Identification of Interpolated Frames by Motion-Compensated Frame-Interpolation via Measuring Irregularity of Optical Flow

Xiangling Ding, Hunan University of Science and Technology, China

Yanming Huang, Hunan University of Science and Technology, China

Dengyong Zhang, Changsha University of Science and Technology, China

Junlin Ouyang, Hunan University of Science and Technology, China

ABSTRACT

Motion-compensated frame-interpolation (MCFI) synthesizes intermediate frames between input frames guided by estimated motion and can be employed to falsify high bit-rate videos or high frame-rate videos with different frame rates. Although existing MCFI identification methods have obtained satisfactory results, they are seriously degraded by stronger compression. Therefore, to conquer this issue, a blind forensics method is proposed to identify the adopted MCFI methods by considering the irregularities of optical flow produced by various MCFIs. In this paper, a set of compact features are constructed from the motion-aligned frame difference-weighted histogram of local binary pattern on the basis of optical flow (MAFD-WHLBP). Experimental results show that the proposed approach outperforms existing MCFI detectors under stronger compression.

KEYWORDS

Identification, Motion-Compensation Frame-Interpolation, Optical Flow, Video Forensics

INTRODUCTION

Using the powerful video editing software, video manipulation can be performed easily and the detection of tampered videos is difficult through human vision. To address the harmful impacts caused by video forgeries, video forensics has attracted wide attentions (Rocha, Scheirer, Boult, & Goldenstein, 2011; Milani, Fontani, Bestagini, Barni, Piva, Tagliasacchi, & Tubaro, 2012). Some inherent traces left by video editing operations can help the detection of video falsification such as the differential energy of residue (Hsu, Hung, Lin, & Hsu, 2008), the motion residual (Feng, Xu, Jia, Zhang, & Xu, 2016), and double compression artifacts (Jiang, Wang,Sun, Shi, & Wang, 2013).

Motion-Compensation Frame-Interpolation (MCFI) is another special frame based video manipulation, which periodically synthesizes intermediate frames to alleviate the motion discontinuity of low frame-rate videos (Yoo, Kang, & Kim, 2013; Li, Gan, Cui, Tang, & Zhu, 2014). Though MCFI is originally proposed to improve the visual quality or increase the frame-rate of low frame

rate videos, it still might be used 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 near-duplicate video detection or the video watermarking system because of the loss of temporal synchronism.

Some studies have been proposed to detect the use of MCFI. Estimating original frame-rate of faked videos was firstly proposed by using video-level artifacts such as prediction error (Bestagini, Battaglia, Milani, Tagliasacchi, & Tubaro, 2013), motion artifact (Jung, & Lee, 2018) and noise variation (Li, Liu, Zhang, Li, & Fu, 2018). They can obtain desirable results, yet cannot locate interpolation frames, let alone identify the adopted MCFI. Then, Yao et al. (Yao, Yang, Sun, & Li, 2016)and Xia et al. (Xia, Yang, Li, Li,& Sun, 2017) located the interpolated frames by the periodicity of edge-intensity and average texture variation, respectively. Subsequently, the localization problem of interpolated frames is discussed under real-world scenarios by employing Tchebichef moments (Ding, Zhu, Li, Li, & Yang, 2018). Recently, Yao et al. (Yao, Ni, & Zhao, 2019), as a pioneer, makes use of the MCFI strategy to invalidate inter-frame continuity based video forensics detection, and then 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 an environment of unknown MCFI techniques (Ding, Li, Xia, He, & Yang, 2019). As far as we know, there is only one work (Ding, Yang, Li, Zhang, Li, & Sun, 2017) about the identification issue of the adopted MCFIs, in which residual signal was firstly exploited as the tampering clue for identification of the adopted MCFIs. But, its performance deteriorates for videos encoded with relatively bigger quantization parameters.

