



# Forgery detection of motion compensation interpolated frames based on discontinuity of optical flow

Xiangling Ding<sup>1</sup>  · Yanming Huang<sup>1</sup> · Yue Li<sup>2</sup> · Jiale He<sup>3</sup>

Received: 17 September 2019 / Revised: 23 June 2020 / Accepted: 13 July 2020 /

Published online: 6 August 2020

© Springer Science+Business Media, LLC, part of Springer Nature 2020

## Abstract

Motion Compensated Frame Interpolation (MCFI), a frame-based video operation to increase the motion continuity of low frame rate video, can be adopted by falsifiers for forging high bitrate video or splicing videos with different frame-rates. For existing MCFI detectors, their performance are degraded by stronger video compression, and noise. To deal with this problem, we propose a blind forensics method to detect the adopted MCFI operation. After investigating the synthetic process of interpolated frames, we discover that motion regions of interpolated frames exist some local slight artifacts, causing the optical flow based inter-frame discontinuity. To capture this irregularities introduced by various MCFI techniques, compact features are designed, which are calculated as Temporal Frame Difference-weighted histogram of Local Binary Pattern computed on Optical Flow field (TFD-OFLBP). Meanwhile, Local Inter-block and Edge-block difference Features (LIEF) are further proposed to detect interpolation frames with stable content. Besides, a set of forensics tools are adopted to eliminate the side effects of possible interferences of the scenes change, sudden lighting change, focus vibration, and some original frames with inherent local artifacts. Experimental results on four representative MCFI software and techniques show that the proposed approach outperforms existing MCFI detectors and also with robustness to compression, and noise.

**Keywords** Video forensics · Identification · Motion-compensation frame-interpolation · Optical flow

---

✉ Xiangling Ding  
xianglingding@163.com

<sup>1</sup> School of Computer Science and Engineering, Hunan University of Science and Technology, Xiangtan, 411004, China

<sup>2</sup> The Computer School, University of South China, Hengyang, 421001, China

<sup>3</sup> School of Information Science and Engineering, Hunan University, Changsha, 410082, China

## 1 Introduction

Using inexpensive and portable video capture devices, digital videos are recorded and then they are enriching our own lives. Meanwhile, the high speed development and spread of powerful video editing software makes it easier to counterfeit digital videos without leaving obvious perceptible traces. These faked videos are very difficult, if not impossible, to be directly distinguished through human vision. This overthrows our traditional views of “seeing is believing”, which gives rise to serious crisis of public confidence. Video forensics techniques, which attempts to verify the genuineness of digital videos, has become a research hotspot in the field of information security. Especially, passive or blind video forensics techniques [28, 34], which expose forging traces without any pre-embedding ancillary information such as watermarking or digital signatures, are extensively researched in recent years.

Digital video can be regarded as a series of successive image frames. Almost all potential image manipulations can be deployed to produce faked video frame. Therefore, most passive image forensic detectors can be extended to video forensics in a direct or indirect way [28, 34]. But, digital video has an additional temporal dimension. Video falsifications include inter-frame and intra-frame manipulations, which are corresponded to temporal-dimension and spatial-dimension tampering, respectively. Video frame-based alterations such as frame inserting, frame deleting, frame repetition, and Group Of Pictures (GOPs) reorganization are temporal-dimension falsifications specific to digital video. By far, there are some well designed forensics methods to disclose inter-frame forgeries including frame inserting, frame deleting, and frame copy-move by exploiting inherent traces left by video editing operations such as optical flow correlation [20], motion compensated residual [13, 18], video’s P-frame prediction error [33], and the similarity of exponential-fourier moments [35].

Motion-Compensation Frame-Interpolation (MCFI) is a special temporal-frame based video manipulation, which periodically synthesizes interpolation frames between each two successive reference frames to increase the frame-rate of video [23, 45]. MCFI can create a slow-motion effect from regular speed video and then polish up the visual quality of low frame-rate videos, which led to its widespread use in the production of video and film. Frame repetition and frame averaging are two simple frame-interpolation approaches which ignored objects’ motions between two reference frames. However, they often cause motion jerkiness and ghosting traces for non-stationary videos, respectively. Therefore, MCFI approaches have been proposed to eliminate these obvious traces by estimating motions as close as possible to true motions between two reference frames. In these approaches, various assumptions and strategies are developed to go after better precision of Motion Estimation (ME) and better visual effects of Motion Compensated Interpolation (MCI). As a results, more natural and realistic faked high frame-rate videos are synthesized without leaving obvious visible artifacts.

Although the original intention of MCFI is not designed for video forgery, it still might be adopted by a falsifier for malicious purposes. First, when faked frame-rate videos are released over video-sharing websites, they will not only waste many storage space but also mislead user’s visit. Second, two videos with different frame-rates might be spliced by up-converting the low frame-rate video to match the higher one. Third, MCFI might invalidate the video watermarking system or near-duplicate video detection because of the loss of temporal synchronism. Forth, MCFI can be deployed as anti-forensic strategy to attack inter-frame continuity based forensics detectors because MCFI operation can eliminate the jump-cut effects stemmed from inter-frame edition such as frame deletion or insertion, by using

two frames on the both sides of the frame deletion or insertion point as reference frames [43].

Until now, the forensic approaches towards detecting MCFI operation can be roughly divided into three categories. Algorithms of the first category exploit inter-frame similarity-analysis [3, 39] or the periodicity of noise level [26] to reveal frame repetition and frame averaging. These methods can obtain high detection performance due to existing motion jerkiness and ghosting traces, but, they cannot directly apply to expose the MCFI operation. Unlike the first category, the algorithms in the second and third category are focus on the detection of MCFI manipulation. The second category is to estimate original frame rate by utilizing video-level's periodic effects, such as the prediction errors between estimated frames and original ones [2], the motion artifact on the objects motion trajectories [21] and the noise level variation [25]. They can effectively estimate original frame rates, yet cannot locate interpolation frames. The third category goes further to locate interpolated frames by frame-wise analysis, instead of video-wise analysis (or periodic analysis). Since there often exists edge discontinuities or over-smoothing artifacts in interpolated frames, Yao et al. [44] and Xia et al. [42] discover MCFI forgery by using the periodic properties of edge intensity and average texture variation, respectively. They achieve desirable localization of interpolation frames for high-quality compressed videos. However, their performance deteriorates seriously when the tampered frame-rate is much higher than the original one, especially more than four times. Subsequently, we discuss the localization problem of interpolated frames under real-world scenarios such as H.264/AVC compression, noise or blur [12]. In this method, the candidate artifact regions is firstly selected after investigating the existing strong correlations between artifact regions and high residual energies. Then, the mean value of absolute high-order Tchebichef moments of selected candidate artifact regions is used to model temporal inconsistencies. Finally, a sliding window is employed to locate interpolated frames. Recently, Yao et al [43], as a pioneer, makes use of the MCFI strategy to invalidate inter-frame continuity based the frame-deletion video detection. Whereafter, they present a global and local joint feature to attack this anti-forensic strategy. Besides, a detector is further proposed to judge the absence or presence of MCFI forgery in a environment of unknown MCFI techniques by utilizing the statistical moments from motion-aligned frame differences and the one-class support vector machine [11].

These existing forensics methods for MCFI tampering obtain satisfactory detection results, but still can not differentiate which MCFI technique has been deployed for the synthesis of interpolated frames. In practical forensics cases, forensic analysts may want to further investigate the specific adopted MCFI technique after detecting a faked high frame-rate video. This is an essential step towards estimating the critical parameters of MCFI operations such as block size participated in ME and MCI, which is actually an in-depth goal of blind video forensics. As far as we know, there is only one discussion [10] about this topic. In this method, residual signal, which is assumed to be the difference between an interpolated frame and its absent original frame, is firstly designed as the clue for MCFI forgery detection. Then, Spatial and Temporal Markov Statistics Features (ST-MSF) are extracted, followed by feed into the Error-Correcting Output Code (ECOC) framework on the basis of Ensemble classifier to identify the adopted MCFI technique. It can identify the adopted MCFI algorithm for both uncompressed videos and compressed videos with high perceptual qualities, but its performance degrades under stronger compression. Its performance also degrades when there are stable content in video frames. Furthermore, candidate videos are usually contaminated by noise. These circumstances are not considered in this work.

Tracking the artifacts left by MCFI editing is an essential task for its forensic analysis. By analyzing the constitutive principle of interpolated frames, we observe that motion

regions of interpolated frames usually contain local slight artifacts, such as blurring or deformed object. This inevitable cause the video inter-frame discontinuity, especially the video motion features, Optical Flow (OF), which gives the motion/flow information in its finest resolution. In the past few years, this property has been employed into some frame-insertion forensics methods. In [5], Chao et al. computed the sum of the optical flow values between two adjacent frames along with the horizontal and vertical direction, and then applied a window based rough detection method with binary searching scheme to detect frame insertion forgery. Wang et al. [40] utilized gaussian distribution to model the probability distributions of optical flow variations for authentic videos. Abnormal flow variations were considered to locate the Frame Manipulated Positions (FMP) in forged videos. After that, velocity field inspired measure was proposed by Wu et al. [41]. The discernible peaks in the relative factor sequences computed by velocity field indicate the FMP. Recently, Singh and Aggarwal [32] presented a frame-insertion forensics method based on the brightness gradient component of optical flow. They used the variance computed from optical flow, and its fast fourier transform coefficients to detect jump-cut effects in tampered video which is not easily perceived for human eyes. As an advantage, all these four methods can be used to expose frame insertion counterfeit, and effectively locate the FMP in a video.

