Comparative Analysis of Feature Matching Algorithms In Region Duplication Detection

Nisha Fule, [2]Vanita Mane

[1]Student, [2]Assistant Professor
Computer Science Department,
Ramrao Adik Institute of Technology Mumbai, India

***Abstract*: Images works as a source of information which can be used as an evidence for any cyber-crime investigation. Because of technological growth, the availability of image morphing software's the true information in images can be modified. Modifying the content from any digital image is called tampering or forgery. Detecting the forgery done on any image is very crucial. Region duplication is one of the kind of digital image forgeries where particular portion is copied and pasted on same image at different location. Many duplication detection methods have been delivered good result. In this work comparative analysis of different block based feature matching algorithms are explained which are used to detect the duplicated regions present in an image.**

***Keywords*: DCT, Forgery, Image Hashing, Kd-tree, Lexicographical sort, Region duplication.**

1. INTRODUCTION

Images are used in varies applications as an important thing from magazines to social media everywhere images plays important role. With the increasing development in digital technologies the forgery on digital images is becoming easy and difficult to detect. Any kind of alteration to an image
without having the rights is called as image forgery. The purpose of forgery is to hide some sensitive information from image which can be used as an important evidence in digital forensics. It is a big challenge to preserve the authenticity of an image.

There are two approaches using which one can perform digital forgery on images.

1) Active Approach
2) Passive Approach

In Active approach the modification to an image is done at the time of its creation, for example adding watermark, digital signatures, etc. In case of Passive approach the alteration done after the image creation for example adding portion of an image into another image. Most crucial types of image forgery fall under passive approach. The commonly known passive techniques are:

1) Splicing
2) Image Retouching
3) Region Duplication (Copy-Move Forgery)

1.1 Types of Image Forgery

Image splicing [2] is a kind of forgery which follows process of creating consolidated image by joining cut portions from two or more photographs. Retouching or image enhancement is a type where manipulation has been done for photo restoration, it applies enhancement like adjusting colors, contrast, white balance, sharpness, noise etc.

Region duplication commonly known as copy move tampering is a kind of cyber attack where any part from image is cloned either at several places or at any particular place in the same image. Aim of the attacker is to mislead the viewer with forged image. The copied portion belongs to the same image, so the image properties like dynamic range and color remains same as the other part of the image. Region
duplication detection aims at detecting extremely similar areas and many methods of detection have been proposed to solve this problem.

Fig. 1. Example of Region Duplication

There are two types of method available to detect forgery;
(1) Block based, (2) Key-point based. In the first method, the process follows by dividing an input image into blocks which overlaps each other and then feature extraction of every block is done followed by matching each block with every other block to detect the forged region.

In the second category of detection method, high intensity points are estimated from the image. Keypoints are then extracted From estimated points these features are collected as a feature descriptors which further compared with each other to detect the matched Key-Points.

The paper is structured as follows: Literature survey is described in section II. Different region duplication techniques are explained in Section III. The comparative analysis of different techniques of feature matching is discussed in Section IV. The work is finally summarized in Section V as the conclusion.

In this paper our contribution is to compare the different matching algorithms used in block based methods and suggest the most appropriate algorithm which can be used as duplication detection method.

2. LITERATURE SURVEY

Literature shows the existing methods for region duplication detection.

A new fast and accurate algorithm for detecting duplication detection in digital images is proposed [], where the resulting features of all blocks are stored in a KD-tree. The block corresponding to each node in the KD-tree is checked with the block corresponding to the nearest neighbor of this node. If the correlation between such blocks is above a pre-specified threshold, the two blocks are considered as clones.

H.B.Kekre[7] introduced Image hashing technique which is used for forgery detection by generating hash value for each image. They have proposed a novel approach for tampering detection in images using hashing technique. Experimentally this technique detects small amount of image tempering in an image.

In [8] author proposed method based on the clustering technique by creating dendrogram, based on statistical descriptors. While making the clusters only the dissimilar pairs which are removed, to ensure the accuracy of similar blocks in each pair. this method reduces the probability of false matching.

