Copy-Move Image Forgery Detection and Classification Using Sift and Birch Approaches

Zhanquan Zhuang

College of Big Data and Internet, Shenzhen Technology University, 3002 Lantian Road, Pingshan District, Shenzhen, Guangdong, 518118, China

Abstract. Copy-move forgery detection (CMFD) continues to pose significant challenges in digital image forensics, particularly under geometric transformations such as rotation and scaling. This study presents a detection pipeline that couples Scale-Invariant Feature Transform (SIFT) keypoints with the Balanced Iterative Reducing and Clustering using Hierarchies (BIRCH) algorithm. Leveraging BIRCH's robustness to noise, computational efficiency and relatively mild hyper-parameter sensitivity, the proposed framework effectively suppresses spurious correspondences and isolates candidate tampered regions with high confidence. Performance is evaluated on the MICC-F220 benchmark and benchmarked against two alternative clustering strategies-Hierarchical Agglomerative Clustering (HAC) and Density-Based Spatial Clustering of Applications with Noise (DBSCAN)-each configured with their best-reported parameter sets. Under the tested settings, the SIFT + BIRCH scheme attains a true-positive rate of 99 % while sustaining a low false-positive rate, outperforming both baselines. The results underscore the suitability of hierarchical clustering for spatially grouping SIFT keypoints and highlight BIRCH as a promising direction for future CMFD research, particularly when integrated with complementary outlier-rejection mechanisms.

1 Introduction

Photographs have been altered since the birth of the medium, and even analogue images are rife with celebrated forgeries. Early cases include the excision of the disgraced Soviet politician Nikolai Yezhov from official portraits, the Iranian press release of 2008 in which a non-existent fourth missile was composited into a launch scene, and the well-known "Lincoln's head on Calhoun's body" montage that briefly damaged President Lincoln's reputation. Such manipulations once required painstaking darkroom work; today, consumer-grade editing software and global distribution channels render copy-move tampering-duplicating a region of an image and pasting it elsewhere-both covert and trivial to disseminate.

202200202104@stumail.sztu.edu.cn

Copy-move manipulation is typically employed to conceal, replicate, or re-contextualise visual content for propaganda or misdirection. Image-forensics technology can be categorized into active and passive approaches [1]. Active schemes embed an authentication signal (e.g. a watermark) at capture time, but they require hardware support and may degrade image quality. Passive schemes rely solely on statistical or geometric traces remaining in the image data and therefore dominate academic research on copy–move detection.

Passive detectors can be grouped into block-based and key-point-based paradigms. Block-based schemes divide the image into overlapping patches, extract hand-crafted features, and flag tampering when similar blocks appear in spatially distant locations. Although straightforward and reasonably robust to mild JPEG compression, block-based methods are computationally demanding and fragile under rotation or scaling. Key-point approaches, by contrast, first locate distinctive local features-such as SIFT or ORB-whose descriptors are intrinsically invariant to many geometric transforms. These methods excel under rotation and scale changes but may fail in low-texture areas where few key points emerge.

Recent studies have applied convolutional neural networks (CNNs) to forensics, replacing manual features with automatically learned representations [2]. Some architectures ingest the RGB image or a transformed domain representation, learn discriminative block descriptors or simulate key-point features, and then integrate back-end matching and clustering to identify tampered regions. Other work divides the image into small blocks and employs a CNN to estimate inter-block similarity. Despite high training cost and data dependence, deep models can, in principle, remain robust to compression, noise, and certain geometric deformations.

In terms of block-based detection, researchers initially used simple features in the time domain or frequency domain, such as partitioning the image into fixed-size blocks and comparing pixels one by one. Such methods turn out to be more precise than the keypoint-based ones [3], but they face the problem of large computational complexity and susceptibility to geometric transformations. Fridrich et al. used DCT to enhance robustness [4]; Cozzolino et al. used PatchMatch to improve matching efficiency [5], but it is still difficult to cope with attacks such as scaling, rotation, and noise addition. Block methods are suitable for smooth images and static scene detection, but they are time-consuming to process large images and often require dimensionality reduction or clustering to accelerate.

Key-point strategies capitalise on the scale- and rotation-invariance of SIFT [6]. Pan and Lyu first used keypoint matching combined with RANSAC for robust copy-move tampering detection [7], their method was shown to be very stable to geometric transformations. Amerini et al. subsequently used hierarchical agglomerative clustering (HAC) to partition key point matches into spatial clusters and applied RANSAC separately to each cluster pair [8,9]. Alternative studies have proposed clustering directly in conceptual space instead of the image plane [10].

