Copy-Move Forgery Detection using Feature Transformation & Hashing Technique

Lovish, Vikas Chawla
Research scholar, Asst. Professor
Electronics & Communication Engineering Department
Galaxy Group of Institutions, Ambala, India

***Abstract*— This work presented a proposed scheme for image forgery based on image segmentation. It applied SIFT densely to make block-based matching possible and cover all the pixels of the image and also combine method with hashing to improve performance of system. The similarity threshold between feature vectors is one of the most important parameters in detecting CMF. The proposed method to compute the automatic thresholding reduces the false positive and decreases the required time to estimate one threshold for different images in the dataset. The proposed method is compared with other CMFD methods and found better in terms of performance with the help of hashing method. All scenarios are implemented with the help of MATLAB tool.**

Keywords – Image Forgery, Static Feature Transform, Copy Move Forgery, Hashing etc .

I. INTRODUCTION

Image may be defined as a two dimensional intensity function f(x, y), where x and y denotes spatial co-ordinates and the value of 'f' at any point is directly proportional to the brightness of the image at that point. Light intensity is a real positive quantity, that is, because intensity is proportional to the modulus squared of the electric field, the image light function is real and nonnegative. Furthermore, in all practical imaging systems, a small amount of background light is always present. The physical imaging system also imposes some restriction on the maximum intensity of an image, for example, film saturation and cathode ray tube (CRT) phosphor heating. The image light function is, therefore, a bounded four-dimensional function with bounded independent variables. As a final restriction, it is assumed that the image function is continuous over its domain of definition. In the past, image forgery detection has been appeared unbelievable in applications of computer observation, criminal investigation, biomedical technology, digital image processing. When image processing powerful software tools are so popular and refined that we can't certify whether an image is manipulate by naked eyes it becomes more attractive and imposing [1]. Without advance information or security codes to perceive detect tampering in given image using blind algorithms image forgery detection is one kind of passive method. By splicing details from itself the image can be forged, or from other images called spliced images, and which is called Copy-Move images. In image the copied regions can be post processed, for Copy-Move images, to hide or remove any information the rotated and scaled before pasting to other places.

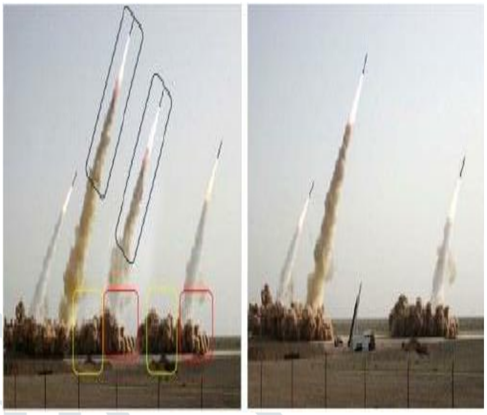

Fig 1: Copy Move Forgery Images

A remarkable array of visual imagery is exposed in an age where we are really living. In the integrity of this imagery we may have historical confidence, this trust has begun erode by today's digital technology. A growing frequency and sophistication are appearing with doctored photographs, tabloid magazines to the fashion industry, scientific journals, political campaign, courtrooms, in mainstream media outlets and in our e-mail boxes the photo hoaxes are landed. In the field of complete bibliography and digital image forgery is presented to analysis the recent developments on the blind techniques for forgery detection an attempt is made. About the image blind technique do not require any explicit prior information.

The set of image forensic tools can be roughly grouped in four categories:

- Introduced by a specific lossy compression scheme the format-based techniques are leverage the statistical correlations.
- Pixel-based techniques in which at the pixel level that detect statistical anomalies.
- Introduced by the on-chip post processing, camera lens, or sensor the camera-based techniques are exploit artifacts.
- Interaction between physical objects, light in the three-dimensional the physically based techniques detect anomalies and explicitly model.

