Al-Mustansiriyah Journal of Science
ISSN: 1814-635X (print), ISSN:2521-3520 (online) Volume 31, Issue 1, 2020 DOI: http://doi.org/10.23851/mjs.v31i1.748

Research Article Open Access

# False Matches Removing in Copy-Move Forgery Detection Algorithms

**Muthana S. Mahdi, Saad N. Alsaad***

Department of Computer Science, Collage of Science, Mustansiriyah University, Baghdad, IRAQ.
*Correspondent author: dr.alsaadcs@uomustansiriyah.edu.iq

**ArticleInfo**

Received
17/09/2019

Accepted
03/11/2019

Published
01/03/2020

## ABSTRACT

Today the technology age is characterized by spreading of digital images. The most common form of transfer the information in magazines, newspapers, scientific journals and all types of social media. This huge use of images technology has been accompanied by an evolution in editing tools of image processing which make modifying and editing an image is very simple. Nowadays, the circulation of such forgery images, which distort the truth, has become common, intentionally or unintentionally. Nowadays many methods of copy-move forgery detection which is one of the most important and popular method of image forgery are available. Most of these methods suffer from the problem of producing false matches as false positives in flat regions. This paper presents an algorithm of the Copy-Move forgery detection using SIFT algorithm with effective method to remove the false positives by rejecting all key-points in matches list that own a neighbor less than the threshold. The accuracy of the proposed algorithm was 95 %. The experimental results refer that the proposed method of false positives removing can remove false matches accurately and quickly.

**KEYWORDS**: Copy-move; Image forgery detection; Features extraction; Digital forensics; Multimedia security; False positive removing.

**الخلاصة**

يعتبر تأمين المعلومات العملية الاكثر أهمية لغرض اتمام التواصل وتخزين المعلومات.من اجل تأمين المعلومات كالتحقق من صحة البيانات وتكامل البيانات وسريتها يتم استخدام خوارزميات التشفير الجزء الاكثر أهمية في أي خوارزمية تشفير هو المفتاح والذي يحدد ما اذا كان النظام قويا كفاية ام لا. نقترح في هذه الورقة طريقة جديدة لإنشاء المفاتيح بالاعتماد على نوعين من نظريات الفوضى من اجل تحسين امن خوارزميات التشفير. اساس هذا الاقتراح هو اكتشاف طريقة جديدة لإنشاء ارقام عشوائية باستخدام خريطة لورنز ثلاثية الابعاد وخريطة هينون ثنائية الابعاد .اجتازت المفاتيح التي تم انشاؤها حديثا بنجاح مجموعة الاختبارات الاحصائية للمعهد الوطني للمعايير والتقنية.

## INTRODUCTION

In general, images are considered effective tools for human communication comparing with texts. The visual system can obtain pictorial information extremely faster than any other type of information. This information forms approximately 75% of information perceived by a visual system [1]. Nowadays different applications like newspapers, social media applications, Journals, courtrooms, and others are dealing daily with thousands of digital images. These images can be easily forged without leaving any obvious signs, due to the advancement of the digital image processing software and editing tools. Sometimes it is very difficult to know if the digital image is forged or not by the naked eyes. In many cases, the purpose of this tampering is to deliberately influence the attention of the recipient, so, it has become very important to confirm the reliability and authenticity of the images [2].

There are many types of image forgery, but the most popular type is Copy-Move forgery (CMF) or cloning, easy to implement and difficult to detect. This type of forger copies part or parts of an image and paste to it again. The copied regions can be in any position or can have any shape, including rotation, translation, scaling, and combining of many types.

This kind of forgery is more difficult to detect than other kinds because the usual methods of detecting incompatibilities that use statistical measures to compare different parts of the image are useless for CMF detection [3].

Many methods of copy-move forgery detection (CMFD) are available. Generally, they can be classified into two main criteria: Block-based and Key-Point-based methods [4].

Copy-Move forgery detection can be performed either using one of these methods or a combination of both. In General, Key-point-based methods work well in terms of robustness, memory usage, and computation time compared to block generation methods, but at the same time, they produce more false positives (false matches) in flat regions [5].