In summary, copy-move detection has evolved from early block methods to key point methods, and finally to an end-to-end detection system based on deep learning. The block method is simple but has poor transformation robustness; the key point method improves flexibility and accuracy; and the deep learning model is expensive but shows the potential to handle complex tampering. Future research will find a better balance between algorithm accuracy and efficiency.

The main contribution of this paper is to use the BIRCH clustering algorithm in the key point matching stage to improve the accuracy of the move-copy tampering detection of images based on the key point method, and conduct comparative experiments with the existing paper methods that use the HAC and DBSCAN clustering algorithms to optimize key point matching. At the same time, this paper will use public datasets and fair evaluation criteria to evaluate the accuracy of the scheme.

This article will first introduce the principles of the algorithm used in this article in Chapter 2, then introduce the data set, evaluation criteria and experimental results used in the test in Chapter 3, and finally Chapter 4 will involve future improvements and expectations of this method.

2 Method

As illustrated in Figure. 1, the proposed pipeline is organised into three sequential stages: (1) feature extraction, (2) clustering, and (3) region-of-interest (ROI) identification.

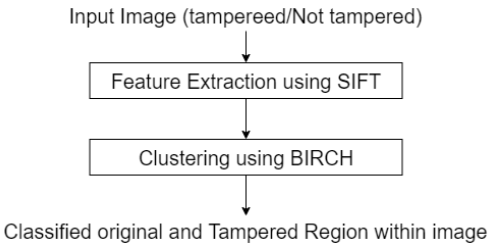

Figure 1. Block Diagram for Proposed Approach. (Picture credit: Original)

In the feature-extraction stage, the approach introduced in [11] is adopted to obtain a sufficiently dense set of SIFT keypoints. In particular, the contrast threshold is lowered so that an adequate number of interest points can still be detected in smooth or spatially limited copy-move regions.

Because no single clustering or segmentation algorithm-and no single set of parameters-performs optimally on all images, Step 2 builds on the DBSCAN-based SIFT optimisation proposed in [11] and introduces the BIRCH algorithm to further improve keypoint matching. The parameters of both DBSCAN and BIRCH are tuned empirically to alleviate mismatching and to provide a fair comparison between the two clustering strategies.

Finally, since green is the color that the human visual system is most sensitive to, and since grayscale images both reduce computational cost and often yield higher recognition accuracy than colour images, every RGB image $I(x, y)$ is converted into a luminance channel $L(x, y)$ using the perceptual weighting scheme described in [12, 13]:

$$L(x, y) = 0.2989 * r + 0.5870 * g + 0.1140 * b \quad (1)$$

2.1 Feature Extraction and Matching Similar Key-Point

Key points are extracted from an image using the SIFT algorithm [6]. Key points (interest points) in a picture that are unaffected by translations, rotations, scale changes, and minor affine transformations can be identified with SIFT. Four steps make up the SIFT algorithm: i) detecting prospective keypoints using DoG's extrema; ii) filtering key points based on contrast and edge thresholds; iii) determining the key point vector's predominant orientation; and iv) creating feature descriptors [14].

An image's scale-space representation is indicated. Formula (1) is subsequently employed to obtain the DoG image:

$$D(x, y, \sigma) = L(x, y, k\sigma) - L(x, y, \sigma) \quad (2)$$

where k is a predetermined parameter. The function $L(m, n, \sigma)$ is the convolution of the Gaussian kernel $G(m, n, \sigma)$ and the input image $I(m, n, \sigma)$ at various scales, which leads to:

$$L(x, y, \sigma) = I(x, y) \otimes G(x, y, \sigma) \quad (3)$$

The Gaussian kernel itself is: $G(m, n, \sigma) = \frac{1}{2\pi\sigma^2} e^{\frac{-(m^2+n^2)}{2\sigma^2}}$。

Stage ii. Candidate key-points that are deemed unstable are removed by means of the contrast and edge thresholds. To ensure adequate preservation of key-points in smooth regions, the present study adopts the strategy proposed in [11] of lowering the SIFT contrast threshold so that extrema with low contrast values are preserved. Nevertheless, an excessively small threshold must be avoided because it produces many spurious points.