Because motion regions of interpolated frames usually exist local slight artifacts, such as blurring or deformed object, the irregularity of optical flow (OF) occur in these local regions. Meanwhile, different local artifacts by various MCFIs produce different degree irregularity of OF. Based on this premise, we propose an effective feature extraction method to characterize the discrepancy among calculated OF from various MCFIs. Firstly, the OF field is calculated; Then, the local binary pattern (LBP) (Ojala, Pietikäinen, & Mäenpää, 2002; Li, Lin, & Fang, 2016) is employed to encode the subtle variance of the calculated OF field; Finally, the motion-aligned frame difference among three successive frames of current frame is calculated, and act as the weighted histogram of LBP. Thus, the proposed features can not only consider the local variation of OF field, but also fuse the change of pixel values along temporal direction. The simulation results demonstrate that our algorithm can achieve satisfactory identification of adopted MCFI for highly compressed videos.

Local Artifact Analysis Of Mcfi

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. The visual quality of MCFIs highly depend on the accuracy of estimated MVs, or adopted MCI. First, we define four neighboring blocks of interpolated frame by V_1, V_2, V_3, and V_4, shown in Figure 1. Subsequently, (v_{1x}, v_{1y}), (v_{2x}, v_{2y}), (v_{3x}, v_{3y}), and (v_{4x}, v_{4y}) are their corresponding motion vectors. As a consequence, the pixels of interpolated block in region A, B or C, which respectively overlaps with four blocks (V_1, V_2, V_3, and V_4), two blocks (V_3, and V_4) or one block V_4, are interpolated by

$$f_t(x,y) = \frac{1}{2} \sum_{i=1}^{\alpha} \theta_i(x,y) \sum_{j=5-i}^{4} \left[f_{t-1}(x + v_{jx}, y + v_{jy}) + f_{t+1}(x - v_{jx}, y - v_{jy}) \right] \tag{1}$$

Where f_t, f_{t-1}, and f_{t+1} are the interpolated frame, the previous frame, and the current frame, respectively, $\alpha \in \{1, 2, 4\}$, $\sum_{i=1}^{\pm} \theta_i (\mathbf{x}, \mathbf{y}) = 1$, and $\theta_i (\mathbf{x}, \mathbf{y})$ is weighting coefficient that determined by pixel relative position of *(x,y)* within interpolated block.

From Figure 1 and Equation (1), we can infer that even if different MCFI techniques adopt the same MCI scheme, their interpolation frames are also different because they might have dissimilar motion searching patterns. Meanwhile, MCFIs generally assume that the estimated MV represent true motion of object (Yoo, Kang, & Kim,2013; Li, Gan, Cui, Tang, & Zhu, 2014). 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 alleviate theirs 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.

