DOI: 10.54254/2755-2721/135/2025.21080

A Review of Feature Matching Based Image Tampering Detection Methods

Zekai Guan[1,a,*]

[1]*Institute of Lanzhou University, School of Information Science&Engineering, Gansu, Lanzhou, China*

a. Guan. 320220900231@lzu.edu.cn

**corresponding author*

Abstract: In recent years, digital image tampering detection techniques play an increasingly important role in dealing with digital image authenticity detection. This review analyzes in detail the feature matching based image tampering detection methods that mainly target the copy-paste tampering problem. Among such methods are subdivided into two subcategories: feature block-based and feature point-based detection. Traditional methods such as feature-point based SIFT and feature-block based Zernike Moments match by extracting local features or global features, but their performance is limited under high noise, low contrast and complex transformations. For this reason, this paper focuses on analyzing a hybrid framework that combines deep learning and traditional methods in addition to traditional methods. This hybrid framework significantly improves the detection efficiency and accuracy. In addition, this paper summarizes the commonly used datasets and their characteristics. It is shown that the hybrid framework demonstrates superiority in dealing with complex geometric transformations and post-processing operations, but there is still room for improvement in detecting small tampered regions. Future research should focus on efficiently fusing multiple methods to adapt to more diverse and complex tampering scenarios.

Keywords: image tampering, image tampering detection, deep learning

1. Introduction

With the rapid development of the Internet and artificial intelligence, making digital images has become a rare thing. This important information carrier is also becoming more and more common in people's lives. The information carried on digital images is very large and dense. In such an environment, digital image tampering has arisen. In order to make others believe in some information that does not exist, people tamper with digital images. One of the classic techniques of image tampering is Copy-Move, also known as Copy-Paste. In the information age where the technology iteration is so fast, the technique of Copy-Move is also growing. In such a background, digital image tampering detection techniques have emerged.

In detection techniques, feature matching based detection methods are mainly for Copy-Move. such methods locate possible tampered regions by extracting feature descriptions of the image and matching these features. Under Copy-Move, the tampered region is highly like the source region in terms of texture, illumination, color, etc., and is rotated, scaled and so on, which makes the detection more difficult. Moreover, the tampered region and the image as a whole may be blurred, noise added, JPEG compressed, etc., and the complex background will increase the difficulty of feature matching in a natural scene; detecting very small tampered regions also requires a more fine-grained In natural scenes, complex backgrounds increase the difficulty of feature matching; detecting very small tampered areas also requires more refined feature extraction and matching methods.

Since 2008, there are mainly nine reviews that have been published in the field of digital image tampering, but literature published in 2008-2014 lack the analysis of new methods in recent years[1-6]. This is outdated in today's rapid technology iteration and will not be elaborated on in detail in this paper. The literature published in 2017 not only summarizes the passive localization methods in natural environment [7], but also reproduces and compares these methods in detail. However, although it is a more complete review of tamper localization methods, it still lacks an analysis of digital image tamper detection methods that incorporate deep learning in recent years. The literature published in 2022 analyzes in detail almost all of the current digital image tamper detection methods [8], but is too comprehensive to be complete under the broad category of feature matching-based detection methods.

The basic process of digital image tampering detection methods based on feature matching is as follows: firstly, local or global features are extracted from the digital image, and the common methods include keypoint matching and block matching; then the similarity between the features is compared to identify the duplicated regions; finally, through geometric constraints, morphology analysis, and other methods, false positives are removed and tampered regions are finally identified. This paper mainly analyzes and compares two major categories of subordinate methods, namely keypoint-based matching and block-based matching, in detail.