Figure 2 demonstrates the sensitivity of SIFT to this parameter: the population of detected key-points varies markedly as the contrast threshold changes. Empirical evaluation indicates that a threshold of 0.01 yields the best trade-off between recall and reliability in the current data set.

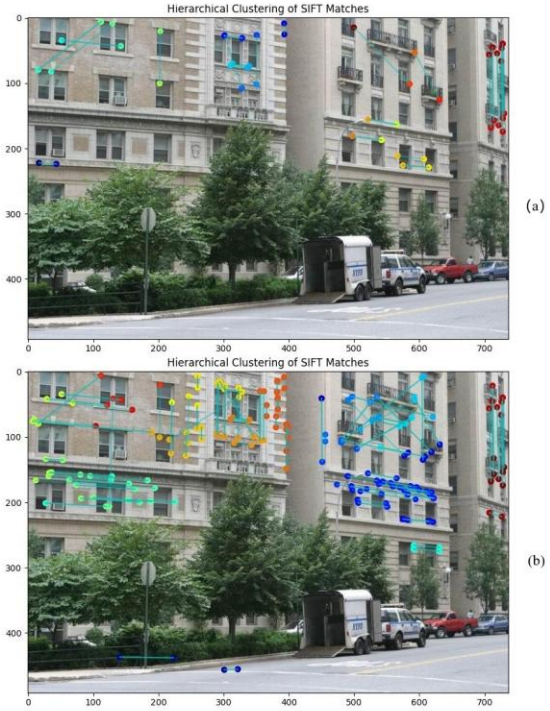

Figure 2. Influence of the contrast threshold on the number of SIFT key-points. (Picture credit: Original) (a) threshold = 0.10 (b) threshold = 0.01

Stage iii. Each candidate key-point is assigned a dominant orientation so that rotational invariance is preserved. A local gradient-orientation histogram is constructed, and the peak of orientation histogram is stored as the key-point's principal direction.

Stage iv. By encoding gradient information inside a 16x16 neighborhood in scale space centered on the key-point, a 128-dimensional descripto $r\{d_1, d_2, \ldots, d_n\}$ is calculated for each SIFT key-point$\{k_1, k_2, \ldots, k_n\}$. For both the source region and the possible copy-moved region, a collection of key points and their descriptions can thus be derived.

When a one-to-one correspondence is observed between two spatially disjoint key-point sets and the correspondences are mutually consistent, copy-move forgery is suspected. Similarity between any two descriptors is evaluated with Lowe's second-nearest-neighbour (2NN) ratio [6]. Let $d = \{d_1, d_2, \ldots, d_{n-1}\}$ be the ascending list of Euclidean distances between a query descriptor and the remaining $(n-1)$ ones, i.e.$d_1 \leq d_2 \leq \ldots \leq d_{n-1}$. A tentative match defined by d_k is considered reliable only if

$$\frac{d_k}{d_{k+1}} < t \qquad (4)$$

where the threshold t usually is set to 0.6 [6].

2.2 Different Clustering

Following feature extraction, similar matches are grouped through hierarchical clustering with the BIRCH algorithm. Two considerations motivate this choice.

First, unlike prototype-based methods such as k-means or conventional agglomerative schemes, BIRCH executes an initial rough clustering in a single pass over the data. Its efficiency is governed by two hyper-parameters: (i) the maximum radius threshold, which bounds the radius of any sub-cluster represented in the CF-tree, and (ii) the branching factor, which limits the number of children that a non-leaf node may possess. In this framework, leaf nodes store the final sub-clusters, whereas internal nodes merely guide the insertion path of incoming samples. Owing to its dynamic, B+-tree-like structure, BIRCH maintains near-linear time complexity and modest memory consumption, even when confronted with large sets of high-dimensional tampering features.

Section 2.1 demonstrated that lowering the SIFT contrast threshold markedly increases the number of keypoints; under this regime, the incremental compression performed by BIRCH further amplifies its runtime and memory advantages. As illustrated in Figure. 3, the algorithm organises data into a Clustering-Feature (CF) tree in which each node compactly stores the sufficient statistics (N, LS, SS) of the points it summarises, enabling rapid updates and queries without repeated scans of the raw feature set.

