IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-6, Issue-7, July 2018 E-ISSN: 2347-2693

Clustering Based Feature Extraction for Image Forgery Detection

Pooja Devi[1], Suman Deswal[2]

[1,2]Dept. C.S.E, Deenbandhu Chhotu Ram University of Science and Technology, Murthal, India

*Corresponding Author: suman_gulia2000@yahoo.co.in

Available online at: www.ijcseonline.org

Accepted: 13/Jul/2018, Published: 31/July/2018

***Abstract*—** Today manipulation of digital images has become easy due to powerful computers, advanced photo-editing software packages and high resolution capturing devices. Verifying the integrity of photo without any kind of special watermark or any prior knowledge is a critical issue. Photograph tampering techniques like copy-paste, which is very easy and effective to use, can extend image forging. The original content of the picture is copied to the desired locations. The increasing image modification software can easily manipulate the digital photo without leaving any visible clue. It's important to study these issues because tempered photographs can cause social chaos, criminal and non-public consequences. It's very important and additionally tough to discover the digital photograph forgeries. The main focus of this paper is to detect picture replica circulate forgery which is depended on SIFT (scale invariant feature transform) descriptors, which are invariant to rotation, scaling etc. Clustering algorithm is used for clustering of key points in images. Results show that, in comparison of existing methods MROGH, SURF-PHA provides consistent precision, recall and F1 score about 98.86%, 99.40%, and 99.13% respectively for the provided dataset. Experimental results indicate that this method is a robust method in detecting the copy-move forgery quickly and withstands certain transformations.

***Keywords*—** Copy-move Image Forgery, Forgery Detection, Feature Extraction, key-points, SIFT, Clustering.

I. INTRODUCTION

It is now a days very easy to forge digital image using high end image editing software. The forgery is so accurate that it's very difficult to differentiate fake from original one. Image forgery is of different types, in which "Copy-move" forgery method is the simplest & usually used forgery method and it is used to copy one part of the image to another. Fig. 1 shows an example for Copy-move forgery. The original image is shown in Fig. 1(a) and the forged image is shown in Fig. 1 (b). There are algorithms to detect these forged images but Copy-move technology and a little touch of smoothness deceives these algorithms. Hence, we should develop such algorithms to detect these forged images.

Figure 1. "Example of Copy-Move forgery:" (a)Original image. (b) forged image [1]

The recently used and effective technique for forgery is "Copy-move" forgery technique. In copy move forgery, some specific part or object is selected and then copied to another location. Various forgery detection methods are proposed in [2], [3]. The methods can either be block based [4], [5] or key point based [6], [7]. Block based method has some limitations like robustness against geometric transformation manipulation & period consuming.

This paper is organized as: section II explain the related work; Section III details the proposed methodology; section IV explain the results; finally, section V explains the conclusion and future work of paper.

II. RELATED WORK

Huang et al [8] proposed a key point-based method called Copy-Move forgery detection (CMFD). In this, native algebraic structures SIFT (Scale Invariant Feature Transform) [9] is used, then SIFT features are matched with the key points to detect the forgery. But its main limitation is that it doesn't show the geometry transformation parameter estimated beside the weak performance in the smooth region. Shiva Kumar et al [7] worked on SURF (Speeded-Up Robust Features) key points, which includes extracting the key points and matching by the KD-tree, but this technique is not able to identify the forgery in minor & smooth areas correctly. Pan and Lyu [10] are proposed a method based on SIFT, but this technique is failed due to weak management of affine transformation. Emam, Mahmoud, et al [11] have proposed a method based on Multi-Support Region Order-based Gradient Histogram (MROGH), to improve the recall rate. P Yadav, et al [12] worked on blocks. These blocks are sorted and duplicated blocks are identified using Phase Correlation as similarity criterion, but this technique is not able to identify forgery in multiple smooth area. M Jaberi, et al [13] worked on SIFT key points, improving the detection and localization of duplicated regions using more powerful key point-based features, but this technique is failed when the similar methods using key point-based features for matching.

Most of the proposed algorithms can't identify the forgery when the smooth areas are recovered on target regions from the image. They fail because of absence of key-points in these regions. In this paper, a new clustering method called Speed up robust feature (SURF) - Potential based Hierarchical Agglomerative (PHA) method to detect Copy move forgery detection is proposed which works by clustering of key-points.

