National conference on Technology innovation in Mechatronics,Energy Management and Intelligent communication(NCTIMEMIC-2017)

International Journal of Advanced Scientific Technologies , Engineering and Management Sciences (IJASTEMS-ISSN: 2454-356X) Volume.3,Special Issue.1,April.2017

Copy-Move Image Forgery Detection Using Expectation Maximization Algorithm

Abhila G. K.
M.Tech Student, Dept of CSE
Lourdes Matha College of Science and Technology
Trivandrum,India
gkabhila@gmail.com

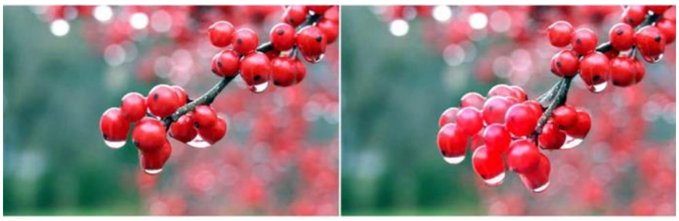
Chithra A. S.
Associate Professor, Dept of CSE
Lourdes Matha College of Science and Technology Trivandrum,India
chithra.as@gmail.com

Abstract—This paper propose a novel forgery technique to detect copy move image forgery by using segmentation based method and expectation maximization algorithm. This is mainly based on by extracting the keypoints for comparison from the forged image region. This scheme first segments the forged images into different segments, each segment is independent and they are called patches. For detecting the forgery two stages of matching take places. The first stage of matching consists of feature extraction, patch matching and transform estimation. Second stage of matching consists of obtaining new correspondance,obtaining new transform matrices and finally repeat above two steps for finding identical regions by using expectation maximization algorithm.

Keywords—copy-move forgery,patches,segmentation,keypoints,overlapping,matching.

I. Introduction

The use of internet become popular nowadays. The internet consists of lots of digital images .The authenticity of digital images need to be ensured because lots of tampering operations can perform on the images.One of most popular tampering operation is copy move forgery(CMF)[1].The copy move digital image forgery means copy a region from the same image and paste it somewhere in the same image region(eg.fig.1).The fig 1(a)[8].shows the original image where fig 1(b)shows tampered image which is not identified by naked eyes. So here is the need for a technology to identify such type of forgery. The digital image forgery is the technology to deal with the way of finding such digital forgeries in images.There exists two main classes of algorithms one is block-wise division algorithm and other is keypoint extraction algorithm. In block-wise division algorithm first divide the image into overlapping blocks then compare each block to find the similarities among them for identifying copy move forgery. The best example of such kind of method[2] is based on DCT (Discrete Cosine Transform),which is used for describing blocks.In DCT matching blocks of tampered image can be find out by comparing the coefficients of individual patches.If two patches has same coefficient then there exist possibility of copy move image forgery. This method use dictionary sort which is used for decreasing the complexities involved for finding the similarities among the matching blocks. In this algorithm descriptor of the block is important. The methods such as discrete wavelet transform(DWT), principal component analysis(PCA) etc where used [3],[4]. In terms of accuracy and detection capability zernike movement is best [4],[5].Same Affine Transformation Selection(SATS)[5] is a post-processing technique which is used for improving the effiency of CMF detection algorithm. The scale invariant features transform SIFT[6] and SURF[7] are most widely used keypoints based algorithms.These two methods are robust to find out forgery with transformations such as rotation,scaling etc,to estimate transform matrixes between copying source regions and pasting target regions[8].To avoid the unwanted outliers RANSAC algorithm[9] is used to improve the effiency of detection algorithms. The gold standard algorithm[10] is used for improving the effiency of RANSAC algorithm.



(a) (b)

fig 1.Shows (a)Original image and (b)Forged image.

This paper propose a new framework for copy move forgery detection.The test image is first segment into independent patches of non-overlapped regions. An extension of the classic registration method iterative closest point (ICP),is used in EM based algorithms for avoiding the problem of transferring partial matches between the obtained patches.This paper propose two stages for finding forgeries in the test image.In first stage find matches and then calculate corresponding transform matrix. The refining of transform matrix help to find out whether there exists copy move forgery or not.The proposed scheme work effectively in block based scheme compared to the keypoint based method.