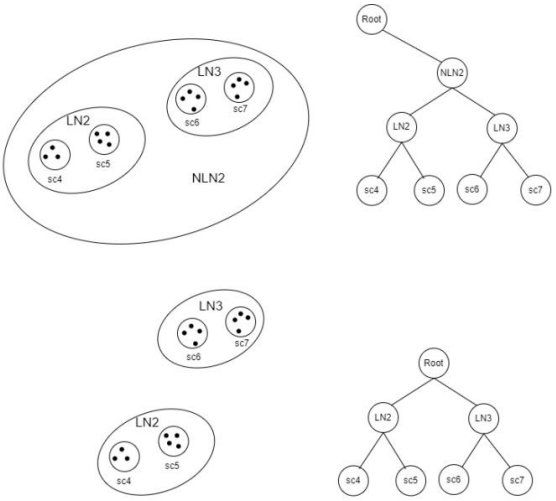

Figure 3. illustrates a Clustering-Feature (CF) tree, a balanced B^+-tree structure that BIRCH maintains dynamically during incremental insertion. (Picture credit: Original)

Secondly, hierarchical clustering basically has the advantages of other clustering methods: it can identify noise points, pre-process the data set for preliminary classification, and support the identification of distribution clusters with "quasi-hypersphere" structures, so it can also better identify image duplication areas.

Hierarchical clustering inherits the usual advantages of density-based methods: it suppresses noise, provides an initial coarse partition of the data, and efficiently reveals clusters with quasi-hyperspherical geometry-properties that facilitate the localisation of duplicated image areas.

By contrast, DBSCAN [12] defines clusters through ε-reachability. Given a data set D, the ε-neighbourhood of a point p is:

$$N_\varepsilon(p) = \{q \in D \mid dist(p, q) \leq \varepsilon\} \quad (5)$$

as well as p is a core point when:

$$N_\varepsilon(p) \geq MinPts \quad (6)$$

Points are accordingly labelled as core, border, or noise. Figure 4 sketches the resulting density-connectivity procedure. Nevertheless, DBSCAN is highly sensitive to its hyper-parameters $(\varepsilon, Minpts)$, and exhaustive experiments revealed that suitable values are image-dependent and non-trivial to select, which ultimately hampers detection accuracy.

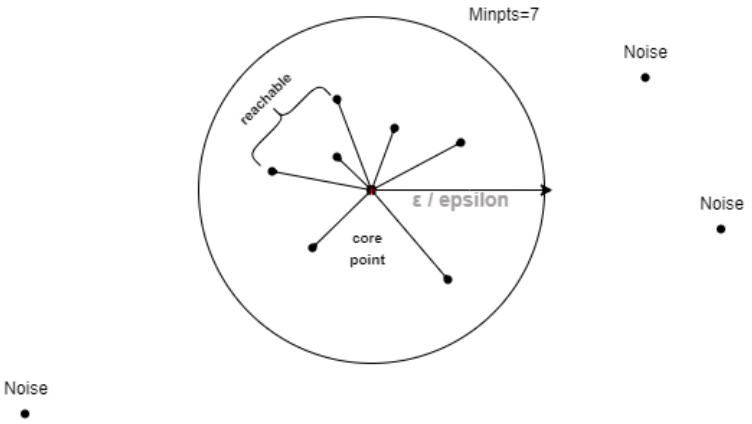

Figure 4. illustrates the ε-neighbourhood employed by DBSCAN. (Picture credit: Original)

Specifically, although the DBSCAN algorithm is robust to cluster shape and noise, and has faster clustering performance than general hierarchical clustering algorithms, based on practical experience, there are no other "non-spherical" structures in two-dimensional images; at the same time, because the DBSCAN algorithm is more sensitive to parameter selection than the hierarchical clustering algorithm, and there is no other simple mechanism to select different parameters according to different images, the DBSCAN clustering effect is not ideal in actual situations.

3 Experimental Results

3.1 Dataset

All experiments were executed on a Windows 11 workstation equipped with a i9-14900K processor and 32 GB of RAM. The implementation was developed in Python 3.13.2 with OpenCV 4.11.0 inside the Visual Studio Code environment. Source code is publicly available at https://github.com/ZZQ323/Sift-Clustering-trail.

The experiment was conducted on the MICC-F220 dataset, a widely recognized public benchmark for copy-move forgery detection. This dataset comprises 220 images (110 authentic and 110 forged), with resolutions varying between 722 × 480 and 800 × 600 pixels. In each manipulated image 1.2% of the total area is copied from the source image and subsequently pasted after undergoing geometric transformations such as translation, rotation, uniform/non-uniform scaling, or their composite operations. The original images are sourced from the Columbia Photographic Image Repository and privately curated photographic collections [13].