The general CMF detection system consists of several main steps. The first step is to pre-process the image, for example, by converting the RGB colour image to a grey scale image. The second step is to extract features from the image. There are two different methods of extracting them: dividing the image into blocks (densely); or detecting interest points in the image (sparsely). With the first method, the image can be divided into overlapping or non-overlapping blocks, which can be either square or circular in shape. The features are extracted from the blocks. In the second, the numbers and the locations of the interest points vary, depending on the method itself. The features are then extracted in the neighbourhood of the interest points. The third step is to find the matches (similarity) between the extracted features. Many methods can be used to locate these similarities. The most common method is either to sort the feature vectors lexicographically and compute the Euclidean distance between adjacent stored vectors.

In this paper, it studies the concept of CMFD scheme. Further, in section II, it represents the related work of system and its various problems. In Section III, It explains Copy move forgery detection schemes with proposed algorithm. Section IV defines the results of proposed scheme. Finally, conclusion is explained in Section V.

II. RELATED WORK

Jian Li et al. [1], to detect the copy-move forgery in an image a scheme was proposed, for evaluation by extracting the key points mainly. To the traditional techniques the major differentiation was the test images into semantically independent patches prior to key point removal the anticipated scheme first segments. As an output, the copy-move regions can be detected by matching among these patches. An Expectation-Maximization-based algorithm was designed to prove the subsistence of copy move forgery and to refine the assumed matrix, in the second stage. On the public databases experimental via comparing it with the state-of-the-art schemes outputs prove the good presentation of the planned scheme.

Cheng-Shian Lin et al. [2], to detect the copy-move forgery tampering attacks a proficient scheme is used. In an image region by a copy of other region in the same image the copy-move forgery attack is defined as a region is replaced. For malicious modifying an image this was useful detection. The anticipated scheme needs fewer calculation time shows by Experimental outputs. Although the overhead of pre-processing is a further load that takes additional time than earlier cluster expanding block scheme, but comparing with earlier study the total computation time is still developed at least 10%. Furthermore, to reduces the false positive rate the block variance is used.

Harpreet kaur et al. [3], had used two copy-move image forgery detection techniques namely 'SURF' and 'PCA combined with SIFT' have been implemented using MATLAB platform. It has also been practical that the recognition correctness of 'PCA combined with SIFT' techniques is superior to 'SURF' and 'DWT combined with SIFT' techniques. On the other hand, this technique is unable to detect image forgery in flat region significantly.

Fredrich et al. [4], a method was proposed to detecting the copy-move forgery. The Discrete Cosine Transform (DCT) of the image blocks was used to avoid the computational burden the Discrete Cosine Transform (DCT) and was considered their lexicographical. Consider block of adjacent identical pair to be copy-moved blocks, one sorted. It cannot detect small duplicate regions this is the disadvantage of this method.

Popescu and Faridat el. [5], using Principal Component Analysis (PCA) a method was suggested. The image was separated into many parts represented into vectors and transformed into gray scale in this method. To represent the dissimilar blocks in a substitute mode these parts or blocks are organized lexicographically and PCA is used. In the noise and wasted compression it is proficient for detecting even minor variations. Moreover, for grey scale images this technique is far efficient. It is superior for gives less number of false positives and detecting copy-move forgeries. The computational cost and the number of calculation are significantly reduced O, where the number of image pixels is N and the dimensionality of the truncated PCA representation is N_t.

Weiqi Luo et al. [6], proposed the statistical examine of pixels in an image chunk computation with use of seven characteristics features and an algorithm for the extraction of image features. The averages of red, green, and blue components were respectively and computing four other features which are based on that block division into two directions parts: horizontal, vertical, and two diagonal directions. The highest frequency of occurrence in the main shift is defined for obtain the correct matching.

III. COPY-MOVE FORGERY DETECTION SYSTEM

In this forgery detection method in which dividing input image into over-lapping rectangular blocks, with which matching quantized DCT coefficients of the blocks for finding the tampered regions. For feature dimensions reduction Principal Component Analysis (PCA) is applied by using the RGB color components and block features as direction information. For image features extraction DWT and SVD are used. In existing systems they have some limitations, although these schemes are effective in forgery detection.