Generally, the main steps in detection of copy-move forgery are pre-processing, features extraction, matching, and post-processing. Figure 1 represents a general framework for detection of copy-move forgery [6].

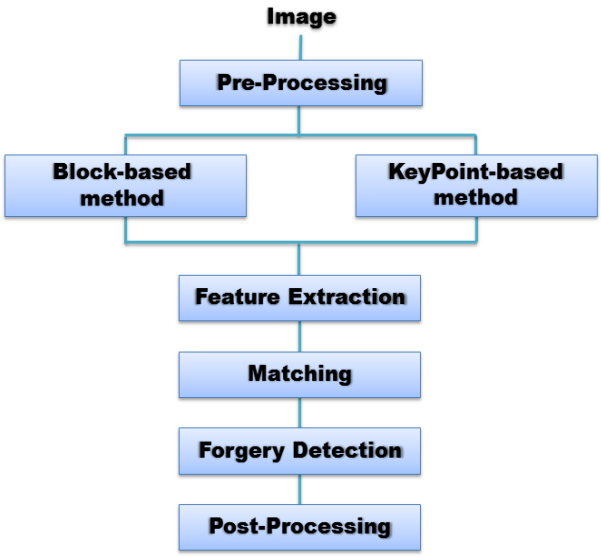


**Figure 1.** General Framework for Copy-Move Forgery detection [4].

**Pre-Processing:** Before features extraction, some operations are performed on the images in order to enhance the structural changes that have occurred in the images because of the forgery. In addition to enhancement image data to better describes the geometrical information of the image texture. This process is applied in both block-based and key-point-based techniques. It depends on the application itself. Different preprocessing functions are applied such as convert the color image into a grayscale, dimension reduction, image resizing, and image filtering [7].

**Feature Extraction:** this process to extract or find the important features of the input image. The goal of feature extraction is to compute the specific representation of the data that can highlight the relevant information. Features must have two basic requirements: The redundancy in the original image should be avoided and dimensionality of data should be reduced. In this step, feature vectors are extracted. The block-based method is used either in overlapping or non-overlapping blocks. The features are extracted corresponding to each block of the image. In the case of the key point based methods, the features corresponding to key-points are extracted [8].

**Matching:** In CMF, different parts are copied and moved to the same image, so there is a strong correlation between these parts. This can be used as evidence to detect the forgery. But the main challenge is to find effective features and matching algorithms to find the associated regions. The feature matching is performed to find a high similarity or matching between feature descriptors. If the similarity between feature descriptors is found then it's interpreted as an indicator for duplicated regions [9]. Many methods can be used to identify these similarities. The common methods either sort the feature vectors lexicographically and calculate the Euclidean distance (using the formula shown in the Equation (1) [10]) between the adjacent stored vectors, or building the k-dimension tree (k is the number of dimensions) contain all the feature vectors and finding the k-Nearest Neighbors for each feature. However, the incorrect matches in some areas of the image can be appearing when the image contains a similar texture such as the sea or the sky, therefore, these erroneous matches should be deleted [11].

$$d(\mathrm{p},\mathrm{q}) = d(\mathrm{q},\mathrm{p}) = \sqrt{\sum_{i=1}^{n}(q_i - p_i)^2} \qquad (1)$$

Where, d: The Euclidean distance, p: First point or vector, q: Second point or vector, n: number of dimensions

**Post-Processing:** The goal of this process is to preserve the matches that exhibit similar behavior. When the image has been classified as non-authentic, the post-processing helps to display the positions of the matched regions with a certain color or shape and reduce the

Al-Mustansiriyah Journal of Science
ISSN: 1814-635X (print), ISSN:2521-3520 (online) Volume 31, Issue 1, 2020 DOI: http://doi.org/10.23851/mjs.v31i1.748

false matches. So, in this paper proposes a method to remove the false matches (false positives) that can remove accurately and quickly the false matches. [12].

## RELATED WORKS

In recent years, many schemes have been proposed to detect a CMF.