2. Manuscript Preparation

Table 1 is several classical datasets commonly used in feature matching based image tampering detection methods, which have been initially organized by the authors, including the explicit procedure, the brief as well as the download address for the readers to take:

Table 1: Dataset Introduction

Data set name	Introduction to the dataset	download address
CASIA V1.0[9]	A total of 1,721 color images are included, of which 800 are real and 921 are fake. The image size is uniformly 384*256 pixels in JPEG format	https://www.kaggle.com/sophatvathana/casia-dataset
CASIA V2.0[9]	A total of 12,323 color images are included, of which 7,200 are real and 5,123 are fake. The image sizes range from 320*240 to 800*600 pixels, and formats include JPEG, BMP and TIFF.	https://www.kaggle.com/ sophatvathana/casia-dataset
CoMoFoD[10]	Contains 260 groups of forged images, each group includes: 1 original image, 1 forged image, and 2 types of forged masks: color mask and black and white binary mask. The images are divided into two categories according to their size: small images (512*512 pixels, 200 groups) and large images (3000*2000 pixels, 60 groups), totaling 13,520 images	http://www.vcl.fer.hr/ comofod/download.html
MICC-F220[11]	Contains 220 images: 110 real and 110 faked. Image resolutions range from 722*480 to 800*600 pixels.	http://www.micc.unifi.it/downloads/MICC-F220.zip
MICC-F600[10]	A total of 600 images are included: 448 real images and 152 fake images. Image resolutions range from 800*533 to 3888*2592 pixels.	http://www.micc.unifi.it/download s/MICC-F600. zip
MICC-F2000[11]	Contains 2,000 images: 1,300 real images and 700 fake images. The image resolution is standardized at 2048*1536 pixels.	http://www.micc.unifi.it/downloads/MICC-F2000.zip

Table 1: (continued).

COVERAGE[12]	Contains 100 pairs of original and forged images. Image resolution averages 400*486 pixels and is saved in lossless TIFF format.	https://github.com/wenbihan/ coverage
Image Manipulation[13]	The dataset contains 48 high-resolution benchmark images with an average resolution of approximately 3000*2300 pixels. Contains 87 snippets of forged regions, which were manually selected from the image and processed to generate forgeries.	https://www5.cs.fau.de/research/data/image-manipulation

3. Generalization of the methodology

3.1. Detection method based on image block matching

Among the detection methods based on image block matching, Zernike Moments is a classical detection method[14], which detects rotated copy-paste tampered regions in an image by utilizing the rotational invariance of Zernike moments. Figure 1 shows the basic operation flowchart of the Zernike Moments algorithm: Zernike Moments first performs unit circle cropping to map the image or tampered region to the unit circle, then performs feature extraction to compute the modulus of the Zernike moments of each order ｜Anm｜, and finally performs similarity detection by comparing the modulus of the Zernike moments of the image blocks to find out the similar region, labeled as the possible tampering regions. However, Zernike Moments feature extraction involves complex mathematical computations, especially time-consuming when dealing with large resolution images, and is mainly designed for rotational invariance, with limited ability to deal with other attacks (e.g., JPEG high compression, Gaussian noise and blurring). In the literature[15], a method combining Fast Fourier Transform, Singular Value Decomposition and Principal Component Analysis is proposed to extract features: the FFT-SVD-PCA cascade method.

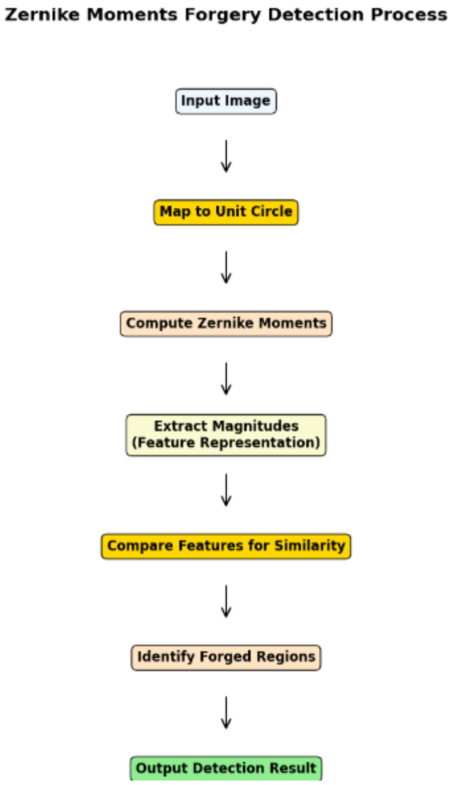

Figure 1: Zernike Moments Forgery Detection Process (Picture credit : Original)

