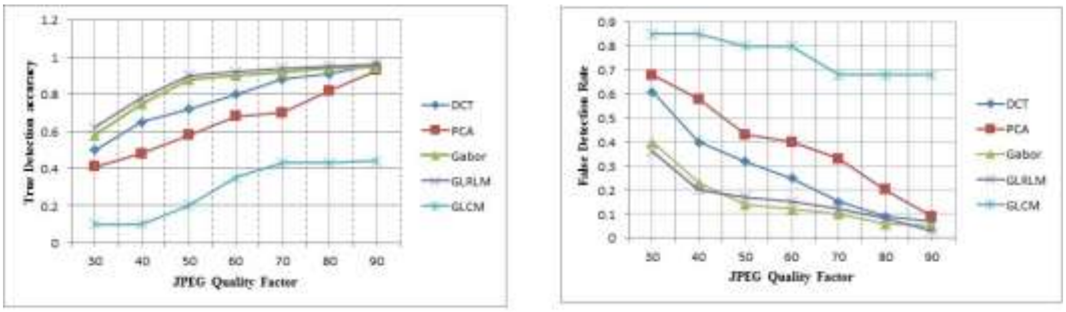

Figure 8. Performance Comparison for DCT, PCA, Gabor Magnitude and Proposed Methods for jpeg Compressed Images with Different Quality Factors

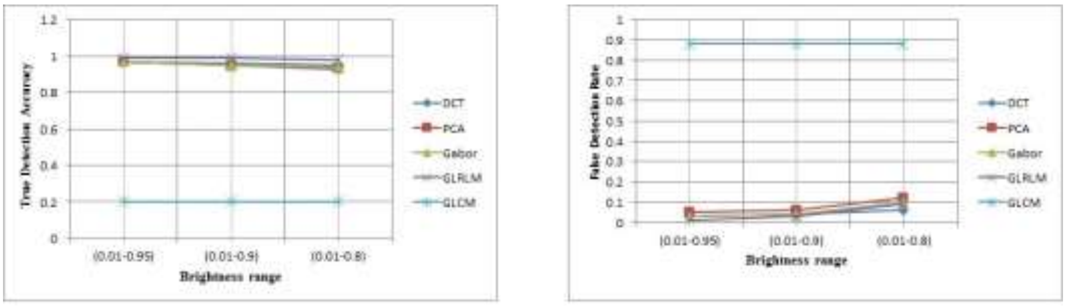

Figure 9. Performance Comparison for DCT, PCA, Gabor Magnitude and Proposed Methods for Brightness Altered Images in Different Ranges of Brightness Changes

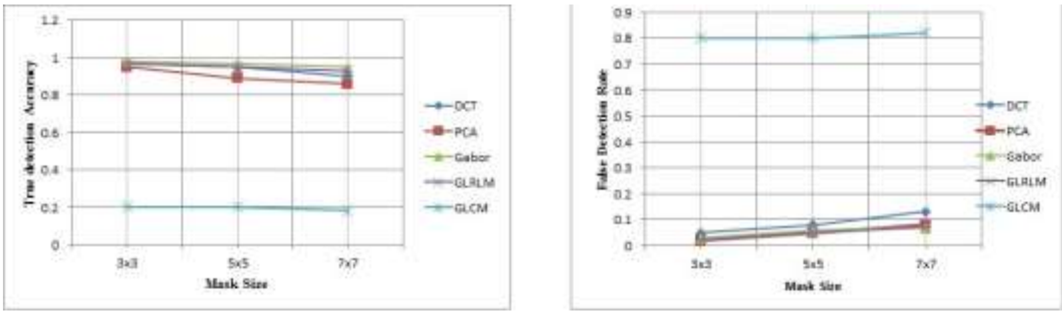

Figure 10. Performance Comparison for DCT, PCA, Gabor Magnitude and Proposed Method for Images Blurred with Different Mask Sizes

Figure 8, Figure 9 and Figure 10 indicate that the proposed method performs well in comparison to methods [8, 19, 25]. The true detection accuracy rate curve for GLRLM features remains close to unity while false detection rate curve remains close to zero. The performance comparison reveals that the GLCM method performs poorly while GLRLM perform well even with post processing operations. Thus establishing the usefulness of the proposed method even in presence of post processing operation carried out on forged images.