Zhao and Guo [13] proposed a method to detect a CMF based on Discrete Cosine Transform (DCT) and Singular Value Decomposition (SVD). The experimental results show that this method can detect a CMF even when an image is distorted by Gaussian Noise or image blurring. A different approach based on Local Binary Pattern (LBP) and neighbor clustering have been presented by Al-Sawadi *et. al.* [14]. Chen [15] proposed a method by extracting Harris corner points as key-points and using step sector statistics to represent the small circle image area around each Harris point. Singh *et. al.* [16] employed DCT and scale invariant feature transform (SIFT) to extract the features from the image and then matching those features to detect the forgery in the image and also perform the localization of the forged regions. Jen-Chun *et. al.*, [17] divided the image into overlapped blocks of size. Afterward, for each block the Gabor filter is applied and the 12 bin "Histogram of Oriented Gabor Magnitude" (HOGM) descriptor is computed. Then, feature vectors are sorted lexicographically and matched blocks are found between adjacent features using Euclidean distance. Although this work can detect the forgery, it produces false matches as false positives in flat regions.

Lee *et. al.* [18] divided the image into overlapping blocks and applied the histogram of orientated gradients of each block. Although this approach is capable of detecting multiple examples of CMF, it is weak with rotation and scaling over large areas. Recently, a different key-point-based method using a speeded up robust features (SURF) and adaptive overlapped segmentation has been presented by Sreelakshmy *et. al.* [19]. Das *et. al.* [20] used the stationary wavelet transform (SWT) to decompose the image and Scale Invariant Feature Transform (SIFT) algorithm to extract features. This method work well, but it produces some false positives. Bin Yang *et. al.* [21] presented a CMFD strategy using a strategy of key-points distribution to key-points selection. This strategy can detect duplicate areas. But it requires an additional cost of computational.

The major drawback of most CMFD methods is producing false matches as false positives in flat regions [5].

In this paper, an effective and robust method to remove the false matches (false positives) in CMFD algorithms has been proposed that can remove accurately and quickly the false matches.

## PROPOSED METHOD

The CMF in digital images can be done in one region or more. The task of detection methods is to determine whether the image contains duplicated areas or not. Since the size and shape of the duplicated areas are unknown, it's definitely computationally impossible attempt to compare pixels by pixels. In order to make a forgery detection algorithm efficient and has less computational complexity, the robust features are used.

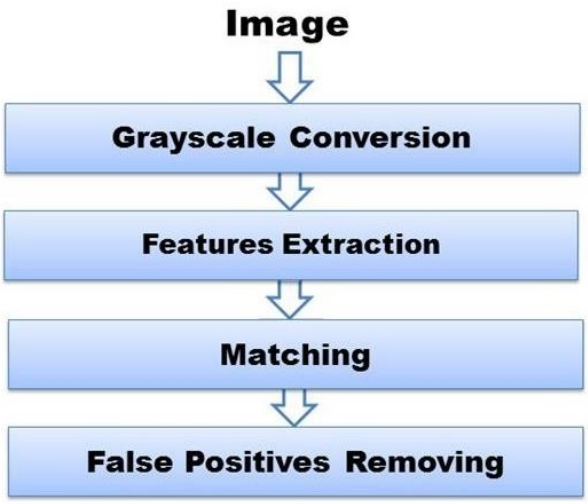


**Figure 2.** The general framework of the proposed algorithm.

This scheme is pictorially shown in figure 2. First, the input image is converted to grayscale. The SIFT algorithm is applied to extract key-points features and their descriptors. Then, the k-nearest neighbors of each key-point are found and the matched key-points which satisfy the conditions are determined. Finally, the

proposed method to removing the false matches (false positives) regions is applied.

## Grayscale conversion

In order to obtain a better representation, the color image is converted to grayscale as a pre-processing step.