The input images is divided into overspreading and regular image blocks; and tampered regions is obtained by image pixels matched blocks or transform coefficients; in existing forgery detection methods which is block based, which extract the image key points and match them to identify the duplicated regions. To find the tampered regions in this forgery detection method in which over-lapping rectangular blocks into which input image is divided matching the quantized Discrete Cosine Transform (DCT) coefficients of the blocks. The applied feature dimensions are reduced by using Principal Component Analysis (PCA). As block features, the components of the RGB color and information direction is used. To extract the image features Discrete Wavelet Transform (DWT) and Singular Value Decomposition (SVD) is used. They have some limitations in existing systems, although in forgery detection these schemes are very effective. Computationally, the host image would be expensive as the image size increases and into over-lapping rectangular blocks it is divided. The forgery regions significant geometrical transformations cannot be addressed in this method.

The recall rate is low when their blocking method is a regular shape. The sophisticated tools can be used for the fact that the images evidence presented in court independently where image is manipulated digitally as the video is in a digital or analog form which diminishes the credibility and video tapes values by creating the threatening non-existing situations. To tamper an analogue video, the analog video stream in computer it can be uploaded in which one can easily digitize, the forgeries is perform, and then on an ordinary video tape the result is save in the NTSC format. As expected by one, the forgeries performed will move from research labs to commercial software as worst the situation will get. Despite the fact that by the research community recognizing the digital forgeries detection need, currently available publications are very few. Digital watermarks has been proposed as a means for fragile authentication, content authentication, tampering detection, localization of changes, and recovery of original content [1]. In the image, the watermark must be present before the tampering occurs, while the image integrity useful information is provided by the digital watermarks and its history processing to controlled environments which include military systems or surveillance cameras due to which their application limits. Unless watermarking chip is what all digital acquisition devices are equipped with, using a watermark unlikely detecting forgery-in-the-wild.

In previous work, the image data is segmented in form of overlapping segments, so which is not properly able to match the copy images. They take static feature which show only one properties of image. These features are not normalized, which is provide many false information at output. Transformation is estimated and not the actual geometric transformation, so matching process given more false positive error. Image segments group by K-nearest method which is not able to tell the relationship between segments.

The goal of research work can be fulfilled to improve the analysis of image by the ORB features in each scale space, the orientated FAST key points to the original image coordinates are reverted and between every two different key points, the ORB features are matched. Finally, remove the false matched key points and reducing problem in the copy-move forgeries classification such as false matching and low robustness.

In this regard, the key point-based methods are faster and more favourable than the block-based ones, because the number of the image key points is smaller than that of the divided blocks. The steps of system is described below:

- *Step1: Input the different types of images.*
- *Step2: Extract different type of features.*
- *Step3: normalize the features by scaling method.*
- *Step4: Matching using ORB features (Oriented FAST & Rotated BRIEF).*
- *Step5: Classification by reducing the false positive error.*
- *Step6: Post processing by analysis precision, recall*

- *Step 1:* Pre-processing: using the standard formula, if the input image “C” is a colored image, in a gray scale image it is converted. Where red channel is R, the green channel is G, and blue channel is B of the image.
- *Step 2:* Overlapping block pixel into a matrix: Over the resulting image “LLL” a “b × b” block is glided and from upper left corner to lower right corner scanning the image. In the matrix “A” the PCA coefficients of one row are stored. The matrix will have b × b columns rows and(M-b+1) × (N–b+1), Where “M” and “N” in input image represent the number of rows and columns respectively. For storing top-left co-ordinates the matrix “B” is formed in the same way as “A” with two additional columns. The block size is based on the image size, b was set dynamically. In the next level of PCA doubling the block size value and continuing the process of block value until final detection is reached in the final image (highest resolution).
- *Step 3:* Low Contrast Elimination: Calculate the contrast for each block in “A”, then ignore blocks where contrast is the least, i.e. the contrast is less than the specified threshold.
- *Step 4:* Phase correlation: For the current row “i” corresponding block with the blocks corresponding to “p” rows above and calculating the phase correlation below the current row. Calculating the preset threshold value “t” as it is exceeded by the maximum phase correlation value, then the corresponding reference block having the top left coordinates is stored in a new matrix new row and the matching block from “B” matrix. The resulting candidate block is progressed to the next phase “Comparison of Reference and Match blocks”:
- *Step 5:* Extract ORB and SURF features: Extract the key point feature based on orientation and capacity of neighbour points, these feature find out pattern in images.
- *Step 6:* Classifier: Use the SVM classifier with EM and RBF kernel model and test the model with analysis of precision, recall, accuracy.

