Non-Overlapping Block-based Parametric Forgery Detection Model

Kusam Sharma
Research Scholar
Department of Computer Science & IT,
University of Jammu, J&K, India

Pawanesh Abrol, PhD
Professor
Department of Computer Science & IT,
University of Jammu, J&K, India

ABSTRACT
Modification of a digital image by adding or removing some of its elements using a wide variety of image processing tools results in image forgery. As a result, authentication of originality of a digital image is becoming a challenging task. Copy-paste forgery is one of the forgeries belonging to context-based forgery. Copy-Paste Forgery Detection (CPFD) aims at finding regions that have been copied and pasted within the same or different images. A small change in the image may change statistical parameters that can be analysed for initial assessment of the forgery. In the present research study, a parametric forgery detection model using non-overlapping block-based technique is developed to ascertain the copy-paste forgery in a given digital image. Statistical parameters of the input image are computed, analysed and compared with those of the forged image. The results show that the proposed model identifies the forged area of the given image and works well with low to moderate copy-paste forgery. The results obtained can be used as the initial verification of the images for forgery and to enhance the forgery detection process by identifying most likely cases of possible image forgeries. The proposed model is tested with large domain of images having different dimensions and for detecting forgery within an image. However, the model has limitations with certain geometrical transformations.

General Terms

Keywords
Copy-paste forgery, Block-based forgery detection techniques, Non-overlapping block-based techniques.

1. INTRODUCTION
Digital image forgery is the process of altering the material elements of an image using several pre-existing image processing tools. Digital image forgeries can be context-based, graphical software based or content-based. The context-based forgery is created by varying the context of an image. Context-based image forgery detection technique locates duplicated image regions with in an image or between the two images or among more than two images. The copy-paste forgery is of two types, forgery in one image and digital splicing. Copy-paste forgery in one image belongs to context-based forgery whereas digital splicing can be done with different images [1]. Since, the copied segments come from the same image, all of its properties, metric and topological, and statistical information will be same with the rest of the image. Detection of such forgeries is a very challenging task.

The copy-paste forgery detection techniques are further classified into two categories based on grouping of evaluated feature sets such as key-point based and block based. These are parametric based techniques. The key-point based methods like SIFT, SURF, LPFT, etc rely on the identification and selection of high entropy image regions instead of blocks. In block-based techniques, the original image is divided into overlapping or non-overlapping blocks. The transformations are applied over the block to generate the feature vectors of the image features like statistical features, geometrical features or textural features. Block-based forgery detection techniques can be non-overlapping or overlapping. Some of the non-overlapping block based techniques may include DCT, DWT, PCA, SVD, PAN’s methods. The overlapping block based techniques include DCT, PCA, DWT, improved SVD, DWT-SVD, DWT & PCA-EVD, FMT, LUO’s method etc. Block-based techniques are invariant to various transformations like flipping, brightness changes and blurring.

Block-based forgery detection techniques can be parametric or non-parametric. The parametric block-based forgery detection technique analyses a wide variety of image parameters like statistical, geometrical and textural. The statistical parameters are very significant and manipulation in an image can be detected by analysing the behaviour of these statistical parameters. These techniques can be applied in various fields like image enhancement, image restoration, image denoising and digital image forgery detection etc. Moreover, these techniques also find their application in the field of edge detection and eye gazing [2]. The analysis of these statistical parameters of an image helps in determining and locating the forged region within an image [3].

The present research work is context based and is carried out for developing a non-overlapping block based parametric forgery detection model for locating duplicated image regions with in an image. Understanding of these parameters may help in optimizing different image processing models especially in the field of forgery detection. There can be different statistical parameters for detecting image forgery. Besides mean and standard deviation there are other statistical parameters i.e. variance, skewness and kurtosis. The mean is used in applications such as noise removal or low pass filtering (smoothing) while variance can be used in identifying sharp details such as edges. Mean and variance of an image is used where pixel variation of images belongs to particular class are same. Variance is normally used to find how each pixel varies from the neighboring pixel or centre pixel and is used to classify them into different regions [4][5]. Standard deviation describes the variability of the data. It indicates how much on an average each of the values in the distribution deviates from the mean or centre of the distribution [6]. Skewness and kurtosis are the shape parameters. The skewness and kurtosis characterize the tails of a probability model rather than the...
central portion. As a result of which, any two probability models with same mean, standard deviation, skewness and kurtosis will have similar shapes [7] [8].