The following sections is summarized as follows. In Section II deals with overview of the proposed CMFD system. Section III deals first stage of matching and section IV describe second stage of matching.Section V deals with experimental results.Section VI deals with conclusion.

II. OVERVIEW OF THE PROPOSED CMFD SYSTEM AND IMAGE SEGMENTATION

In this section describe about CMFD Revisiting and the Framework of the Proposed Scheme then describe the reason behind why segmentation is used. In CMFD

A. CMFD Revisiting and the Framework of the Proposed Scheme

The detection result accuracy increases if more information is acquired from the forged image. The mission of copy move forgery detection involved both identifying and locating tampered regions.In CMFD scheme a set of image patches is considered and finding the similarity by comparing these patches coordinate values.This is for both block based and keypoint based method.This comparison process is time consuming if the number of patches in the test image is high. So it is better to decrease the number of patches in the test image.In case of keypoint based method number of patches is less compared to block based method.

The problem regarding the keypoint based method is that keypoints which are closed spatially are not compared to each other because they are spatially similar in nature.It is tricky to determine the shortest distance between two comparable keypoints in the forged image.The keypoints are not concentrated together. Due to this it is very difficult to identify from where the image is copied and to where it is pasted in case of keypoint based detection scheme.For avoiding this problem clustering of keypoints based method is introduced. The clustering matched scheme improve the accuracy of CMFD.In case of block based scheme the image is divided into different patches then the similarity can be identified between copied and pasting regions by comparing these patches.

B. Image Segmentation

The image is segmented into different small regions called patches which help to distinguish between copied and pasted regions in forged images.This can be done by experts who are focused in digital image forgery detection. This method use four segmentation methods to detect the forgery but these are not influenced by CMFD detection efficiency. Fig 2 shows the segmentation.In this figure two tower are in CMF region.SLIC algorithm is used for segmentation.In this figure one CMFD region is divided into several different patches. This can be done in manually or automatically. In this method use automatic scheme. In this method the image is divided into not more than 100 patches. Finally the detection phase find out that two or more patches are similar in nature i.e., copy move forgery exists in that regions. This is shown in fig 3.If use keypoint based method use only less number of keypoints approximately four number is enough. The segmentation time is 15 s for an average sized image.The segmentation process hopefully help to increase the speed of detection process in test image.

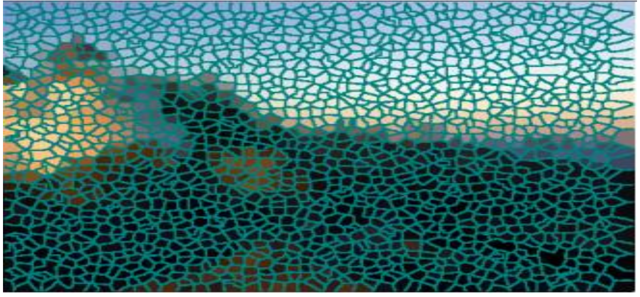


fig 2.Image segmentation

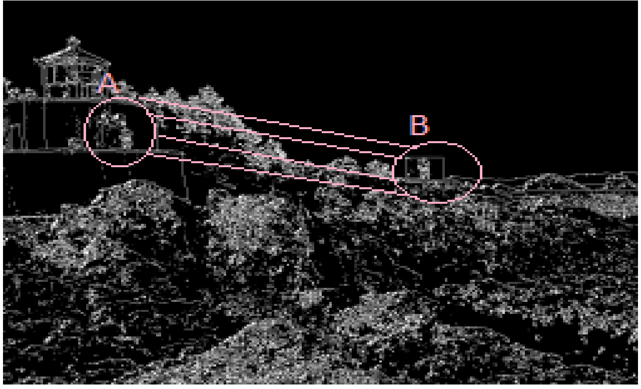


fig 3.Matched keypoint pairs in patch A and patch B are represented by line connecting the points in the patch regions.