Although both frame-interpolation by MCFI and frame-insertion operation intercalate a series of frames between the FMP, they possess the essential differences in the process of fabrication of tampered videos. First of all, MCFI algorithms utilized estimated motions between two adjacent reference frames to synthesize sets of interpolated frames in the middle of them. Thereby, interpolated frames and reference frames share similar characteristics, such as contrast, brightness, and saturation levels. On the contrast, the forged videos for validating the frame-insertion operation are constructed by inserting sets of frames from other videos into the target videos, as reported in aforementioned four methods. This process of forgery creation tends to hard maintain constant frame characteristics between the source and target videos, causing the jump-cut effects at the FMP. As a result, the strength of discontinuity in OFs for frame-insertion is much stronger than that for interpolated frames by MCFI. Secondly, the detection capability of these four methods mainly depend on the discernible peaks of OFs in the FMP, which are originated from the nature of the frame insertion process, the jump-cut effects. But, because MCFI techniques can reduce abnormal peaks to preserve the continuity of objects' motion as far as possible, they are used as anti-forensic strategy to attack OF-discontinuity based frame-insertion forensics detectors [43]. Therefore, when these methods directly extended to MCFI detection, their performance degrades.

In light of the drawback of the existing OF-discontinuity based forensics methods for MCFI detection, we propose an effective feature extraction method by using the Temporal Frame Difference-weighted histogram of Local Binary Pattern (LBP) [24, 29] calculated on the OF field (TFD-OFLBP) to characterize the discrepancy among calculated OF. Subsequently, since different local artifacts by various MCFIs produce different degree irregularity of OF, an Error-Correcting Output Code (ECOC) [9] framework on the basis of Ensemble classifier [22] is adopted to identify the adopted MCFI technique. Moreover, since the video frames with stable content will be ignored because of the very weak artifacts, Local Inter-block and Edge-block Features (LIEF) are designed as the complementary of the TFD-OFLBP to address this situation. In addition, three another forensics algorithms, including ACE forensics, Mean Gradient Judgement (MGJ), and Interpolation Periodicity Correction (IPC), are presented to eliminate four possible interferences, such as the scenes change, sudden lighting change, focus vibration, and some original frames with inherent

local artifacts due to acute object motion. The contributions of the proposed approach are four-aspects:

- The subtle effects in optical flow sequence introduced by MCFI operations are analyzed.
- To characterize the temporal change of OFs caused by MCFI, a compact global TFD-OFLBP features are extracted.
- To conquer the side effect of possible interferences, a set of forensics tools are integrated into our proposed system.
- The evaluated results on two frequently-used MCFI software and two representative MCFI techniques clearly indicate the efficacy of the proposed approach, and its robust against lossy compression, and noise.

The remainder of this letter is organized as follows. In Section 2, we formally analyzes the local artifact. Section 3 presents the proposed approach, followed by the evaluation results in Section 4. Conclusion is made in Section 5.

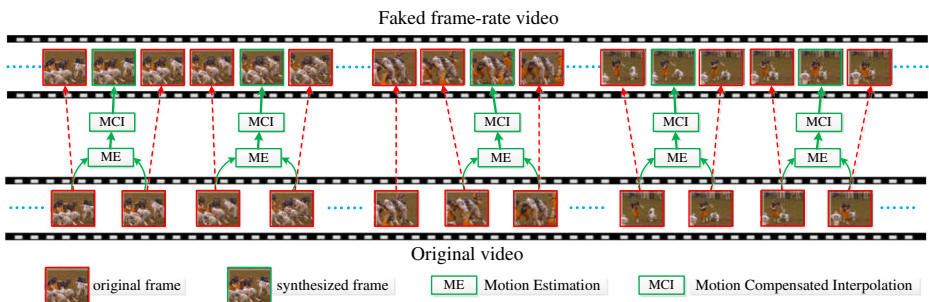
## 2 Local artifacts analysis

Since visual artifacts are important for blind forensics, we firstly investigate the local artifacts caused by various MCFI operations.

### 2.1 MCFI analysis

The MCFIs usually have two key steps: Motion Estimation (ME), and Motion Compensated Interpolation (MCI). ME is to estimate Motion Vectors (MVs) as close as possible to true motions of object, and MCI is to synthesize frames by estimated MVs. As illustrated in Fig. 1, a serial of synthesized frames from two adjacent original frames are periodic intercalated into a video shot. The visual quality of MCFIs highly depend on the accuracy of estimated MVs, or adopted MCI. For a block  $b_t(x, y)$  in interpolated frame  $f_t$ , it is often generated based on the estimated MV  $(v_x, v_y)$  between the previous frame  $f_{t-1}$ , and the following one  $f_{t+1}$  as:

$$b_t(x, y) = \frac{1}{2}[f_{t-1}(x + \frac{1}{2}v_x, y + \frac{1}{2}v_y) + f_{t+1}(x - \frac{1}{2}v_x, y - \frac{1}{2}v_y)] \tag{1}$$



**Fig. 1** The illustration of faked frame-rate video. Synthesized frames (marked in green) are produced from two adjacent original frames (marked in red), and then inserted between them. The faked frame-rate video is obtained after using MCFI with the periodicity of an synthesized frame after every two original frames. Our purpose is to determine the position of the synthesized frames, and then identify the adopted MCFI technique

Generally, MCFIs assume that the estimated MV represent true motion of object [23, 45]. However, due to the complexities of video appearance and motion, this assumption does not always hold. Though various strategies are exploited by MCFIs to achieve better ME, a few MVs may still be unreliable for the use of frame interpolation. Besides, to reduce their side effects, MCI uses weighted averaging for those regions guided by unreliable MVs, which have similarities with spatial-domain image smoothing. Therefore, this will inevitably lead to some local artifacts in interpolated frames.

Actually, different MCFIs may have disparate ME or distinct MCI. As a results, interpolated blocks have different reference blocks or weighting coefficients [10]. Therefore, there inevitably exists some differences among interpolated blocks by various MCFIs, accompanied by different local artifacts. Fig. 2a-e shows original frame and its corresponding interpolated frames. From it, we observe that local artifacts do exist, especially for red circle, and have different shapes and intensities from each other.

### 2.2 Derivation of the periodic variation of MVs

In this subsection, we deduce the periodicity of the MVs by the MCFIs. First, the operated process of the MCFI (1) can be represented in a simplified model as below:

$$P_{mcfi}(t) = \sum_{\chi=t-1}^{t+1} P_{origin}(\chi)\omega(\frac{t}{\Delta} - \chi) \tag{2}$$

where  $P_{mcfi}$ ,  $P_{origin}$ ,  $\omega$ , and  $\Delta$  denote the position of the interpolated block in interpolated frame, the position of reference block in adjacent frames, the weighting coefficient, and the period of interpolated.

When the video follows stationary signal, the periodic variation of position can be obtained after  $n_{th}$  derivation [8, 21, 27]:

$$D^n[P_{mcfi}(t)] = \frac{\partial P_{mcfi}(t)}{\partial t^n}, \text{ for } n > 0 \tag{3}$$

Substituting (2) into (3), we obtain:

$$D^n(P_{mcfi})(t) = \sum_{\chi=t-1}^{t+1} P_{origin}(\chi)D^n(\omega)(\frac{t}{\Delta} - \chi) \tag{4}$$

Suppose the original video follows a variance  $\sigma^2$  [8, 21, 27], (4) can be expressed as

$$\text{var}\{D^n(P_{mcfi})(t)\} = \sigma^2 \cdot \sum_{\chi=t-1}^{t+1} D^n(\omega)(\frac{t}{\Delta} - \chi)^2 \tag{5}$$

Further, for  $\lambda \in Z$ , above equation can be derived as

$$\text{var}\{D^n(P_{mcfi})(t + \lambda \cdot \Delta)\} = \sigma^2 \cdot \sum_{\chi=t-1}^{t+1} D^n(\omega)(\frac{t}{\Delta} - (\chi - \lambda))^2 \tag{6}$$



**Fig. 2** The 14<sup>th</sup> original frames of “Football” sequence and its interpolated frames [12] by MSU, MVTools, DSME [45] and MCMP [23]



Therefore, the variance of position is periodic with  $\Delta$

$$\text{var}\{D^n(P_{mcfi})(t)\} = \text{var}\{D^n(P_{mcfi})(t + \lambda \cdot \Delta)\} \quad (7)$$

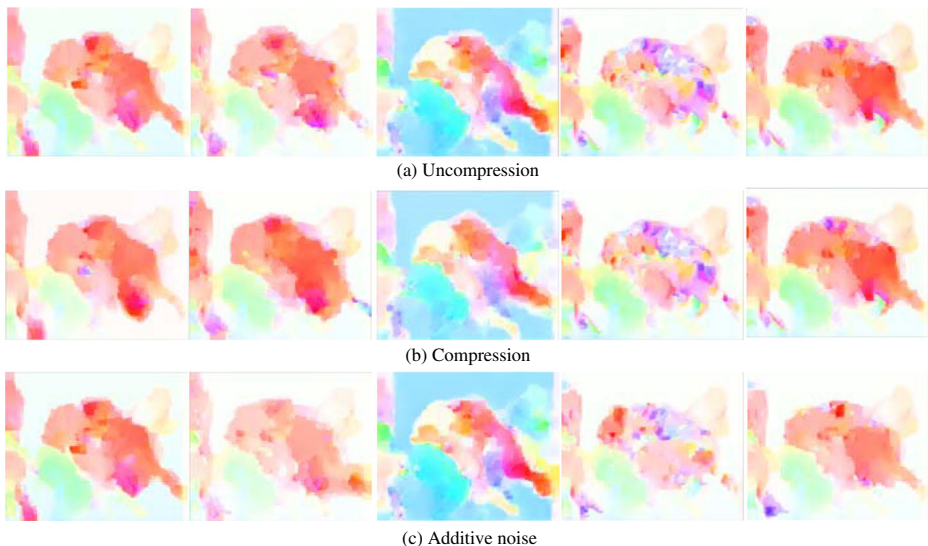
This imply that the position of interpolated block occur period change. Further, we can utilize MVs to reflect the change of position. Besides, due to different  $\omega$  for distinct MCFI, the MVs will be different for various MCFIs.