Comparative results obtained with the proposed method and competing approaches are reported in Section 3.2 and Section 3.3.

3.2 Evaluation Criteria

The performance of the proposed pipeline is assessed primarily through the true-positive rate (TPR) and the false-positive rate (FPR), defined as:

$$\begin{aligned} \mathrm{TPR} &= \frac{\#\text{images detected as forged being forged}}{\#\text{forged images}} \\ \mathrm{FPR} &= \frac{\#\text{images detected as not being forged}}{\#\text{not forged images}} \end{aligned} \tag{7}$$

The "accuracy" here can demonstrate the robustness and detection effect of the proposed method; of course, in order to consider the effect more comprehensively, a low "false positive rate" (FPR) should be guaranteed.

Therefore, Table 1 shows the parameters used by the author for the comparative experiment.A high TPR indicates strong sensitivity to copy-move manipulation, whereas a low FPR ensures that genuine images are rarely misclassified-both properties are essential for a practical forensic detector.

All hyper-parameters employed in the comparative study are summarised in Table 1. Unless otherwise stated, the point-matching ratio threshold is fixed at $t = 0.6$, following Lowe's recommendation [6]. For DBSCAN, ε and MinPts adopt the values suggested in [13]. An image is deemed manipulated whenever at least two clusters are connected by three or more consistent SIFT matches, implying a copy-move operation.

Table 1 Hyper-parameters used in the experiments.

Module	Parameter	Setting
SIFT	contrastThreshold	0.01
	edgeThreshold	50
BIRCH	distance metric	Euclidean
	n_clusters	7
	threshold	0.10
	branching_factor	10

	distance metric	Euclidean
DBSCAN	ε (epsilon)	40
	MinPts	2

The threshold tof the point matching stage is 0.6 as recommended by Lowe [6]; the DBSCAN algorithm uses the value recommended by [13].

Let me reiterate here: if two or more clusters with each at least three pairs of points match found similar, the image is considered to be altered.

3.3 Test Results

This paper reproduces and adjusts the parameters of the three algorithms, and then selects the one with the highest TPR and the lowest FPR as the representative of the algorithm performance. As shown in Table 2, when using the hierarchical clustering algorithm for matching, it is easy to maintain the FPR at around 10% and improve the TPR as much as possible. This shows that as far as the hierarchical clustering algorithms such as BIRCH clustering are concerned, they are better at clustering feature points on two-dimensional coordinates than DBSCAN, which is more sensitive to parameters. The paper [12], who reported that hierarchical clustering combined with additional filtering achieved up to 100 % TPR at ≈ 5 % FPR.

The HAC baseline reported in Table 2 is re-implemented from [15] with only minor modifications; default OpenCV parameters are adopted. Although [15] identified centroid-based linkage (“ward”) as the optimal choice, the present reproduction indicates that single-linkage offers a marginally higher TPR under the current experimental protocol.

Table 2 Performance of different clustering algorithms on MICC-F220.

Clustering algorithm	TPR	FPR
DBSCAN	90.91%	10.91%
HAC (single linkage)	95.45%	9.09%
BIRCH	99.09%	15.45%

Parameter-sensitivity analysis was conducted for the BIRCH algorithm. When no post-processing was applied, large errors were observed across most configurations. Table 3 reports the results obtained under several representative settings. A clear positive correlation emerges between TPR and the FPR: lowering the branching factor or the cluster-diameter threshold generates a greater number of smaller CF-tree sub-clusters. During the subsequent RANSAC stage, the abundance of clusters increases the likelihood that sufficient correspondences are available to estimate a homography, thereby boosting TPR (Figure. 5a). At the same time, many grossly inconsistent matches are filtered out, which prevents FPR from deteriorating catastrophically; nevertheless, FPR still rises in tandem with TPR (Figure. 5b).

