Journal of Ambient Intelligence and Humanized Computing (2023) 14:1647–1658
https://doi.org/10.1007/s12652-021-03386-4

ORIGINAL RESEARCH

A fast forgery frame detection method for video copy-move inter/intra-frame identification

Jun-Liu Zhong[1] · **Yan-Fen Gan**[2] · **Ji-Xiang Yang**[3,4]

Received: 16 March 2021 / Accepted: 6 July 2021 / Published online: 20 July 2021

Abstract

Digital video is critical visual evidence in various fields and is easily manipulated under different techniques such as the popular video copy-move forgery. In the past decades, although machine intelligence has been widely adopted to detect the forgery in digital images automatically, It still remains a very challenging detection task for carefully-crafted copy-move forgery in digital video for three reasons: (i) A video of medium length containing hundreds of frames already incurs a prohibitive computational cost; (ii) Similar backgrounds in contiguous frames are easily mistakenly detected as copy-move forgery regions, resulting to a large number of false alarms; (iii) Most state-of-the-art methods cannot detect video copy-move inter-frame or intra-frame forgeries; To effectively address these issues, a fast forgery frame detection method for video copy-move inter/intra-frame identification is proposed: (i) The sparse feature extraction and matching speed-up the algorithm processing and reduce the time cost greatly (Defect (i)); (ii) The adaptive two-pass filtering and copy-move frame-pair matching can address the similarity problem (Defect (ii)) to locate truly forgery frame-pairs (FFP); (iii) Based on the results of these FFP, the type of video copy-move forgery detection can be identified (Defect (iii)). Furthermore, the copy-move frame-pair matching algorithm locates truly FFP, thus further reducing the computation cost and false alarm for detecting the inter/intra-frame forgery efficiently and effectively (Defect (i)). Finally, based on the truly FFP, the video can be checked for forgery or original. If there is no truly FFP, the video is considered as the original one. Otherwise, the video is checked if the forgery is inter-frame (i.e., truly FFP frames are two different frames) or intra-frame (the same frame). The experimental results show that our proposed algorithm achieves higher detection accuracy and higher robustness (*false alarm* = 2 and $F_1 = 0.90$) in the whole GRIP dataset than the existing state-of-the-art methods under various adverse conditions.

Keywords A fast forgery frame detection · Video copy-move inter/intra-frame identification · Sparse feature extraction and matching · Two-pass filtering · Copy-move frame-pair matching

✉ Yan-Fen Gan
fannygyf@foxmail.com

✉ Ji-Xiang Yang
jansonyoung@foxmail.com

Jun-Liu Zhong
junliuzhong@foxmail.com

[1] Department of Information and Communication Engineering, Guangzhou Maritime University, Guangzhou 510725, China

[2] Department of Information Science and Technology, South China Business College, Guangdong University of Foreign Studies, Guangzhou 510545, China

[3] Department of Electronics and Communications, Guangdong Mechanical and Electrical College, Guangzhou 510515, China

[4] Faculty of Information Technology, Macau University of Science and Technology, Macao 999078, China

1 Introduction

For the past two decades, video content can be easily modified or falsified (called video forgery) with many commercial multimedia editing tools (Singh and Aggarwal 2015). Such falsification on video content can lead to severe results. For example, voters can be misled for elections with video forgery of politicians; video forgery in the military field may lead to a war crisis. Practically, such carefully crafted video forgery may not be easily distinguishable even for human experts, leading to the issues of authenticity, originality, and integrity of video contents. For these reasons, effective forensic techniques are urgently demanded.

Digital video forgery is mainly divided into two categories: *whole frame forgery*, and *object forgery*. Whole frame forgery (Li et al. 2016; Liu and Huang 2017; Zhang et al.

2015) is the modification of video contents using an image frame as the forgery unit. Existing techniques in this category include frame deletion, frame insertion, and frame duplication. On the other hand, object forgery is the insertion or deletion of objects in the video content, e.g., video splicing forgery (Chen et al. 2016), video copy-move inter/intra-frame forgery (hereinafter referred to as inter/intra-frame forgery).

The construction of the whole frame forgery is relatively simple, and its forgery result is usually imperfect, and the visual effect always looks unnatural. Therefore, most state-of-the-art detection methods can achieve satisfactory results for whole frame forgery, including scene dependency (Li et al. 2016; Liu and Huang 2017; Zhang et al. 2015), optical flow (Bidokhti and Ghaemmaghami 2015; Jia et al. 2018), compression artifacts exploitation (Aghamaleki and Behrad 2016; Yu et al. 2016), and deep learning (Bakas and Naskar 2018; Long et al. 2017, 2019). A coarse-to-fine detection strategy (Jia et al. 2018) based on Optical Flow (OF) is designed to address the frame copy-move forgery, namely, frame duplication. The coarse detection analyzes the consistency of OF sum between the consecutive frames to find the suspected tampered positions (start-points or end-points of the duplicated frame sequences). The fine detection matches the duplicated frame pairs based on OF correlation.

The other forgery type, object forgery, can achieve a realistic result because it requires more sophisticated and finer forgery techniques such as splicing forgery and copy-move forgery. In splicing forgery, a splicing object and the background elements are firstly shot with different surveillance cameras and then synthesized together (Chen et al. 2016; Davino et al. 2017). For detecting splicing forgery, a machine learning method (Chen et al. 2016) and a deep learning method (Davino et al. 2017) were proposed to identify the inconsistency of statistical properties between the splicing object and real background. Their effective and efficient performance was reported in (Chen et al. 2016; Davino et al. 2017).

Video copy-move forgery achieves an excellent visual effect but requires relatively complex manipulation, that can be done with inter-frame and intra-frame (Zhong et al. 2020). Inter-frame forgery pastes the copied objects from one frame to other corresponding frames in the video, while intra-frame forgery involves successive operations of pasting one or some copied objects from one frame into the same frame. When a video copy-move forgery aims to confuse the frame by adding some objects, it is called additive manipulation. Oppositely, it is called occlusive manipulation when aiming at hiding some objects. Figure 1 shows some examples of inter/intra-frame forgeries with additive/occlusive manipulations. Noteworthy, it is very difficult to detect a carefully crafted inter/intra-frame forgery using the above-mentioned machine learning or deep learning methods under the consistency of statistical properties. It is because the copied objects and the background of the pasted frame are shot under the same surveillance camera, both of which have the same statistical properties and therefore indistinguishable. For this reason, video copy-move forgery detection is currently the most challenging technique for video forensics.