### 2.3 Measure of local artifacts

To detect the local artifacts, we can resort to the inherent correlation of these regions in pixel domain. However, this property is vulnerable for video compression [10]. Therefore, we want to adopt OF features to detect the interpolated frames. The reasons behind this are summarized as follows.

First, it is reported that because OFs are more resilient to video compression, and noise [20, 21], even though these post-operations introduces differences, the OF features will change little. Second, we know that when one or both of the frames used in the OF computation contain local artifacts, the data cost and smoothness cost will be violated, then to minimize the error of these costs, the OF algorithm will attempt to warp the current frame to match the appearance of the reference frame [30]. This will inevitably produce irregular OF maps in these regions where there have local artifacts. Third, there exist the periodic variation of MVs for up-conversion videos, the detailed derivation of which is given in above subsection. Forth, due to different intensities of local artifacts for interpolated results by MCFI operations, the degree of irregularity of OF maps in these regions may also different.

Figure 3 shows OF maps from original frame and interpolated frames by MSU, MVTools, DSME [45] and MCMP [23] under different scenario. From it, the shapes and colors of OFs are different among original frames and interpolated frames to a certain degree.



**Fig. 3** The OF maps under **a** uncompression; **b** compression; **c** additive noise. From the first column to the last column, the OF maps are respectively computed from the 14<sup>th</sup> original frame of “Football” sequence, and its corresponding interpolated frames by MSU, MVTools, DSME [45] and MCMP [23]

This implies local artifacts affect the regularity of OFs. Meanwhile, the shapes and colors of OFs are also different from each other. Moreover, post-processing on videos may be accompanied by MCFI forgery, making the falsifying hard to detect, such as lossy compression, and adding noise. As illustrated in (b) and (c) of Fig. 3, although the video shot undergoes lossy compression, and noise, the diversities among the original frame OFs and the interpolated frame OFs are still apparent, which motivates us to exploit it as forgery trace to identify the adopted MCFI techniques. While Fig. 3 is an demonstrative example, we also observe the phenomenon that different local artifacts lead to disparate intensity of OF work consistently well for interpolated frames by various MCFIs.

### 3 Proposed MCFI identification algorithm

In this section, we demonstrate the details of the proposed identification strategy for MCFI forgery. Figure 4 is the block diagram of the proposed approach. Firstly, OF maps are extracted from each frame, and then OFLBP is constructed by computing LBP on OF maps to characterize the subtle change among them. Secondly, the temporal frame difference-weighted histogram of OFLBP (TFD-OFLBP) is calculated by simultaneously considering the variation of pixel domain and OF domain. Thirdly, due to weak local artifact in the video frames with stable content, a set of Local Inter-block and Edge-block Features (LIEF) are proposed to conquer this issue. Finally, a localization classifier is firstly executed with joint extracted TFD-OFLBP and LIEF, then a multi-class classifier is exploited for the identification of various MCFI techniques on the basis of the previous detected interpolation frames with TFD-OFLBP features. In addition, to exclude possible interferences in the form of the sudden lighting change, acute object motion, or focus vibration, and interpolation frames generated in the process of scene change, two auxiliary pre-processing tools, and one post-processing algorithm are integrated to reduce these false detection.

#### 3.1 OFLBP construction

Due to local artifacts of interpolated frames, there exist the irregularity of OF. Here, an existing OF method [4] is used in our work to extract low-level features. The magnitude of OF (MOF) are computed as

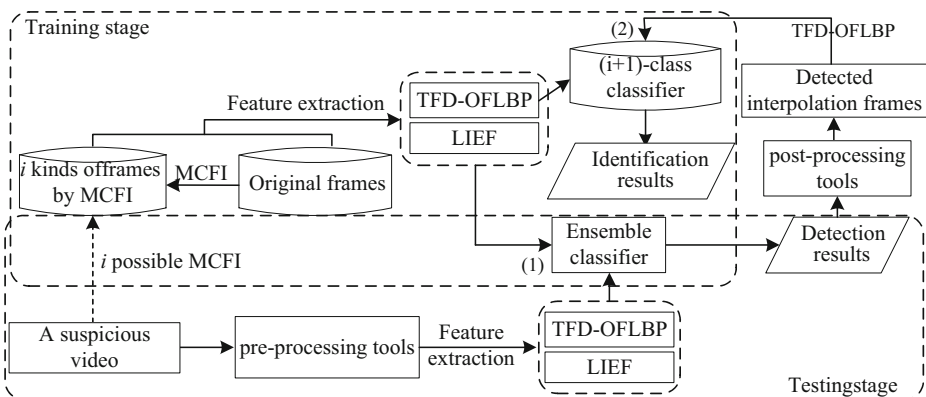


Fig. 4 Flowchart of the proposed approach includes (1) localization step, and (2) identification step



$$MOF(x, y) = \sqrt{v_x^2 + v_y^2} \quad (8)$$

Whereafter, the LBP is calculated for each pixel in the MOF. The LBP descriptor [24, 29] is designed to reflect the correlation between the center pixel and its adjacent pixel. Due to it has effective representation capability of micro-structure, we believe that it can be used to discriminate the subtle difference among MOFs by various MCFIs. By applying the LBP descriptor on the MOF, the OFLBP at each pixel is inferred as

$$OFLBP_{N,R} = \sum_{i=0}^{N-1} \delta(MOF_i - MOF_c) \cdot 2^i \quad (9)$$

$$\delta(MOF_i - MOF_c) = \begin{cases} 1, & MOF_i - MOF_c \geq 0 \\ 0, & MOF_i - MOF_c < 0 \end{cases}$$

where  $N$  and  $R$  are the number of adjacent pixel, and the radius of the neighborhood (Please refer to [29] for details),  $MOF_c$ , and  $MOF_i$  are the MOF at center pixel and its neighbor.

OFLBP represents the inter-MOF relationship in an MOF adjacent pixel, and such micro-structural patterns are effective to capture the subtle difference among MOFs caused by different local artifact from various MCFIs.

Figure 5 shows five  $8 \times 8$  blocks, for which we enlarge 4 times for convenient observation, extracted from the head area of the second player from the right in Fig. 2a-e, and their corresponding MOFs, and OFLBPs. It is observed that although various MCFIs change the pixel values to different degree, the discrepancy among the MOFs of various MCFIs, especially for MVTools, DSME, and MCMP, do exist but they are not obvious. As a results, different irregularity of OF change the OFLBP patterns in their own characteristic ways, making it an effective measure to discriminate the adopted MCFIs.

### 3.2 Temporal frame difference-weighted OFLBP histogram

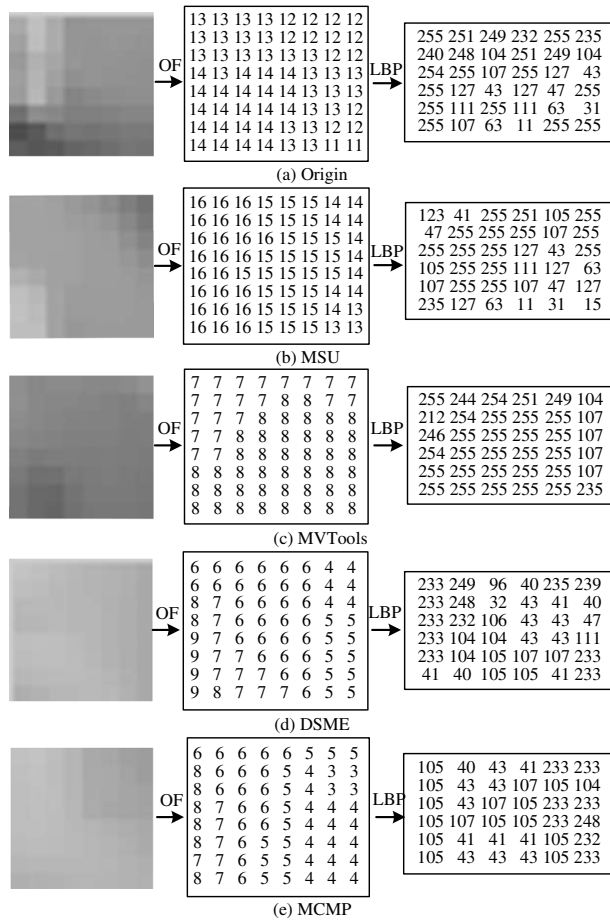
OFLBP only consider the irregularity of OF, and excludes the change of pixel value in local artifact regions. However, the change of pixel value by various MCFIs is beneficial to identify the adopted MCFIs, which is verified in method [10]. To effectively capture the difference from pixel domain, we proposed to fuse these discrepancy of local artifact into a single representation. We propose to accumulate the values of temporal frame difference with same OFLBP pattern, which can be treated as temporal frame difference-weighted OFLBP histogram, the calculated process is described as.

$$\varphi_t(x, y) = \frac{1}{2} [(f_t(x, y) - f_{t-1}(x, y)) + (f_t(x, y) - f_{t+1}(x, y))] \quad (10)$$

$$h_{oflbp}(\vartheta) = \sum_{i=1}^W \sum_{j=1}^H \varphi_t(i, j) \cdot \delta(OFLBP_{N,R}(i, j), \vartheta) \quad (11)$$

$$\delta(\alpha, \beta) = \begin{cases} 1, & \alpha = \beta \\ 0, & otherwise \end{cases}$$

where  $W$  and  $H$  denotes the width and the height of a video frame,  $\vartheta \in [0, K]$  is the possible OFLBP patterns,  $\varphi_t$  is the weight assigned to the OFLBP pattern,  $\alpha$  and  $\beta$  are the OFLBP pattern. Here, we use the temporal frame difference calculated by (10) to embody the discrepancy among motion regions by various MCFIs as the OFLBP weight of each pixel. In this way we highlight the irregularity of OF, and also take into account the change from pixel domain. During experiments, for OFLBP, the  $N$  and  $R$  are respectively set as 8 and 1. Thus, the extracted features are 256 dimensions. Fig. 6a-e show the average histograms from some original frames, and corresponding interpolated frames by MSU, MVTools, DSME,