## Features Extraction

There are a lot of key-points in each image, which can be extracted to provide a description of the image. With the implementation of the SIFT technique, a large number of the featured key-points can be withdrawn which are invariant to different factors such as scaling or rotation and robust to change in illumination and noise. So, in this phase, the SIFT algorithm is used to extract the important key-points of the image. The details of the SIFT algorithm are illustrated in [22].

## Matching

A matching procedure that is a generalization of the third neighbor is used to be able to deal with multiple copies. This means the two key-points are considered matched only if the following constraint is satisfied:

$$\frac{d1}{d3} < 0.5 \qquad (2)$$

Where:

- d1 is the Euclidean distance between the key-point descriptor and the k-nearest neighbors.
- d3 is the Euclidean distance between the key-point descriptor and the k+2 nearest neighbors.
- k is 1, 2…10.

To giving more accuracy to the matching process, the mini distance condition has been added. The distance between the locations of each two matching key-points must be greater than the mini distance threshold, as follows:

$$\text{The distance between key-point location \& KNN location} > \text{mini distance} \qquad (3)$$

Where:

- KNN is K-nearest neighbors.
- K is 1, 2…10.
- mini distance is 30.

Finally, by iterating this KNN and mini distance strategy on all key points, the set of candidate matches can be obtained.

## False Positives Removing

An image can contain regions with a very similar texture. This leads to show false matches (false positives) in some area of the image. A method to remove the false positive matches from the image is proposed by rejecting all key-points in matches list that own a neighbor less than N (the default value of N is 2). All threshold values used in this method have been determined experimentally.

The main steps of this work illustrated in figure 3 and algorithm 1 respectively.

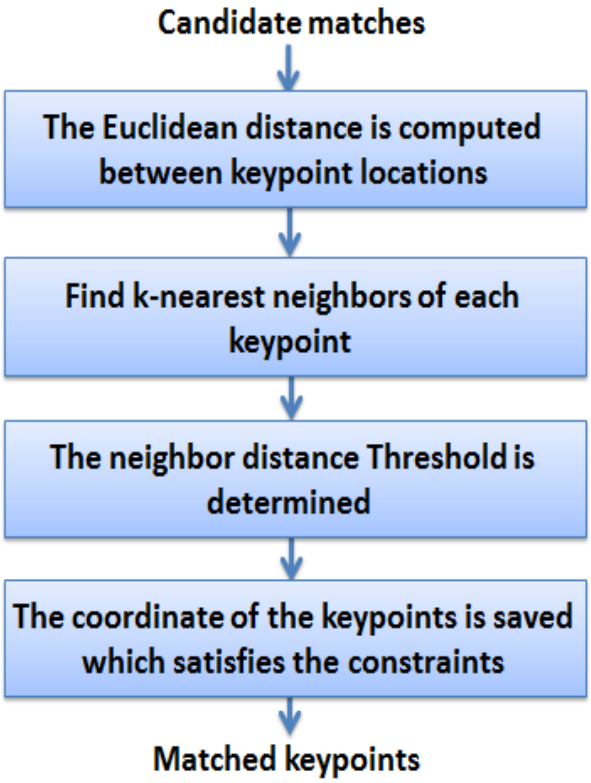


**Figure 3.** The main steps of the proposed method to remove the false positive.

The idea of this method can be summarized as follows: In general, the copied region has a very similarity with the original, this leads to production of several matched key-points between them, due the Euclidean distances are low between them. Therefore, any key-point in the matches list will have many neighbors because these key-points describe a specific region, so their locations are close to each other, thus can be considered robust key-points.

Al-Mustansiriyah Journal of Science
ISSN: 1814-635X (print), ISSN:2521-3520 (online) Volume 31, Issue 1, 2020 DOI: http://doi.org/10.23851/mjs.v31i1.748

On the other hand, in the case of false matches (false positives). These key-points will be scattered in different regions of the image, due they consist of the chaos of background. This means, their locations are not close to each other, thus they will not have multiple neighbors in the matches list.

So, the number of neighbors of each key-point in the matches list can be computed. Then, each key-point which has neighbors less than N value will be removed.