To overcome the above drawbacks, it proposes a method where we use both block-based and the feature point-based algorithm. (DWT) is used for the image segmentation, where we use fourth level (DWT) to find the frequency energy coefficients. Considering the coefficients it calculates the initial size of the super pixel. This super pixel is used in SLIC algorithm to form the non-overlapping irregular blocks. These non-overlapping irregular blocks give more accurate results for the high resolution images. To extract the features (SIFT) algorithm is applied to the irregular blocks. The feature are extracted in every irregular blocks, they are matched by calculating the Dot products between unit vectors.

As (DWT) is concentrated on both time and frequency, this transform gives good frequency and high temporal resolution for low and high frequency components. According to proposed algorithm (DWT) is applied using ’Haar’ wavelet on to the image to verify whether the image is smooth or detailed image. This is done by calculating the low frequency energy. This process gives the appropriate frequency energy coefficients to calculate the initial size of the super pixel.

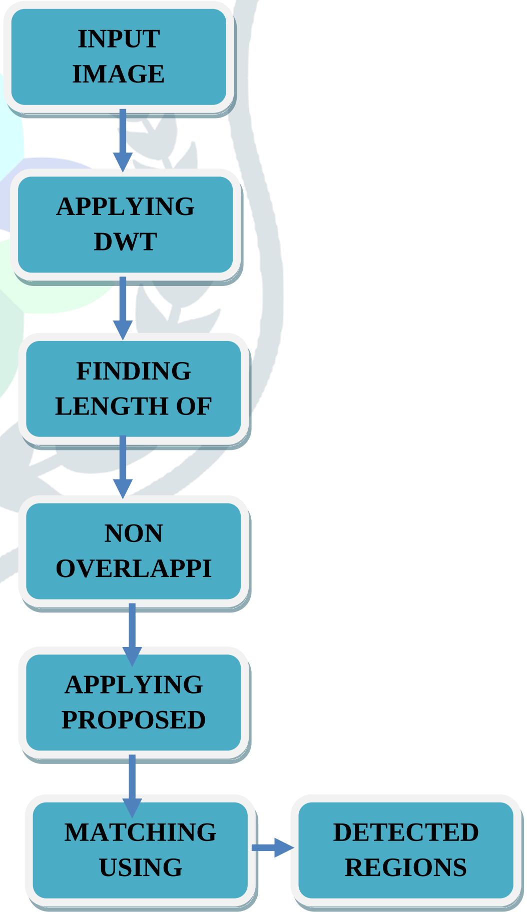

Fig 2: Proposed Steps of System

Deep Hashing

Let $X=[x_1,x_2,\cdots,x_N] \in \mathbb{R}\ d\times N$ be the preparation set which contains N samples, where $x_n \in \mathbb{R}\ d$ $(1\leq n\leq N)$ is the n_{th} test in X. Learning-based hashing strategies intend to look for various hash capacities to delineate quantize each example into a minimized double vector. Accept there are K hashing capacities to be realized, which delineate x_n into a K-bit paired codes vector $b_n=[b_{n1},\cdots,b_{nK}] \in \{-1,1\}$ $K\times1$, and the k_{th} two fold piece b_{nk} of x_n is registered. At that point, the mapping of x_n can be processed as: $g(x_n)=W_{T}x_n$, which can be further binarize to acquire the paired codes.