Figure 1. The architecture (Ding, Yang, Li, Zhang, Li, & Sun, 2017) of MCFI for a synthesized block. (Green, Red, Brown, and Blue denote reference blocks; Dashed box is interpolated block)
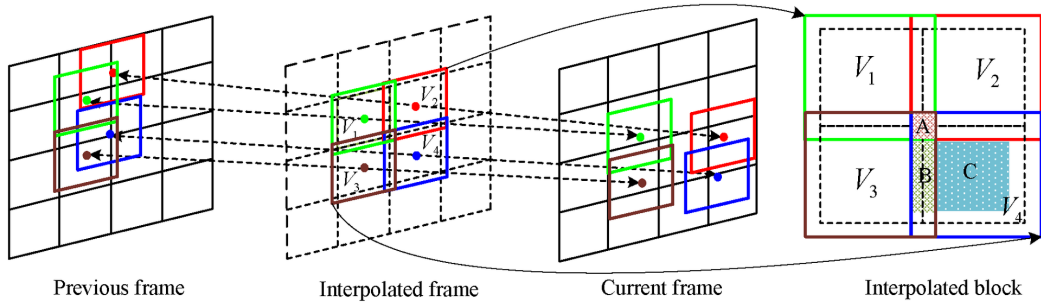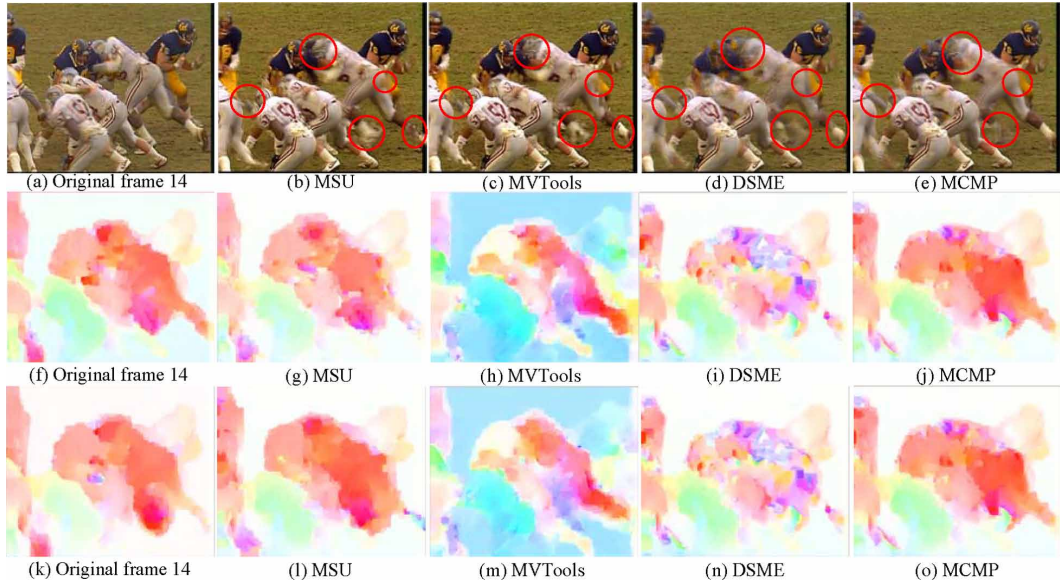

Due to different MCFIs with disparate ME or distinct MCI, synthesized blocks have different reference blocks or weighting coefficients. As a results, some differences are really existed among these synthesized blocks generated by various MCFIs, the same to their local artifacts. We can observe from the original frame and its corresponding synthesized frames, shown in Figure 2(a)- 2(e), that the local differences embodied by local artifacts are demonstrated as expected, such as red circle.

For forensics purpose, these inherent correlation of local artifacts region scan be spontaneously resorted in pixel domain. But, it is vulnerable for video compression (Ding, Yang, Li, Zhang, Li, & Sun, 2017). As a consequence, the OFs features can be employed to recognize the synthesized frames. Four reasons are illustrated as follows.

- Because OFs are more resilient to video compression (Jung,& Lee,2018), the features extracted from OFs may occur little changes.
- When one or both of the frames used in the OF computation contain some difference regions with local artifacts, the data cost and smoothness cost in OF algorithms will be violated, then to minimize the error of these costs, the OF algorithms will attempt to warp the current frame to match the appearance of the reference frame (Black & Anandan, 1993; Sun, Roth, & Black, 2010; Portz, Zhang, & Jiang, 2012). This process will produce irregular OFs in these regions with local artifacts.
- There exist the periodic variation of OFs for faked high frame-rate videos. The detailed derivation is given in Appendix 1.

• For different intensities of regions with local artifacts, their irregular degree of OFs in these regions may be also different.

Figure 2. Original frames of ``Football" sequence and its synthesized frames by MSU, MVTools, DSME (Yoo, Kang, & Kim, 2013) and MCMP (Li, Gan, Cui, Tang, & Zhu, 2014). From the first row to the last row, they correspond to video frames, OF maps from uncompressed video frames, and compressed video frames, respectively.