Figure 2 shows the basic flowchart of the algorithm: firstly, the input image is divided into overlapping blocks of fixed size, each representing a candidate detection unit, and secondly, features are extracted using Fast Fourier Transform (FFT), Singular Value Decomposition (SVD) and Principal Component Analysis (PCA) for each block of image respectively. The extracted features are then matched using a cascade filtering strategy to identify possible tampering regions. Finally, by filtering the false matches, the tampered regions are identified and the detection results are outputted as black and white images. The FFT-SVD-PCA cascade method not only makes a threshold-free design that does not require multiple thresholds to be set manually and avoids performance degradation due to improper threshold selection, but also significantly improves its computational efficiency compared to Zernike Moments. Robustness compared to Zernike Moments also maintains high detection accuracy under multiple attacks (e.g., JPEG compression, noise, blurring) in a highly coupled FFT, SVD, and PCA process, achieving more than 97% accuracy even when the JPEG quality factor is as low as 20.

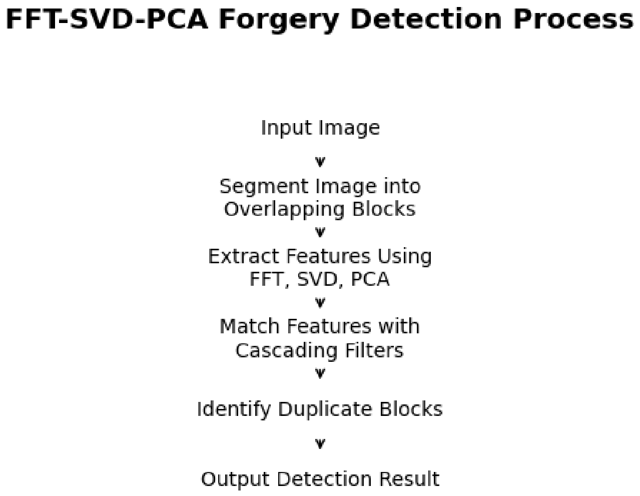
FFT-SVD-PCA Forgery Detection Process
Input Image
Segment Image into Overlapping Blocks
Extract Features Using FFT, SVD, PCA
Match Features with Cascading Filters
Identify Duplicate Blocks
Output Detection Result

Figure 2: FFT-SVD-PCA Forgery Detection Process (Picture credit : Original)

3.2. Detection method based on keypoint matching

In keypoint-based, SIFT[16] is a classical detection method which extracts the SIFT descriptors of keypoints in the image which are robust to scale, rotation, noise and illumination variations, and then locates the tampered region by matching between the descriptors to achieve the effect of detecting the tampered region in the image.

Figure 3 brackets the process of SIFT algorithm: firstly, find the extreme points at different scales, secondly, filter out the unstable keypoints and keep only the keypoints with high contrast and accurate localization, then assign the direction to each keypoint to ensure the rotation invariance, and finally generate 128-dimensional feature vector based on the local gradient distribution of keypoints. And then feature matching is performed: the similarity between feature descriptors is calculated using Euclidean distance, and a distance ratio threshold is set to filter out the optimal matching pairs. If multiple matching pairs are found to be concentrated in a certain region and match the pattern of tampering, it is determined that the image may have been copy-paste tampered.