Jie Hu[9] proposed a block matching detecting method based on discrete cosine transform (DCT) to detect duplicate region forgery. Firstly, the image is split into specified squared blocks, and then the image data undergoes a DCT. Then the DCT coefficients were grouped to reduce the dimension according to the frequency property.

3. REGION DUPLICATION DETECTION

In region duplication forgery[10] the intention is to hide something in the original image with some other part of the same image. In digital image forensics several general techniques can be applied to detect duplicated regions.

All region duplication detection methods follow a common process [6], as shown in Fig. 3.1. In both cases keypoint based and block-based the process start with preprocessing which includes conversion of RGB image to grayscale. For feature extraction both methods work differently to extract features. Similar features obtained from image are matched afterwards. Matching Techniques differs with methods used for extraction. To consolidate the matches which follow same transformation pattern, post-processing on detected duplicated regions can be performed.

In general, Many Key-Point matching based techniques are fast but their performance is not robust against small cloned regions and is sensitive to little variations in pixel intensities. Block-matching based algorithms are robust compared to keypoint matching.

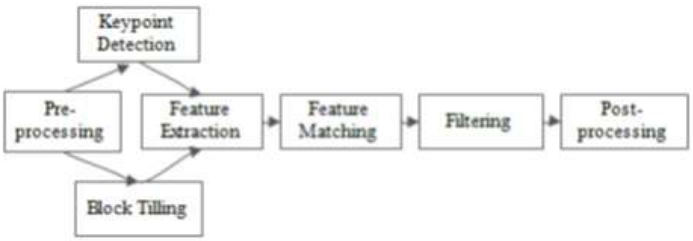

Fig. 2. Region Duplication Detection Process [6]

3.1. Block Based Methods

In block based detection, the image is first split into overlapped blocks, for every such block, a feature vector is computed. This features are extracted by using different feature extraction technique available for block based method such as DWT(Discrete Wavelet Transform), DCT(Discrete Cosine Transform), PCA(Principal component analysis), etc.

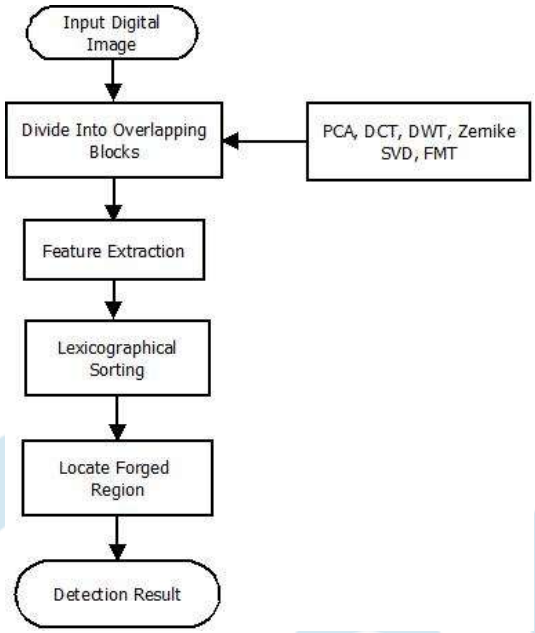

Fig. 3. Block Based Detection Common Pipeline

The extracted features are subsequently matched using different matching algorithms and the most matched descriptors are predicted as a clue for a region with duplication [6]. For block based methods, most authors proposed lexicographic sorting to identify similar feature vectors, technique like image hashing, KD-tree, clustering also work for matching the feature. Figure shows the common pipeline followed by the block based region duplication detection methods.

There are many techniques available in block based method for matching the features. Our contribution in this work is to give the comparative analysis of techniques presented in previous studies so as to get the most efficient technique which deliver best results.

The algorithm [5] is presented by Abdullah M. Moussa which belongs to the block matching based class of region duplication techniques. The algorithm starts with converting the suspicious image SI into greyscale and segmented to overlapping square blocks with a predefined side length η. The side length of the block is used as one of the algorithm parameters. At each pixel in SI the block is split into a grid of square, equally spaced k sub-blocks. Feature vectors are then produced from blocks these feature vectors are used to build a KD-tree and then near duplicated nodes in the KD-tree are estimated and apply a block matching process to detect the forgery.