(a) Original frame 14 (b) MSU (c) MVTools (d) DSME (e) MCMP

(f) Original frame 14 (g) MSU (h) MVTools (i) DSME (j) MCMP

(k) Original frame 14 (l) MSU (m) MVTools (n) DSME (o) MCMP

From Figure 2(f)- 2(j), the shapes and colors of OFs are different. This implies local artifacts introduced by MCFIs affect the regularity of OFs. As a results, the shapes and colors of OFs are also different from each other. We know that video compression may be executed after MCFI forgery, attacking the MCFI detector. However, we can observe from Figure 2(k)- 2(o) that the diversities among the OF from original frame and that from synthesized frames under the video compression are still apparent, which motivates us to exploit OFs to design features for the identification of the adopted MCFIs.

THE PROPOSED SCHEMA

MOFLBP Construction

In order to capture the irregularity of OF caused by the local artifacts of synthesized frames, a popular OF algorithm(Black,& Anandan,1993) is utilized here to extract the subtle changes. First, the Magnitude of OF (MOF) are calculated by

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

Second, the LBP is computed on each pixel of MOF. The LBP descriptor (Ojala, Pietikäinen, & Mäenpää, 2002; Li, Lin, & Fang, 2016) is employed to reflect the correlation between the OF value

of center pixel and that of its adjacent pixel. Due to its effective representation capability of micro-structure, we can infer that that the subtle difference among MOFs introduced by various MCFI technologies can be effectively represented by LBP. The MOFLBP (LBP of the MOF) at each pixel position is defined as

$$MOFLBP_N = \sum_{i=0}^{N-1} \delta(MOF_i - MOF_c) \cdot 2^i$$

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

(3)

where N is the number of adjacent position determined by the radius of the neighborhood, i.e. R (Please refer to (Ojala, Pietikäinen, & Mäenpää, 2002) for details), MOF_c, and MOF_i are the MOF at center pixel position and its neighbor.

FIGURE 3. Extracted Blocks (First Row), MOFS (Second Row). and MOFLBPS (Third Row) for Original Frame and its Corresponding Synthesized Frames.
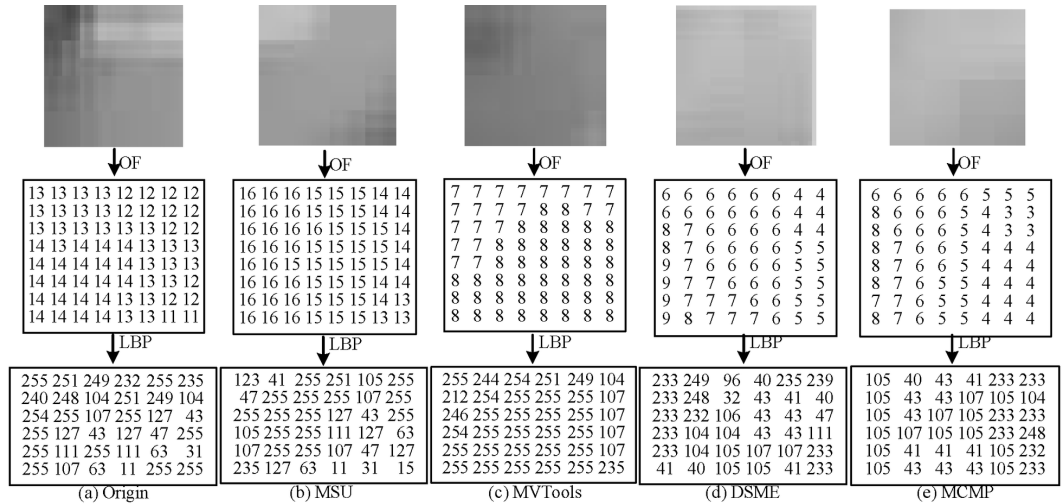
The inter-MOF relationship in an MOF adjacent pixel position can be represented by MOFLBP, and then the subtle difference among MOFs caused by different local artifact from various MCFIs can be effectively captured by such micro-structural patterns of LBP. Here, an experiment is conducted to further illustrate this issue. The effects are demonstrated in Figure 3, in which five 8×8 blocks are extracted from the head area of the second player from the right in Figure 2(a)- 2(e), and theirs corresponding MOFs, and MOFLBPs are simultaneously computed. From it, we can see from the first row that although various MCFIs change the pixel values to different degree, the discrepancy among their MOFsdo exist but they are not obvious. Furthermore, different irregularity of OF change the MOFLBP patterns in their own characteristic ways, making it an effective measure to discriminate the adopted MCFIs.