7. Discussion

The proposed techniques proves to be robust against post processing operation. However the inbuilt drawback with block matching algorithms is the computation complexity which arises due to the sorting and matching of blocks. In the proposed technique lexicographical sorting is used. Sorting the rows in the feature vector matrix requires comparing the rows with each other. This in turn depends on the number of rows i.e. the blocks of image and within each row the number of features which are to be compared. Both these affect the computation complexity. The best case complexity with lexicographical sorting is $O(\tau\omega \log_2 \omega)$ where τ is number of features in the feature vector and ω is number of blocks. Equation (11) calculates the feature distance to perform matching process. This is repeated for all rows of the matrix. Thus our proposed method requires$O(2\tau\omega\varepsilon)$, where ε is the number of ith rows upto which feature distance comparison is made. Thus total complexity for the proposed technique is the sum of these two.

$$\text{Computation Complexity} = O(\tau\omega \log_2 \omega) + O(2\tau\omega\varepsilon) \tag{25}$$

It is worth mentioning that number of blocks depends on image size M×N. Large size images will generate more number of blocks and will add to computation complexity by a factor of $O(M \times N)$. The proposed approach has $\tau = 7$ in case of GLRLM which is considerably less than for methods based on DCT [8,20], PCA [19] , Blur invariant moment[22] and Gabor Magnitude[25] indicated in Table 7 and thus as per equation (25) computation complexity of the proposed method is less than these methods.

Table 7. Feature Dimensionality Comparison

Method/Technique		Sorting Algorithm	Feature Vector Dimension
Fridrich[8]	DCT	Lexicographical Sort	64
Popescu[19]	PCA	Lexicographical Sort	32
Mahdian[22]	Blur Invariant moment	Lexicographical Sort	24

Haung[20]	DCT	Lexicographical Sort	16
Lee[25]	Gabor Magnitude	Lexicographical Sort	12
Proposed	GLCM	Lexicographical Sort	**9**
	GLRLM		**7**

The results obtained using GLRLM and GLCM features establish GLRLM may be suitably used for duplicate region detection. However, GLCM results are not promising.

8. Conclusion

Duplicate region detection in image is a hot topic of research. We have proposed a robust method for the detection of this forgery using GLRLM and GLCM texture features. The experimental results demonstrate the usefulness of GLRLM features in precisely detecting duplicate regions in an image besides being robust against various post processing operations like jpeg compression, blurring, color adjustment and contrast adjustment operations. The proposed feature extraction technique has lesser number of features as compared to most existing techniques making the techniques computationally less complex. However, image forgeries can still be concealed using rotation and scaling operations making detection much more difficult. This technique can thence be used in the field of forensics for verifying the authenticity of images against copy-move forgery and future work will address the issues of rotation and scaling.

Acknowledgements

We thank Information Security Education and Awareness (ISEA) projects for funding this research that we carried out in NIT Srinagar.

References

[1] O.M. Al-Qershi and B.E. Khoo, “Passive detection of copy-move forgery in digital images: State-of-the-art”, Forensic science international, vol. 231, no. 1, **(2013)**, pp.284-295.
[2] S. Mushtaq, and A.H. Mir, “Digital image forgeries and passive image authentication techniques: A survey”, International Journal of Advanced Science and Technology, vol. 73, **(2014)**, pp.15-32.
[3] S. Bayram, H.T. Sencar and N. Memon, “A survey of copy-move forgery detection techniques”, IEEE Western New York Image Processing Workshop, **(2008)**, pp. 538-542.
[4] A. Haouzia and R. Noumeir, “Methods for image authentication: a survey”, Multimedia tools and applications, vol. 39, no. 1, **(2008)**, pp.1-46.
[5] M.F. Hashmi, V. Anand and A.G. Keskar, “Copy-move image forgery detection using and efficient and robust method combining un-decimated wavelet transform and scale invariant feature transform”, AASRI Procedia, vol. 9, **(2014)**, pp.84-91.
[6] J. Zhao, and J. Guo, “Passive forensics for copy-move image forgery using a method based on DCT and SVD”, Forensic science international, vol. 233, no. 1, **(2013)**, pp.158-166
[7] G.K. Birajdar and V.H. Mankar, “Digital image forgery detection using passive techniques: A survey”, Digital Investigation, vol. 10, no. 3, **(2013)**, pp.226-245.
[8] A.J. Fridrich, B.D. Soukal, and A.J. Lukáš, “Detection of copy-move forgery in digital images”, Proceedings of Digital Forensic Research Workshop, **(2003)**.
[9] I. Amerini, L. Ballan, R. Caldelli, A. Del Bimbo and G. Serra, “A sift-based forensic method for copy–move attack detection and transformation recovery”, IEEE Transactions on Information Forensics and Security, **(2011)**, vol. 6, no. 3, pp.1099-1110.
[10] H. Huang, W. Guo and Y. Zhang, “Detection of copy-move forgery in digital images using SIFT algorithm”, Computational Intelligence and Industrial Application vol. 2, **(2008)**, pp. 272-276.
[11] M. Jaberi, G. Bebis, M. Hussain and G. Muhammad, “Improving the detection and localization of duplicated regions in copy-move image forgery. In Digital Signal Processing (DSP)”, 2013 18th International Conference, **(2013)**, pp. 1-6.
[12] E. Ardizzone, A. Bruno and G. Mazzola, “Detecting multiple copies in tampered images”, Image Processing (ICIP), **(2010)**, pp. 2117-2120.
[13] X. Bo, W. Junwen, L. Guangjie and D. Yuewei, “Image copy-move forgery detection based on SURF. In Multimedia information networking and security (MINES)”, **(2010)**, pp. 889-892.