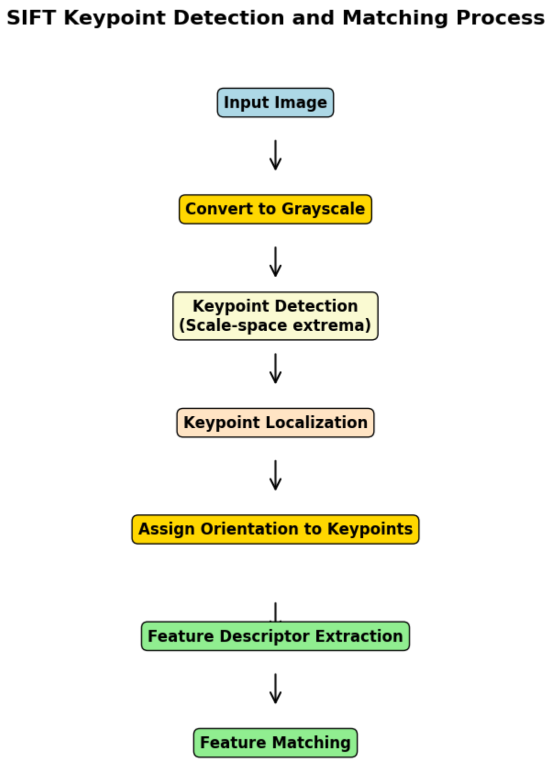

Figure 3: SIFT Keypoint Detection and Matching Process (Picture credit : Original)

However, the computational complexity of SIFT is high, the process of keypoint detection and matching is computationally intensive, especially on high-resolution images, and the accuracy of keypoint extraction and matching decreases in high-noise and low-contrast environments, and the matching effect of SIFT does not show high performance for subtle tampered regions. However, in the literature[17] the authors design another detection method based on keypoint matching: BusterNet.The network architecture of BusterNet is divided into three parts: tamper detection branch, similarity detection branch and fusion module. Compared with the traditional SIFT, this structure not only avoids the complexity of manual feature design and matching and saves most of the time, but also works better for low-texture regions or scenes with weak tampering traces. The tamper detection branch uses the first four convolutional blocks of VGG16 to extract features, which are restored to the original resolution by the up-sampling module (inverse convolutional layer or interpolation operation), and outputs a binary mask indicating the tampered region. The similarity detection branch extracts features using VGG16 and computes the similarity of each pixel to other pixels in the feature space, generating an autocorrelation feature map. This branch also adds quantile pooling for extracting statistical information about the similarity distribution to enhance robustness, and finally outputs a binary mask representing the similar region as in the tamper detection branch. The fusion module is used to combine the outputs of the tamper detection branch and the similarity detection branch to generate the final triple classification mask. The triple classification result is generated: 0: real region. 1: source region. 2: target region. Figure 4 shows the basic flowchart of the BusterNet algorithm.

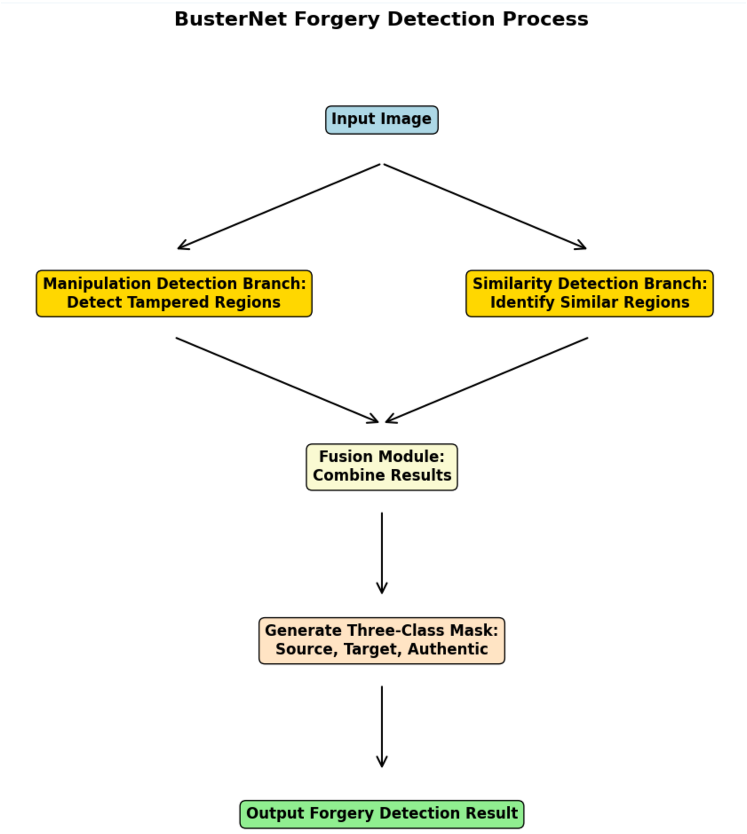
BusterNet Forgery Detection Process
Input Image
Manipulation Detection Branch: Detect Tampered Regions
Similarity Detection Branch: Identify Similar Regions
Fusion Module: Combine Results
Generate Three-Class Mask: Source, Target, Authentic
Output Forgery Detection Result