Motion-Aligned Frame-Difference-Weighted MOFLBP Histogram

The change of pixel value by various MCFIs is beneficial to identify the adopted MCFIs, which is verified in method (Ding, Yang, Li, Zhang, Li, & Sun, 2017). However, MOFLBP only reflect the irregularity of OF. To effectively capture the difference from pixel domain and OF domain simultaneously, we proposed to integrate these discrepancy of local artifact into a single representation way. Here, the values of motion-aligned frame-difference with same MOFLBP pattern are accumulated, which can be referred as motion-aligned frame-difference-weighted MOFLBP histogram, and the calculated process is defined as.

$$MAFD_t(x,y) = \frac{1}{2}\{[f_t(x,y) - f_{t-1}(x + OF_x, y + OF_y)] + [f_t(x,y) - f_{t+1}(x - OF_x, y - OF_y)]\}$$

(4)

$$h_{moflbp}(\rho) = \sum_{i=1}^{M}\sum_{j=1}^{N} MAFD_t(i,j) \cdot \delta(MOFLBP_N(x,y),\rho)$$
$$\delta(\alpha,\beta) = \begin{cases} 1, & \alpha = \beta \\ 0, & otherwise \end{cases}$$

(5)

Where m and n denotes the width and the height of a video frame, $\acute{A} \in [0, K]$ is the possible moflbp patterns, $mafd_t$ is the weight assigned to the moflbp pattern, A and B are the moflbp pattern, of_x and of_y are the horizontal and vertical values of optical flow in the position of *(x,y)* among frame f_t, f_{t-1}, and f_{t+1}, respectively. In this paper, the simple motion-aligned frame-difference calculated by equation (4) to reflect the discrepancy among motion regions as the moflbp weight of each pixel position. Specially, the motion information between three consecutive frames is considered to reduce the interference of redundant information via calculated ofs, i.e. (of_x, of_y). In this way we highlight the irregularity of of in motion aspect, and also take into account the change from pixel domain in temporal direction. During experiments, for moflbp, the *n* and *r* are respectively set as 8 and 1. Thus, the extracted features are 256 dimensions.

EXPERIMENTAL RESULTS AND DISCUSSION

Experimental Settings

Twenty known YUV videos in CIF format (352× 288) with 15fps are selected (OnlineVD). Two popular softwares (MSU and MVTools) and two progressive MCFIs (DSME (Yoo, Kang, & Kim, 2013) and MCMP (Li, Gan, Cui, Tang, & Zhu, 2014)) are used. Four target frame-rates (20, 30, 60, and 120) *fps* are involved. Subsequently, we use the popular video compression standards H.264/AVC to provide video compression for the experiments with configurations: 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 (Black, & Anandan,1993).

Synthesized frames and original frames are denoted as positive samples (S_p) and negative samples (S_n), respectively. The *F1* (Ding, Yang, Li, Zhang, Li, & Sun, 2017; Rijsbergen, 1979) is adopted for performance evaluation.

$$F_1 = \begin{cases} \dfrac{(\gamma^2 + 1) \cdot precision \times recall}{\gamma^2 \cdot precision + recall} & if \sum S_{tp} > 0 \\ 0 & if \sum S_{tp} = 0 \end{cases} \quad (6)$$

$$where \quad precision = \frac{\sum S_{tp}}{\sum S_{tp} + \sum S_{fp}}, recall = \frac{\sum S_{tp}}{\sum S_p}$$

And γ controls the balance between *precision* and *recall*. Normally, γ is set to 1. S_{tp} and s_{fp} are true positive and false positive samples, respectively.