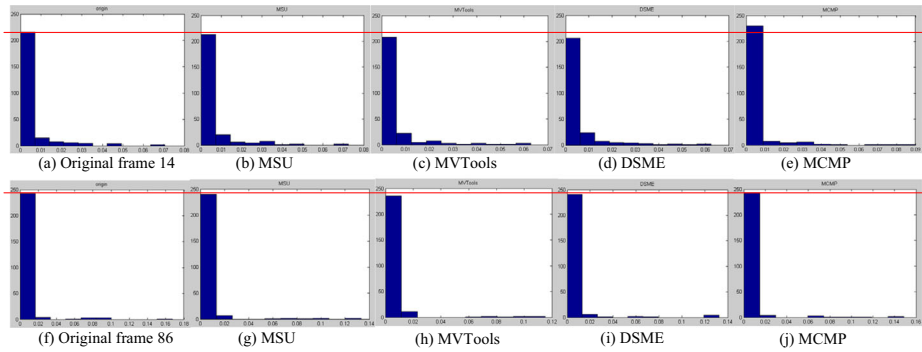
**Fig. 5** Extracted blocks(first row), MOFs(second row), and OFLBPs (third row) for original frame, and its corresponding interpolated frames

and MCMP, respectively. From it, we can infer that the TFD-OFLBP have more strong discriminative power by fusing OF domain, and pixel domain.

### 3.3 Local inter-block and edge-block difference feature

When the video frames exists stable content, the artifacts left by MCFI interpolation will be very weak. We can infer that the performance of the TFD-OFLBP will degrade in this case. To confirm this inference, we randomly selected several pristine frames with stable content, and their corresponding interpolated frames to observe the distribution of the TFD-OFLBP. The average histogram are shown in the Fig. 6f-j, one can see that the difference is not obvious for classification, even identification. Therefore, we design another sets of local block feature to address this problem.

Since the process of interpolation may add entropy to the quantized Discrete Cosine Transform (DCT) coefficients of predicted residuals, the dependence between two neigh-



**Fig. 6** The TFD-OFLBP maps of 14<sup>th</sup> frame (a)-(e), and 86<sup>th</sup> frame (f)-(j) of “Football” sequence corresponding to original frames, and its interpolated frames by MSU, MVTools, DSME, and MCMP, respectively

bouring block could change. Thus, we utilize the horizontal and vertical difference between two  $n \times n$  adjacent block to characterize this relation, calculated as follow:

$$\begin{aligned}
 f_{inter} = & \frac{\sum_{x=1}^{n \cdot \lceil \frac{W}{n} \rceil - n} \sum_{y=1}^{n \cdot \lceil \frac{H}{n} \rceil} |b_{dct}(x, y) - b_{dct}(x+n, y)|}{n^2 \cdot (\lceil \frac{W}{n} \rceil - 1) \cdot \lceil \frac{H}{n} \rceil} \\
 & + \frac{\sum_{x=1}^{n \cdot \lceil \frac{W}{n} \rceil} \sum_{y=1}^{n \cdot \lceil \frac{H}{n} \rceil - n} |b_{dct}(x, y) - b_{dct}(x, y+n)|}{n^2 \cdot \lceil \frac{W}{n} \rceil \cdot (\lceil \frac{H}{n} \rceil - 1)} \\
 n = & 4, 8, 16, \text{ or } 32
 \end{aligned} \tag{12}$$

where  $b_{dct}$  is the  $n \times n$  coefficient matrix after DCT, and  $W$  and  $H$  denotes the width and the height of the prediction residual of a video frame.

Because the interpolated process is mainly focus on the motion regions, we firstly extract the motion regions by using the extracted OF. Since any MCFI techniques inevitably produce some blurring effects or deformed structures in motion regions of interpolated frames, the features of edge-block in motion regions is design to represent the discontinuity dependency of boundary of block. As expected, interpolation indeed increases the discontinuity of block’s edge. We measure it as the sum of the difference between two adjacent blocks’boundary of motion regions in the spatial domain.

$$\begin{aligned}
 f_{edge} = & \frac{\sum_{b_t \in L_1} |b_t(n \cdot x, y) - b_t(n \cdot x + 1, y)|^\tau}{N_{L_1}} \\
 & + \frac{\sum_{b_t \in L_1} |b_t(x, n \cdot y) - b_t(x, n \cdot y + 1)|^\tau}{N_{L_1}} \\
 n = & 4, 8, 16, \text{ or } 32;
 \end{aligned} \tag{13}$$

where  $b_t$  is the block of the  $t^{th}$  frame in pixel domain, and  $L_1$  is the set of no-zero MOF blocks, in which the percent of no-zero MOF in a block is large than 50%,  $n$  is the size of block,  $\tau \in 1, 2$ . Finally, the 268-D global and local joint feature, which is combination of the 12-D of Local Inter-block and Edge-block Feature (LIEF) and the 256-D of the TFD-OFLBP, is used in this paper to expose interpolation forensics.

### 3.4 The pre-processing and post-processing

The purpose of above designed features is to find the interpolated frames with local artifacts. However, the output candidate interpolation frames may still contain possible interferences when the scenes change very rapidly, video frames to be tested may exist sudden lighting change, and focus vibration, or the video itself has some inherent local artifacts for acute object motion. Therefore, we propose a set of tools including two pre-processing, and one post-processing, i.e., ACE forensics [13, 19], mean gradient judgement, and interpolation periodicity correction for removal of the interfering frames.

For the case of scene change, since new object appears, or old object disappears, irregular OF may emerge, causing a false alarming. Meanwhile, because the luminance of the lighting vibration frame suddenly changes, OF maps in this case are also occur irregular. Furthermore, we observe that between the deviation of the scene change frame, or sudden lighting change frame, and its adjacent frames might have a large difference. Thus, we deploy the algorithm of ACE to measure the candidate interpolation frame  $f_t$  and its previous frame  $f_{t-1}$ . We calculate the histogram difference before, and after the equalization in luminance component of video frame, and set them as  $h_t$ , and  $h_{t-1}$ , respectively. If  $|\bar{h}_t - \bar{h}_{t-1}| > \tau$ , where  $|\cdot|$ , and  $\tau$  represent the operation of absolute value, and an empirical value, then, the candidate frame is judged as a frame of scene change, or lighting vibration.

Since focus vibration is mostly stemmed from camera focus adjustment, it does not change with the video content. Thus, its detection may be easy to implement with a simple metric. We observe that the sharpness in a normal frame is usually keep in a constant range, but, the focus vibration destroys this phenomenon. Thus, we can resort to the measurement of sharpness to judge whether focus vibration. Here, we use the mean gradient to measure it. When the mean gradient of a video frame is less than an empirical value  $\epsilon$ , it is judged as a focus vibration.

Because of acute object motion, sometimes original frames may burst out some inherent local artifacts, causing irregular OF. This might bring out erroneous judgment. Besides, since interpolated frames are often intercalated into low frame-rate video in a periodic way, the candidate interpolated frames should also be periodic. Therefore, the periodicity of interpolation should be employed to address this issue. Here, we take a example to illustrate the periodicity of interpolation. Suppose the frame rate is up-converted from 15 *fps* to 20 *fps*, there could exist an interpolation frame after every three original frames. When the periodicity of frame interpolation does not follow this inferred periodicity, it will be adjusted as an original frame, and vice versa.

As far as we known, the number of four interferences are generally few, and may exist simultaneously or not in one video shot. It is reported that a Set of Forensics Tools (FTS) [13, 14] may improve the final detection results. Motivated by this work, we treat these three algorithms as a FTS. In the process of practical blind forensics, these algorithms can be selectively used in accordance with the properties of the video shot. Thus, if other kinds of interferences appear, corresponding forensic tools can be integrated in similar ways into this FTS.

### 3.5 The identification strategy

After the detection of interpolation frames, and the elimination of interference frames, extracted TFD-OFLBP from those detected interpolated-frames are employed for the identification of various MCFI techniques. Because the Ensemble Classifier (EC) [22] compromises well between computational complexity and detection accuracy, and the

Error-Correcting Output Code (ECOC) strategy [9] is an excellent multi-class categorization tool as well, EC with its default settings is utilized to distinguish the pristine frames and interpolation frames, and then an Error-Correcting Output Code (ECOC) strategy built on EC is exploited to transform a multi-class problem into binary sub-problems.

In this paper, four MCFI techniques are involved. Therefore, a five-class (covering the pristine video without MCFI as a special class) classifier is made, and a scheme of pairwise coupling [17] is used. In ECOC strategy, a discrete decomposition matrix (codematrix) is firstly denoted as the five-class classification problem. Then, it is disassembled into  $N = 5 \times 6/2 = 15$  binary subproblems, i.e., dichotomies, in terms of 0s and 1s sequence of columns of the codematrix. After training the EC on these dichotomies, the extracted TFD-OFLBP are tested to generate a binary vector. The final identified MCFI method is designated as the class with the smallest Hamming distance between this vector and the codewords.

## 4 Experimental results and discussion

### 4.1 Experimental settings

Twenty known YUV videos in CIF format ( $352 \times 288$ ) with  $15\text{ fps}$  are selected [37]. To increase the quantity of samples, each original videos are partitioned into non-overlapped video clips with 100 frames. Two popular softwares [38] (MSU and MVTools) and two progressive MCFIs (DSME [45] and MCMP [23]) are used. Four target frame-rates are involved (20, 30, 60, and 120)  $\text{fps}$ . The faked high frame-rate videos are adopted as the basic data to conduct all the experiments in this paper, and denoted as **DS1**. In the constructive process of these tampered videos, optimized parameters are employed for both DSME and MCMP to acquire high-quality videos, and MSU only favor integer times interpolation.