[14] Shivakumar, B.L. and Baboo, L.D.S.S., 2011. Detection of region duplication forgery in digital images using SURF. IJCSI International Journal of Computer Science Issues, 8(4).
[15] Mishra, P., Mishra, N., Sharma, S. and Patel, R., 2013. Region duplication forgery detection technique based on SURF and HAC. The Scientific World Journal, 2013.
[16] Kakar, P. and Sudha, N., 2012. Exposing postprocessed copy–paste forgeries through transform-invariant features. IEEE Transactions on Information Forensics and Security, 7(3), pp.1018-1028.
[17] Chen, L., Lu, W., Ni, J., Sun, W. and Huang, J., 2013. Region duplication detection based on Harris corner points and step sector statistics. Journal of Visual Communication and Image Representation, 24(3), pp.244-254.
[18] Yu, L., Han, Q. and Niu, X., 2016. Feature point-based copy-move forgery detection: covering the non-textured areas. Multimedia Tools and Applications, 75(2), pp.1159-1176.
[19] Popescu, A.C. and Farid, H., 2004. Exposing digital forgeries by detecting duplicated image regions. Department Computer Science, Dartmouth College, Technology Report TR2004-515.
[20] Huang, Y., Lu, W., Sun, W. and Long, D., 2011. Improved DCT-based detection of copy-move forgery in images. Forensic science international, 206(1), pp.178-184.
[21] Cao, Y., Gao, T., Fan, L. and Yang, Q., 2012. A robust detection algorithm for copy-move forgery in digital images. Forensic science international, 214(1), pp.33-43.
[22] Mahdian, B. and Saic, S., 2007. Detection of copy–move forgery using a method based on blur moment invariants. Forensic science international, 171(2), pp.180-189.
[23] Zhang, J., Feng, Z. and Su, Y., 2008, November. A new approach for detecting copy-move forgery in digital images. In Communication Systems, 2008. ICCS 2008. 11th IEEE Singapore International Conference on (pp. 362-366). IEEE.
[24] Muhammad, G., Hussain, M. and Bebis, G., 2012. Passive copy move image forgery detection using undecimated dyadic wavelet transform. Digital Investigation, 9(1), pp.49-57.
[25] Lee, J.C., 2015. Copy-move image forgery detection based on Gabor magnitude. Journal of Visual Communication and Image Representation, 31, pp.320-334.
[26] Sekeh, M.A., Maarof, M.A., Rohani, M.F. and Motiei, M., 2011, July. Efficient image block matching algorithm with two layer feature extraction. In Information Technology in Asia (CITA 11), 2011 7th International Conference on (pp. 1-5). IEEE.
[27] Sekeh, M.A., Maarof, M.A., Rohani, M.F. and Mahdian, B., 2013. Efficient image duplicated region detection model using sequential block clustering. Digital Investigation, 10(1), pp.73-84.
[28] Mir, A.H., Hanmandlu, M. and Tandon, S.N., 1995. Texture analysis of CT images. IEEE Engineering in Medicine and Biology Magazine, 14(6), pp.781-786.
[29] Mougiakakou, S.G., Valavanis, I.K., Nikita, A. and Nikita, K.S., 2007. Differential diagnosis of CT focal liver lesions using texture features, feature selection and ensemble driven classifiers. Artificial Intelligence in Medicine, 41(1), pp.25-37.
[30] Shi, Y.Q., Chen, C., Xuan, G. and Su, W., 2007, December. Steganalysis versus splicing detection. In International Workshop on Digital Watermarking (pp. 158-172). Springer Berlin Heidelberg.
[31] Dong, J., Wang, W., Tan, T. and Shi, Y.Q., 2008, November. Run-length and edge statistics based approach for image splicing detection. In International Workshop on Digital Watermarking (pp. 76-87). Springer Berlin Heidelberg.
[32] Russ, J.C., 1999. The image processing handbook. CRC press 3rd Edition.
[33] Bharati, M.H., Liu, J.J. and MacGregor, J.F., 2004. Image texture analysis: methods and comparisons. Chemometrics and intelligent laboratory systems, 72(1), pp.57-71.
[34] Srinivasan, G.N. and Shobha, G., 2008, December. Statistical texture analysis. In Proceedings of world academy of science, engineering and technology (Vol. 36, pp. 1264-1269).
[35] Nailon, W.H., 2010. Texture analysis methods for medical image characterisation. INTECH Open Access Publisher.
[36] Haralick, R.M., 1971, December. On a texture-context feature extraction algorithm for remotely sensed imagery. In Decision and Control, 1971 IEEE Conference on (Vol. 10, pp. 650-657). IEEE.
[37] Baraldi, A. and Parmiggiani, F., 1995. An investigation of the textural characteristics associated with gray level cooccurrence matrix statistical parameters. IEEE Transactions on Geoscience and Remote Sensing, 33(2), pp.293-304.
[38] Franke, K., Bunnemeyer, O. and Sy, T., 2002. Ink texture analysis for writer identification. In Frontiers in Handwriting Recognition, 2002. Proceedings. Eighth International Workshop on (pp. 268-273). IEEE.
[39] Eleyan, A. and Demirel, H., 2011. Co-occurrence matrix and its statistical features as a new approach for face recognition. Turkish Journal of Electrical Engineering & Computer Sciences, 19(1), pp.97-107.
[40] Ferrer, M.A., Vargas, J.F., Morales, A. and Ordóñez, A., 2012. Robustness of offline signature verification based on gray level features. IEEE Transactions on Information Forensics and Security, 7(3), pp.966-977.
[41] Zulpe, N. and Pawar, V., 2012. GLCM textural features for brain tumor classification. Int. J. Comput. Sci, 9(3), pp.354-359.
[42] Chhikara, R.R. and Bansal, D., 2014, September. GLCM based features for steganalysis. In Confluence The Next Generation Information Technology Summit (Confluence), 2014 5th International Conference- (pp. 385-390). IEEE.