In our experiments, the error-correcting output code(Dietterich & Bakiri, 1994) based on ensemble classifier (Kodovsky, Fridrich, & Holub, 2011) with its default settings is employed to identify the adopted MCFIs (Ding, Yang, Li, Zhang, Li, & Sun, 2017). 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.

Localization Results Under Lossy Compression

Table 1. 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	-	92.14	94.08	97.54	-	81.27	84.67	84.72	-	76.94	78.02	79.95
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	91.09	93.72	95.67	96.44	76.17	78.25	84.63	86.37	74.36	78.82	80.40	84.35
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	89.36	93.08	95.84	98.67	72.13	75.62	81.67	85.95	65.35	70.95	77.07	84.95
MCMP	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	89.34	91.67	95.52	97.17	69.41	73.08	76.17	84.35	62.67	65.08	73.87	82.27
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	89.93	92.65	95.27	97.45	72.57	77.05	81.78	85.34	67.46	72.94	77.34	82.98

We employed Yao et al. (Yao, Yang, Sun, & Li, 2016), and ST-MSF (Ding, Yang, Li, Zhang, Li, & Sun, 2018) for comparison. Table 1 gives the localization accuracies under different QP values. Conclusions can be drawn as follow:

First, with the increase of frame-rate, the localization accuracies increase steadily for the proposed method 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 *120fps*, the periodicity of edge intensity exists aliasing, causing performance degradation.

Second, the ST-MSF and the proposed method are better than the method Yao for considering the scene change and the periodicity of frame interpolation.

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.

Identification Results Under Lossy Compression

The experiment is a 5-class (including the original videos without MCFI as a special class) classification problem due to four known MCFIs involved.

Table 2 gives the average identification accuracies acquired by using ST-MSF, and proposed MAFD-MOFLBP, in which ``mix'' means the mixture of tampered datasets with the whole target frame-rate. The proposed method is superior to the ST-MSF for better capacity to classify the original frames and interpolated ones by effectively fusing the changes in the pixel domain and MOF domain.

Table 2. The average identification accuracies (%) along the diagonal direction in the corresponding confusion matrices.

DETECTOR	Uncompressed					QP12					QP30					QP42				
	20	30	60	120	mix	20	30	60	120	mix	20	30	60	120	mix	20	30	60	120	mix
ST-MSF	86.76	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	87.80	88.35	89.73	92.05	88.54	69.85	74.87	80.33	86.98	84.12	50.54	51.86	55.07	68.23	66.67	47.33	48.05	53.66	62.33	60.53

Table 3 reports the average accuracies of ``mix'' tampered videos under QP12 or QP42. From it, we observe that the MAFD-MOFLBP 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 frames because from ME and MC points of view, H.264 compression has a similar character with MCFIs. Besides, because original frames with acute object motion also occur some inherent local artifacts, they are incorrect classified as synthesized frames, causing the identification accuracy degrade. Although the MAFD-MOFLBP can identify the adopted MCFIs, it cannot classify the unknown MCFIs as a new type. This is a limitation for the proposed approach.

Table 3. Confusion matrix on ``mix'' tampered Dataset under QP12/QP42 status. The asterisks ``*'' denote that the corresponding values are below 1%

QP12/QP42	Classified as				
MCFIs	MSU	MVTools	DSME	MCMP	Compressed
MSU	85.30/60.83	*/*	*/*	*/*	*/*
MVTools	*/*	89.75/71.67	*/*	*/*	*/*
DSME	*/*	*/*	82.84/62.44	15.67/24.08	1.45/6.67
MCMP	9.63/11.21	5.65/6.07	6.56/7.07	71.54/42.33	8.08/30.17
Compressed	4.22/26.67	4.54/21.59	9.05/29.53	10.67/32.87	89.48/62.33

Discussions

From the experimental results, we can conclude that the proposed MAFD-MOFLBP 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 MAFD-MOFLBP which combines the change from pixel domain along temporal direction and OF domain can provide better performance. Next, we conduct an experiment to illustrate these two aspects.