IV. RESULTS & DISCUSSION

The proposed algorithm uses Difference of Gaussians as a scale-space filter to make the SIFT scale invariant. The Difference of Gaussian, $D(x, y, \sigma)$, is found as the difference of Gaussian blurring of an image with two different standard deviations. This method was applied for different octaves of the image in the Gaussian Pyramid. Each pixel in an image is compared with its 8 neighbors as well as 9 pixels in the previous scale and 9 pixels in the next scales. If it is a local extrema, it is a potential key point. This basically means that the key point is best represented in that scale.

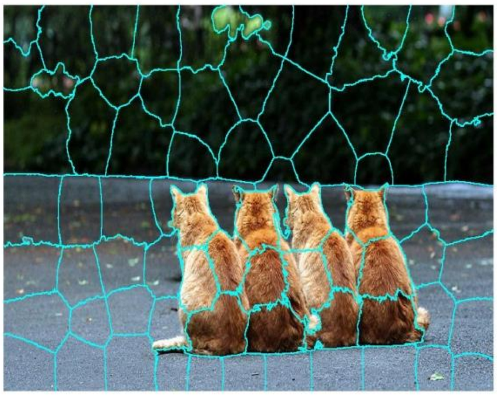

Fig 3: Original Image

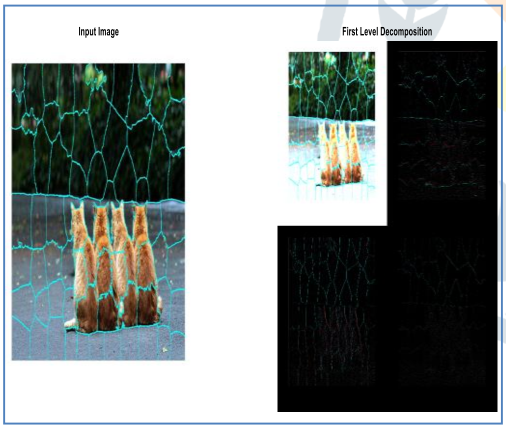

Fig 4: level Decomposition by DWT Method

In order to evaluate the proposed hashing approach, quantitative criteria is used. It follows two popular search procedures, i.e., hash lookup and Hamming ranking. Hash lookup first constructs a lookup table for the binary codes of all data samples, and then returns the percentage of the data samples falling into a small Hamming radius centred at the query sample. For example, HAM2 (i.e., the Hamming radius is 2) means that the Hamming distance between the training samples and the query sample is smaller than 2. In these experiments, it employed HAM2 to evaluate the results of hash lookup. Hamming ranking ranks all the samples in the database based on their Hamming distances to the query sample and then returns the top ones. Basically it performed with high resolution images giving more accuracy where comparatively above table provides some equal recall rates with block-based techniques but at high precision rates. This proposed scheme giving more accurate results and performances proves to be a novel technique to be implemented. Precision (**P**) is explained as the number of true positives (**Tp**) over the number of true positives plus the number of false positives (**Fp**). Recall (**R**) is the number of true positives (**Tp**) over the number of true positives plus the number of false negatives (**Fn**) as shown in fig 5 & 6. Table 1 shows the performance comparison of system with other methods.

Table 1: Performance Comparison of Different Methods

Classifier	Precision	Recall
SVMROC	0.883	0.882
CMFD	0.6596	0.500
Proposed Method	0.9535	0.916