Subsequently, two most popular video compression standards H.264/AVC, and H.265/HEVC, i.e., JM18.6 [15], and HM16.9 [16] are used to encode video for the experiments with configurations: Group-Of-Picture(GOP) sets as 8, and Quantization Parameter(QP) is within 12, 30, and 42. Other parameters use the default settings of the baseline profile. All the parameters for the calculation of MOF are set to the default values as reported in [4]. Since the influence of GOP length for detection results is relatively small [10], the results under different GOP length are not reported here. The H.264/AVC, and H.265/HEVC encoded **DS1** datasets are denoted as **DS2**, and **DS3**, respectively.

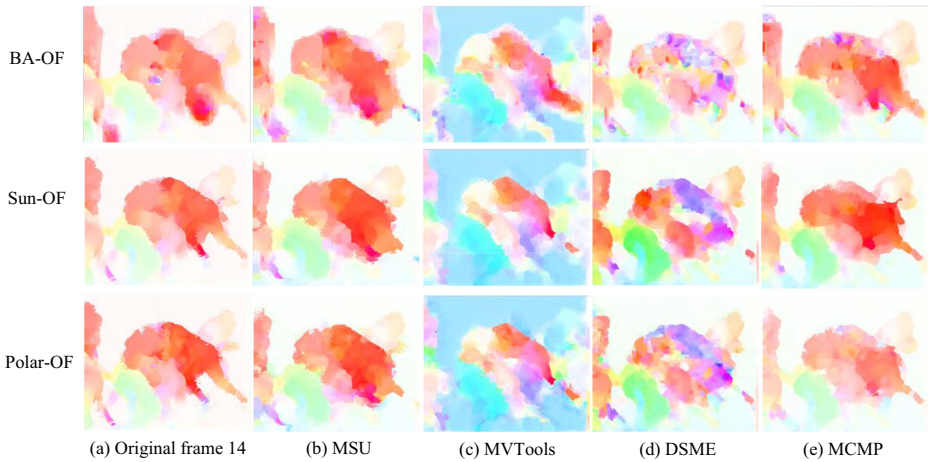
We also generated additional datasets to evaluate the robustness of the algorithm. For the test videos in **DS1**, each frame is attacked with Gaussian white noise, keeping their Signal Noise Ratio (SNR) be 33dB and 36dB, respectively. The videos with additive noises are denoted as **DS4**. Moreover, the **DS2** and **DS3** datasets are further added with Gaussian white noise following the above parameters, and the resultant videos are expressed as **DS5**, and **DS6**, respectively.

Interpolated frames and original frames are denoted as positive samples ( $S_p$ ) and negative samples ( $S_n$ ), respectively. The  $F_1$  [10, 31] is adopted for performance evaluation.

$$F_1 = \begin{cases} \frac{(\gamma^2+1) \cdot \text{Precision} \times \text{Recall}}{\gamma^2 \cdot \text{Precision} + \text{Recall}} & \text{if } \sum S_{tp} > 0 \\ 0 & \text{if } \sum S_{tp} = 0 \end{cases} \quad (14)$$

where  $\text{Precision} = \frac{\sum S_{tp}}{\sum S_{tp} + \sum S_{fp}}$ ,  $\text{Recall} = \frac{\sum S_{tp}}{\sum S_p}$

and  $\gamma$  controls the balance between *Precision* and *Recall*. Normally,  $\gamma$  is set to 1.  $S_{tp}$  and  $S_{fp}$  are true positive and false positive samples, respectively.



**Fig. 7** The MOF maps. **a–e** correspond to the OF maps from 14<sup>th</sup> original frames of “Football” sequence and its interpolated frames by MSU, MVTools, DSME, and MCMP, respectively. From the first row to the last row, they correspond to the OF maps by BA-OF, Sun-OF, and Polar-OF, respectively

In following experiments, the Error-Correcting Output Code [9] based on Ensemble classifier [22] with its default settings is employed to identify the adopted MCFIs [10]. Videos are randomly divided into two categories: 50% for training and the rest 50% for testing. The training and testing are repeated for 10 times, and the average results are reported.

#### 4.2 The effects of different optical flow algorithms

Some experiments have been conducted to prove the effect of accuracy of OF in the performance of the proposed method by choosing another two popular, and representative OF methods, including Sun-OF [36], and Polar-OF [1]. Firstly, the effects of different OF methods are evaluated on the MOF results. Figure 7 shows the MOF results on the 14<sup>th</sup> frame of the “Football” sequence, and its corresponding interpolated frames by MSU, MVTools, DSME, and MCMP by different OF algorithms. From it, we observe that there are only a few differences among them. Thus, any OF methods can be utilized for the generation of MOFs.

Subsequently, we conducted another experiment on the **DS2** with target frame-rate 30 *fps* under QP12. The average *F1*, and identification accuracy are reported in Table 1. From it, we can draw a conclusion that the performance of our proposed TFD-OFLBP features is not sensitive to the accuracy of OF method.

**Table 1** Average *F1*, and identification accuracy with different OF methods on the subset of **DS2** with target frame-rate 30 *fps* under QP12

OF methods	<i>F1</i>	identification rate
BA-OF [4]	87.67	70.11
Sun-OF [36]	88.04	70.35
Polar-OF [1]	87.7	70.18



### 4.3 The effects of different components of the proposed approach

Our proposed system is mainly originated from the discontinuity of OF, and the exclusion of some possible interferences. Thus, we conduct an experiment to evaluate the gains by LBP, temporal frame difference-weighted (TFD), LIEF, and three processing operations, i.e., ACE forensics, mean gradient judgement (MGJ), and interpolation periodicity correction (IPC), respectively. We employ the Histogram of Oriented Optical Flow (HOOF) [6] to directly extract OF features, in which we set 256 bins for better performance. Table 2 reports the average performance gain of  $F_1$ , and identification accuracy for tampered videos dataset DS2 with the whole target frame-rate under different QP values.

From it, we can conclude that each component promotes some performance, and the proposed system can achieve promising performance. First, although OF can embody the subtle difference among various MCFIs, the discrimination of OF is still weak. Therefore, LBP operator is used to enhance the discriminative power at a micro-level. Second, since there also exist the change of pixel value in local artifact regions, it is important to emphasize the difference from pixel domain. The TFD-OFLBP which combines the change from pixel domain and OF domain can provide better performance. Third, contrast with the number of scene change, sudden lighting change, and focus vibration, the number of original video frames with some inherent local artifacts, and stable content is much higher. Thus, they contribute a lot to the final results. Fourth, since the LIEF features and a Set of Forensics Tools (FTS), i.e., ACE, MGJ, and IPC, exclude the abnormal frames, the identification accuracy obtains enhancement.

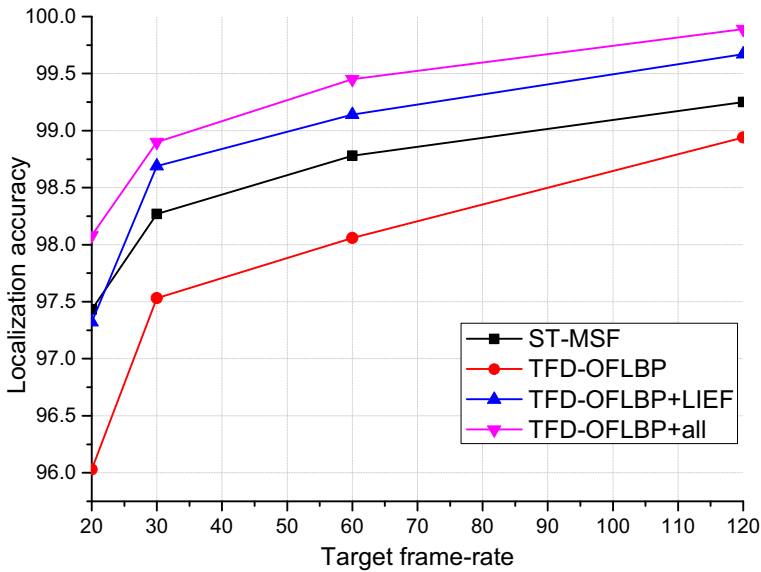
Furthermore, we also can observe that the average gains of the LIEF, ACE, MGJ, and IPC for  $F_1$ , and identification accuracy are 1.30%, 0.21%, 0.05%, 0.67% and 0.28%, 0.07%, 0.02%, 0.32%, respectively. Then, we can infer that the whole average gain of above four components for  $F_1$ , and identification accuracy are 2.22%, and 0.69%, respectively. This further verifies that the percentage of four interferences are relatively low, and they may exist simultaneously or not in one video shot. Therefore, the side effects of these four situations have little impact on the final detection accuracy of the proposed system.

### 4.4 Localization results for uncompressed video

Dataset DS1 is chosen for test, and comparisons are made between the proposed approach and our previous work [10] (hereinafter referred as ST-MSF). Figure 8 reports the average

**Table 2** Average performance gain of  $F_1$ , and identification accuracy for each component.(%)

	QP12		QP30		QP42	
	$F_1$	identification rate	$F_1$	identification rate	$F_1$	identification rate
HOOF (Only OF)	62.43	37.16	62.45	35.59	58.55	31.26
OFLBP	24.04	28.34	11.54	15.28	10.65	14.1
with TFD	4.56	12.34	3.14	6.4	4.15	7.83
with LIEF	1.55	0.36	1.25	0.29	1.09	0.2
with ACE	0.22	0.08	0.2	0.07	0.21	0.07
with MGJ	0.05	0.02	0.05	0.02	0.05	0.02
with IPC	0.97	0.39	0.55	0.3	0.48	0.26
final results	93.82	78.69	79.18	57.95	75.18	53.74



**Fig. 8** The average  $F_1$  on the forgery dataset **DS1**

localization accuracies, in which “TFD-OFLBP + all” represents the combination of TFD-OFLBP, LIEF, and FTS.