The organization of this paper is as follows. Literature review and objective are discussed in the next two subsequent sections II and III. The proposed statistical model is presented in section IV followed by experimental results and discussion in section V. In section VI, inferences and conclusion is discussed and section VII presents limitations and future work.

2. LITERATURE REVIEW

The digital image forgeries can be content based, context based or creation based. Several researchers have adopted different methods for detection of such forgeries. Copy-move forgery being the context based forgery is detected by two broadly classified techniques key-point based and block-based. Key-point based methods basically rely on the identification and selection of high entropy image regions instead of blocks. Some of the widely accepted and used key-point based copy-move forgery detection techniques are discussed below:

A novel approach to detect copy-move forgery which works even in the presence of rotation and scaling that took place before copying is presented by Wu et al. This method is based on LPFT (Log-Polar Fourier Transform). Similarity between the original and forged regions is revealed by comparing cross-spectrum coefficients of the LPFT magnitude spectra [9]. Another method based on an efficient key-point and feature computation algorithm known as Scale Invariant Feature Transform (SIFT) is proposed in [10]. In this method, the image key-points will be located and image features at the detected key-points will be collected. This method is robust to distortions of the duplicated regions. This method fails to detect reliable SIFT key-points in regions with little visual structures. Another automatic and robust forgery detection system based on SURF (Speeded Up Robust Features), which are fast detectors and descriptors and KD-Tree is used to identify the duplicated regions is proposed by Shivakumar et al [11]. This method sometimes fails in successfully detecting the few small sized copied regions. An efficient and reliable passive-blind detection method in which block matching procedures are used which first divides the image into the same size x b blocks is suggested in [12]. In this improved singular value decomposition (improved SVD) is applied to all the image blocks. This algorithm has strong detection as well as anti-noise capabilities. An SVD based sorted neighborhood approach for detection in which an image is decomposed into four sub-bands by applying DWT (Discrete Wavelet Transform) is proposed by Li et al [13]. However, the computation of SVD takes a lot of time and the time complexity of sorting is reduced to O(7k log k) then that discussed in PCA method given by Popescu and Li. Another SVD based method for investigating the extent of noise for detecting forgery in digital images is given in [14]. An improved detection algorithm based on Discrete Wavelet Transform (DWT) and Principal Component Analysis – Eigen Value Decomposition (PCA-EVD) has been presented by Zimba in [15]. Another efficient non-intrusive method for copy-move forgery detection is based on image segmentation and a new denoising algorithm does not require prior knowledge about the camera used to capture the image [16]. An FMT based method is used to extract image feature from image blocks of size 16 to reduce the detection time [17]. Counting Bloom Filters are used instead of lexicographical sorting to improve the efficiency of detecting duplicated regions. The complexity is O(length (MN)). Lin et al. suggested a method in which radix sort is used for sorting the feature vectors of the divided overlapping sub-blocks as an alternative to lexicographic sorting, which is commonly used by the existing copy-move forgery detection schemes. The medium filtering and connected component analysis are performed on the tentative detected result to obtain the final result. Even though the proposed technique reduced the time complexity to O(k log k) with the help of radix sort, the method fails to detect all copied region of small size. This scheme performs well when the degree of rotation is 90°, 180° and 270° [18].

Another forgery detection method involves block-based copy-move forgery detection techniques which are further classified into non-overlapping and overlapping based techniques. Some of the recently developed and used non-overlapping and overlapping based forgery detection techniques are discussed below:

A passive forensic method for detecting copy-move attacks based on DCT and DWT is suggested by Wang et al. The test image is segmented into non-overlapping 8x8 blocks. DCT and DWT are applied to each image block to extract features and then compare the statistical parameters of each image block to detect replicated regions. The time complexity of their method for sorting is further reduced to O(k log k) then that given in method proposed by Luo, Li and Popescu. This method is robust to compression up to JPEG quality level 20 and against additive noise [19]. Another improved DCT based method for detection of copy-paste forgery in digital images is given by Huang et al. This method is capable of detecting the duplicated regions even in the presence of blurring, additive white Gaussian noise and JPEG compression. A new two-phase detection method to effectively locate image forgeries based on inconsistency of noise levels in different regions of the image is given in [21]. In this method, the image is first segmented into non-overlapping image blocks for initial noise estimation using an effective noise estimation method. The kurtosis of the original natural image and the variance of the added noise are computed for forgery detection. It not only improves the detection accuracy but reduces the computation complexity and detects image forgeries both quantitatively and qualitatively. This method works only for grayscale images. The advantage of this method is that no previous information about the original image and the imaging device is required. Another segmentation method based on inconsistencies of noise for digital image forgery detection is given by Mahdian et al [22]. In this method noise standard deviation of each block is estimated using median-based method.

Some of the overlapping block-based feature extraction approaches for forgery detection includes a direct approach of applying an exhaustive search as proposed by Fridrich et al. This approach is computationally very expensive. It is simple and effective for small sized images. The second approach is of block-matching by using quantized DCT coefficients where the image is divided into overlapping blocks. The features of these blocks are extracted and then these blocks are vectorized and sorted lexicographically for further detection. This approach is robust to retouching operations. The computational time depends upon factors such as number of blocks, sorting techniques and the number of features [23]. Another DCT-SVD based method used for detection of copy-move forgery based on overlapping and non-overlapping technique is given by Zhao et al [24]. This method is quite efficient in detecting duplicated regions even in the presence...
of Gaussian blurring, AWGN, JPEG compression. Later on, Luo et al. proposed a new method based on pixel block characteristics. This approach firstly divides an image into small overlapped blocks, and then the similarity of these blocks is compared for identifying duplicated regions. The time complexity of their method for sorting is further reduced to $O(7k \lg k)$ then that given by Li and Popescu. This method is robust to compression up to JPEG quality level 30 and Gaussian blurring and additive noise with SNR 24 dB [25]. A block-based approach which exploits texture as feature to be extracted from blocks is given in [26]. This method is tested on both JPEG compressed and un-compressed images. Another DWT based technique to get a reduced dimension representation is robust to common post-processing operations and has lower computational complexity [27]. A novel image hashing method used for image forgery detection which employs both the local and global features of an image is given by Sundaram et al. The local features include position and texture features of each significant region of the image whereas the global features based on the complex Zernike moments represent the chrominance and luminance characteristics of the image [28].

There are certain widely accepted, latest and authenticated parametric and non-parametric copy-move forgery detection approaches.

An efficient technique for detecting and localizing duplicated regions in an image by first applying a Principal Component Analysis (PCA) on small fixed-size image blocks to get a reduced dimension DCT block representation is given by Popescu et al. The time complexity of sorting is $O(32k \lg k)$ [29]. Dattatherya et al. has given an image authentication method for grey and coloured images having different dimensions and formats for hybridization of colour histogram. It is associated with first four statistical moments i.e. mean, standard deviation, skewness and kurtosis to achieve the objectives of low cost and high speed. The authentication code is used to analyze the characteristics of received image from tampering point of view [30]. The technique proposed by Zhang et al. works by applying DWT (Discrete Wavelet Transform) to the input image to get a reduced dimension representation [31]. For the most common blur operations of composite forged image, a localization approach of forged region based on detection of image edge is proposed [32]. This technique can accurately detect the blur operation traces of composite forged images and can precisely locate the forged region. But the high complexity of the texture and the low robustness of edge detection algorithm may cause some false detection.

Several copy-paste digital image forgery detection approaches are existing at present and can be distinguished on the basis of their domain of working. Key-point based techniques include SIFT, SURF, LPFT, FMT, etc rely on the identification and selection of high entropy image regions instead of blocks. Block-based techniques include non-overlapping and overlapping techniques which include DCT, DWT, PCA, etc. In non-overlapping, there are several parametric and non-parametric techniques for forgery detection. Parametric techniques are simple and efficient in detecting copy-paste forgery. In this research work a parametric non-overlapping block-based copy-paste technique is developed for forgery detection. The objectives of the present research study are discussed in the next section.