Figure 4: BusterNet Forgery Detection Process (Picture credit : Original)

BusterNet's detection method comparing to the traditional SIFT keypoint extraction not only realizes end-to-end learning of tampering features automatically by deep learning without manual parameter tuning, but also distinguishes between source/target regions, provides fine-grained tampering detection results, and is highly adaptive to complex geometric transformations and post-processing operations. However, it does not demonstrate high performance in detecting small tamper regions and detailed tampering, but in literature[18] a method of applying a two-stage detection framework is proposed: the Hybrid Forgery Detection Framework. Hybrid Forgery Detection Framework is based on the traditional keypoint detection method and the deep learning detection method, which divides the detection into two stages to improve the efficiency, as shown in Figure 5: In the first stage, a deep learning network (e.g., VGG16) is used to extract the image multiscale features, analyzes them by correlation and quickly generates a coarse mask of the tampered region, providing a preliminary prediction of the tampered region and narrowing the scope of subsequent processing. The second stage optimizes keypoint matching, extracts local keypoint features using traditional methods, refines tampered region localization through keypoint matching and region verification, eliminates false positives using morphological operations, and improves pixel-level accuracy.

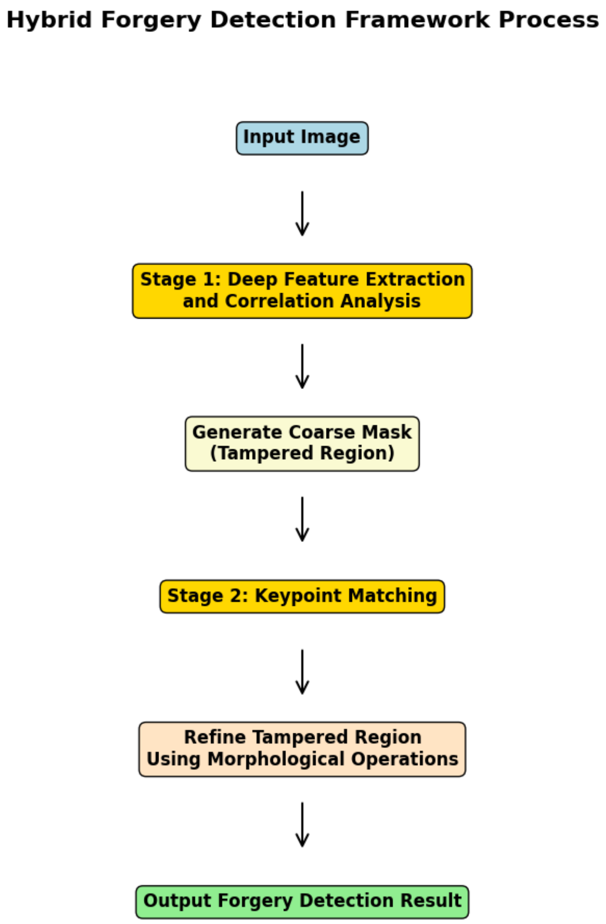
Hybrid Forgery Detection Framework Process
Input Image
Stage 1: Deep Feature Extraction and Correlation Analysis
Generate Coarse Mask (Tampered Region)
Stage 2: Keypoint Matching
Refine Tampered Region Using Morphological Operations
Output Forgery Detection Result

Figure 5: Hybrid Forgery Detection Framework Process (Picture credit: Original)