[43] Tang, X., 1998. Texture information in run-length matrices. IEEE transactions on image processing, 7(11), pp.1602-1609.
[44] Galloway, M.M., 1975. Texture analysis using gray level run lengths. Computer graphics and image processing, 4(2), pp.172-179.
[45] Chu, A., Sehgal, C.M. and Greenleaf, J.F., 1990. Use of gray value distribution of run lengths for texture analysis. Pattern Recognition Letters, 11(6), pp.415-419.
[46] Tralic, D., Zupancic, I., Grgic, S. and Grgic, M., 2013, September. CoMoFoD—New database for copy-move forgery detection. In ELMAR, 2013 55th international symposium (pp. 49-54). IEEE.
[47] Popescu, A.C. and Farid, H., 2004, May. Statistical tools for digital forensics. In International Workshop on Information Hiding (pp. 128-147). Springer Berlin Heidelberg.

Authors

Saba Mushtaq received her B.E. degree in Electronics and Communications Engineering from Kashmir University, India in 2008. She obtained her M. Tech. degree in Communication and Information Technology from National Institute of Technology, Srinagar, India in 2012. She joined NIT Srinagar in September 2012, as a faculty member. Presently she is a research scholar at NIT Srinagar in Department of Electronics and Communication.Her research interests are Image Processing and Biometrics. Saba can be reached at sab.mushtaq@gmail.com

Ajaz Hussain Mir has done his B.E in Electrical Engineering with specialization in Electronics & Communication Engineering (ECE) .He did his M.Tech in Computer Technology and Ph.D both from IIT Delhi in the year 1989 and 1996 respectively. He is Chief Investigator of Ministry of Communication and Information Technology, Govt. of India project: Information Security Education and Awareness (ISEA). He has been guiding Ph.D and M.Tech thesis in Security and other related areas and has a number of International publications to his credit Presently he is working as Professor in the Department of Electronics & Communication Engineering at NIT Srinagar, India. His areas of interest are Biometrics; Image processing, Security, Wireless Communication and Networks.