III. METHODOLOGY

This section discusses a new clustering based method for forgery detection in images. The steps involved in the algorithm are as follows:

i. Key point Detection

Once a key point candidate has been found by comparing a pixel to its neighbours, the next step is to perform a detailed fit to the nearby data for location, scale, and ratio of principal curvatures. This information allows points to be rejected that have low contrast (and are therefore sensitive to noise) or are poorly localized along an edge. To determine the interpolated location of the maximum, Brown has established a technique (Brown and Lowe, 2002) for fitting a 3D quadratic purpose to the local sample ideas. His approach uses the Taylor expansion (up to the quadratic terms) of the scale-space function, D (y, z, σ), shifted so that the origin is at the model power point:

$$D(y) = D + \frac{\partial D^T}{\partial y} Y + \frac{1}{2} Y^T \frac{\partial^2 D}{\partial y^2} Y \qquad (1)$$

where evaluation of D & its derivation occur at same point & the offset from this fact is y = (y, z, $\sigma)^T$ w.r.t x by taking & setting it to 0. The location of extremum $\hat{Y}$ is determined as :

$$\hat{Y} = -\frac{\partial^2 D^{-1}}{\partial Y^2} \frac{\partial D}{\partial Y} \qquad (2)$$

The estimation of the Hessian & derivative of D is done by the using changes of neighboring model points. By the use of minimal cost resulting 3x3 linear system can be resolved. In any dimension if the value of offset $\hat{Y}$ is higher than 0.5 than it means closer to the different sample point extremum is present. To get the interpolated estimate for the location of the extremum the final offset ˆx is added to the location of its sample point. At the extremum function value D($\hat{Y}$) is useful. This can be found by substituting equation (2) into (1), giving the equation as follows:

$$D(\hat{Y}) = D + \frac{1}{2} \frac{\partial D^T}{\partial y} \hat{y} \qquad (3)$$

where D is scale space feature description.

ii. Key point matching

Second closest neighbor is usually defined as the closest neighbor coming out from the object that is different from the first one, in case there are more than one keeping fit pictures of the same object. This procedure works well because for a match to be appropriate require the nearby neighbor to be suggestively earlier than the closest improper match in order to reach a level of reliability in matching.

The second closest match is usually perceived as the estimation of bulk of incorrect matches within the portion of feature galaxy and identifying similar instances of feature ambiguity at the similar time.

iii. Potential-based Hierarchical Agglomerative

First of all, a hypothetical potential field of all records points is constructed and the close relationship between this potential field & nonparametric estimate of the global probability density function of the data points is established. This process is then followed by the proposal of a new parallel metric inheriting both the potential field representing global data distribution information and the distance matrix representing local data distribution information. Ultimately, development of another equivalent similarity metric build on the basis of an edge weighted tree of all the data points leading to a fast agglomerative clustering algorithm with time complexity O(N^2) is taken out.

Clustering is a process in which data points are divided into groups or clusters on the basis of similarity measures. This method is classified into two types which are partitional (single partition of data points) and hierarchical (nested clustering and cluster tree form). A novel Potential-based Hierarchical Agglomerative (PHA) clustering technique is proposed here which develops theoretical potential field of the considerable number of information focuses and demonstrates the identification of potential field according to nonparametric estimation. Clustering below a speculative potential field is a new & interesting idea. The potential field is demonstrated in [14]. For fulfillment, the key substance of the model is described here. For focuses m & n , if $r_{mn}$is the separation between them, the potential at power point m and n is described as

$$\Phi_{mn}(r_{mn}) = \begin{cases} -\frac{1}{r_{mn}} \;\; if\; r_{mn} \geq \delta \\ -\frac{1}{\delta} if\; r_{mn} < \delta \end{cases} \qquad (4)$$

where a parameter δ is utilized to maintain a strategic distance from the issue of peculiarity when r_{mn} winds up 0 and Φ is the number of data points.
Fig 2 explains the complete set of steps used in the proposed SURF PHA algorithm for copy move forgery detection.

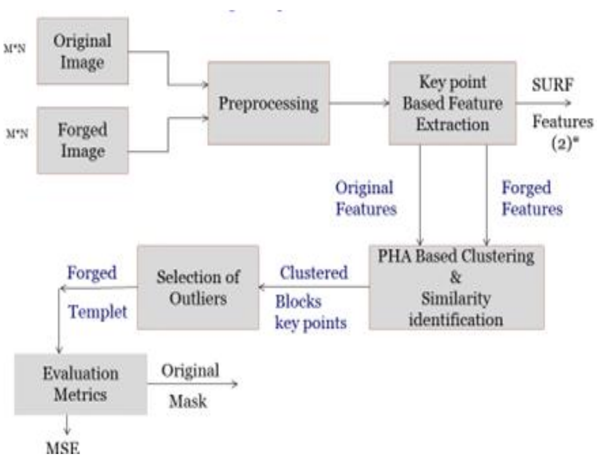
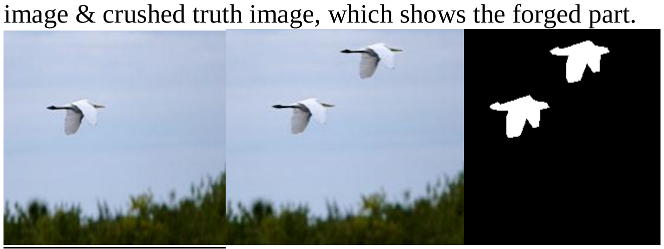


Figure 2. Single step of SURF PHA

IV. RESULTS AND DISCUSSION

This section presents the results related to the comparison of proposed algorithm SURF PHA and MROGH with SIFT for image forging and clustering. The two algorithm have been compared on three parameters namely: Precision, Recall and F1 score.

i. Dataset used
The picture dataset used for testing the proposed method contains 67 high resolution color images. Forged image has variations after smooth to high surfaced regions & various ordered revolution are applied like scaling, rotation. This one varies in size as well. Fig 3 shows the original image, forged image & crushed truth image, which shows the forged part.



(a) (b) (c)

Figure 3. Example of Copy-Move forgery from the dataset: (a) original image. (b) forged image. (c) crushed truth image.

Table 1 shows the number of images used and the manipulations types used to study the proposed technique.

Table 1: Total images and type of manipulations used to study the proposed method

Manipulation type	Control parameters	Number of Images
Plain Copy-Move	No manipulation applied	67
Scaling	Scaling factor s = 0.91: 1.09 with stage of 0.02	670
Rotation	Rotation edge θ = 0 : 10 with stage of 2	402
Additive noise	Standard Deviation σ = 0 : 0.1 with stage of 0.02.	402

ii. Precision
Following the vast majority of current CMFD techniques, performance of the proposed method is likewise tried at pixel level by Precision-Recall rates. Precision means the quality, condition, or truth of being definite and precise. Precision rate is described as shown in Eq. 5

$$\text{Precision} = \frac{\text{Tp}}{\text{Tp+Fp}} \qquad (5)$$

where Tp is the quantity of different pixels named altered (True Positive), Fp is the quantity of real pixels named altered (False Positive).
Using this formula, from equation (5) a graph is plotted between precision and rotation angle and compared between

existing algorithm and proposed algorithm as shown in Figure 4

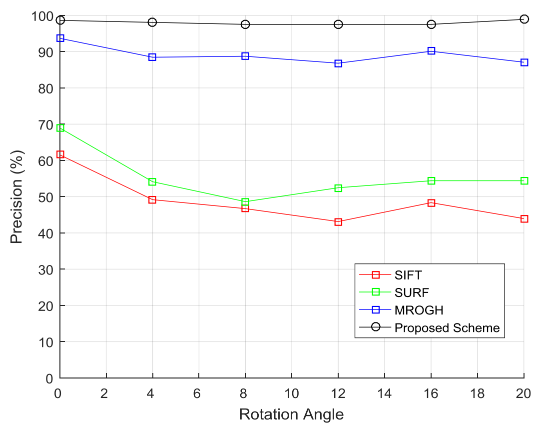


Figure 4: Precision (%) Comparison at various rotation angles of MROGH and proposed Scheme

The precision value of the proposed scheme and MROGH scheme is 94.51% and 84% respectively as shown in figure 4. The proposed scheme performs better than MORGH scheme.

iii. Recall

Recall is the activity of recollecting something educated or experienced. Recall rate is defined as shown in equation (6):

$$\text{Recall} = \frac{\text{Tp}}{\text{Tp+Fn}} \qquad (6)$$

where Tp is the quantity of different pixels named altered (True Positive) and Fn (False Negative)

Using the formula from equation (6) a graph is plotted between recall and rotation angle and compared between existing algorithm and proposed algorithm as shown in figure 5.

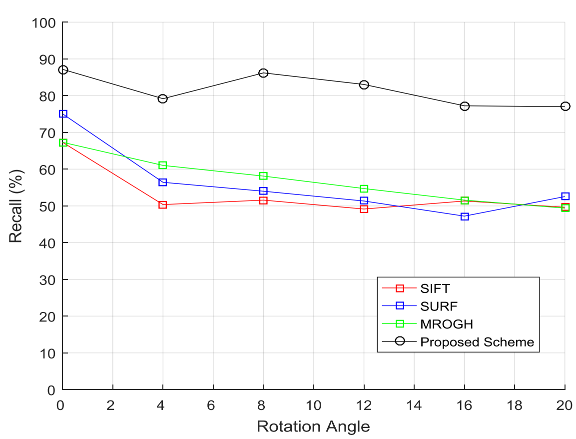


Figure 5: Recall (%) Comparison at various rotation angles of MROGH and proposed Scheme

This recall value of the proposed scheme and MROGH scheme is 81% and 56% respectively as shown in figure 5 The proposed scheme performs better than MORGH scheme.

iv. F1 score

The performance parameter F1 score is calculated in terms of precision and recall as shown in equation 7:

$$F_1\text{score} = 2.\frac{Precision.Recall}{Precision+Recall} \qquad (7)$$

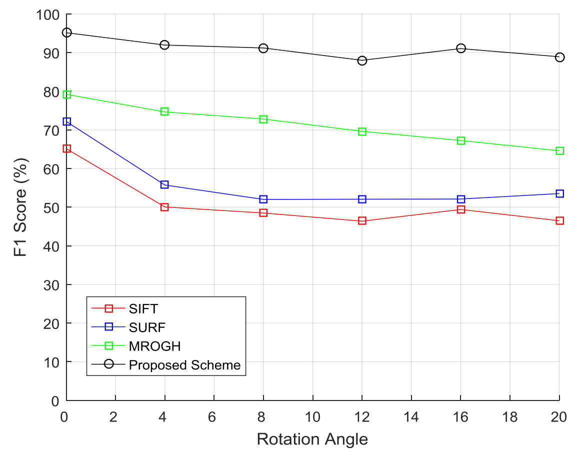


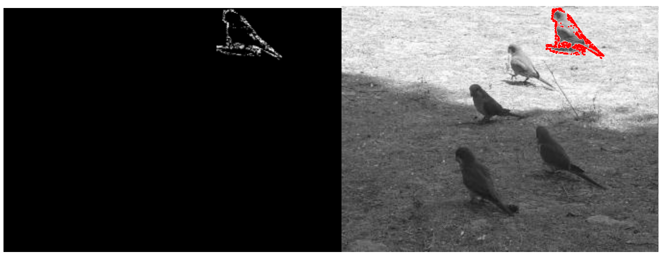
Figure 6: F1- score (%) Comparison at various rotation angles of MROGH and Proposed Scheme

The F1-score value of the proposed scheme and MROGH scheme is 92% and 52% respectively as shown in figure 6. The proposed algorithm SURF-PHA outperforms the existing algorithm MROGH. This is because the existing method MROGH is based on identification of key points which are in the form of points. This results into error computed in term of MSE (Mean squared error). The proposed algorithm PHA reduces the MSE by clustering the points. Thus making regions of similar points as one cluster.

v. Effect of using Proposed algorithm

Fig. 7 demonstrates a case meant for Copy-move forgery & its identification by the proposed technique. Original image is shown in Fig. 7(a) & the forged image shown in Fig. 7 (b) In the former, a parrot on the top was copied and pasted on its top right side. Clustered Outline of Copy-Paste Forgery using PHA clustering is shown in Fig. 7(c) Super Imposed Clustered Feature points on Original Image is shown in Fig.7(d) without using any Image editing software.

(a) (b)



(c) (d)

Technique	Evaluation metric (%)		
	(a)	(b)	(c)
	Precision	Recall	F1 Score
SIFT	62.26%	69.56%	65.71%
SURF	69.43%	74.85%	72.04%
MROGH	98.70%	67.01%	79.83%
SURF-PHA	98.86%	99.40%	99.13%

Figure 7. (a) original image. (b) forged image. (c) Clustered Outline of Copy-Paste Forgery using PHA clustering (d) Super Imposed Clustered Feature points on Original Image

vi. Comparative Analysis

Comparative analysis of different approaches such as SIFT, SURF, MROGH and SURF-PHA using Precision, Recall and F1-score is shown in table 2

Table 2: Execution assessment under revolution controls: (a) Precision (b) recall. (c) F1score

V. CONCULSION AND FUTURE SCOPE

Duplicate move is the most well-known picture altering procedure used because of its straightforwardness and adequacy. In this method parts of the picture are replicated, moved to a coveted area & glued. To this end, it's extremely vital and difficult to discover the successful strategies to identify computerized picture phonies. In this work, a quick key point based strategy to identify picture duplicate move phony is proposed in light of the SIFT (scale-invariant component change) descriptors. It is demonstrated that the proposed strategy is substantial in identifying the picture district duplication and very vigorous to added substance commotion and obscuring. Likewise contrasted with existing strategies proposed strategy gives better predictable Precision, recall and F1 Score for the considered dataset. Thus algorithm is robust method to detect image copy-move forgery. In future work, the detection speed of the proposed scheme by means of parallel programming can be improved and applications of proposed technique in fake currency recognition can be studied.

References

[1] https://www.researchgate.net/figure/Example-of-copy-move-forgery-a-Original-image-b-Forged-image-duplicated-object-h_fig1_317495890.

[2] V. Christlein, C. Riess, J. Jordan, C. Riess, and E. Angelopoulou, “An evaluation of popular copy-move forgery detection approaches,” Information Forensics and Security, IEEE Transactions on, vol. 7, no. 6, pp. 1841–1854, 2012.

[3] O. M. Alqershi, B. E. Khoo, “Passive detection of copy-move forgery in digital images: State-of-the-art,” Forensic Science International, vol. 231, no. 1, pp. 284–95, 2013.

[4] M. Emam, Q. Han, and X. Niu, “PCET based copy-move forgery detection in images under geometric transforms,” Multimedia Tools and Applications, vol. 75, no. 18, pp. 11513-11527, 2016.

[5] J. C. Lee, “Copy-move image forgery detection based on Gabor magnitude,” Journal of Visual Communication and Image Representation, vol. 31, pp. 320-334, 2015.

[6] I. Amerini, L. Ballan, R. Caldelli, A. Del Bimbo, and G. Serra, “A siftbased forensic method for copy–move attack detection and transformation recovery,” Information Forensics and Security, IEEE Transactions on, vol. 6, no. 3, pp. 1099–1110, 2011.

[7] B. Shivakumar and L. D. S. S. Baboo, “Detection of region duplication forgery in digital images using surf,” IJCSI International Journal of Computer Science Issues, vol. 8, no. 4, pp. 199-205, 2011.

[8] H. Huang, W. Guo, and Y. Zhang, “Detection of copy-move forgery in digital images using sift algorithm,” in Computational Intelligence and Industrial Application, 2008. PACIIA’08. Pacific-Asia Workshop on, vol. 2. IEEE, 2008, pp. 272–276.

[9] D. G. Lowe, “Distinctive image features from scale-invariant keypoints,” International journal of computer vision, vol. 60, no. 2, pp. 91–110, 2004.

[10] X. Pan and S. Lyu, “Detecting image region duplication using sift features,” in Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on. IEEE, 2010, pp. 1706–1709.

[11] Emam , Mahmoud, et al. "A robust detection algorithm for image Copy-Move forgery in smooth regions." Circuits, System and

Simulation (ICCSS), 2017 International Conference on. IEEE , 2017.
[12] Yadav, Preeti, and Yogesh Rathore. "Detection of copy-move forgery of images using discrete wavelet transform." International Journal on Computer Science and Engineering4.4 (2012): 565.
[13] Jaberi, Maryam, et al. "Improving the detection and localization of duplicated regions in copy-move image forgery." Digital Signal Processing (DSP), 2013 18th International Conference on. IEEE, 2013.
[14] Y. Lu, Y. Wan, Clustering by sorting potential values (CSPV): A novel potential-based clustering method, Pattern Recognition, 45(9) (2012), pp. 3512–3522.

Authors Profile

Suman Deswal, is pursuing her Ph.D from DCR University of Science & Technology, Murthal, India. She has completed her M.Tech (Computer Science & Engg.) from Kurukshetra University, Kurukshetra, India and B.Tech (Computer Science & Engg.) from CR State College of Engg., Murthal, India in 2009 and 1998 respectively. She possesses 16 years of teaching experience and is presently working as Assistant Professor in the department of Computer Science and Engg. at DCR University of Science and Technology, Murthal, India. She has published more than 30 papers in various international journals.

Pooja Devi , is pursuing her M.Tech (Computer Science & Engg) from DCR University of Science & Technology, Murthal, India. She has completed her B.Tech (Computer Science & Engg.) from IITM, Gannaur, India in 2015