**Algorithm 1:** (False Positives Removing).

**Input:** Candidate matches.

**Output:** Matched key-point.

Begin

**Step1:** *The Euclidean distance is computed between locations of each key-point with others.*

**Step2:** *The first k-nearest neighbors of each key-point are determined, (K= N+1).*

**Step3:** *The Threshold of a neighbor distance (TND) is determined based on the size of the inputting image, as follow:*
*If image size >= (2000*1500) pixel then TND = 40 else TND =30.*

**Step4:** *The coordinate of the key-points are saved which satisfy:*
*The Euclidean distances between key-point location and each one of the k-nearest neighbors < TND*

End

## THE EXPERIMENTAL RESULTS

The false matches can be appearing in any image contain a very similar texture, that leads to show the false positives in some area of the image.

The proposed method has been implemented using Windows 7 and Matlab R2016a. VLFeat library (written in C++ language) is exploited, to increase the implementation speed of the SIFT algorithm.

The testing and evaluating process has been performed using 100 images determined randomly from dataset "MICC" [23].

This collection is composed of images that have different size in JPEG images format.

The number of original images that are exactly detected as original was 47 from 50. The number of forged images that are detected as forged was 48 from 50. The processing time average for all tested images was 3.02 second. Table 1 presents the final results of testing.

**Table 1.** Final results of the test.

| TP | TN | FP | FN |
|---|---|---|---|
| 48 | 47 | 3 | 2 |

Where TP (true positive) is the number of forged images that are detected as forged, TN (true negative) is the number of original images that are detected as original, FP (false positive) is the number of original images that are detected as forged, FN (false negative) is the number of forged images that are detected as original. The accuracy of the proposed algorithm was:

$$\text{Accuracy} = \frac{TP + TN}{TP + FN + TN + FP} = 0.95 \qquad (4)$$

The proposed method of false positives removing has been able to remove all false matches accurately and quickly.

The average time to remove the false matches for all images was 0.7 seconds.

Figure (4) and Figure (5) illustrates the visual results of the proposed method.

In the figure above, the proposed algorithm took about 2 s, 3 s, and 4 s respectively from left to right to determine the tampered areas.

To evaluate the accuracy of the proposed algorithm, it has been compared with three algorithms. Figure (6) illustrates the overall accuracy of these forgery detection algorithms.

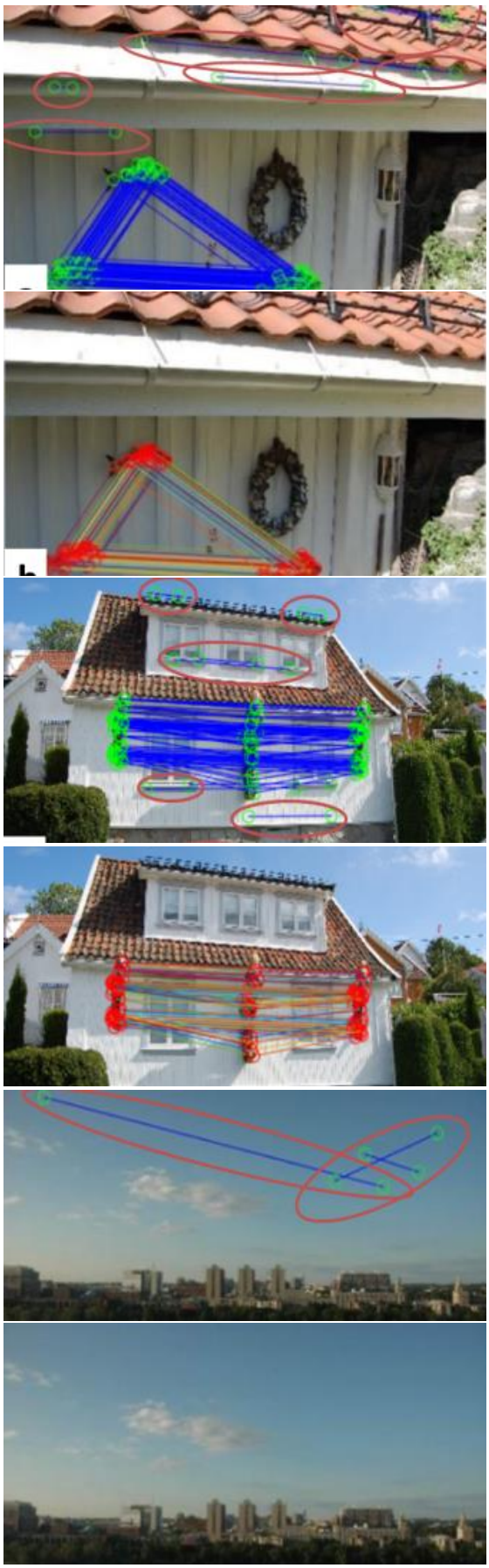

**Figure 4.** The final result of the proposed CMFD algorithm before and after applying the removal of the false positives.

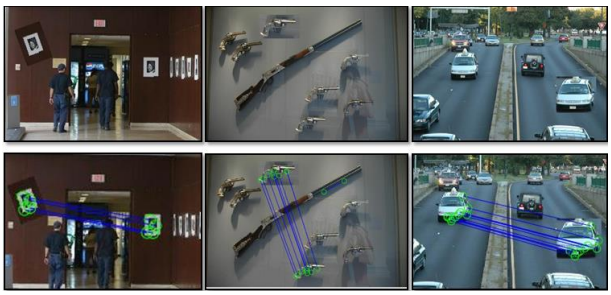

**Figure 5.** The final result of the proposed CMFD algorithm, images in the top row before execution and in the bottom after execution.

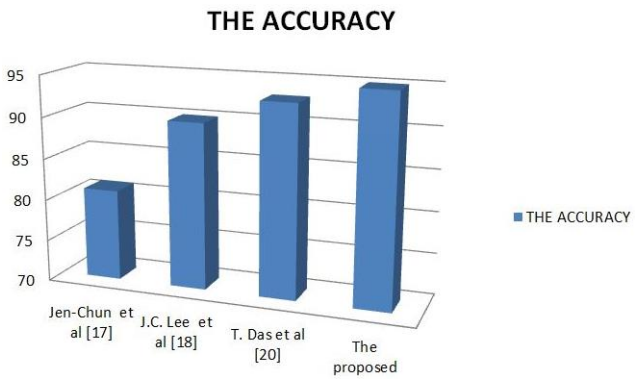


**Figure 6.** The comparison process between the proposed method and other related works.

## CONCLUSIONS

The detection of forgery in the digital image is an interesting research topic in forensic science. The specific type of image tampering (copy-move forgery) can be considered one of the emerging problems in the field of digital image forensics. There is a large number of published papers on copy-move detection can be found in the literature. The number of these papers is increasing. However, most of these algorithms produce false matches in flat regions because the image can contain regions with very similar texture. In this paper, an algorithm of CMFD is proposed with effective method to remove the false matches (false positives). Many experiments have been performed to suggest suitable values of all thresholds used in the algorithm. The experiments and analysis proved this method has lower computational complexity and can remove accurately and quickly the false matches.

## REFERENCES

[1] N. K. Gill, R. Garg, and E. A. Doegar, "A review paper on digital image forgery detection techniques," in 2017 8th International Conference

on Computing, Communication and Networking Technologies (ICCCNT), 2017, pp. 1-7: IEEE.

[2] M. Zandi, A. Mahmoudi-Aznaveh, A. J. I. T. o. I. F. Talebpour, and Security, "Iterative copy-move forgery detection based on a new interest point detector," vol. 11, no. 11, pp. 2499-2512, 2016.

[3] M. Kumar and S. J. A. J. o. F. S. Srivastava, "Image forgery detection based on physics and pixels: a study," vol. 51, no. 2, pp. 119-134, 2019.

[4] S. Sadeghi, S. Dadkhah, H. A. Jalab, G. Mazzola, D. J. P. A. Uliyan, and Applications, "State of the art in passive digital image forgery detection: copy-move image forgery," vol. 21, no. 2, pp. 291-306, 2018.