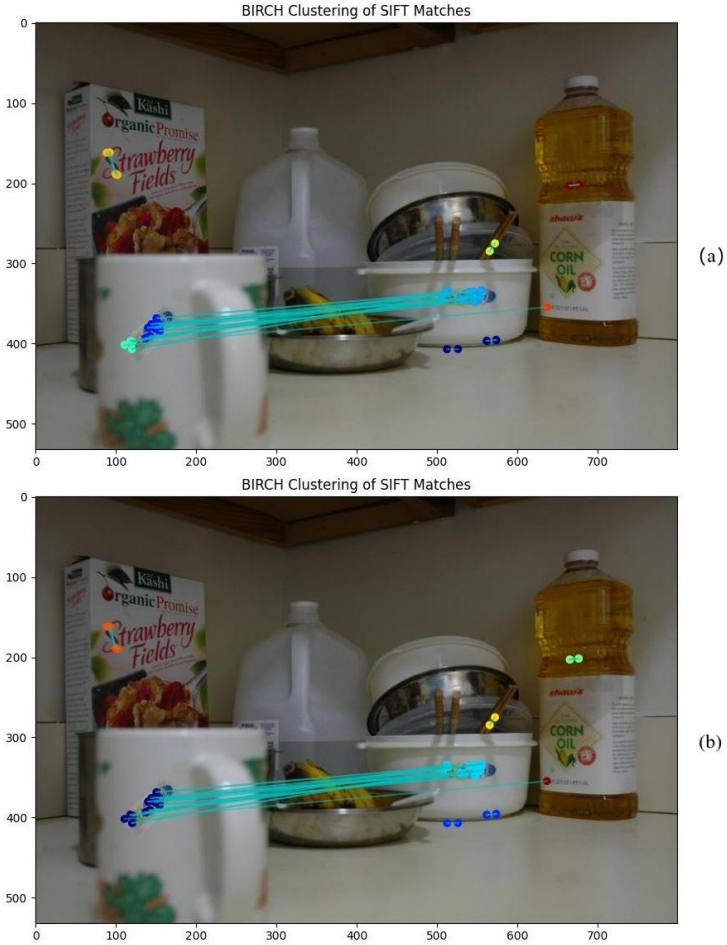

Figure 5. Impact of cluster granularity on copy-move matching. (Picture credit: Original).

Table 3 below shows the results of this method using different parameters. In the evaluation protocol of Section 3.2, the detector declares an image forged whenever at least min_cluster_pts clusters are linked by three or more inlier matches. Even without additional refinements, BIRCH consistently maintained FPR at ≈ 15 %, confirming its suitability for scenarios with dense key-point sets.

Table 3 Effect of BIRCH hyper-parameters on detection accuracy (MICC-F220).

Configuration	branching factor	diameter thr.	min_cluster_pts	TPR [%]	FPR [%]
Cfg-1	5	0.05	2	96.36	14.55
Cfg-2	10	0.10	2	99.09	15.45
Cfg-3	20	0.15	2	92.73	12.73

Table 4 Performance comparison with representative copy-move detection schemes.

Author (year)	Methodology	TPR in %
Umair A. Khanet.al. (2018) [16]	Sift Hybrid	81
	SURF Hybrid	89
	MinEigen Hybrid	90
Gulivindala et al,(2016) [17]	DWT + SIFT + PCA	64
G.G.Rajput et.al. (2023) [18]	SIFT+DBSCAN	92
Dabhole, S. D., et al. (2024) [11]	SIFT+DBSCAN+ELA	96
Amerini, I., et al. (2011) [15]	HAC+SIFT	98
Proposed Method	SIFT+BIRCH	99

Table 4, adapted from the overview in [13], contrasts the proposed SIFT + BIRCH pipeline with several established copy-move detection strategies. The new scheme attains the highest true-positive rate (TPR, 99 %) among the surveyed methods. This result confirms that incorporating the BIRCH hierarchical clustering stage can deliver consistently accurate correspondence grouping at the point-matching phase. In particular, the improvement over density-based clustering (SIFT + DBSCAN) illustrates the potential of BIRCH for reliable large-scale key-point aggregation on the MICC-F220 dataset.

4 Conclusion

A copy-move forgery-detection pipeline that fuses SIFT keypoints with BIRCH hierarchical clustering is presented. Comparative experiments against a DBSCAN-based baseline confirm the superior reliability of the proposed clustering strategy for image forensics. Lowering the SIFT contrast threshold markedly increases the number of keypoints detected in smooth or small manipulated regions; the subsequent BIRCH stage then aggregates this large keypoint set with modest computational overhead, and RANSAC verification further boosts detection accuracy.