In the literature, only a few works can achieve satisfactory detection results for video copy-move forgery while requiring a high computational cost. Moreover, most existing work is only designed for a single type of video copy-move forgery detection (either inter-frame or intra-frame) but not both. Subramanyam et al. (Subramanyam and Emmanuel 2012) proposed a Histogram of Oriented Gradients (HOG) feature matching and video compression properties to address only intra-frame forgery in MPEG4 format. However, this work takes unacceptably high computational cost, and hence unsuitable for long video clips. In (Bestagini et al. 2013), a detection algorithm is proposed that allows a forensic analyst to reveal and locate the inter-frame forgeries, but fails in resisting the geometrical manipulations, such as rotation. Su et al. (Su et al. 2018) presented the extraction of Exponential-Fourier Moments features in each frame to find the potential matching pairs of intra-frame forgery. However, this work can only detect the inter-frame forgery and lacks robustness to resist compression. The deep neural network (DNN) schemes, e.g., Motion Residual and Parasitic Layers (MRPL) (Saddique et al. 2020)are proposed to address the video copy-move forgery issue. However, MPRL only detect the differences between the forgery frame start and end, and the adjacent frames. Due to the rich forgery objects, DNN for detecting inter/intra-frame forgeries is in an infant stage.

To summarize, the existing detection methods for video copy-move forgery suffer from three defects:

(i) Most state-of-the-art methods cannot make a good trade-off between accuracy and A video with a medium length may contain hundreds of frames that already require a prohibitive computational cost.
(ii) It is almost impossible to identify the true video copy-move forgery regions from similar backgrounds of the adjacent frames based on statistical properties.
(iii) Most state-of-the-art methods cannot detect video copy-move forgery only suitable for a single type of video copy-move forgery: either inter-frame or intra-Furthermore, most methods cannot achieve satisfactory results while detecting forgery regions under post-processing and geometrical transformation.

A fast forgery frame detection method is proposed for both inter and intra-frame video copy-move forgery identification to address the defects. The contributions of this proposed method are as follows:

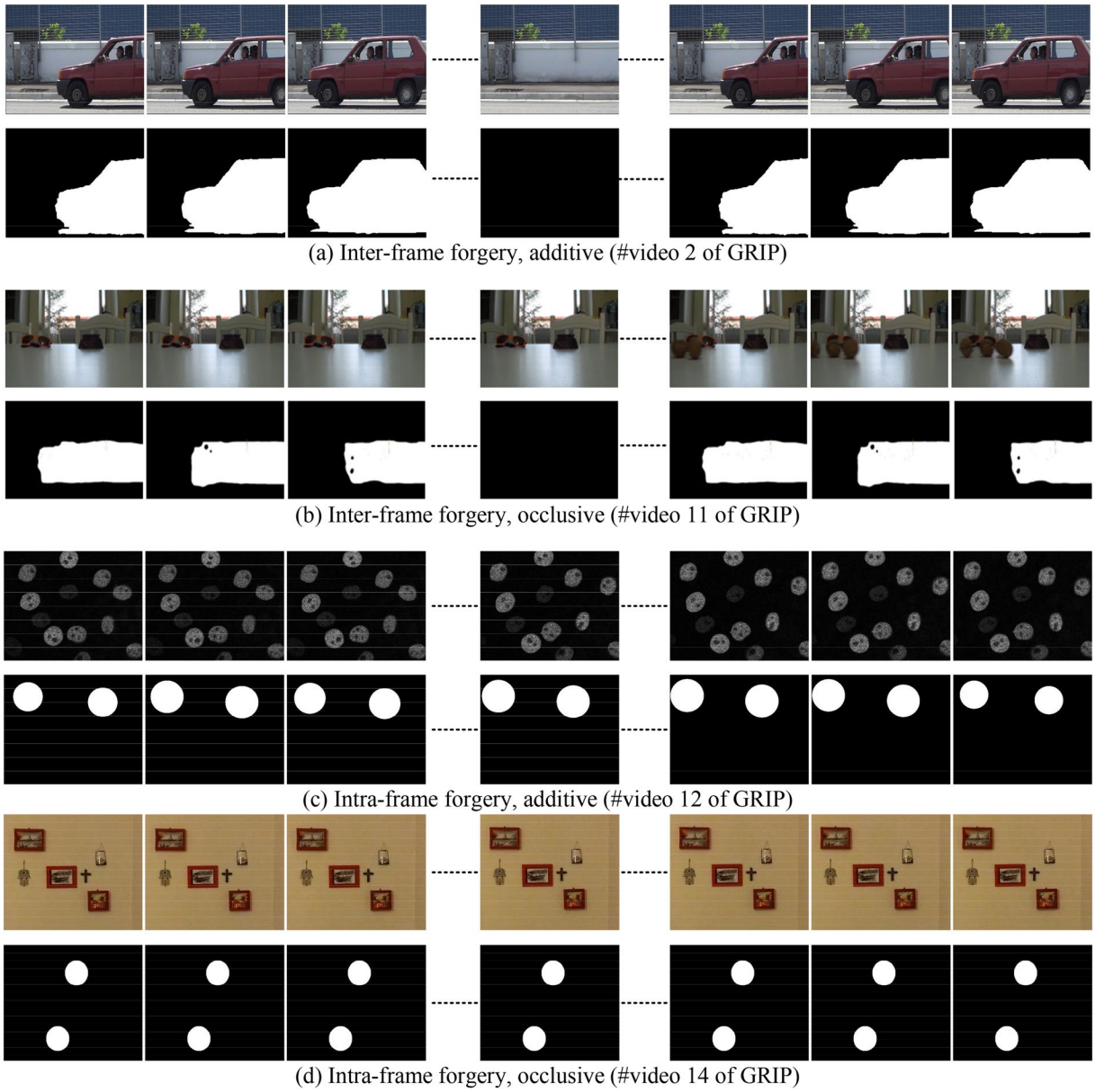


Fig. 1 The sample clips of the video copy-move forgery with additive/occlusive manipulations. **a**, **b** show the additive (a red car) and the occlusive (the background floors) samples of inter-frame forgeries, respectively; **c**, **d** show the additive (a cell) and the occlusive (the background wall) samples of intra-frame forgeries, respectively; The 1st and the 2nd rows of **a**–**d** show the forgery frame clips and the corresponding ground-truth clips, respectively. The black color indicates the background, and the white color indicates the forgery region

(i) The sparse feature extraction and matching (in Section III-speed-up the algorithm processing and reduce the time cost greatly (Defect (i)).
(ii) A newly adaptive *two-pass filtering* algorithm (in Sect. 3-B) is proposed to remove the outlier-pairs for locate truly forgery frame-pairs (FFP) effectively and address the similarity problem (Defect (ii)) both in the inter and intra-frame forgery.
(iii) Based on the results of these frame-pairs, the type of video copy-move forgery detection can be identified (Defect (iii)). Furthermore, the copy-move frame-pair matching algorithm (in Sect. 3-C)) locates truly FFP, thus further reducing the computation cost and false alarm for detecting the inter/intra-frame forgery efficiently and effectively (Defect (i)).