In order to provide a fast algorithm and maintain an accurate performance, the feature vectors are extracted from blocks should not be complicated and it should provide a sufficient condition for distinguishing between the potentially matched blocks and the blocks that have no chance to be within a duplicated region. So, it calculates the sum of pixel intensities with the help of a sliding window within each sub block and stores the resulting nine values as a feature vector for that block. At the end of this procedure, we should have a nine values vector for each pixel in SI. These vectors are stored in nine-dimensional tree **T** using the KD-tree data structure with 1-norm distance.

For each node in T, every pixel in the corresponding block is checked with the corresponding pixel in the block associated with its nearest neighbor node in the tree according to the following condition:

$$Cf \leq \eta^2 * tm \qquad (1)$$

Where Cf is a correlation factor represents as the total number of pixel pairs, and tm is a problem dependent threshold to specify the correct match which varies from 0 to below 1. If the condition (1) is satisfied, the two blocks are considered as duplicate ones.

The algorithm developed by H.B.Kekre[7] which uses a concept of hashing into image region duplication works by computing both original image hash value and forged image hash value, those computed hash values are compared for getting match results. When perfect match has not found it concludes that the image has been forged. Hash value delivers good result even if there is a very little amount of forgery performed on the image. The algorithm works as follows:

Step 1: Compute Fast Fourier Transform(FFT) of an image
Step 2: After getting Fourier complex number, extract real and imaginary part from that complex number.
Step 3: Convert Cartesian coordinate to polar coordinate
Step 4: Generate Feature vector based on the angles of complex plane on which complex numbers plotted using Eulers formula.
Step 5: Generate 32 bit feature vector by combining 8 bits with 4 feature vector.
Step 6: From 32 bit feature vector generate the gray code to get final hash.

The technique explained by Jie Hu [9] uses DCT which converts signal component to frequency, widely used for image compression. In this paper, image features refers to DCT coefficients which are then sorted by using lexicographic sorting algorithm. This technique uses a concept of comparing the eigenvectors of two regions. Because the eigenvectors of two regions remains

completely the same if the tampered region is not processed manually. However in forgery, attacker performs image processing operations to cover up the tampering traces. They have assumed that the eigenvector has only two eigenvalues x and y. The two eigenvalues can form two kinds of eigenvector by different sorting mode.

Mode 1: Eigenvector is [x, y] where x is having higher priority over y

Mode 2: Eigenvector is [y, x] where y is having higher priority over x

Motasem S. [6] used a concept of clustering to match the duplicated regions which classify data over a variety of scales by making a cluster tree or dendrogram. A cluster tree is like a multilevel hierarchy of different clusters which gets joined at each level to form a tree. A dendrogram is the representation of the results of hierarchical clustering.

As the candidate of blocks are isolated and do not form any cluster it gets removed from the block to reduce chances of false predictions. The proposed system involves statistical features and dendrogram clustering, as follows:

3.1.1. Steps of Hierarchical Clustering

M*M matrix produced by using set of M candidate blocks to be clustered: 1) Block assignment to a cluster 2) Merge most similar pair of clusters into a single cluster 3) Similarity calculation using distance between the new and old cluster 4) Repeat above two steps until we get single cluster formed by all items from different cluster 5) If all required objects covered in single cluster, stop. Else, go to step 2.

According to distance i.e. from least to most it rearrange the block pairs (from least to most), so that the first pair of blocks has the least distance. For each iteration new ranking list of matching pair is calculated. As the copy moved region or block should not be adjacent it excludes the neighboring blocks from the ranking list to deliver more accurate and correct results. This reduces the chances of having false matches in final result.

4. COMPARISON

We have discussed few feature matching algorithms in block based approach and compared these techniques with respect to different parameter.

True Positive (T_P): Count of images which are detected as tampered.
False Positive (F_P): Count of images which are original but detected as duplicates.
False Negative (F_N): Count of images which are originally forged but detected as original.

Recall *r* and Precision *p* will get calculated by using above results as follows:

$$p = \frac{T_P}{T_P + F_P}$$

$$r = \frac{T_P}{T_P + F_N}$$

The comparison below shows all the architectural approaches and their respective features.

Sr. No.	Technique	False Positive
1.	Lexicographical Sort Algorithm	With higher threshold it cannot detects false matches accurately
2.	Kd-Tree	It is accurate but
3.	Dendrogram Clustering	Low false positive rate as adjacent blocks are removed at clustering step.
4.	Image Hashing	Low false positive rate

TABLE I. COMPARISON OF BLOCK BASED FEATURE MATCHING TECHNIQUES

4.1 COMPARATIVE ANALYSIS

We have compared block based feature matching techniques to handle region duplication forgery. We have considered the false positive parameter to see which technique delivers better performance. According to our analysis, dendrogram clustering technique achieves more accurate results compare to others. As it exclude neighboring blocks because duplicated blocks should not be adjacent there should have some physical distance between them. Therefore it minimizes the rate of false matching which leads to deliver accurate results.

CONCLUSIONS

Region duplication forgery is one of the powerful techniques widely used in digital image forgeries. Images created with duplicated regions are challenging to detect visually. An efficient system is required to be developed that can detect such forgeries.We have compared different block based feature matching techniques to handle region duplication forgery. According to our analysis, using dendrogram clustering technique we can achieve more accurate results compare to other techniques. Based on the performance of dendrogram clustering
for region duplication detection in digital images, in
future by detecting small target area and big size image some improvement can be achieved.

REFERENCES

1. Snigdha K. Mankar, Prof. Dr. Ajay A. Gurjar. ”Image Forgery Types and Their Detection: A Review” International Journal of Advanced Research in Computer Science and Software Engineering April- 2015
2. Parameswaran Nampoothiri V , Dr. N Sugitha Digital Image Forgery - A threaten to Digital Forensics, International Conference on Circuit,Power and Computing Technologies [ICCPCT] 2016
3. H. T. Sencar and N. Memon, ”Overview Of State of-the-art in Digital Image Forensics”, Department of Computer and Information Science, Polytechnic University, Brooklyn, NY, USA, 2008.
4. Salam A Thajeel and Ghazali Sulong, A Survey of copy-move forgery detection techniques, Journal of Theoretical and Applied Information Technology, 2014, Vol.70, No.1, Pp.25-35.
5. Abdullah M. Moussa ”A Fast and Accurate Algorithm for Copy-Move Forgery Detection” Tenth International Conference on Computer Engineering Systems (ICCES), 2015
6. Vincent Christlein; Christian Riess; Johannes Jordan; Corinna Riess; Elli Angelopoulou, ”An Evaluation of Popular Copy-Move Forgery Detection Approaches”, IEEE Transactions on Information Forensics and Security, 2012
7. H.B.Kekre, Dhirendra Mishra, Pallavi N. Halarnkar, Prajakta Shende, and Sukriti Gupta Digital Image Forgery Detection using Image Hashing International Conference on Advances in Technology and Engineering (ICATE), 2013
8. Motasem S. Al-sawadi Automatic Detection of Copy-Move Image Forgery Based on Clustering Technique Dec 2013
9. Jie Hu, Huaxiong Zhang, Qiang Gao, Hai Huang ”An Improved Lexicographical Sort Algorithm of Copy-Move Forgery Detection”,Second International Conference on Networking and Distributed Computing, 2011
10. Xunyu Pan, Siwei Lyu, Region Duplication Detection Using Image Feature Matching, ieee transactions on information forensics and security, vol. 5, no. 4, december 2010
11. A. Langille; Minglun Gong, ”An Efficient Match-based Duplication Detection Algorithm” The 3rd Canadian Conference on Computer and Robot Vision (CRV’06), 2006
12. G. Li, Q. Wu, D. Tu and S .Sun, ”A sorted neighborhood approach for detecting duplicated regions in image forgeries based on DWT and SVD”, International Conference on Multimedia and Expo (ICME) 2007, pp.1750-1753.
13. M. Bashar, K. Noda, N. Ohnishi, and K. Mori, Exploring Duplicated Regions in Natural Images, IEEE Transactions on Image Processing, Mar. 2010.
14. Jessica Fridrich, David Soukal, and Jan Lukas. Detection of Copy- Move Forgery in Digital Images, Proceedings of Digital Forensic Research Workshop, Cleveland, USA: OH, 2003. 1-10.