Table 4. Average F_1, and identification accuracy(IA).(%)

Detectors	QP12		QP30		QP42	
	F_1	IA	F_1	IA	F_1	IA
HOOF	62.43	37.16	62.45	35.59	58.55	31.26
MOFLBP	86.47	65.50	73.99	50.87	69.20	45.36
MAFD-MOFLBP	93.82	79.23	79.18	58.47	75.18	54.38

We employ the Histogram of Oriented Optical Flow (HOOF) (Chaudhry, Ravichandran, Hager, & Vidal, 2009) to directly extract OF features, in which we set 256 bins for better performance. Table 4 reports the average F_1, and identification accuracy for tampered videos with the whole target frame-rate under different QP values. From it, we observed that LBP, and motion-aligned frame difference-weighted raise the accuracy.

CONCLUSION

In this paper, a blind forensics approach have been proposed to identify the adopted MCFIs by measuring irregularity of optical flow. To capture the irregularity of OFs, an effective compact feature, MAFD-MOFLBP, is designed to reflect the discrepancy of calculated OF for multi-class classification. Experimental results have represented that for tampered videos in compressed format, the proposed strategy can not only locate synthesized frames, but also identify the used MCFIs. In the future, we will fuse convolutional neural network into our detection framework to obtain better identification performance for its excellent ability of feature representation.

ACKNOWLEDGMENT

REFERENCES

Bestagini, P., Battaglia, S., Milani, S., Tagliasacchi, M., & Tubaro, S. (2013, May). Detection of temporal interpolation in video sequences. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing* (pp. 3033-3037). IEEE.

Black, M. J., & Anandan, P. (1993, May). A framework for the robust estimation of optical flow. In *1993 (4th) International Conference on Computer Vision* (pp. 231-236). IEEE.

Chaudhry, R., Ravichandran, A., Hager, G., & Vidal, R. (2009, June). Histograms of oriented optical flow and binet-cauchy kernels on nonlinear dynamical systems for the recognition of human actions. In *2009 IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1932-1939). IEEE.

Dar, Y., & Bruckstein, A. M. (2015). Motion-compensated coding and frame rate up-conversion: Models and analysis. *IEEE Transactions on Image Processing*, *24*(7), 2051–2066.

Dietterich, T. G., & Bakiri, G. (1994). Solving multiclass learning problems via error-correcting output codes. *Journal of Artificial Intelligence Research*, *2*, 263–286.

Ding, X., Li, Y., Xia, M., He, J., & Yang, G. (2019). Detection of motion compensated frame interpolation via motion-aligned temporal difference. *Multimedia Tools and Applications*, *78*(6), 7453–7477.

Ding, X., Yang, G., Li, R., Zhang, L., Li, Y., & Sun, X. (2017). Identification of motion-compensated frame rate up-conversion based on residual signals. *IEEE Transactions on Circuits and Systems for Video Technology*, *28*(7), 1497–1512.

Ding, X., Zhu, N., Li, L., Li, Y., & Yang, G. (2019). Robust localization of interpolated frames by motion-compensated frame-interpolation based on artifact indicated map and tchebichef moments. *IEEE Transactions on Circuits and Systems for Video Technology*.

Feng, C., Xu, Z., Jia, S., Zhang, W., & Xu, Y. (2016). Motion-adaptive frame deletion detection for digital video forensics. *IEEE Transactions on Circuits and Systems for Video Technology*, *27*(12), 2543–2554.

Hsu, C. C., Hung, T. Y., Lin, C. W., & Hsu, C. T. (2008, October). Video forgery detection using correlation of noise residue. In *2008 IEEE 10th workshop on multimedia signal processing* (pp. 170-174). IEEE.

Jiang, X., Wang, W., Sun, T., Shi, Y. Q., & Wang, S. (2013). Detection of double compression in MPEG-4 videos based on Markov statistics. *IEEE Signal Processing Letters*, *20*(5), 447–450.