Experimental results demonstrate that our proposed method achieves better performance (in accuracy and time) than the existing state-of-the-art methods, even under post-processing manipulations and geometric attacks.

The rest of the paper is organized as follows. Section II briefly overviews the related work. Section III gives the novel video copy-move forgery detection. The experimental discussions and the conclusion are presented in Sects. IV, V, respectively.

2 Related work

Only a few state-of-the-art methods (Bestagini et al. 2013; Lowe 2004; Saddique et al. 2020; Su et al. 2018; Subramanyam and Emmanuel 2012; Zhang et al. 2015) can address video copy-move forgery detection. Subramanyam et al. (Subramanyam and Emmanuel 2012), propose a Histogram of Oriented Gradients (HOG) feature matching and video compression properties to address only intra-frame forgery in MPEG4 format. However, this work is not sufficiently robust to resist rotation manipulation and also takes unacceptably high computational cost, and hence unsuitable for long video clips. Su et al. (Su et al. 2018). presented the extraction features of Exponential-Fourier Moments (EFMs) in each frame to find the potential matching pairs of intra-frame forgery. An adaptive parameter-based fast compression tracking is applied to track the above forgery object in the subsequent frames if any suspicious forgery object is found. However, this work can only detect the inter-frame forgery and lacks robustness to resist compression. Even worse, the EFMs relying on the block-based feature, is similar to the other block-based methods which fail in detecting scaling forgeries.

Recently, the local descriptors with the superiority of geometrical invariances and high efficiency present good solutions to the above defects for video copy-move forgery identification. Therefore, our proposed method uses local descriptors with the geometrical invariances instead of the block features to extract useful keypoints.

The popular and effective local descriptors contain Scale Invariant Feature Transform (SIFT) (Lowe 2004), and speeded up robust features (SURF) (Bay et al. 2006). Each local descriptor for keypoint extraction has its own characteristics, e.g., the simple feature bit and sparse keypoints of ORB descriptor for fast matching, or the abundant features and dense keypoints of SIFT and SURF for accurate matching. Considering the localization of the copy-move forgery frames, the relatively sparse ORB keypoints 0with the binary bits (0/1) can greatly speed up the matching for the frame localization, and the relatively dense SURF keypoints with abundant features (0–255) can find more keypoint matches for the fine pixel indication. Different local descriptors are suitable for different stages that can strike a balance between efficiency and effectiveness. In our proposed method, we aim at to obtain a near real-time processing speed. Therefore, we prefer sparse ORB feature extraction to other local descriptors and the following feature matching to speed-up the algorithm processing.

In the matching stage, there are many effective matching and filtering algorithms, such as FLANN (Muja and Lowe 2014), KNN matching (Abeywickrama et al. 2016), and Random Sample Consistency (RANSAC) (Fischler and Bolles 1981). However, some of them are not well-designed for keypoint matching, especially while addressing a huge amount of the keypoints with high-dimensional descriptors. These methods will generate a large number of false-positive matches. In literature, the Nearest-Neighbor (2NN) test (Amerini et al. 2011) and GMS are respectively demonstrated to be an effective technique for keypoint matches, and a good solution to address a number of false-positive matches.

3 Our proposed method

The pre-processing operation of the proposed method is used to transform the RGB video into a gray-scale composite image. The sparse features are extracted in the composite image and matched to find the best matching keypoint-pairs (Sect. 3-A). If any best matching keypoint-pair is found, a newly adaptive two-pass filtering algorithm is applied to remove the outlier-pairs (Sect. 3-B). The statistical information of the remaining best matching keypoint-pairs (namely, the inter/intra-frame keypoint-pairs) in all the frame-pairs is used to locate the best matching frame-pairs (Step 1 in Sect. 3-C). Then, the successive best matching frame-pairs are preserved as the truly FFP, which contributes to identifying if the video is the original or inter/intra-frame forgery (Step 2 in Sect. 3-C).

The proposed method consists of three subsections:

(A) sparse feature extraction and matching for finding the best matching keypoint-pairs;
(B) an adaptive two-pass filter for removing the outlier-pairs from the best matching keypoint-pairs to obtain inter/intra-frame keypoint-pairs;
(C) copy-move frame-pairs matching algorithm locates the best frame-pairs (Step 1), the successive best matching frame-pairs are preserved as the truly FFP (Step 2) (Fig. 2).

- *A. Sparse feature extraction and matching*.

  ORB is a combination of the FAST keypoint detector and the BRIEF descriptor generation algorithm. ORB with the inherent orthogonality and geometrical invariances, can effectively resist post-processing and geometrical manipulation. Arguably, ORB performs nearly as well as SIFT and SURF in the geometrical invari-