III. FIRST STAGE OF MATCHING

The first stage of matching involved three stages of matching.This is shown in fig 4.

A. Keypoint Extraction and Description

In this method use vlFeat3[11] software to detect keypoints. The Difference of Gaussian (DoG),Harris-affine and Hessian-affine etc algorithms are commonly used for keypoint detection and description.The good detection results need more than 2000 keypoints in the test image.

B. Matching Between Patches

In this section the matches between the keypoints is find out by comparing each patch with rest of the patches. Consider the fig 3.The patch A is compared with K nearest neighbors.In this method k value is set as 10.In this implementation if difference is less than threshold say 0.004 then there exists a match between keypoints present in the two different patches.In fig 3 patches A and B are said to be suspicious pair of patches. Because they consists of copy move image regions as result the tampered regions are related to line connecting them in the figure.

C. Affine Transform Estimation

In this method find affine transform matrix after finding the suspicious pair of matched pairs.After that calculate the matrix for that regions.This is for finding the relationship between two identical regions.The relationship between the transform matrix H is find out by the following way.

$$x = H \dot{x}$$

.

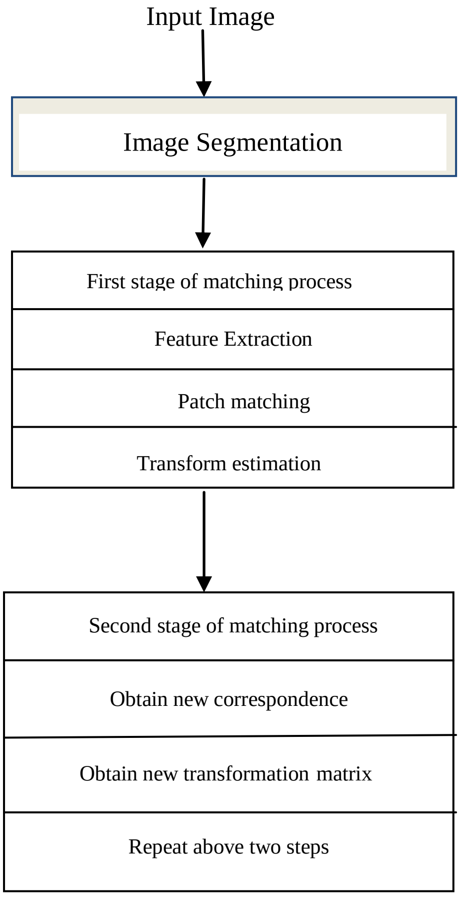

Input Image
Image Segmentation
First stage of matching process
Feature Extraction
Patch matching
Transform estimation
Second stage of matching process
Obtain new correspondence
Obtain new transformation matrix
Repeat above two steps

Fig 4.CMFD framework

The x and ẋ are coordinates in copying and pasting regions.Todays existing CMFD schemes only focuses on identifying the copying and pasted regions and do not identify the relationship between the two regions.This method identify the relationship between these regions by calculating affine transformation matrix which help to identify which transformation is applied to the regions such rotation, scaling etc.Doing this estimation helps to avoid false detections if forgery is not present in that region.The forgers do not change the location of copying source regions for avoiding the additional forgery traces.The classical method is used for the estimation of transformation between source and target regions. For minimizing the geometrical distance use three random non-collinear matched keypoints.The RANSAC is used for estimating the transformation matrix in case of existence of noise in keypoint detection.In some detection processes small regions with arround 5 keypoints affect the detection accuracy.If size of the forged region is smaller than 32×32 then there exists difficulty in detection accuracy.Because in small regions only limited number of keypoint exists this makes error in keypoint extraction process.Second stage of detection is introduced to improve the accuracy of the result.

IV. SECOND STAGE OF MATCHING

In first stage of matching process find suspicious pair of matching patches and then calculate affine transformation matrix between them to find which transformation is applied to it.For this purpose use RANSAC algorithm. It provides a robust transform estimation but sometimes it is not accurate. It ring false alarm even if no forgery is present in it. So there is a need for second stage of matching.In second stage transform matrix is refined by using EM based algorithm.This helps to avoid false alarms.

A. CMF Determination Based on Probability

In second stage of matching all the pixels in matched patches is used for finding the transformation matrix H.The pixels in the matched regions should be distinguished from the background region.In some test image consists of only small region of forgery in this situation only small amount of keypoint is present.In this case the result of first stage is not convenient due to limited number of keypoints.In second stage of matching pixels the matched patches is used to find out accurate transformation matrix.The pixels in the copying and pasted target regions can be distinguished from each other. This increases the accuracy of the result.The relationship can be written as follows.

$$f (x) = f (H-1\, \dot{x})$$

The image characteristic function is represented by function

f (·).The SHIFT descriptors are used for robust and efficient estimations.

B. Obtaining the New Correspondences of the Pixels

The H0 denotes the transformation matrix obtained from the first stage of matching .Second stage of matching requires more accurate confirmation regarding the forgery. So there is a need for reestimation of transformation matrix for pixel location at x. The correspondence obtained is more similar to old correspondence at location x then there exist a copy move forgery.The new correspondence estimation of pixels help to confirm forgery and also avoid false alarm.The probability of occurrence of error rate became decreased if

use SHIFT based calculations for keypoint based method.

C. Iterative Re-Estimation of the Transform Matrix

The newly matched pixel pairs are used for re-estimation of transform matrix.The re-estimated transformation matrix which is similar to old transformation matrix in case of copy move forgery exists if not otherwise.In case of smooth region the correspondants are not accurate.For avoiding the situation

RANSAC algorithm is used.The drawback of RANSAC algorithm is that it uses more time for processing.

Each patch consists of two classes of pixels.Here the problem is that distinguishing two pixels is difficult.The EM algorithm is helpful in this situation. The EM algorithm is used for statistical parameter estimation.The algorithm repeat until target value is reached. The algorithm consists of E-step and M-step. That's why it is called EM algorithm.

V. EXPERIMENTAL RESULTS

A. Test Image Databases and Segmentation Settings

To examine the performance of proposed system use the database namely Benchmark database for CMFD evaluation. The database consists of 48 original images and 48 images with CMF. The original size of the image is set as 3264 × 2488.In case of internet and multimedia applications the size of the image should be small. So the images size should be resized to not larger than 800.The resizing need to reduce the number of keypoints in the image. So resizing is not suitable for keypoint based method.The images in the database can be segmented by vlFeat software.The function vl_quickseg with certain parameters is used in case of resized image segmentation process.This function is used because it can implement quick shift image segmentation algorithm.In this algorithm set the value for two control parameters ratio and kernelsize to 0.7 and 1.

B. Error Measures

The error in image occured at two levels,image level and pixel level.In image level false negative rate and false positive rate are used for error detection.In false positive rate the ratio of missing detection to forged one give the error detection rate.The false alarm to the original image give error ratio in case of false positive rate.The precision and recall criteria are used in error detection rate at pixel level.

C. Results on the Database

The detection error rate is represented as false negative rate and false positive rate in database.The SURF and SHIFT schemes are used for forgery detection if keypoint based method is used.The snippet from the image is selected and paste it to somewhere in the same image so that it is unnoticable. Then pass this test image as input. In the first stage of process generate corresponding histogram of the test image which is imputed.This is shown in fig 5. Then the input image is passed to next stage of process. The next stage is edge detection phase. The cany edge detector is used for it. After that it is passed to filtration process for smoothing and noise reduction by using Gaussian filter. Then pass the image to segmentation process and pass to transform estimation. Then pass the image to second stage of matching processes for finalizing the image forgery with the help of expectation maximization algorithm. Next stage shows how much forgery is detected.The proposed scheme is good for detecting the tampering operation with smallest false negative rate.The false positive rate is larger in this method.Consider the robustness of the scheme against several attacks such as four attacks 1)JPEG compression2)adding noise3)rotation4)scaling down and up perform effective detection capabilities.The result shows that the computational complexity of this method is low and they also provide accurate results. The method efficiently detect which transformation operation is applied to the forged regions.

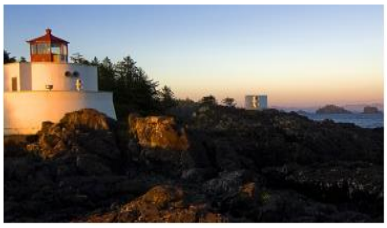
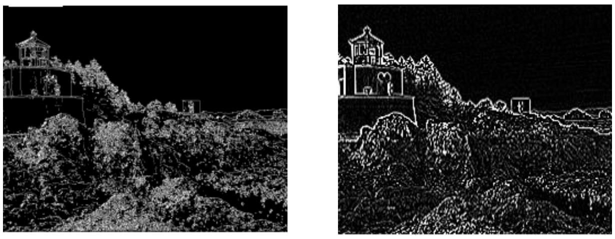


Fig 5.(a)Input image



(b)Edge detection (c)Filtration

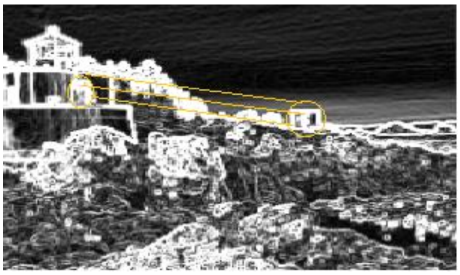


Fig 6.The regions inside the circle are similar in nature and the relationship between the matched regions are connected by the lines connecting between two circles. This regions shows copy move image forgery.

VI. CONCLUSION

In this paper a new method for copy move image forgery detection is proposed. The detection process consists of two stages of matching.First stage of matching consists of feature extraction,patch matching and transform estimation.The transform estimation help to identify which transformation is applied to forged region.Second stage of matching includes obtaining new correspondance,obtaining new transformation matrix. Finally with the help of expectation maximization algorithm forgery can be confirmed. The advantage of using this method is that the chance of occurring false alarm if copy move forgery is present can be eliminated. Because here checking take place twice.This method also detect which transaction is applied to the forged region.

REFERENCES

[1] Jian Li, Xiaolong Li, Bin Yang, and Xingming Sun" Segmentation-Based Image Copy-Move Forgery Detection Scheme" IEEE transactions on information forensics and security, vol. 10 ,pp.507-518,march 2015.

[2] W. Luo, J. Huang, and G. Qiu, "Robust detection of region duplication forgery in digital image", International Conference on Pattern Recognition. vol. 4. 2006, pp. 746–749.

[3] A. J. Fridrich, B. D. Soukal, and A. J. Lukáš, "Detection of copy-move forgery in digital images," Digital Forensic Research Workshop, 2003

[4] S.-J. Ryu, M. Kirchner, M.-J. Lee, and H.-K. Lee, "Rotation invariant localization of duplicated image regions based on Zernike moments," IEEE Transactions on information Forensics Security, vol. 8, no. 8,2013, pp. 1355–1370,Aug. 2013.

[5] V. Christlein, C. Riess, and E. Angelopoulou, "On rotation invariance in copy-move forgery detection," in Proc. IEEE Workshop International Information Forensics Security. (WIFS), Dec. 2010, pp. 1–6.

[6] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," International Journals in Computer Vision., vol. 60, no. 2, pp. 91–110, Nov. 2004.

[7] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "SURF: Speeded up robust features," Computer Vision Image Understand., vol. 110, no. 3, pp. 346–359, Jun. 2008.

[8] X. Pan and S. Lyu, "Region duplication detection using image feature matching",IEEE Transactions on Information Forensics Security, vol. 5, no. 4, pp. 857–867, Dec. 2010.

[9] M. A. Fischler and R. C. Bolles, "Random sample consensus:A paradigm for model fitting with applications to image analysis and automated cartography," ACM, vol. 24, no. 6, pp. 381–395, Jun. 1981.

[10] R. Hartley and A. Zisserman," Multiple View Geometry in Computer Vision", 2nd ed. New York, NY, USA: Cambridge Univ. Press, 2004.