From it, we can see that the performances of our proposed TFD-OFLBP are approximately the same with ST-MSF because the global feature, TFD-OFLBP, and ST-MSF can effectively capture these interpolated frames with local artifacts. Since interpolated frames may generated from original frames with stable content, ST-MSF can not seize the distinction well. Meanwhile, TFD-OFLBP may not address the issue of scene change, sudden lighting change, focus vibration, and original video frames with some inherent local artifacts. With the integration of LIEF, and FTS, our proposed system can capture the anomalies between the pristine frame and the interpolated frame more effectively. As a result, our proposed system is gradually better than ST-MSF after combining LIEF, and FTS step by step. In the next experiments, we only take the proposed TFD-OFLBP features to compare with the state-of-the-art forensics methods for fairness.

#### 4.5 Localization accuracy of interpolated frame against different lossy compression with different configurations

We employed ST-MSF, and Yao et al. [44], or ST-MSF, and Chao et al. [5] for comparison under H.264/AVC or H.265/HEVC compression. Since there exist subtle irregular of OF maps, we set small thresholds (0.004) in Chao’s method to detect interpolated frames. Tables 3 and 4 give the localization accuracies of **DS2**, and **DS3** with AVC and HEVC under different QP values, respectively. The experimental results will be analyzed from the following different perspectives.

First, with the increase of frame-rate, the localization accuracies increase steadily for the proposed TFD-OFLBP and ST-MSF, whereas the accuracies firstly increase and then deteriorate for Yao. The reason is the intensity of local artifact is enhanced, which contribute to all MCFI detectors, however, when the target frame-rate reach to 120 *fps*, the periodicity of

**Table 3**  $F_1$  under H.264/AVC compression.(%)

MCFI	Detector	QP=12			QP=30			QP=42					
		20	30	60	120	20	30	60	120	20	30	60	120
MSU	Yao	–	62.27	80.32	75.40	–	61.09	70.55	59.63	–	57.22	63.51	59.63
	ST-MSF	–	82.19	91.81	93.63	–	62.52	65.01	78.21	–	58.09	60.52	69.20
	Proposed	–	91.45	93.39	96.71	–	80.48	83.06	84.15	–	76.68	77.47	79.72
MVTools	Yao	61.40	62.10	80.32	75.40	56.98	60.73	69.62	63.21	53.85	57.56	68.23	64.36
	ST-MSF	90.56	92.75	90.46	93.11	58.03	61.29	68.11	72.38	60.08	61.35	62.87	66.13
	Proposed	90.2	90.79	94.87	95.87	73.95	77.44	82.07	85.14	73.97	78.15	79.62	82.88
DSME	Yao	65.76	60.03	84.85	79.61	60.13	59.21	69.88	65.31	56.13	56.57	70.51	61.65
	ST-MSF	88.67	92.74	94.79	97.91	64.57	68.78	80.91	84.11	56.39	58.81	67.73	84.31
	Proposed	86.34	87.45	92.57	97.11	70.87	73.12	75.10	83.06	63.20	65.72	71.82	80.73
MCOMP	Yao	60.26	61.01	82.27	76.15	62.15	61.16	66.35	59.85	53.64	57.56	64.59	63.02
	ST-MSF	88.73	90.82	94.94	96.34	60.69	63.66	75.32	82.41	52.85	55.60	64.86	72.79
	Proposed	84.67	85.94	90.75	96.17	68.23	71.70	73.96	80.71	60.93	64.23	72.01	80.45
Average	Yao	62.47	61.35	81.94	76.64	59.75	60.55	69.10	62.00	54.54	57.23	66.71	62.17
	ST-MSF	89.32	89.63	93.00	95.25	61.10	64.06	72.34	79.28	56.44	58.46	64.00	73.11
	Proposed	87.07	88.91	92.90	96.47	71.02	75.69	78.55	83.27	66.03	71.20	75.23	80.95

**Table 4**  $F_1$  under H.265/HEVC.(%)

MCFI	Detector	QP=12			QP=30			QP=42					
		20	30	60	120	20	30	60	120	20	30	60	120
MSU	Chao	–	75.00	64.56	51.22	–	65.39	50.00	40.22	–	55.67	42.13	38.00
	ST-MSF	–	81.76	88.94	92.17	–	61.30	65.01	76.91	–	57.49	60.30	67.89
	Proposed	–	87.00	89.96	94.00	–	78.30	81.11	83.56	–	74.11	76.44	78.39
MVTools	Chao	75.00	75.67	65.33	55.44	65.11	64.35	55.15	45.96	53.85	52.89	45.68	40.33
	ST-MSF	90.33	90.89	92.30	92.96	58.03	60.87	66.39	71.98	58.11	60.54	62.31	65.11
	Proposed	89.05	90.67	93.53	95.00	71.33	74.85	79.83	81.78	70.01	71.56	76.44	79.56
DSME	Chao	78.10	77.00	61.34	58.45	66.10	65.00	50.22	44.67	55.11	54.00	40.87	35.17
	ST-MSF	86.39	89.87	92.17	95.83	64.56	66.85	80.33	83.00	54.90	56.30	65.57	80.11
	Proposed	85.22	87.13	91.45	95.89	68.67	72.39	78.96	83.11	62.52	63.76	70.19	80.67
MCMP	Chao	70.00	70.11	64.00	50.56	68.34	66.67	51.00	42.04	51.56	50.11	38.35	31.33
	ST-MSF	86.67	89.35	92.45	94.18	60.66	62.35	72.17	80.35	52.04	54.22	63.70	70.13
	Proposed	83.33	85.87	90.56	95.52	65.35	70.00	73.89	80.56	60.61	63.13	70.00	78.67
Average	Chao	74.37	74.45	63.81	53.92	66.52	65.35	51.59	43.22	53.51	53.17	41.76	36.21
	ST-MSF	87.80	87.97	91.47	93.79	61.08	62.84	70.98	78.06	55.02	57.14	62.97	70.81
	Proposed	85.87	87.67	91.38	95.10	68.45	73.89	78.45	82.25	64.38	68.14	73.27	79.32

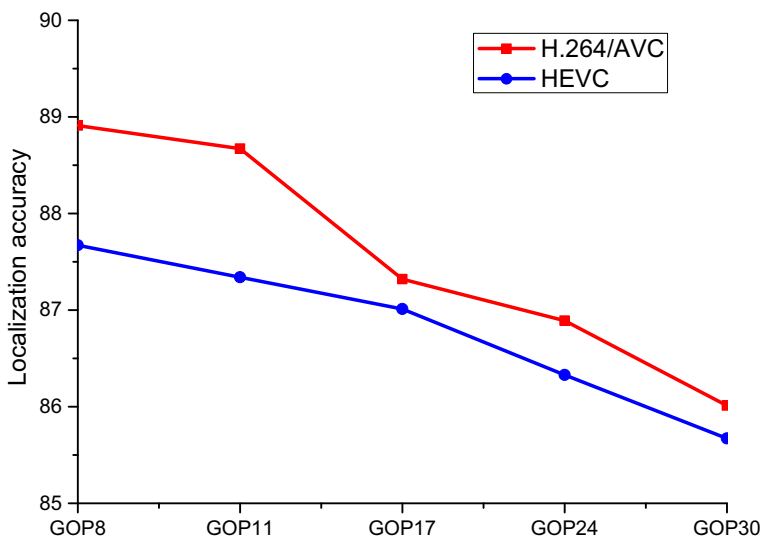
edge intensity [44] exists aliasing, causing performance degradation. For Chao's method, when the number of interpolation become large, the irregular fluctuate between interpolated frame and pristine frames or interpolated frame will be relatively small, causing performance degradation. This further verify that MCFI can be deployed as anti-forensic strategy to attack inter-frame continuity based forensics detectors.

Second, the proposed TFD-OFLBP is not always the best one in some cases because the ST-MSF considers the detection of original interference frames by using the tool sets including static frame detection, scene change detection and interpolation periodicity detection. However, since the percentage of original interference frames is relatively low compared to the entire tampered video dataset, the advantages of measurement of original interference frames die down with the increase of frame-rate, or with enlargement of QP value. Consequently, our proposed method is gradually better than ST-MSF. Meanwhile, if our proposed TFD-OFLBP is also integrated with our proposed forensics tool sets, including LIEF, ACE, MGJ, and IPC, its performance can be better than ST-MSF.

Third, with enlargement of QP value, motion regions become smoother, which gradually degrades the location accuracy of all detectors. However, the proposed method has small fluctuation for considering both pixel domain and MOF domain.

Forth, the localization results of H.264/AVC encoded videos are obviously higher than that of H.265/HEVC compressed videos in the same QP values after comparing Tables 3 and 4. The reason behind it is that H.264/HEVC does well in motion estimation, and motion compensation, the residual gets weaker followed by the difference of local artifacts between pristine frames, and interpolation frames smaller. In summary, the experimental results demonstrate that the proposed TFD-OFLBP is effective for differentiating the interpolated frames and original frames.

Subsequently, although the influence of GOP length for detection results is relatively small, which is similar to the comparable method ST-MSF, we still verify this conclusion with another experiment. Another database with target frame-rate 30 *fps* under QP = 12, and the GOP is within 11, 17, 24, 30 under H.264/AVC, or HEVC is selected. Figure 9 reports



**Fig. 9** The average detection accuracy (%) with different GOP length under H.264/AVC, and HEVC

the average detection results. From this figure, we can see that with the increment of GOP length, the proposed approach has a very slow decline of detection accuracy as expected. Therefore, our proposed approach is suitable for different GOP under *H.264/AVC*, and *HEVC*.

#### 4.6 Identification of the adopted MCFI techniques

The experiment is a 5-class (including the original videos without MCFI as a special class) classification problem due to four known MCFIs involved. Datasets **DS2**, and **DS3** are chosen for experiments. “mixed” means the mixture of target frame-rate *20 fps*, *30 fps*, *60 fps* and *120 fps*.

Tables 5 and 6 give the average identification accuracies, which are the elements of confusion matrix along diagonal direction, acquired by using *ST-MSF* [10], and proposed *TFD-OFLBP*. Apparently, the accuracies also increase with the increment of target frame rate. Since the proposed *TFD-OFLBP* integrated with *LIEF*, and *FTS* can exclude more possible interferences in contrast with the *ST-MSF*, this helps to improve the performance. Meanwhile, the proposed *TFD-OFLBP* is superior to the *ST-MSF* for better capacity to classify the pristine frames and interpolated ones by effectively fusing the changes in the pixel domain and *OF* domain. This further represents the important of correct classification of interpolated frames. Actually, this is in accordance with our expectation. Moreover, the loss of accuracy is also caused by strong compression, since it smoothes local artifacts left by various MCFIs. We can also observe that the result of *H.264/AVC* videos are better than *H.265/HEVC*. This is because as the most advanced video coding strategy, the block measurement scheme of *H.265/HEVC* is more refined, and motion estimation is more precise, leading to more fewer residuals. Whereafter, the difference of local artifacts among various MCFIs become small, causing the accuracy degrade.

Tables 7 and 8 report the average accuracies of “mixed” tampered videos under *QP12* or *QP42*. From it, we observe that our proposed *TFD-OFLBP* can effectively identify the adopted MCFIs. Meanwhile, with the increase of *QP* value, we find that it became gradually difficult to distinguish interpolated frames and original compressed frames because from *ME* and *MC* points of view, *H.264/AVC* or *H.265/HEVC* compression has a similar character with MCFIs. Besides, when only slight differences occur in the motion search pattern or weighted average of two MCFI techniques, it is extremely difficult to recognize them by investigating irregularities of *OF* maps.

#### 4.7 Identification of unknown MCFI methods

To evaluate the identification ability of the proposed *TFD-OFLBP* for the unknown MCFI methods, we conduct another database, in which the fake frame-rate videos generated by *YUVsoft*, and *AOBMC* [7] are not adopted in the training stage, but they are adopted in the testing stage. The subset of **DS1** with target frame-rate *30 fps* is trained. Table 9 demonstrates the experimental results. For the tampered videos by *YUVsoft* or *AOBMC*, the proposed *TFD-OFLBP* can detect them as forgery videos, and their types are judged as open *MC-FRUC* software or *OBMC*-based algorithms. The detection accuracies of *YUVsoft*, and *AOBMC* are calculated by summing up the numbers except the “Pristine” column. They are 95.76%, and 97.84%, respectively. The proposed *TFD-OFLBP* can correctly detect the counterfeited videos by unknown MCFI methods as the forged ones, but it still can not correctly judge the unknown MCFI methods with totally different *ME* or *MCI* strategies as a new type. This is a limitation for the proposed approach.



**Table 5** The average identification accuracies (%) with **DS2** along the diagonal direction in the corresponding confusion matrices

Detector	QP12			QP30			QP42													
	Uncompressed	20fps	30fps	60fps	120fps	mixed	20fps	30fps	60fps	120fps	mixed	20fps	30fps	60fps	120fps	mixed				
ST-MSF	86.93	87.39	88.17	81.27	68.29	70.11	75.44	83.17	75.99	35.02	38.98	53.21	64.81	56.86	32.44	33.98	40.99	52.06	49.98	
Proposed	87.25	87.76	89.52	91.17	88.09	69.73	73.24	79.93	86.87	83.67	49.67	51.37	54.75	67.88	66.07	46.73	47.89	52.45	61.84	59.78

**Table 6** The average identification accuracies (%) with **DS3** along the diagonal direction in the corresponding confusion matrices

Detector	QP12				QP30				QP42											
	Uncompressed	20 fps	30 fps	60 fps	120 fps	mixed	20 fps	30 fps	60 fps	120 fps	mixed	20 fps	30 fps	60 fps	120 fps	mixed				
ST-MSF	85.67	86.12	87.11	87.89	80.89	66.67	68.23	73.56	82.11	74.35	34.00	37.11	51.34	62.78	55.67	30.44	33.11	38.20	48.78	45.76
Proposed	86.56	86.98	88.17	90.00	86.35	68.04	70.11	76.56	85.22	80.04	47.11	50.35	54.35	66.22	64.83	45.11	46.30	50.56	60.56	58.91

**Table 7** Confusion matrix on “mixed” tampered Dataset of **DS2** under QP12/QP42 status. The asterisks “\*” denote that the corresponding values are below 1%

QP12/QP42 FRUCs	Classified as				
	MSU	MVTools	DSME	MCMP	Compressed
MSU	85.23/60.67	*/*	*/*	*/*	*/*
MVTools	*/*	89.67/71.23	*/*	*/*	*/*
DSME	*/*	*/*	82.65/62.28	16.87/24.41	1.34/6.72
MCMP	9.71/11.02	5.78/7.15	6.56/7.59	71.35/42.09	8.34/30.53
Compressed	4.23/27.39	4.23/21.87	9.05/29.61	10.89/33.36	89.56/62.57

#### 4.8 Detection results under additive noise

Since the proposed TFD-OFLBP on **DS2**, or **DS3** has an approximative performance, Dataset **DS5** are exploited to evaluate the robustness of the proposed TFD-OFLBP under additive noise. At last, **DS4**, and **DS5** are selected as datasets in next experiments. Here, Matlab function “imnoise” is used to produce Gaussian-white-noise polluted videos.

Figure 10 reports the evaluative results, in which “None” denotes that tested videos are only encoded without any noise pollution. As one can see, the average  $F_1$ , and identification accuracy of the proposed method are above 87% and 75% for various situations. With the decrement of the SNR, the proposed method has a slight decline both of localization and identification accuracy. For those faked frame-rate videos with noise, holding SNRs with 33db and 36db, the  $F_1$  values, and identification accuracy enhance about 11.87%, 9.73% and 15.13%, 12.6% than those of ST-MSF, respectively. As a consequence, the proposed method precedes ST-MSF in terms of the robustness against noise. The reason behind this is summarized as: though adding noise leads to the change of local artifact regions, the OF methods are more resilient to noise, thereby changing TFD-OFLBP little. But, ST-MSF with the help of with the first-order Markov process are blindfold to a certain extent.

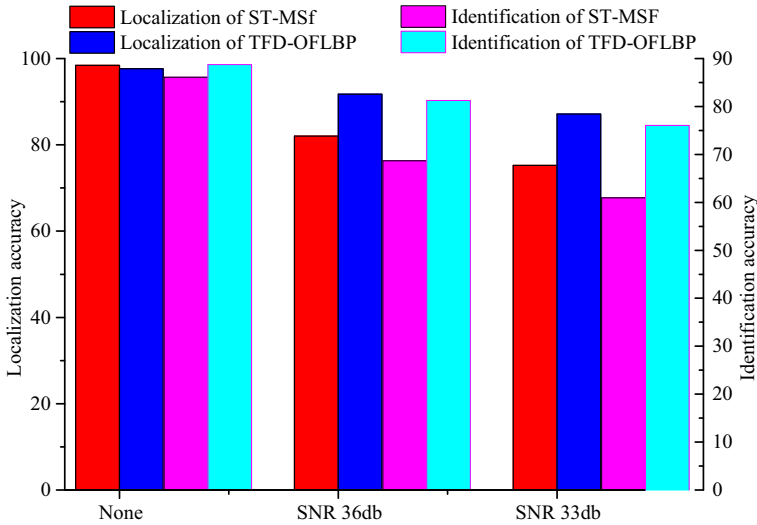
**DS5** is further introduced for experiments. Figure 11 reports the detection results. We can find that when the QP value is the same, the  $F_1$  values, and identification accuracy gradually decrease with decrement of SNR. Meanwhile, when the value of SNR is the same, the proposed method obtains similar trend for encoded videos, the reason of which have been analysed in above subsection.

**Table 8** Confusion matrix on “mixed” tampered Dataset of **DS3** under QP12/QP42 status. The asterisks “\*” denote that the corresponding values are below 1%

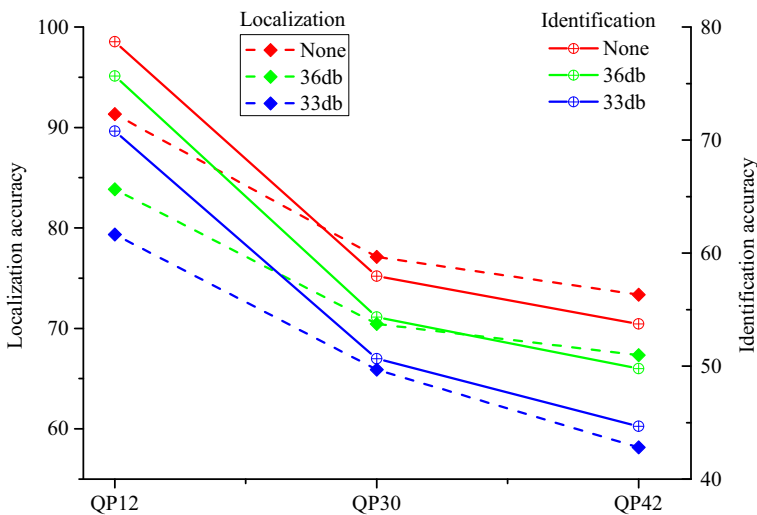
QP12/QP42 FRUCs	Classified as				
	MSU	MVTools	DSME	MCMP	Compressed
MSU	80.01/60.35	*/*	*/*	*/*	*/*
MVTools	*/*	82.65/64.79	*/*	*/*	*/*
DSME	*/*	*/*	78.87/64.56	15.35/22.57	2.02/5.7
MCMP	8.52/10.56	6.22/6.87	5.89/7.13	72/44.3	9.44/32.67
Compressed	7.56/28.43	7.78/23.67	10.33/27.33	12.11/31.7	86.71/66.61

**Table 9** Test results for unknown MCFI methods. The asterisks “\*” denote that the corresponding values are below 1%

unknown MCFI	Classified as				
	MSU	MVTools	DSME	MCMP	Pristine
YUVsoft	65.67	30.09	*	*	3.56
AOBMC [7]	*	*	18.28	79.56	0.86



**Fig. 10** The average localization and identification accuracy (%) on the **DS4**



**Fig. 11** The average localization and identification accuracy (%) on the **DS5**

## 5 Conclusion

In this paper, we have proposed a blind forensics approach to identify the adopted MCFIs by measuring irregularity of optical flow. Specifically, an effective compact feature, TFD-OFLBP, is designed to characterize the discrepancy of calculated OF for multi-class classification. Meanwhile, in order to remit the effect of possible interferences, a set of forensics tools, comprised by the features of LIEF, and the operations of ACE, MGJ, and IPC, are constructed. Experimental results have illustrated that for tampered videos in compressed format, or contaminated by noise, the proposed method can not only locate interpolated frames, but also identify the adopted MCFIs. In this work, we relied on the default parameters for the adopted OF method, therefore, further study of the effects of adjusting the parameters of OF methods remains an area of further exploration.

**Acknowledgments** This work was supported in part by Doctoral research foundation of Hunan University of Science and Technology under E51974, the Scientific Research Foundation of Hunan Provincial Education Department of China under 19B199, and the Natural Science Foundation of Hunan Province of China under Grant 2020JJ4029.

## References

1. Adato Y., Zickler T., Ben-Shahar O. (2011) A polar representation of motion and implications for optical flow. In: *Proceeding IEEE conference computer vision and pattern recognition (CVPR)*, pp 1145–1152
2. Bestagini P., Battaglia S., Milani S., Tagliasacchi M., Tubaro S. (2013) Detection of temporal interpolation in video sequences. In: *Proceeding International Conference Acoustics, Speech Signal Process. (ICASSP)*, May, pp 3033–3037
3. Bian S., Luo W., Huang J. (2014) Detecting video frame-rate up conversion based on periodic properties of inter-frame similarity. *Multimed. Tools Appl.* 72(1):437–451
4. Black M. J., Anandan P. (1993) A framework for the robust estimation of optical flow. In: *Proceeding fourth international conference computer vision (ICCV)*, pp 231–236
5. Chao J., Jiang X., Sun T. (2012) A novel video inter-frame forgery model detection scheme based on optical flow consistency. In: *International workshop on digital watermarking, (IWDW)*, October, pp 267–281
6. Chaudhry R., Ravichandran A., Hager G., Vidal R. (2009) Histograms of oriented optical flow and binet-cauchy kernels on nonlinear dynamical systems for the recognition of human actions. In: *Proceeding IEEE conference computer vision pattern recognition (CVPR)*, June, pp 1932–1939
7. Choi B. D., Han J. W., Kim C. S., Ko S. J. (2007) Motion-compensated frame interpolation using bilateral motion estimation and adaptive overlapped block motion compensation. *IEEE Trans. Circuits Syst. Video Technol.* 17(4):407–416
8. Dar Y., Bruckstein A. M. (2015) Motion-compensated coding and frame rate up-conversion: models and analysis. *IEEE Trans. Image Process.* 24(7):2051–2066
9. Dietterich T. G., Bakiri G. (1995) Solving multiclass learning problems via error correcting output codes. *J. Artif. Intell. Res.* 2:263–286
10. Ding X., Gaobo Y., Li R., Zhang L., Li Y., Sun X. (2018) Identification of motion-compensated frame rate up-conversion based on residual signal. *IEEE Trans. Circuits Syst. Video Technol.* 28(7):1497–1512
11. Ding X., Li Y., Xia M., He J., Gaobo Y. (2019) Detection of motion compensated frame interpolation via motion-aligned temporal difference. *Multimed. Tools Appl.* 78(6):7453–7477
12. Ding X., Zhu N., Li L., Li Y., Gaobo Y. (2019) Robust localization of interpolated frames by motion-compensated frame-interpolation based on artifact indicated map and tchebichef moments. *IEEE Trans. Circuits Syst. Video Technol.* 29(7):1893–1906
13. Feng C., Xu Z., Jia S., Zhang W., Xu Y. (2017) Motion-adaptive frame deletion detection for digital video forensics. *IEEE Trans. Circuits Syst. Video Technol.* 27(12):2543–2554
14. Fridrich J., Soukal D., Lukas J. (2003) Detection of copy-move forgery in digital images. In: *Proceeding digital forensics research workshop*
15. H.264/AVC software[Online]. Available: <http://iphome.hhi.de/suehring/tml/>
16. H.265/HEVC software[Online]. Available: : <https://hevc.hhi.fraunhofer.de/>

17. Hastie T., Tibshirani R. (1998) Classification by pairwise coupling. *Ann. Statist.* 26(2):451–C471
18. Hsu C. C., Hung T. Y., Lin C. W., Hsu C. T. (2008) Video forgery detection using correlation of noise residue. In: *Proceeding IEEE International Workshop Multimedia Signal Process (MMSP)*, Oct. pp 170–174
19. Jia S., Feng C., Xu Z., Xu Y., Wang T. (2014) ACE algorithm in the application of video forensics. In: *Proceeding international conference multimedia, communication computing application (MCCA)*, pp 177–184
20. 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
21. Jung D. J., Lee H. K. (2017) Frame-rate conversion detection based on periodicity of motion artifact. *Multimed. Tools Appl.*, pp 1–22
22. Kodovsky J., Fridrich J., Holub V. (2012) Ensemble classifiers for steganalysis of digital media. *IEEE Trans. Inf. Forensics Secur.* 7(2):432–444
23. Li R., Gan Z., Cui Z., Tang G., Zhu X. (2014) Multi-channel mixed-pattern based frame rate up-conversion using spatio-temporal motion vector refinement and dual weighted overlapped block motion compensation. *J. Disp. Technol.* 10(12):1010–1023
24. Li Q., Lin W., Fang Y. (2016) No-reference quality assessment for multiply-distorted images in gradient domain. *IEEE Signal Process. Lett.* 23(4):541–545
25. Li R., Liu Z., Zhang Y., Li Y., Fu Z. (2016) Noise-level estimation based detection of motion-compensated frame interpolation in video sequences. *Multimed. Tools Appl.*, pp 1–26
26. Li Y., mei L., Li R., Wu C. (2018) Using noise level to detect frame repetition forgery in video frame rate up-conversion. *Future Internet* 10(9):84(1–11)
27. Mahdian B., Saic S. (2008) Blind authentication using periodic properties of interpolation. *IEEE Trans. Inf. Forensics Secur.* 3(3):529–538
28. Milani S., Fontani M., Bestagini P., Barni M., Piva A., Tagliasacchi M., Tubaro S. (2012) An overview on video forensics. *APSIPA Trans. Signal Inf. Process.* 1:1–18
29. Ojala T., Pietikainen M., Maenpaa T. (2002) Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. Pattern Anal. Mach. Intell.* 24(7):971–987
30. Portz T., Zhang L., Jiang H. (2012) Optical flow in the presence of spatially-varying motion blur. In: *Proceeding IEEE conference on computer vision and pattern recognition (CVPR)*, June, pp 1752–1759
31. Rijsbergen C. J. V. (1979) *Information Retrieval*. Newton, MA, USA: Butterworth-Heinemann
32. Singh R. D., Aggarwal N. (2017) Optical flow and prediction residual based hybrid forensic system for inter-frame tampering detection. *J. Circuit Syst. Comp.* 26(7):1750107(1–37)
33. Stamm M. C., Lin W., Liu K. J. (2012) Temporal forensics and anti-forensics for motion compensated video. *IEEE Trans. Inf. Forensics Security.* 7(4):1315–1329
34. Stamm M. C., Wu M., Liu K. J. R. (2013) Information forensics: an overview of the first decade. *IEEE Access.* 1:167–200
35. 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
36. Sun D., Roth S., Black M. J. (2014) A quantitative analysis of current practices in optical flow estimation and the principles behind them. *Int. J. Comput. Vis.* 106(2):115–137
37. The online video databases[Online]. Available: <http://media.xiph.org/video/derf/#>
38. The MCFI softwares[Online]. Available: <http://www.wondershare.com/multimedia-tips/slow-motion-software.html>
39. Wang W., Farid H. (2007) Exposing digital forgeries in interlaced and deinterlaced video. *IEEE Trans. Inf. Forensics Secur.* 2(3):438–449
40. Wang W., Jiang X., Wang S., Wan M., Sun T. (2013) Identifying video forgery process using optical flow. In: *International workshop on digital watermarking (IWDW)*, October, pp 244–257
41. Wu Y., Jiang X., Sun T., Wang W. (2014) Exposing video inter-frame forgery based on velocity field consistency. In: *Proceeding international conference acoustics speech signal process. (ICASSP)*, May, pp 2674–2678
42. Xia M., Yang G., Li L., Li R., Sun X. (2017) Detecting video frame rate up-conversion based on frame-level analysis of average texture variation. *Multimed. Tools Appl.* 76(6):8399–8421
43. Yao H., Ni R., Zhao Y. (2019) An approach to detect video frame deletion under anti-forensics. *J. Real-Time Image Proc.* 1–14
44. Yao Y., Yang G., Sun X., Li L. (2016) Detecting video frame-rate up-conversion based on periodic properties of edge-intensity. *J. Inf. Secur. Appl.* 26:39–50
45. Yoo D. G., Kang S. J., Kim Y. H. (2013) Direction-select motion estimation for motion-compensated frame rate up-conversion. *J. Display Technol.* 9(10):840–850