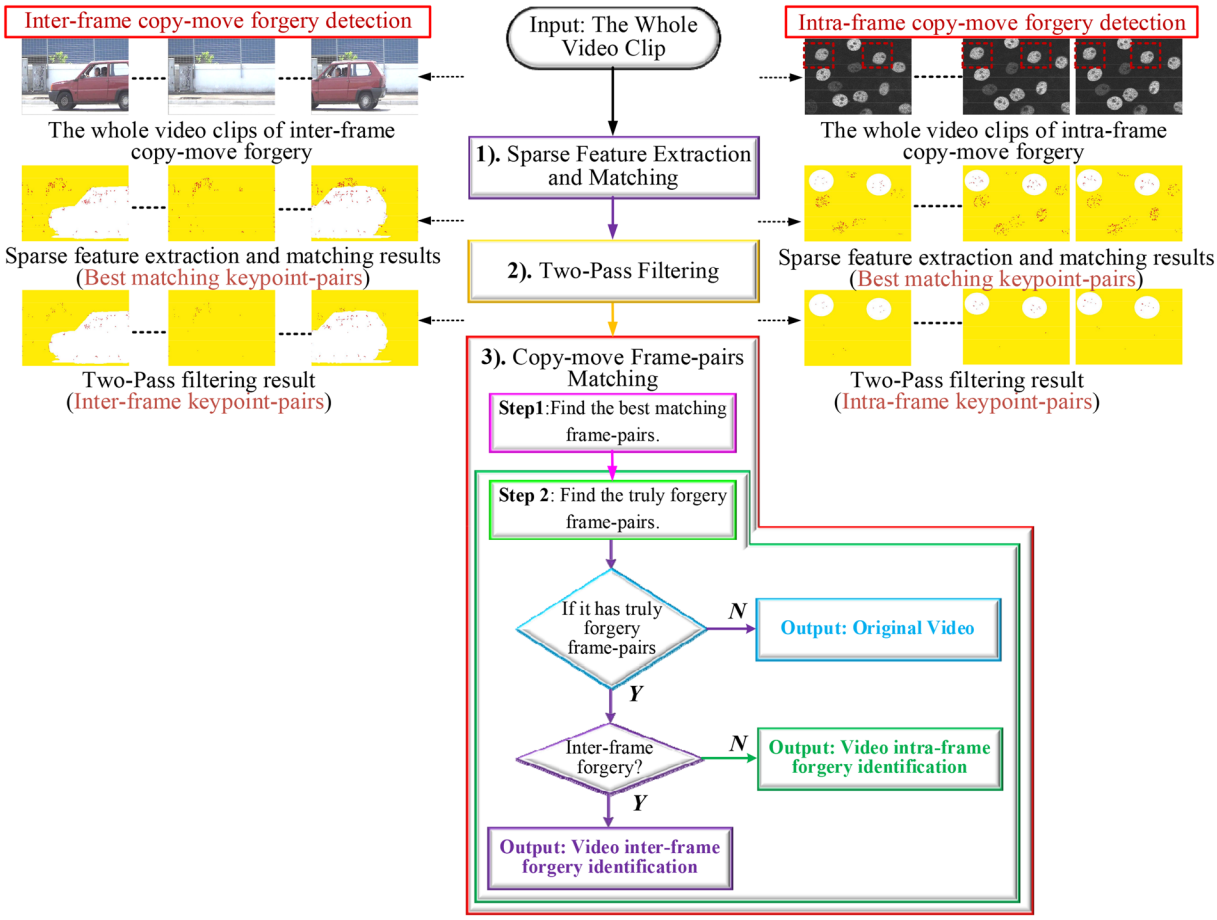


Fig. 2 The framework of a fast forgery frame detection method for video copy-move inter/intra-frame identification

ances but faster in almost two orders of magnitude. However, it is well known that feature matching takes much higher computation than feature extraction. For this reason, only 128-dimensional features of binary bits (0/1) are used in the extracted ORB descriptor in order to speed up the descriptor matching and lower the matching cost. Compared to SIFT and SURF, the relatively low dimension of the descriptor can also improve the matching efficiency. On the basis of assurance efficiency, ORB provides a sufficient number of keypoints for fast frame-pair matching.

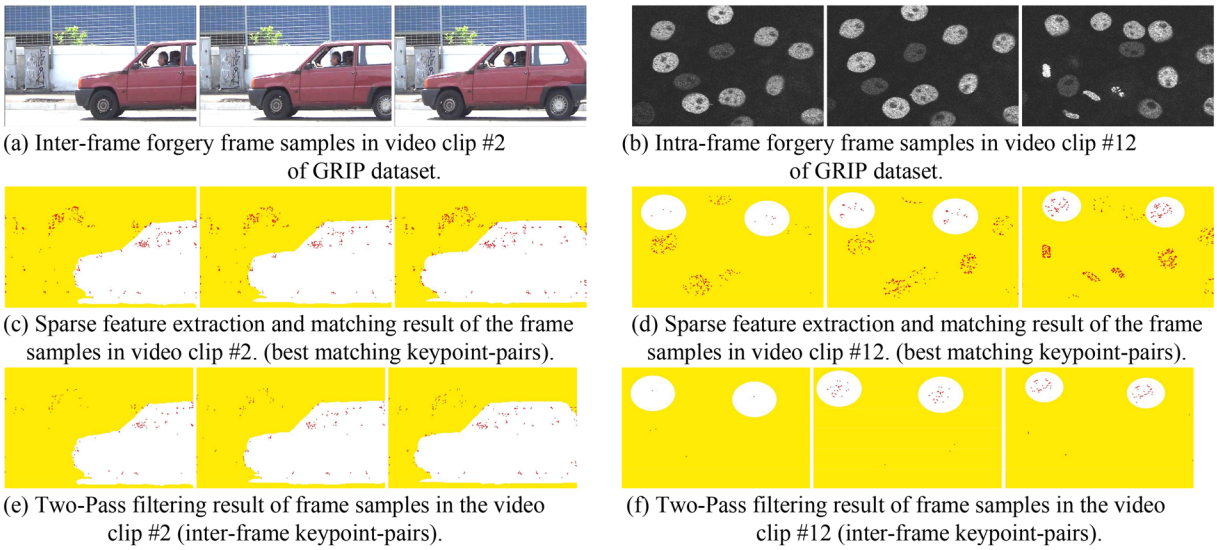
Then, the Nearest-Neighbor (2NN) test and Euclidean distance (Amerini et al. 2011) are used to match the keypoint-pairs with similar local descriptors as the keypoint-pairs. Given a vector, $d = \{d_1, d_2, d_3, \ldots, d_{n-1}\}$ records the 128-dimensional Euclidean distances between the local descriptors of keypoint kp_i and the remaining (n-1) keypoints, where n is the keypoint number. Then, the vector ds is sorted in increasing order to obtain $ds = \{ds_1, ds_2, ds_3, \ldots, ds_{n-1}\}$. The 2NN matching procedure is conducted by evaluating the ratio of the 1st closest distance ds_1 to the 2nd closest one ds_2. While the ratio of the Euclidean distance, namely the correlation coefficient, satisfies the following:

$$ds_1/ds_2 < t \tag{1}$$

where threshold $t = 0.6$ is demonstrated as an effective hyperparameter for keypoint matching in CMFD (Li and Zhou 2019). Others as the false match will be filtered out.

- *B. Two-pass filtering in inter/intra-frame forgery.*

After 2NN test matching, each keypoint can find its best matching keypoint as the best matching keypoint-pairs. However, there are some disturbed keypoint-pairs in the result of the best matching keypoint-pairs. In particular, many disturbed keypoint-pairs belong to the same object with small spatial-distance in the intra-frame forgery. Besides, the similar background of adjacent frames in inter-frame forgery may also generate many disturbed keypoint-pairs. Figure 3c, d shows the sparse feature extraction and matching result (best matching keypoint-pairs), which contains disturbed



(a) Inter-frame forgery frame samples in video clip #2 of GRIP dataset.

(b) Intra-frame forgery frame samples in video clip #12 of GRIP dataset.

(c) Sparse feature extraction and matching result of the frame samples in video clip #2. (best matching keypoint-pairs).

(d) Sparse feature extraction and matching result of the frame samples in video clip #12. (best matching keypoint-pairs).

(e) Two-Pass filtering result of frame samples in the video clip #2 (inter-frame keypoint-pairs).

(f) Two-Pass filtering result of frame samples in the video clip #12 (inter-frame keypoint-pairs).

Fig. 3 Sparse keypoint filtering results of different kinds of forgeries. For better indication, yellow and white respectively indicate the background and the forgery region of the ground truths; the red points indicate the keypoints

keypoint-pairs in the adjacent frames of Fig. 3a, b. Therefore, an adaptive two-pass filter consisting of low-pass and high-pass filters is proposed.