4. Conclusion

In this paper a detailed categorization of feature matching based methods for detecting image tampering and a detailed analysis and comparison of their subordinate classical and latest methods based on feature point matching and block matching are presented. In this paper, it is found that the latest and high-performance methods are all hybrid detection frameworks of deep learning and traditional methods. Deep learning techniques have significant advantages in automated feature extraction, detection efficiency and fine-grained tampering recognition, but small tampered region detection is still a major challenge in image tampering detection at present. Detection techniques that efficiently fuse traditional and deep learning methods should be further explored in the future to cope with more complex and fine-grained tampering scenarios.

References

[1] *Wang, W., Dong, J. and Tan, T. (2009). A survey of passive image tampering detection. International Workshop on Digital Watermarking.*

[2] *Piva, A. (2013). An Overview on Image Forensics. ISRN Signal Proc essing.*

[3] *Qazi, T., Hayat, K., Khan, S. U., et al. (2013). Survey on Blind Image Forgery Detection. IET Image Processing.*

[4] *Birajdar, G. K., Mankar, V. H. (2013). Digital Image Forgery Detection Using Passive Techniques: A Survey. Digital Investigation.*

[5] *Liu, L. (2009). Survey on Passive Digital Image Authenticity Check Tech niques. Computer Engineering and Applications.*

[6] Wu, Q., Li, G. H., Tu, D., et al. (2008). A Survey of Blind Digital Image Forensics Technology for Authenticity Detection. Acta Automatica Sinica.

[7] Zampoglou, M., Papadopoulos, S., Kompatsiaris, Y. (2017). Large-Scale Evaluation of Splicing Localization Algorithms for Web Images. Multimedia Tools and Applications.

[8] Zhang, Y. X., Zhao, X. F., Cao, Y. (2022). A Survey on Blind Detection of Tampered Digital Images. Journal of Cyber Security.

[9] Dong, J., Wang, W., Tan, T. N. (2013). CASIA Image Tampering Detection Evaluation Database. 2013 IEEE China Summit and Interna tional Conference on Signal and Information Processing.

[10] Tralic, D., Zupancic, I., Grgic, S., et al. (2013). CoMoFoD—New Database for Copy-Move Forgery Detection. Proceedings ELMAR-2013.

[11] Amerini, I., Ballan, L., Caldelli, R., et al. (2011). A SIFT-Based Forensic Method for Copy–Move Attack Detection and Transformation Re covery. IEEE Transactions on Information Forensics and Secu rity.

[12] Wen, B. H., Zhu, Y., Subramanian, R., et al. (2016). COVERAGE—A Novel Database for Copy-Move Forgery Detection. 2016 IEEE International Conference on Image Processing.

[13] Christlein, V., Riess, C., Jordan, J., et al. (2012). An Evaluation of Popular Copy-Move Forgery Detection Approaches. IEEE Transactions on Information Forensics and Security.

[14] Seung-Jin, R., Min-Jeong. L. and Heung-Kyu, L. (2010). Detection of Copy-Rotate-Move Forgery Using Zernike Moments. Korea Advanced Institute of Science and Technology.

[15] Huang, D. Y., Huang, C. N., Hu, W. C., Chou, C. H. (2015). Robustness of copy-move forgery detection under high JPEG compression artifacts. Springer Science+Business Media New York.

[16] Huang, H. L., Guo, W. Q., Zhang, Y. (2008). Detection of Copy-Move Forgery in Digital Images Using SIFT Algorithm. IEEE Pacific-Asia Workshop on Computational Intelligence and Industrial Application.

[17] Yue, W., Wael. A. and Prem, N. (2018). BusterNet: Detecting Copy-Move Image Forgery with Source/Target Localization. Springer Nature Switzerland.

[18] Liu, T., Yuan, X. C., Xie, Z. Y., Zhao, K. Q., Pun, C. M. (2024). A Two-Phase Scheme by Integration of Deep and Corner Feature for Balanced Copy-Move Forgery Localization. IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS.