A Fast DCT Based Method for Copy Move Forgery Detection

Sunil Kumar, Jagannath Desai
Faculty of Engineering & Technology
Mody Institute of Technology & Science
Laxmangarh, India
skvasistha@ieee.org, jagandesai@yahoo.com

Shaktidev Mukherjee
Dept. of Electrical Engineering
Indian Institute of Technology
Roorkee, India
mukherjee.shaktidev@gmail.com

***Abstract*— Copy move forgery detection is emerging as one of the hot research topic among researchers in the area of image forensics. Many techniques have been suggested to detect such type of tampering with the original image, but the problem is far from being solved. Some issues still remained either unsolved or there is a lot of scope for performance improvement. Block matching algorithm or block tiling algorithm is the most commonly used method to detect the duplication in the image. One of the major challenges is the time complexity of such algorithms. As the image size increases the number of overlapping blocks increases rapidly and feature collection and matching takes relatively long time. In the proposed method this issue has been addressed without compromising the quality of the method. Discrete Cosine Transform (DCT) is used to represent the features of overlapping blocks. Also effort has been made to automate the threshold for separating the stray matches from the authentic matches.**

Keywords*— *Copy move forgery detection techniques; image forensics; image tampering; DCT based algorithm; passive forgery detection

I. Introduction

Digital images play very important role in the areas like criminal investigation, journalism and insurance processing. In all such areas the digital images are used as proof of some event or as a corroboratory evidence. Many a times people involved in the profession try to forge the original image to influence the outcome of the case. Manipulation of the digital image has become very easy nowadays with the ever improving hardware and availability of sophisticated software. So detecting such type of manipulations is not only important but also necessary. Digital image forgery detection may be categorized as in Fig. 1. Watermarks are already being used to protect digital images and any kind of tampering can be detected by change in the watermark [1] [2]. This type of forgery detection is termed as active forgery detection, as access to original image is required to insert watermarks. In many practical situations this may not be feasible. So there is a need for blind forgery detection method, where no prior information about the image is required. Forgery may be performed in many different ways [3].Out of these; the commonly used method is copy move forgery, where parts of the same image are used to hide some other parts of the image. Existence of two same regions is not common in natural images and this property is exploited to detect copy move forgery. Even after applying some post processes, like edge smoothing, blurring and noise adding to remove the visible clues, there exist two extremely similar regions in the manipulated image. One of the most frequently used methods to detect such type of forgery is to use block matching algorithm. In the block matching algorithm the image is divided into overlapping blocks and the blocks are matched to find the duplicated region. Many people have used it to find duplication of the region with different features representing one block of the image [4] [5]. Fridrich [6] first suggested the method using exhaustive search and then proposed a more efficient block matching detecting method based on DCT. Popescu [7] proposed a method using Principal Component Analysis (PCA) instead of DCT. Due to the characteristics of PCA, the number of features required to represent a block was reduced by about half as compared to the number of features used by Fridrich. So the method using PCA has better time complexity. But the method lacks robustness against small rotations of copy moved regions. There are some methods

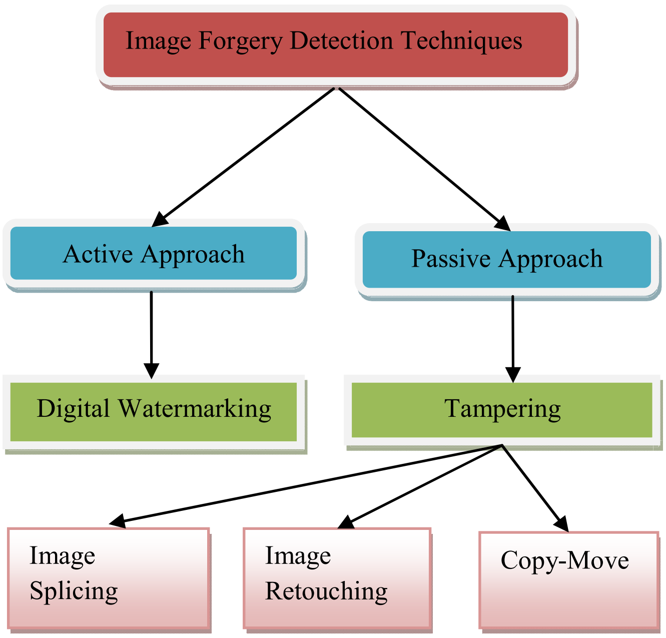

Fig. 1. Forgery detection classification

suggested for detecting copy move forgery in the presence of post processing operations, like Mahdian's method [8] can detect the duplicated region in the presence of blurring and added noise. However, one of the major issues reported is the time taken by such methods to detect the matching blocks. For example in [8] and [9] execution time is reported in several minutes. Also there are certain thresholds which need to be set manually either by hit and trial or by experience and the thresholds are dependent on the image texture. In the present paper the issues of time complexity with the robustness against post processing operations such as noise adding, JPEG compression are addressed and the experimental results show that the performance of the method has improved in terms of execution time without affecting the robustness against post processing operations. There are three major approaches which may be employed to reduce time complexity of block matching methods namely decreasing number of instances, reducing feature vector dimension and improving block matching algorithm.

DCT has been widely used to represent the image in frequency domain due to its ability to represent most of the intensity distribution details with fewer coefficients. The proposed forgery detection algorithm is based on DCT. The intensity of image at pixel (x, y) is I(x, y) and the block size is 'b'. The overlapping blocks of the image are represented by corresponding DCT coefficients as D (u, v) as per equation (1).

$$D(u,v) = \frac{2}{b} C(u)C(v) \sum_{x=0}^{b-1} \sum_{y=0}^{b-1} I(x,y) \cos\frac{\pi u(2x+1)}{2b} \cos\frac{\pi v(2y+1)}{2b} \tag{1}$$

$$Where\ C(u) = \begin{cases} \frac{1}{\sqrt{2}} & if\ u = 0 \\ 1 & otherwise \end{cases}$$

In comparison with Fridrich's technique [6] the proposed algorithm is quicker due to dimension reduction and modified matching algorithm without compromising the robustness. The rest of the paper is organized as follows. The proposed copy move forgery detection method is described in section II. The experimental setup and results are shown in Section III. The conclusion and future work is discussed in Section IV.

II. PROPOSED METHOD

A. Algorithm Structure

The proposed method uses the DCT coefficients to represent the overlapping block. The DCT coefficients are ordered in zigzag manner to keep the low frequency coefficients together and before the high frequency coefficients in the row vector. According to the framework shown in Fig. 2, the proposed algorithm works as follows.

(1) The input image is a gray scale image 'I' of the size m x n. If it is a color image, it can be converted to a grayscale image using the standard formula, I = 0.299R + 0.587G + 0.114B. R,G,B represents the three color components of RGB color model.

(2) Slide a fixed-sized b x b square window by one pixel from the upper left corner to the bottom right of the image 'I' to divide it into (m-b +1)(n-b +1) overlapping blocks.

(3) Apply DCT to every block and reshape the b x b quantized coefficient matrix to a row vector by ordering DCT coefficients in zigzag order. To reduce the size of the vector and to retain only low frequency coefficients, the vector is truncated to only p x b^2 elements to retain only low frequency coefficients. The parameter 'p' decides the number of coefficients retained.

(4) All vectors are sorted lexicographically and form a (m-b +1) (n-b +1) x pb^2 matrix 'A'.

(5) For each row a_i in 'A', test its neighboring rows a_j which satisfy the condition that the first 'q' quantized DCT coefficients are same. As these DCT coefficients are sufficient to represent the major intensity distribution over the block.

(6) If a_i and a_j come out to be similar, the distance between two should be more than the block size i.e. 'b'.

(7) If distance between similar blocks is greater than' b', then calculate the shift vector 's' and increase the count for 's'.

Where s= $(s_1, s_2) = (i_1\text{-}j_1, i_2\text{-}j_2)$ where similar block coordinates are (i_1, j_1) and (i_2, j_2).

(8) The highest count of 's' is taken to be threshold frequency. Also, it should be more than b x b to represent some significant duplication.

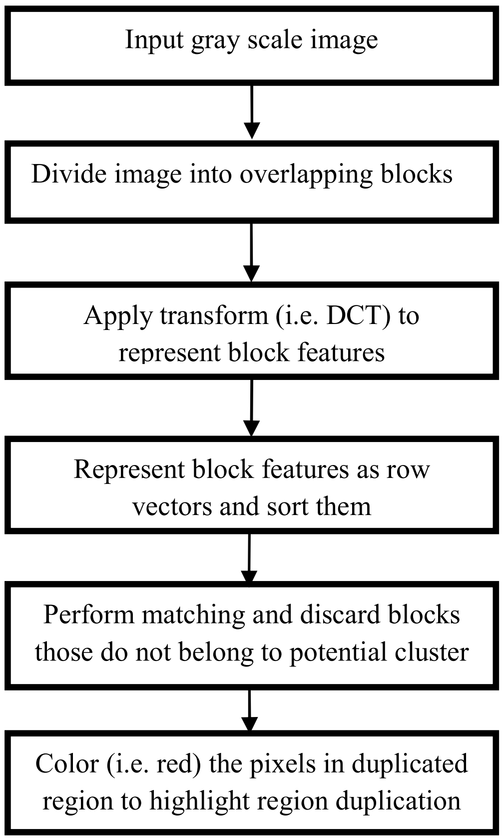

Fig. 2. Algorithm framework

(9) For all the blocks having shift value greater than the threshold value, mark the regions in the image with red color to represent copy moved regions.

B. Analysis of the method

The proposed method removes limitations of popular block matching algorithm [6] by modifying the structure of matching algorithm. In the matching step performed after sorting the feature vector array, not all row vectors within a fix range are considered to be similar but a more stringent criterion is used to establish similarity. As DCT is used to represent the features of a block, characteristics of DCT coefficients are exploited. The high frequency coefficients are susceptible to noise, so the row vectors are truncated. Further first few coefficients represent the major intensity distribution of the block [10]. Therefore the low frequency coefficients must be either same or should be very close for the copy moved regions. Only the blocks satisfying this condition are used further to update the frequency count and hence the numbers of potential similar blocks are significantly reduced.

Another modification is to get rid of the manual setting of some thresholds. First one is the threshold for number of neighboring row vectors (N_n) tested for similarity. Due to the use of stricter similarity criteria there is hardly any need for this parameter. Another is the threshold for minimum distance between matching blocks (N_d). Experiments have shown that it can be taken as the size of the block as it will avoid the neighboring overlapping blocks to be tagged as similar. Larger threshold has a risk of missing potential similar blocks. So it has been fixed to 'b'. Lastly the threshold count for valid shifts (N_f) will also vary from image to image and hence difficult to set prior to applying on arbitrary image. A more intelligent approach is being suggested. The maximum frequency of shift vectors is taken as threshold frequency as this will be corresponding to the largest cluster of similar blocks. To detect multiple duplicated regions lower value may be considered. A minimum of b x b shift frequency should be there to represent at least single block redundancy. However, in some cases of post processing operations of rotation and scaling lowering the threshold may help. Experimental results have confirmed a marked improvement in the execution time compared with the existing method.

III. EXPERIMENTAL SETUP AND RESULTS

To compare the speed and robustness of the proposed algorithm with the existing method, a dataset of 100 images is prepared. The dataset consists of images with different, resolutions and contrasts. The size of the forged region is also varied in the dataset. Some images in the dataset are taken from the benchmark dataset [11]. For testing robustness of the algorithm against added noise, zero mean Gaussian noise is added to the forged images with Signal to Noise Ratio (SNR) ranging from 90db to 40db. Also JPEG compressed images of quality factor up to 7 are included. Double duplication is performed on some of the images. Block size is also varied from 4 to 16. The algorithm is coded in MATLAB 2012a on a machine equipped with Intel i5 2.5 GHz processor with 8GB DDR3RAM. The test image in Fig. 3 and Fig. 4 are of low contrast and two types of forgery are performed. In first case, the bulb is copied at multiple locations and in the second case in Fig. 4, the bulb is hidden using multiple clusters from the neighborhood area. In Fig. 5, high contrast image is used. Fig. 6 contains a low resolution image and in the last case in Fig. 7, high resolution image is taken. In all the cases the proposed algorithm has better execution time as shown in Table I. Also for all block sizes the performance is better as shown in Fig. 8. Table I justifies the lesser time taken as the number of blocks stored as similar pair and to be processed for shift vector frequency are much lesser in case of the proposed method. Block size of 8x8 found to give the best results. The success rate of algorithm against post processing operations is measured as per equation (2).

$$Success\ Rate = \frac{T_P}{T_P+F_P+F_N} \tag{2}$$

Where T_P is number of images correctly categorized as forged and F_P are the number of images falsely categorized forged and F_N are the number of forged images where forgery could not be established. The proposed method detected with 100% success rate for SNR above 50db. The success rate comes down to 90% for SNR down to 40db. Below 40db, success rate comes down to less than 50%. Robustness against JPEG compression is also checked and 100% success rate is observed up to quality factor 8. The success rate falls down to 50% for JPEG quality factor 7.5.The proposed algorithm is robust against minor scaling and rotation up to 2%. Also the efficiency of detection algorithm found to be highly dependent upon size of the copy moved region in case of rotation and scaling. The robustness decreases with the decrease in the size of copy moved region.

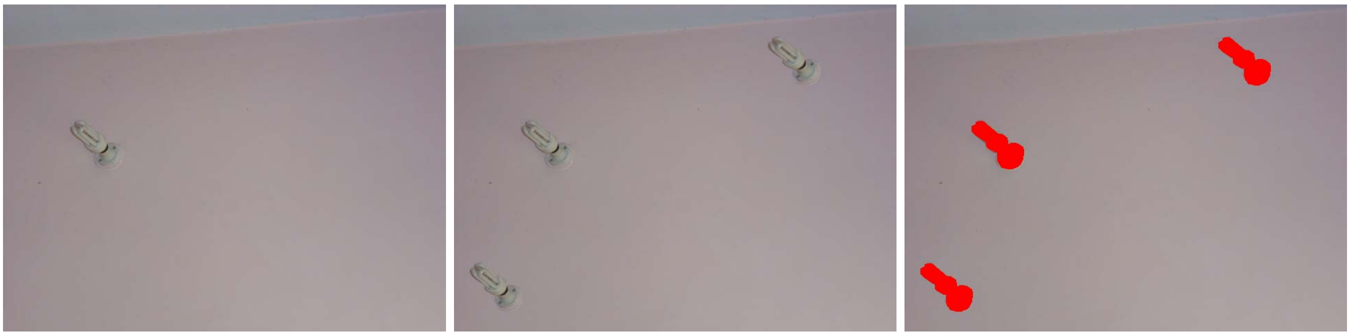

Fig. 3. Low contrast image (standard deviation=0.0367). Left: Original image of a bulb on the wall, middle: forged image with the bulb region duplicated at two places, right: result of the proposed method showing the similar regions highlighted.

Fig. 4. Low contrast image (standard deviation=0.0367). Left: Original image of a bulb on the wall, middle: forged image with the bulb region is hidden using multiple surrounding regions and blurred edges, right: result of the proposed method showing the similar regions highlighted.

Fig. 5. High contrast image (standard deviation=0.2992). Left: Original image of a card lying on the bag , middle: forged image with the card hidden by the bag region, right: result of the proposed method showing the similar regions highlighted.

Fig. 6. Low resolution image (256x256). Left: Original image of a park with a dustbin, middle: forged image with the dustbin hidden by the grass, right: result of the proposed method showing the similar regions highlighted.

Fig. 7. High resolution image (800x533). Left: Original image of a giraffe, middle: forged image with small portion of the pattern duplicated, right: result of the proposed method showing the similar regions highlighted.

TABLE I. PERFORMANCE COMPARISION

Image name/Image size	*Block size*	Execution Time in Seconds		Number of potentially similar blocks	
		Existing method [6]	*Proposed method*	*Existing method [6]*	*Proposed method*
bulb.bmp (640x480)	4x4	23.34	14.47	870409	12796
	8x8	25.24	17.8	848976	4685
	16x16	33.20	25.10	821736	1560
wall.bmp (640x480)	4x4	23.68	14.23	869631	9730
	8x8	24.69	16.47	847532	1697
card.bmp (640x480)	4x4	23.35	14.19	846302	2569
	8x8	25.88	16.24	826650	984
park.bmp (256x256)	4x4	4.95	3.5	183492	5391
	8x8	5.32	3.92	174812	4664
	16x16	6.4	5.43	158623	3469
giraffe.bmp (800x533)	4x4	31.66	19.55	1252903	9702
	8x8	36.03	22.43	1236590	3455
	16x16	46.94	34.62	1203396	2243

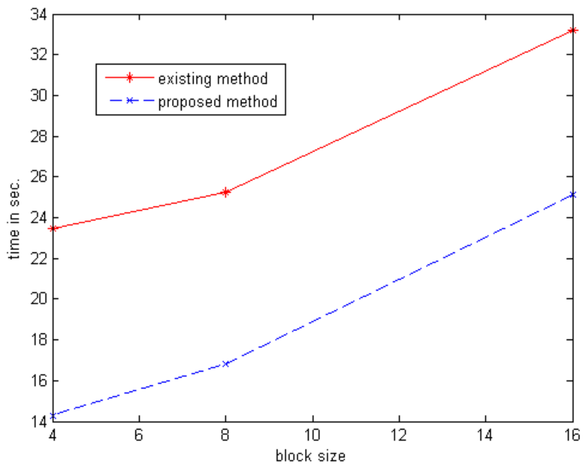

Fig. 8. Execution time comparison for image size 640x480

To compare the proposed method with the existing method [6], threshold values of the parameters have been set in the existing algorithm. N_n (The number of neighboring rows in the sorted array of row vectors for the calculation of shift vector.) is set to 5. N_f (Threshold frequency of shift vector 's') is taken 250. N_d (The minimum distance between similar blocks) is 16. p (The ratio of high frequency DCT coefficients truncated in the row vectors) is set to 0.25. Selection criterion of 'p' is purely experimental and the chosen value of 'p' is found to give best results. 'q' (The number of quantized low frequency coefficients matched to test the similarity of neighboring rows) in the proposed algorithm is set to 3. A more comprehensive criterion for measuring the efficiency of the algorithm has been introduced namely success rate.

IV. CONCLUSION AND FUTURE WORK

The method of block matching is one of the most frequently used for detecting copy move forgery. As there is no single method to establish the authenticity of given image, high computational requirement hinders researchers for checking the given image using different techniques. Hence any improvement in execution time will help in checking the given image with multiple techniques. The proposed method has addressed the issue successfully and is considerably faster than the existing method. It has detected forgery with good success rate in the image dataset. Also, it has shown robustness against added Gaussian noise, JPEG compression and small amount of scaling and rotation. However, robustness against more post processing operations like flipping, shearing and local intensity variations may be studied further.

REFERENCES

[1] E. Lin, C. Podilchuk, E. Delp, "Detection of image alterations using semi-fragile watermarks," Proc. SPIE, Security and Watermarking of Multimedia Contens II, vol. 3971, 2000, pp. 152–163.

[2] D. Kundur, D. Hatzinakos, "Digital watermarking for tell-tale tamper proofing and authentication," Proc. IEEE ,vol 8,no. 7, 1999, pp. 1167–1180.

[3] Gajanan K. Birajdar, Vijay H. Mankar, "Digital image forgery detection using passive techniques: A survey," Digital Investigation, vol. 10, no. 3, 2013, pp. 226-245.

[4] S. Kumar, P. Das, and S. Mukherjee, "Copy-Move Forgery Detection in Digital Images: Progress and Challenges," International Journal on computer Science and Engineering, vol. 3, no. 2, 2011, pp. 652-663.

[5] 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.

[6] J. Fridrich, D. Soukalm, J. Luka ´s ˇ, “Detection of copy-move forgery in digital images,” Digital Forensic Research Workshop, Cleveland, OH, 2003, pp. 19–23.

[7] A.C. Popescu, H. Farid, “Exposing Digital Forgeries By Detecting Duplicated Image Regions,” Tech. Rep. TR2004-515, Dartmouth College, 2004.

[8] B. Mahdian and S. Saic, “Detection of copy-move forgery using a method based on blur moment invariants,” Forensic science international, vol. 171, no. 2, 2007, pp. 180–189.

[9] K. S. Bacchuwar and K. Ramakrishnan, “A Jump Patch-Block Match Algorithm for Multiple Forgery Detection,” International Multi-Conference on Automation, Computing, Communication, Control and Compressed Sensing,2013, pp. 723-728.

[10] R.C. Gonzalez, R.E. Woods, “Digital Image Processing”, 2nd edition, Addison- Wesley, 2003.

[11] V. Christlein, C. Riess, J. Jordan, C. Riess, E. Angelopoulou: "An Evaluation of Popular Copy-Move Forgery Detection Approaches," IEEE Transactions on Information Forensics and Security, vol. 7, no. 6, 2012, pp. 1841-1854.