1. Low-pass filtering in intra-frame forgery.

   The low-pass filter uses a relatively small spatial-distance to remove the outlier-pairs and obtain the intra-frame keypoint-pairs. As a matter of fact, every frame of the intra-frame forgery can be regarded as the copy-move image forgery. Therefore, we have referred to the filtering distance L_1 of the copy-move image forgery detection (Zhong and Pun 2019) as shown in Eq. (2). In intra-frame forgery, the copy and paste regions are both in the same frame so that the distances of the best matching keypoint-pair k_{d1} must be smaller than the frame width W:

$$L_1 \leqslant k_{d1} < W \tag{2}$$

   Here$L_1 = \frac{H+W}{\sqrt{min(H,W)}}$where H and W are respectively the height and the width of a video frame.

2. High-pass filtering in inter-frame forgery.

   Firstly, forgery frames must be of a certain length in the inter-frame forgery. Based on the persistence of vision, it requires 0.4 s, namely, 10 frames per second (fps), for the human eye to better discern the continuous contents of the video. It means that the required number of copy clips and paste clips of an inter-frame video forgery is no less than 10 frames. Secondly, the backgrounds of the adjacent frames taken by the same surveillance cameras are so similar that the 2NN test generates many disturbed keypoint-pairs. Therefore, a high-pass filter is designed to remove the disturbed keypoint-pairs with relatively long spatial distance. In inter-frame forgery, the high-pass filtering distances k_{d1} between the best matching keypoints-pair is given in Eq. (3).

$$k_{d1} \geq L_2 \cdot W \tag{3}$$

   where L_2 is the number of filtering frames, the smaller L_2 is, the more disturbed keypoint-pairs preserve. The persistence of vision determines that the number of forgery frames is no less than 10 frames. Therefore, the filtering number L_2 is set in 1–9 frames.

   To determine the best number L_2, we have conducted an extensive test on our available dataset. Figure 4 shows that the percentage of the remaining keypoint-pairs on the distance of 1 to 9 filtering frames. Noted that, the best matching keypoint-pairs of the inter-frame forgery contain inter-frame keypoint-pairs and disturbed keypoint-pairs. While the filtering number L_2 increases, the more disturbed keypoint-pairs are removed, and the number of remaining keypoint-pairs is rapidly decreased. When the filtering number L_2 of the keypoint-pair increases from 1 to 7, the remaining keypoint-pairs

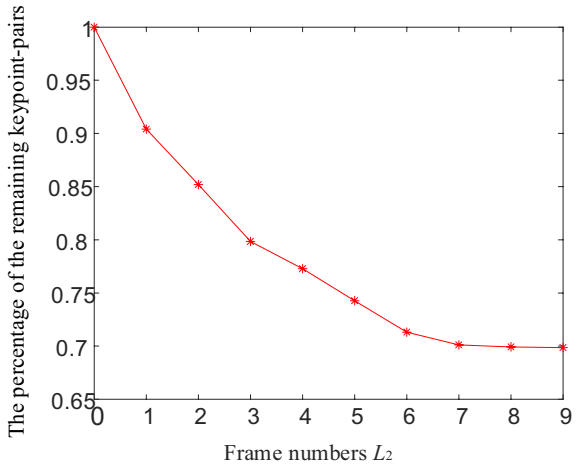


Fig. 4 The percentage of remaining on different frame distance

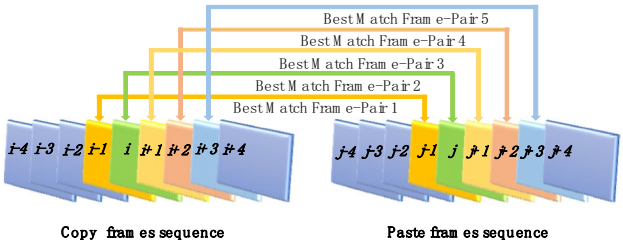


Fig. 5 The example of truly forgery frame-pairs.keypoint-pairs

decrease from 90.41 to 70.12 %. When L_2 is more than 7, the total number of the keypoint-pairs is unchanged essentially. It means that the disturbed keypoint-pairs are almost filtered out, and the remaining are inter-frame keypoint-pairs. Therefore, the L_2 is set to 7 based on the analysis of Fig. 4.

Combining two-pass filtering analysis of the inter/intra-frame forgery, we finally set the distances k_{d1} of the best match keypoint-pair as follows for filtering the disturbed keypoint-pairs: if $L_1 \leqslant k_{d1} < W$, the remaining best matching keypoint-pairs belong to intra-frame keypoint-pairs; if $k_{d1} \geqslant L_2 \cdot W$, the remaining best matching keypoint-pairs belong to inter-frame keypoint-pairs. To summarize,

$$\text{The remaining best matching keypoint - pairs} \in \left\{ \begin{array}{l} \text{Intra} - \text{frame keypoint - pairs, } w.r.t.\ L_1 < k_{d1} < W \\ \text{Inter} - \text{frame keypoint - pairs, } w.r.t.\ k_{d1} > (L_2 \times W) \end{array} \right\} \quad (4)$$

Figure 3e and f show that the inter/intra-frame keypoint-pairs marked in red mainly exist in the forgery region after the two-pass filtering. Based on the above analysis, finding the copy-move frame-pairs in the next step is very beneficial.

- *C. Copy-move frame-pairs matching*.

After two-pass filtering, the preserved keypoints are inter/intra-frame (remaining best matching) keypoint-pairs in the forgery frame-pairs. The frame-pairs with the maximum number of remaining best matching keypoint-pairs are regarded as the potential best matching frame-pairs. Therefore, we use this characteristic as the index to find the potentially best matching frame-pairs. However, the best matching frame-pair only represents the frame-pair with the strongest correlation. It is not necessarily the truly FFP. Based on the persistence of vision, there is no sense for forgery in an isolated frame-pair. In other words, the truly FFP must be successive best matching frame-pair. For this reason, our goal is to find the successive best frame-pairs as the truly FFP.

Given the total number N of video frames, a combined set of all candidate best matching frame-pairs of a video $U=\{u_{1,1}, u_{1,2}, \ldots u_{1,N}, u_{2,1}, \ldots, u_{i,j}, \ldots, u_{N,N}\}$, where $i, j \in \{1, 2, 3, \ldots, N\}$. The total number of the keypoint-pairs of each frame-pair $u_{i,j}$ is denoted as $s_{i,j}$. The steps and the pseudocode of the proposed Algorithm 1 are given in the following.

Step 1 Find the best matching frame-pairs. Given any frame i, search all frames except frame i to find frame j with the maximum number $s_{i,j}$ of the best matching keypoint-pairs. Then, return the best matching frame-pair $u_{i,j}$.

Step 2 Find the truly forgery frame-pairs (FFP). Set the number of the successive frame-pairs in order to filter out the isolated best matching frame-pairs, and obtain the truly FFP.

Figure 5 shows an example of a truly FFP $u_{i,j}$ with successive frame-pairs, i.e., if $\tau > 5$, the best matching frame-pair $u_{i,j}$ is a truly FFP.