[5] S. Panda and M. Mishra, "Passive Techniques of Digital Image Forgery Detection: Developments and Challenges," in Advances in Electronics, Communication and Computing: Springer, 2018, pp. 281-290.

[6] A. Doegar, M. Dutta, and G. Kumar, "A Review of Passive Image Cloning Detection Approaches," in Proceedings of 2nd International Conference on Communication, Computing and Networking, 2019, pp. 469-478: Springer.

[7] A. Dixit, R. J. I. J. o. S. P. Gupta, Image Processing, and P. Recognition, "Copy-Move Image Forgery Detection using Frequency-based Techniques: A Review," vol. 9, no. 3, pp. 71-88, 2016.

[8] K. Asghar, Z. Habib, and M. J. A. J. o. F. S. Hussain, "Copy-move and splicing image forgery detection and localization techniques: a review," vol. 49, no. 3, pp. 281-307, 2017.

[9] M. A. Qureshi and M. J. S. P. I. C. Deriche, "A bibliography of pixel-based blind image forgery detection techniques," vol. 39, pp. 46-74, 2015.

[10] G. Dougherty, Pattern recognition and classification: an introduction. Springer Science & Business Media, 2012.

[11] Y. J. F. s. i. Li, "Image copy-move forgery detection based on polar cosine transform and approximate nearest neighbor searching," vol. 224, no. 1-3, pp. 59-67, 2013.

[12] N. B. A. Warif et al., "Copy-move forgery detection: survey, challenges and future directions," vol. 75, pp. 259-278, 2016.

[13] J. Zhao and J. J. F. s. i. Guo, "Passive forensics for copy-move image forgery using a method based on DCT and SVD," vol. 233, no. 1-3, pp. 158-166, 2013.

[14] M. AlSawadi, G. Muhammad, M. Hussain, and G. Bebis, "Copy-move image forgery detection using local binary pattern and neighborhood clustering," in 2013 European Modelling Symposium, 2013, pp. 249-254: IEEE.

[15] L. Chen, W. Lu, J. Ni, W. Sun, J. J. J. o. V. C. Huang, and I. Representation, "Region duplication detection based on Harris corner points and step sector statistics," vol. 24, no. 3, pp. 244-254, 2013.

[16] R. Singh, A. Oberoi, and N. J. I. J. o. C. A. Goel, "Copy move forgery detection on digital images," vol. 98, no. 9, 2014.

[17] J.-C. J. J. o. V. C. Lee and I. Representation, "Copy-move image forgery detection based on Gabor magnitude," vol. 31, pp. 320-334, 2015.

[18] J.-C. Lee, C.-P. Chang, and W.-K. J. I. S. Chen, "Detection of copy-move image forgery using histogram of orientated gradients," vol. 321, pp. 250-262, 2015.

[19] I. Sreelakshmy and J. Anver, "An improved method for copy-move forgery detection in digital forensic," in 2016 Online International Conference on Green Engineering and Technologies (IC-GET), 2016, pp. 1-4: IEEE.

[20] T. Das, R. Hasan, M. R. Azam, and J. Uddin, "A robust method for detecting copy-move image forgery using stationary wavelet transform and scale invariant feature transform," in 2018 International Conference on Computer, Communication, Chemical, Material and Electronic Engineering (IC4ME2), 2018, pp. 1-4: IEEE.

[21] B. Yang, X. Sun, H. Guo, Z. Xia, X. J. M. T. Chen, and Applications, "A copy-move forgery detection method based on CMFD-SIFT," vol. 77, no. 1, pp. 837-855, 2018.

[22] D. G. J. I. j. o. c. v. Lowe, "Distinctive image features from scale-invariant keypoints," vol. 60, no. 2, pp. 91-110, 2004.

[23] I. Amerini, L. Ballan, R. Caldelli, A. Del Bimbo, G. J. I. t. o. i. f. Serra, and security, "A sift-based forensic method for copy-move attack detection and transformation recovery," vol. 6, no. 3, pp. 1099-1110, 2011.