3. OBJECTIVES

It is evident from the literature review that the analysis of image parameters may indicate some important aspects of image region duplication, tampering or forgery within an image. Further non-overlapping block-based parametric image analysis is one of the significant techniques that are used for image analysis. Based on these observations this study has been conducted with the following research objectives:

- Study of forgery detection algorithms for digital images.
- Understanding copy-paste forgery detection models and identification of significant statistical parameters.
- Developing a non-overlapping block-based model for detecting copy-paste forgery within the same image.
- Analysis of the parametric result for ascertaining the extent and location of the forgery within an image using non-overlapping block-based parametric forgery detection model.

In the next section, methodology of the proposed non-overlapping block-based parametric forgery detection model for detecting copy-paste forgery has been presented.

4. METHODOLOGY

In order to detect copy-paste forgery, a statistical parameter based forgery detection model using non-overlapping block-based technique has been proposed. The block diagram of the proposed non-overlapping block based forgery detection model is presented in Figure 1.
image, the parameters mean (μ), standard deviation (SD), variance (σ), skewness (Sₖ) and kurtosis (K) of the original image are computed. Copy-paste forgery is done in the original image resulting into a forged image (Iₜ). This copy-paste forged image is again segmented into non-overlapping blocks and all parameters under study are computed. Each corresponding block of Iₜ are analysed and compared with those of I₁ after further forgery detection. The statistical variation \( \Delta P = |P_1 - P_1'| \) is computed and analysed for ascertaining the forgery.

Threshold \( t \) defines the permissible proportion of variation in the parameter under study. The value \( l \) of threshold \( t \) is set after testing a wide range of image sets and analysing the behaviour of parameters. Based on the variation, the forgery status \( DT \) (detected) or \( ND \) (not detected) is established. Further analysis is done to find the specific location, \( l \) where the forgery is actually been done. The proposed interface for forgery detection is developed in MATLAB version 7.6.0.324 (R2008a) and is tested for more than 200 gray scale bmp images.

**Fig 2: Parametric non-overlapping block-based copy-paste image forgery detection stages**

Figure 2 shows the stages of the proposed parametric non-overlapping block-based copy-paste image forgery detection model. The original image \( I_{OT} \) after normalization is divided into four non-overlapping blocks of size \([m \times n]\). Then, copy-paste forgery is done in the original image resulting into a forged image \( I_{FT} \). This forged image is again segmented into four non-overlapping blocks of same dimension and all statistical parameters under study are computed and compared for both \( I_{OT} \) and \( I_{FT} \), thus giving the forged block \( B_k \).
Variation in there is significant variation of the different parameter is given by distribution the distribution is far from symmetrical. greater than 1.0, variation for showing maximum variation for variation for all statistical parameters ∂ into four non specific location of the forgery in the image after division status is established. Further analysis is done to find the wide range of image sets and analysing the behaviour of parameter is set. The value variation in the parameter under study, for each statistical images and their blocks is computed for each parameter. The statistical variation in each parameter is given by ∂µ, ∂SD, ∂v, ∂Sk and ∂K.

The threshold t, the permissible proportion of statistical variation in the parameter under study, for each statistical parameter is set. The value t of threshold t is set after testing a wide range of image sets and analysing the behaviour of statistical parameters. Based on the variation, the forgery status is established. Further analysis is done to find the specific location of the forgery in the image after dividing it into four non-overlapping blocks. The value of threshold t for ∂µ, t2 for ∂SD, t3 for ∂v, t4 for ∂Sk, t5 for ∂K is set to 0.3, 0.2, 3, 0.6 and 3 respectively. The image I8 is showing least variation for all statistical parameters whereas I9 and I6 are showing maximum variation for K, I9 is showing maximum variation for v.

Moreover, the ∂Sk of the images I7, I4, I6, I8 and I9 are greater than 1.0, which signifies that the ∂Sk is substantial and the distribution is far from symmetrical. These asymmetrical distributions will have long tail to the right and a positive skew.

Figure 3 shows the original test image IOT and the corresponding forged image IFT. The copy-paste forgery has been induced manually in the original image. A set of six gray-scale bmp images with corresponding forged images have been shown in it.

The experimental results thus obtained are further analysed and discussed in the next section.

5. RESULTS AND DISCUSSIONS
The experimental results are obtained after the implementation of the proposed parametric non-overlapping block-based model for copy-paste forgery detection. The cases shown in Table 1 are selected cases out of 200 test results where there is significant variation of the different statistical parameters of original and forged images depicting copy-paste forgery detection based on threshold t. One of the cases of I10 shows the Not Detected i.e. ND status since all the threshold values are below t. The statistical variation ∂ in the given above said parameters for original and forged images and their blocks is computed for each parameter. The statistical variation in each parameter is given by ∂µ, ∂SD, ∂v, ∂Sk and ∂K.

Also, it is deduced from the results generated above that the ∂Sk of all the images except I10 are greater than 0. These positive kurtosis images would have a fairly uniform distribution of gray levels.

Table 1. Parametric differences of original and forged images depicting copy-paste forgery detection based on threshold t, (t=3)

<table>
<thead>
<tr>
<th>Test image I_T</th>
<th>Parametric variation (t=0.3)</th>
<th>Forgery status</th>
</tr>
</thead>
<tbody>
<tr>
<td>t1=0.3</td>
<td>0.19</td>
<td>DT</td>
</tr>
<tr>
<td>t2=0.2</td>
<td>0.00</td>
<td>DT</td>
</tr>
<tr>
<td>t3=0.6</td>
<td>0.04</td>
<td>DT</td>
</tr>
<tr>
<td>t4=0.01</td>
<td>0.00</td>
<td>DT</td>
</tr>
<tr>
<td>t5=0.02</td>
<td>0.00</td>
<td>DT</td>
</tr>
<tr>
<td>t6=0.01</td>
<td>0.00</td>
<td>DT</td>
</tr>
<tr>
<td>t7=0.01</td>
<td>0.00</td>
<td>DT</td>
</tr>
<tr>
<td>t8=0.03</td>
<td>0.00</td>
<td>DT</td>
</tr>
<tr>
<td>t9=0.02</td>
<td>0.00</td>
<td>DT</td>
</tr>
<tr>
<td>t10=0.00</td>
<td>0.00</td>
<td>ND</td>
</tr>
</tbody>
</table>

Also, it is deduced from the results generated above that the ∂Sk of all the images except I10 are greater than 0. These positive kurtosis images would have a fairly uniform distribution of gray levels.

Figure 4 depicts the relationship of two statistical parameters µ and SD of original and forged images (I_FT-I_T10). Forgery can be observed in all the test images except I10. Also, maximum variation is seen in µ for test image I_Final and in test image I_T10 for SD. The statistical variation in these two parameters is noticeable.

Fig 3: Set of gray-scale copy-paste forged images (IFT) along with their corresponding gray-scale original images (IOT)

Fig 4: Parametric difference in μ and SD of original and forged images (IFT-I_T10)
in which forgery table also shows the corresponding blocks of both the images considered as a wide range of images. Any for each parametric variation has been computed after testing in their statistical parameters and their corresponding graphs.

These two images depict the presence of forgery in all the images except I_T4 and in test image I_T8 for skewness. For maximum cases, the statistical variation in variance is significant.

Further in Table 2, blocks B_2, B_3, B_4 of test image I_T4 and blocks B_7, B_8, B_9 of test image I_T7 show statistical variation in all their parameters thus confirming the existence of forgery. The block B_1 of test image I_T4 and block B_4 of test image I_T7 show no statistical variation in any of its parameters thus depicting that no forgery has been done.

Further the table also shows the corresponding blocks of both the images in which forgery has been observed.

Table 2. Block-wise parametric differences of original and forged images depicting copy-paste forgery detection based on threshold \( t_i (i \geq l) \)

<table>
<thead>
<tr>
<th>Image blocks</th>
<th>Parametric variation ((i \geq l))</th>
<th>Forger status &amp; location, loc</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \hat{\sigma}_\mu )</td>
<td>( \hat{\sigma}_{SD} )</td>
</tr>
<tr>
<td>I_T4</td>
<td>03.89</td>
<td>00.23</td>
</tr>
<tr>
<td>B_1</td>
<td>00.00</td>
<td>00.00</td>
</tr>
<tr>
<td>B_2</td>
<td>05.82</td>
<td>00.10</td>
</tr>
<tr>
<td>B_3</td>
<td>06.39</td>
<td>02.38</td>
</tr>
<tr>
<td>B_4</td>
<td>03.34</td>
<td>01.07</td>
</tr>
<tr>
<td>I_T7</td>
<td>01.41</td>
<td>00.31</td>
</tr>
<tr>
<td>B_1</td>
<td>01.54</td>
<td>00.59</td>
</tr>
<tr>
<td>B_2</td>
<td>03.04</td>
<td>00.78</td>
</tr>
<tr>
<td>B_3</td>
<td>01.05</td>
<td>00.61</td>
</tr>
<tr>
<td>B_4</td>
<td>00.00</td>
<td>00.00</td>
</tr>
</tbody>
</table>

Figure 7 represents the column chart for parametric variation of original and forged image I_T4 thus giving the evidence of forgery. The graph shows maximum variation in kurtosis and insignificant variation in standard deviation for test image I_T4.

Fig 5: Parametric difference in \( v \) and \( S_D \) of original and forged images (I_T1..I_T10)

Variance and skewness of original and forged images (I_T1..I_T10) shows forgeries in all the images except I_T10 shown in Figure 5. The graph shows maximum variation in variance for test image I_T8 and in test image I_T9 for skewness. For maximum cases, the statistical variation in variance is significant.

Fig 6: Parametric difference in \( K \) of original and forged images (I_T1..I_T10)

Variation in kurtosis for original and forged images (I_T1..I_T10) can be seen in Figure 6 shows the behaviour of another parameter kurtosis of original and forged images (I_T1..I_T10) thus depicting the presence of forgery in all the images except I_T10. The images I_T8 and I_T4 show maximum variation whereas image I_T10 shows no variation in kurtosis.

Fig 7: Parametric variation in original and forged image I_T4
The proposed model for copy-paste forgery detection within an image is based on non-overlapping block based techniques. Five statistical parameters have been analysed by using specific threshold values. The threshold values have been ascertained by performing different tests for each of the selected parameters. A domain of more than 200 images taken from different sources in bmp format has been tested using MATLAB interface. The forgery has been induced manually at different locations within the given image.

The results observed for different statistical parameters and for different images along with their blocks depict the forgery status, its extent and location as per careful selection of $t$ and $l$ values. Out of five statistical parameters, variance and kurtosis are the ideal parameters showing maximum variation in original and forged image. The results obtained can be used as the initial verification of the images for forgery. This may enhance the forgery detection process by identifying the most likely cases of possible image forgeries.

7. LIMITATIONS AND FUTURE WORK

Parametric block-based copy-paste forgery detection model may fail when the copied region undergoes some geometrical transformations like scaling, rotation, translation, etc. The proposed model generates satisfactory results and is further explored for certain other features and parameters over a wider range of images having different types, formats and dimensions.

8. REFERENCES


Fig 8: Block-wise parametric variation in $\mu$, $SD$, $\nu$, $S_2$ and $K$ of original and forged image $I_{T4}$

Figure 8 shows the relationship among four statistical parameters mean, standard deviation, skewness and kurtosis for blocks of original and forged image $I_{T4}$. The statistical variation in these parameters shows forgery in three blocks $B_2$, $B_3$, $B_4$. Block $B_1$ shows no variation for any of these parameters whereas $B_2$ shows maximum variation for mean, skewness and kurtosis and least for standard deviation.

Fig 9: Block-wise parametric variation in $\nu$ of original and forged image $I_{T4}$

Figure 9 shows the block wise variation in variance of original and forged image $I_{T4}$. Maximum variation in variance for blocks $B_2$ and $B_3$ can be seen in the graph.

The inferences and conclusion about the proposed experimental model are presented in the next section.

6. INFERENCEs AND CONCLUSION

This parametric model for copy-paste forgery detection within an image is based on non-overlapping block based techniques. Five statistical parameters have been analysed by using specific threshold values. The threshold values have been ascertained by performing different tests for each of the selected parameters. A domain of more than 200 images taken from different sources in bmp format has been tested using MATLAB interface. The forgery has been induced manually at different locations within the given image.