Algorithm 1. Copy-Move Frame-Pairs Matching.

Input: all candidate frame-pairs $U=\{u_{1,1}, u_{1,2}, \ldots u_{1,N}, u_{2,1},\ldots, u_{i,j},\ldots u_{N,N}\}$ and the corresponding matched keypoint-pairs.

where $i, j, l \in \{1, 2, 3, \ldots, N\}$.

1: Initialization: Calculate $s_{i,j}$ = the total number of the matched keypoint-pairs of each frame-pair $u_{i,j}$.

2: Step 1:

3: For i=1 to N do

4: j =1;

5: For l=1 to N do

6: If $|s_{i,l}| > |s_{i,j}|$ then

7: $j = l$;

8: End

9: End

10: End

11: Return the best matching $u_{i,j}$

12: Step 2:

13: For i=1 to N do

14: If the frame i and its best matching frame j have at least τ successive frame-pairs

15: $u_{i,j} \in$ *truly FFP*

16: End

17: End

Output: The truly FFP.

In fact, inter/intra-frame forgeries have different properties on their truly FFP. For inter-frame forgery, the copied and the pasted regions are found in different frames, as shown in Fig. 6a. Therefore, its truly FFP are two different frames. For intra-frame forgery, the copied and the pasted regions are in the same frame as shown in Fig. 6b. As a result, its truly FFP are the same frame. Based on the truly FFP, the video can be checked for forgery or original. If there is no truly FFP, the video is considered as original. Otherwise, the video is checked if the forgery is inter-frame (i.e., truly FFP frames are two different frames) or intra-frame (the same frame). In this way, the video forgery frames can be identified accurately with truly FFP.

4 Experiments

The proposed method is compared with several state-of-the-art methods through many experiments under various adverse conditions. This section presents the datasets, the performance metrics, and finally, the experiment results comparisons and analysis.

A. *Datasets for video copy-move forgery detection*.

Three benchmark datasets (GRIP) are employed to evaluate the proposed method and the state-of-the-art methods. The GRIP dataset[1] comprises 15 short videos and 93 derivative forgeries of inter/intra-frame videos. There are very little or even no traces to raise suspicion on the forgery videos. All the 15 base videos suffered from JPEG compression (compression factor is 10, 15, 20), 8 of them suffered from rotation (5°, 25°, 45°), and 9 of them suffered from flipping, which makes more difficult to detect the forgeries. Table 1 shows the synthetic statistics of the GRIP dataset. On the left column of Table 1 (Original Video) the properties of the original base videos are shown. On the right column (Copy-Move Video), the properties of the forgery videos are shown, including additive (Add.) or occlusive (Occ.), inter or intra-frame forgery, JPEG compression (Com.), rotation (Rot.), flipping (Flip.).

B. *Unified performance metrics*.

We have presented detection accuracy (*det.*), false alarm (*f.a.*), performance metrics (F_1) and processing time (*time*). If a forgery video is correctly detected in

[1] http://www.grip.unina.it/web-download.html.

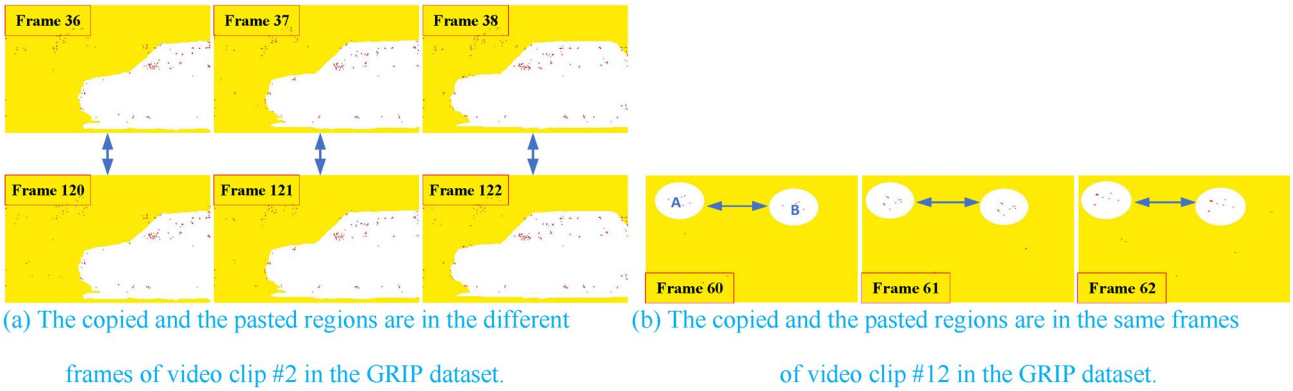


Fig. 6 The copied and the pasted regions in the different/same frames of the video clip

the proposed method, it is marked with "✓" in "*det.*" column in Table 2. If an original video is falsely detected as a copy-move one, we mark it with "✖" in "*f.a.*" column in Table 2. To accurately measure the detection performance, the evaluation criteria *TP* (True Positive), *FP* (False Positive), and *FN* (False Negative) are used, where *TP* represents the detected forgery frames as true forgery frames, *FP* represents the original frames that are falsely detected as forgery frames, and *FN* represents the forgery frames that are missed in detection. The combination of *TP*, *FP*, and *FN*, constitutes F_1 indicator as Eq. (5).

$$F_1 = \frac{2TP}{2TP + FP + FN} \tag{5}$$

It is known that a higher F_1 score denotes performance better. Finally, the experiments are implemented on a computer with an Intel (R) Core i7-8700 @3.20 GHz CPU and 16 GB RAM and GPU RTX2070. The efficiency is measured in terms of normalized CPU time in s/Mpixel.

C. *Experimental results*.

There are several state-of-the-art methods, including the Dense moment feature index and best match algorithm with radial-harmonic-Fourier moments (DMFIBM) (Zhong et al. 2020), Bestagini et al. (Bestagini et al. 2013), MRPL method (Saddique et al. 2020). However, the methods in the literatures of (Subramanyam and Emmanuel 2012) cannot be applied to real datasets like GRIP because they are with very restrictive assumptions on forgery videos. The Bestagini (Bestagini et al. 2013) method can only detect the inter/intra-frame forgery at the frame level. DMFIBM method can detect both the inter-frame and intra-frame forgeries. However, it is based on block feature extraction, and the following block feature matching will lead to expensive

Table 1 Statistics of the GRIP dataset

Original video				Copy-move video					
#	Name	Frame size	Frames	Add./occ.	Inter	Intra	Com.	Rot.	Flip.
1	TV screen	576×720	141	Add	✓		✓	✓	✓
2	Fast Car	370×720	140	Add	✓		✓		✓
3	Felt-Tip Pen	550×720	100	Add	✓		✓	✓	✓
4	Rolling Can	480×660	125	Add	✓		✓	✓	✓
5	Falling Can	480×720	174	Add	✓		✓		
6	Walnuts	480×720	221	Occ	✓		✓		
7	Can 1	520×720	201	Add	✓		✓		✓
8	Can 2	720×720	210	Add	✓		✓	✓	✓
9	Lamp	390×465	455	Add	✓		✓	✓	
10	Tennis Ball	640×360	200	Add	✓		✓		✓
11	Student	400×380	340	Occ	✓		✓		
12	Cell 1	400×500	92	Add		✓	✓	✓	✓
13	Cell 2	512×512	92	Occ		✓	✓	✓	✓
14	Wall Frame	500×570	200	Occ		✓	✓	✓	
15	Statue	590×480	100	Occ		✓	✓		

Table 2 Detection and efficiency performance for plain copy-moves on the GRIP dataset

	DMFIBM (Zhong et al. 2020)				Bestagini (Bestagini et al. 2013)				MPRL (Saddique et al. 2020)				Proposed			
video	*det.*	*f.a.*	F_1	*time*	*det.*	*f.a.*	F_1	*time*	*det.*	*f.a.*	F_1	*time*	*det.*	*f.a.*	F_1	*time*
1	✓		0.98	15.42	✓		0.78	8.9	✓	✖	0.48	1.52	✓		0.99	0.86
2	✓		0.94	15.45	✓		0.83	7.3	✓		0.47	1.94	✓		0.98	1.08
3	✓		0.76	16.39	✓		0.74	6.7	✓	✖	0.35	1.87	✓		0.98	0.86
4	✓		0.93	14.92	✓		0.80	7.2	✓		0.44	1.94	✓		0.96	0.96
5	✓		0.91	16.70	✓	✖	0.90	14.9	✓		0.40	1.45	✓		0.81	1.12
6	✓		0.85	16.50			–	11.7	✓	✖	0.39	1.65	✓		0.79	1.01
7	✓		0.91	18.45	✓	✖	0.80	11.5	✓		0.42	1.78	✓		0.96	1.01
8	✓		0.96	19.73	✓	✖	0.79	15.2	✓		0.41	1.54	✓		0.98	0.91
9	✓		0.99	17.80	✓	✖	0.84	14.4	✓		0.41	1.68	✓		0.99	3.10
10	✓		0.99	15.69	✓	✖	0.86	6.3	✓		0.48	1.61	✓		0.98	1.49
11	✓		0.98	14.14			–	7.7		✖	–	2.13	✓		0.99	1.31
12	✓		0.93	16.23			–	2.6		✖	–	1.37	✓		0.97	1.52
13	✓		0.99	15.43			–	4.4		✖	–	1.74	✓		0.94	1.10
14	✓		0.83	16.66			–	8.8			–	1.65	✓		0.79	1.16
15	✓	×	0.85	16.05			–	3.8		✖	–	1.58	✓		0.77	1.12
$\sum, \mu$	15	1	0.92	16.37	9	5	0.49	8.8	10	7	0.28	1.70	**15**	**0**	**0.93**	**1.23**

computation costs. MPRL, based on the residual signal between the adjacent frames, is suitable for identifying the forgery frame start and end, but fails in addressing the static forgeries.

The results of plain copy-move forgeries for GRIP are shown in Table 2, where$\sum$ represents the total number of variables, including the *det. or f.a.*, andμrepresents the average number of variables F_1 or *time*. Noted that, the plain manipulations involve translations on forgery objects without other geometrical attacks and post-processing transformations.

Table 2 shows that our proposed method can detect all the plain forgery videos and gets the best performance of $F_1 = 0.93$. The DMFIBM method also detects all the forgery videos with an F_1 score of 0.92. The followed method, the Bestagini method, misses six videos (#videos 6, 11, 12, 13, 14, and 15) and gets a bad F_1 score of 0.49. The CNN model, the MPRL model, misses five videos (#videos 11, 12, 13, 14, and 15) and gets the weakest performance with an F_1 score of 0.28. This because MPRL is only competent in searching the difference and coherence between the adjacent frames. Nevertheless, the copied object and its source frame belong to the genuine sources that do not appear the forgery traces. The MPRL has already missed half of the copy-move frames. Therefore, MPRL gets the weakest performance.

In terms of the false alarm, our proposed method can identify all of the original videos accurately, namely, *f.a.*=0, while DMFIBM, Bestagini, and MPRL respectively get the *f.a.*=1, 5, and 7. In comparing the average time cost, our proposed method gets an average of 1.23 s/Mpixel, which is

Table 3 Detection results on the whole grip dataset

			DMFIBM (Zhong et al. 2020)			Bestagini (Bestagini et al. 2013)			MPRL (Saddique et al. 2020)			Proposed		
Dataset	Cases	#video	*det.*	*f.a.*	F_1	*det.*	*f.a.*	F_1	*det.*	*f.a.*	F_1	*det.*	*f.a.*	F_1
GRIP	Plain	15	15	1	0.92	9	5	0.49	10	7	0.28	15	0	0.93
GRIP	QF = 10	15	15	0	0.91	9	5	0.50	9	5	0.26	15	0	0.93
	QF = 15	15	14	1	0.89	9	4	0.37	9	5	0.25	14	0	0.90
	QF = 20	15	14	1	0.81	9	5	0.45	8	6	0.21	13	2	0.86
GRIP	$\theta = 5^\circ$	8	8	–	0.92	2	–	–	4	–	0.22	8	–	0.97
	$\theta = 25^\circ$	8	8	–	0.92	2	–	–	4	–	0.22	7	–	0.87
	$\theta = 45^\circ$	8	8	–	0.91	2	–	–	4	–	0.21	7	–	0.82
GRIP	flipping	9	8	–	0.91	3	–	–	5	–	0.24	9	–	0.93
$\sum, \mu$		93	**90**	3	0.87	45	19	0.18	45	23	0.14	88	**2**	**0.90**

much faster than 16.37 s/Mpixel in DMFIBM, 8.8 s/Mpixel in Bestagini, and 1.70 s/Mpixel in MPRL. In summary, Table 2 shows that the proposed method achieves the best performances, namely, detection accuracy (*det.*), false alarm (*f.a.*), comprehensive performance (F_1), and the lowest computational costs in the plain copy-moves of the GRIP dataset.

Subsequently, the comparisons under the challenging forgery attacks of the whole GRIP dataset are presented, including JPEG compression, rotation, and flipping attacks. For simplicity, experimental results are simplified and reported in Table 3. In these comparisons, the DMFIBM method achieves the best detection performance at the video level (*det.*=90/93), and the followed method is our proposed method with *det.*=88/93, which is only slightly lower than the DMFIBM method. Nevertheless, the proposed method and DMFIBM method obtain the best performance at the frame level (the identical score $F_1 = 0.90$). The proposed method also achieves the best score to identify the original video (*f.a.*=2/93). It is a similar case to Table 2 that the Bestagini method comes third place, and the MPRL gets the last place. In Table 3, the total statistics of the mean μ_{F1} score are based on the F_1 score of each case, as listed in Eq. (6).

$$\mu_{F_1} = \frac{\sum_{i=1}^{8} w_i F_{1,i}}{93} \tag{6}$$

where $i = 1, 2, 3, 4, 5, 6, 7, 8$, respectively represent the cases of Plain, QF = 10, QF = 15, QF = 20, $\theta = 5°$, $\theta = 25°$, $\theta = 45°$, flipping, w_i represents the number of video in the corresponding 8 cases mentioned above, namely, w_i = 15, 15, 15, 15, 8, 8, 8, and 9. The scores $F_{1,i}$ represent the F_1 scores of the corresponding 8 cases.

5 Conclusion

This paper proposes a fast forgery frame detection method for video copy-move inter/intra-frame identification. It consists of sparse feature extraction and matching, two-pass filtering, and copy-move frame-pairs matching can address three issues:

(i) A video of medium length containing hundreds of frames incurs a prohibitive computational cost;
(ii) Similar backgrounds in contiguous frames are easily mistakenly detected as copy-move forgery regions, resulting in a large number of false alarms;
(iii) Most state-of-the-art methods cannot detect video copy-move inter-frame or intra-frame forgeries, simultaneously.

The proposed method makes a good trade-off between efficiency and effectiveness. Our proposed method achieves the best false alarm 2/93 and the best performance $F_1 = 0.90$ in the whole GRIP (Table 3) dataset, and the lowest computational costs of 1.23 s/Mpixel. In future work, we plan to develop novel and efficient techniques, e.g., CNN, for video copy-move forgery detection for higher computation efficiency.

Acknowledgements This work was supported in part by Guangdong basic and applied basic research foundation under Grant No. 2020A151501783 (2020A1515010700), the 2020 characteristic innovation project of general universities in Guangdong province (natural science) under Grant 2020KTSCX205, the 2019 youth project of general universities in Guangdong province (natural science) under Grant 2019GKQNCX028.

References

Abeywickrama T, Cheema MA, Taniar D (2016) K-nearest neighbors on road networks: a journey in experimentation and in-memory implementation. arXiv preprint arXiv:1601.01549

Aghamaleki JA, Behrad A (2016) Inter-frame video forgery detection and localization using intrinsic effects of double compression on quantization errors of video coding. Sig Process Image Commun 47:289–302

Amerini I, Ballan L, Caldelli R, Del Bimbo A, Serra G (2011) A sift-based forensic method for copy–move attack detection and transformation recovery. IEEE Trans Inf Forensics Secur 6(3):1099–1110

Bakas J, Naskar R (2018) *A* digital forensic technique for inter–frame video forgery detection based on 3D CNN. In: Proc. the international conference on information systems security, pp. 304–317

Bay H, Tuytelaars T, Van Gool L (2006) Surf: speeded up robust features. In: Proc. the European conference on computer vision, pp. 404–417

Bestagini P, Milani S, Tagliasacchi M, Tubaro S (2013) Local tampering detection in video sequences. In: Proc. IEEE 15th international workshop on the multimedia signal processing, pp. 488–493

Bidokhti A, Ghaemmaghami S (2015) Detection of regional copy/move forgery in MPEG videos using optical flow. In: Proc. the international symposium on artificial intelligence, pp. 13–17

Chen S, Tan S, Li B, Huang J (2016) Automatic detection of object-based forgery in advanced video. IEEE Trans Circuits Syst Video Technol 26(11):2138–2151

Davino D, Cozzolino D, Poggi G, Verdoliva L (2017) Autoencoder with recurrent neural networks for video forgery detection. Electron Imaging 2017(7):92–99

Fischler MA, Bolles RC (1981) Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun ACM 24(6):381–395

Jia S, Xu Z, Wang H, Feng C, Wang T (2018) Coarse-to-fine copy-move forgery detection for video forensics. IEEE Access 6:25323–25335

Li Y, Zhou J (2019) Fast and effective image copy-move forgery detection via hierarchical feature point matching. IEEE Trans Inf Forensics Secur 14(5):1307–1322

Li Z, Zhang Z, Guo S, Wang J (2016) Video inter-frame forgery identification based on the consistency of quotient of MSSIM. Security Communication Networks 9(17):4548–4556

Liu Y, Huang T (2017) Exposing video inter-frame forgery by Zernike opponent chromaticity moments and coarseness analysis. Multimed Syst 23(2):223–238

Long C, Smith E, Basharat A, Hoogs A (2017) A C3D-based convolutional neural network for frame dropping detection in a single video shot. In: Proc. IEEE conference on computer vision and pattern recognition, pp. 1898–1906

Long C, Basharat A, Hoogs A, Singh P, Farid H, Rafi (2019) A coarse-to-fine deep convolutional neural network framework for frame duplication detection and localization in forged videos. In: Proc. IEEE conference on computer vision and pattern recognition WS, pp. 1–10

Lowe DG (2004) Distinctive image features from scale-invariant keypoints. Int J Comput Vis 60(2):91–110

Muja M, Lowe DG (2014) Scalable nearest neighbor algorithms for high dimensional data. IEEE Trans Pattern Anal Mach Intell 36(11):2227–2240

Saddique M, Asghar K, Bajwa UI, Hussain M, Aboalsamh HA, Habib ZJIA (2020) Classification of authentic and tampered video using motion residual and parasitic layers. IEEE Access 8:56782–56797

Singh RD, Aggarwal N (2015) Detection of re-compression, transcoding and frame-deletion for digital video authentication. In: Proc. 2nd international conference on recent advances in engineering and computational sciences, pp. 1–6

Su L, Li C, Lai Y, Yang J (2018) A fast forgery detection algorithm based on exponential-fourier moments for video region duplication. IEEE Trans Multimed 20(4):825–840

Subramanyam AV, Emmanuel S (2012) Video forgery detection using HOG features and compression properties. In: Proc. IEEE 14th international workshop on the multimedia signal processing

Yu L, Wang H, Han Q, Niu X, Yiu S, Fang J, Wang Z (2016) Exposing frame deletion by detecting abrupt changes in video streams. Neurocomputing 205:84–91

Zhang Z, Hou J, Ma Q, Li Z (2015) Efficient video frame insertion and deletion detection based on inconsistency of correlations between local binary pattern coded frames. Secur Commun Netw 8(2):311–320

Zhong J-L, Pun C-M (2019) Copy-move forgery detection using adaptive keypoint filtering and iterative region merging. Multimed Tools Appl 78(18):26313–26339

Zhong JL, Pun CM, Gan YF (2020) Dense moment feature index and best match algorithms for video copy-move forgery detection. Inf Sci 537:184–202