On the MICC-F220 benchmark, careful hyper-parameter tuning enables the SIFT + BIRCH configuration to achieve a true-positive rate (TPR) exceeding 99 % while keeping the false-positive rate (FPR) within a manageable range. This performance surpasses that of density-based clustering and highlights several promising optimisation avenues: (i) adaptive parameter scheduling for the clustering algorithm, (ii) secondary clustering or filtering based on multi-scale or depth-aware features, (iii) dominant-orientation screening, and (iv) deep cross-validation using complementary cues such as texture and illumination.

Future work will extend the evaluation to additional attack scenarios, including JPEG compression and additive noise, in order to assess the generalisability of the proposed scheme.

References

1. Al-Qershi, O.M., Khoo, B.E.: 'Passive detection of copy-move forgery in digital images: State-of-the-art', Forensic Sci. Int., 2013, 231, (1–3), pp. 284–295
2. Warif, N.B.A., Wahab, A.W.A., Idris, M.Y.I., et al.: 'Copy-move forgery detection: survey, challenges and future directions', J. Netw. Comput. Appl., 2016, 75, pp. 259–278
3. Christlein, V., Riess, C., Jordan, J., et al.: 'An evaluation of popular copy-move forgery detection approaches', IEEE Trans. Inf. Forensics Security, 2012, 7, (6), pp. 1841–1854
4. Fridrich, J., Soukal, D., Lukas, J.: 'Detection of copy-move forgery in digital images'. Proc. Digit. Forensic Res. Workshop (DFRWS), Baltimore, USA, Aug. 2003, pp. 652–663
5. Cozzolino, D., Poggi, G., Verdoliva, L.: 'Efficient dense-field copy-move forgery detection', IEEE Trans. Inf. Forensics Security, 2015, 10, (11), pp. 2284–2297
6. Lowe, D.G.: 'Distinctive image features from scale-invariant keypoints', Int. J. Comput. Vis., 2004, 60, pp. 91–110
7. Pan, X., Lyu, S.: 'Region duplication detection using image feature matching', IEEE Trans. Inf. Forensics Security, 2010, 5, (4), pp. 857–867
8. Amerini, I., Ballan, L., Caldelli, R., et al.: 'A SIFT-based forensic method for copy–move attack detection and transformation recovery', IEEE Trans. Inf. Forensics Security, 2011, 6, (3), pp. 1099–1110
9. Hastie, T., Tibshirani, R., Friedman, J.: 'The elements of statistical learning' (Springer, New York, NY, USA, 2009)
10. Amerini, I., Ballan, L., Caldelli, R., et al.: 'Copy-move forgery detection and localization by means of robust clustering with J-Linkage', Signal Process. Image Commun., 2013, 28, (6), pp. 659–669
11. Li, Y., Zhou, J.: 'Fast and effective image copy-move forgery detection via hierarchical feature point matching', IEEE Trans. Inf. Forensics Security, 2019, 14, (5), pp. 1307–1322
12. Dabhole, S.D., Rajput, G.G., Ullat, V., et al.: 'Copy-move image forgery detection and classification using SIFT and DBSCAN approaches'. Proc. 15th Int. Conf. Comput. Commun. Netw. Technol. (ICCCNT), Kharagpur, India, Jul. 2024, pp. 1–7
13. Ng, T.-T., Chang, S.-F., Hsu, J., Pepeljugoski, M.: 'Columbia photographic images and photorealistic computer graphics dataset' (Columbia University, 2004)
14. Kanan, C., Cottrell, G.W.: 'Color-to-grayscale: does the method matter in image recognition', PLoS One, 2012, 7, (1), pp. 1–7
15. Amerini, I., Ballan, L., Caldelli, R., et al.: 'A SIFT-based forensic method for copy–move attack detection and transformation recovery', IEEE Trans. Inf. Forensics Security, 2011, 6, (3), pp. 1099–1110
16. Khan, U.A., Feng, K., Arif, M.S., et al.: 'A hybrid technique for copy-move image forgery detection'. Proc. 3rd Int. Conf. Comput. Commun. Syst. (ICCCS), Nagoya, Japan, Apr. 2018, pp. 1–6
17. Suresh, G., Srinivasa Rao, C.: 'RST invariant image forgery detection', Indian J. Sci. Technol., 2016, 9, (22), pp. 1–8
18. Rajput, G.G., Dabhole, S.D.: 'Keypoint-based copy-move area detection'. Proc. Int. Conf. Innovative Comput. Commun. (ICICC), Singapore, Feb. 2023, pp. 1–7