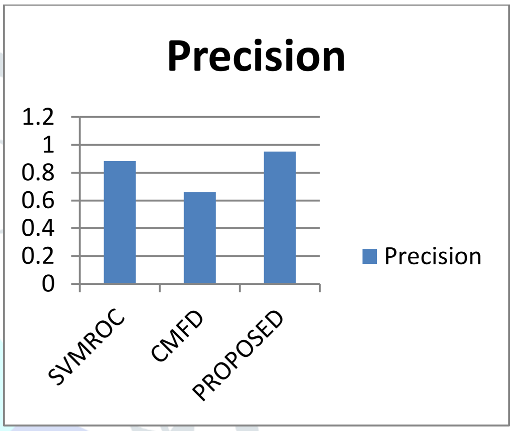

Fig 5: Performance Comparison of Precision with Different Methods

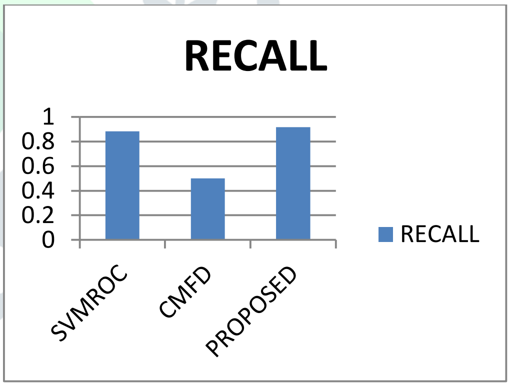

Fig 6: Performance Comparison of Recall with Different Methods

V. CONCLUSION

This work presented a proposed scheme for image forgery based on image segmentation. Although the CMF regions are detected mainly by comparing the key points extracted in the image, it cannot simply classify the proposed scheme as a key point-based one. It can be seen as a combination of both existing schemes because in the two stages of matching process both key points and pixel features are employed. It applied SIFT densely to make block-based matching possible and cover all the pixels of the image and also combine method with hashing to improve performance of system. The similarity threshold between feature vectors is one of the most important parameters in detecting CMF. The proposed method to compute the automatic thresholding reduces the false positive and decreases the required time to estimate one threshold for different images in the dataset. The proposed method is compared with other CMFD methods and found better in terms of performance.

REFFERENCES

[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 ,10(3), pp. 507-518, 2015.

[2] Chien-Chang Chen, Yi-Cheng Chang,"An Efficient Enhanced Cluster Expanding Block Algorithm For Copy-Move Forgery Detection," International Conference on Intelligent Networking and Collaborative System, pp. 228-231,2015.

[3] Harpreet Kaur, Joyti Saxena and Sukhjinder Singh, "Key-point based Copy-Move Forgery Detection and their Hybrid" Journal of the International Association of Advanced Technology and Science," 16(2), pp. 1-7, 2015.

[4] Fridrich, et al.," Detection of Copy-move Forgery in Digital Images," 2003.

[5] A. Popescu and H. Farid, "Exposing digital forgeries by detecting duplicate image regions," Dept. Computer. Sci. Dartmouth College, Tech.Rep. TR2004 515, 2004.

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

[7] Zhang, J., Feng, Z., & 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

[8] Bayram, S. Sencar, T. Memon, N. (2009), An Efficient And Robust Method For Detecting Copy-Move Forgery," in Proceeding of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '09), pp. 1053–1056, Taipei, Taiwan.

[9] Frank Y. Shih and Yuan Yuan,"A Comparison Study on Copy-Cover Image Forgery Detection," The Open Artificial Intelligence Journal, 4, 49-54,2010.

[10] Vincent Christlein," An Evaluation of Popular Copy-Move Forgery Detection Approaches," IEEE Transactions On Information Forensics And Security, 2012.

[11] B.L.Shivakumar1 and Lt. Dr. S.SanthoshBaboo," Detection of Region duplication Forgery in Digital Images Using SURF," IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 4, No 1, July 2011.

[12] HieuCuong Nguyen and Stefan Katzenbeisser, "Detection of Copy-Move forgery in digital images using Radon transformation and phase correlation", IEEE 18th International Conference on Intelligent Information Hiding and Multimedia Signal Processing, Piraeu, pp. 418-449, 2012.

[13] Gonapalli Ramu, S.B.G. Thilak Babu, " Image forgery detection for high resolution images using SIFT and RANSAC algorithm", IEEE International Conference on Communication and Electronics Systems, 2017.

[14] Cao, Y. Gao, T. Fan, L. Yang, Q. , "A Robust Detection Algorithm For Copy-Move Forgery in Digital Images, Forensic Science International, vol. 214, No. 1–3,pp. 33–43,2012.

[15] Li Jing, Chao Shao, "Image Copy-Move-Forgery Detecting Based on Local Invariant feature Journal of Multimedia," Vol.7,No.1, Februrary 2012.