Jung, D. J., & Lee, H. K. (2018). Frame-rate conversion detection based on periodicity of motion artifact. *Multimedia Tools and Applications*, *77*(5), 6095–6116.

Kodovsky, J., Fridrich, J., & Holub, V. (2011). Ensemble classifiers for steganalysis of digital media. *IEEE Transactions on Information Forensics and Security*, *7*(2), 432–444.

Li, Q., Lin, W., & Fang, Y. (2016). No-reference quality assessment for multiply-distorted images in gradient domain. *IEEE Signal Processing Letters*, *23*(4), 541–545.

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. *Journal of Display Technology*, *10*(12), 1010–1023.

Li, R., Liu, Z., Zhang, Y., Li, Y., & Fu, Z. (2018). Noise-level estimation based detection of motion-compensated frame interpolation in video sequences. *Multimedia Tools and Applications*, *77*(1), 663–688.

Mahdian, B., & Saic, S. (2008). Blind authentication using periodic properties of interpolation. *IEEE Transactions on Information Forensics and Security*, *3*(3), 529–538.

Milani, S., Fontani, M., Bestagini, P., Barni, M., Piva, A., Tagliasacchi, M., & Tubaro, S. (2012). An overview on video forensics. *APSIPA Transactions on Signal and Information Processing*, ●●●, 1.

Ojala, T., Pietikäinen, M., & Mäenpää, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (7), 971–987.

Online, V. D. (n.d.). Retrieved from http://media.xiph.org/video/derf/\#

Portz, T., Zhang, L., & Jiang, H. (2012, June). Optical flow in the presence of spatially-varying motion blur. In *2012 IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1752-1759). IEEE.

Rijsbergen, C. J. V. (1979). *Information Retrieval*. Butterworth-Heinemann.

Rocha, A., Scheirer, W., Boult, T., & Goldenstein, S. (2011). Vision of the unseen: Current trends and challenges in digital image and video forensics. *ACM Computing Surveys*, *43*(4), 26. doi:10.1145/1978802.1978805

Sun, D., Roth, S., & Black, M. J. (2010, June). Secrets of optical flow estimation and their principles. In *2010 IEEE computer society conference on computer vision and pattern recognition* (pp. 2432-2439). IEEE.

Xia, M., Yang, G., Li, L., Li, R., & Sun, X. (2018). Detecting video frame rate up-conversion based on frame-level analysis of average texture variation. *Multimedia Tools and Applications*, *76*(6), 8399–8421.

Yao, H., Ni, R., & Zhao, Y. (2019). An approach to detect video frame deletion under anti-forensics. *Journal of Real-Time Image Processing*, *16*(3), 751–764.

Yao, Y., Yang, G., Sun, X., & Li, L. (2016). Detecting video frame-rate up-conversion based on periodic properties of edge-intensity. *Journal of Information Security and Applications*, *26*, 39–50.

Yoo, D. G., Kang, S. J., & Kim, Y. H. (2013). Direction-select motion estimation for motion-compensated frame rate up-conversion. *Journal of Display Technology*, *9*(10), 840–850.

APPENDIX A. DERIVATION OF THE PERIODIC VARIATION OF OFS

We can model the MCFI operated process (formula (1))as below:

$$P_{mcfi}(t) = \sum_{\chi=t-1}^{t+1} P_{origin}(\chi)\varpi(\frac{t}{\Delta} - \chi) \tag{7}$$

where P_{mcfi}, P_{origin}, ϖ and Δ denote the position of the synthesized block in synthesized intermediate frame, the position of reference block in current frames, the weighting coefficient, and the synthesizedperiod.
When videos follow stationary signal, the periodic variation of position can be obtained after nth derivation (Jung,& Lee,2018; Dar,& Bruckstein,2015; Mahdian,& Saic,2008):

$$D^n[P_{mcfi}(t)] = \frac{\partial P_{mcfi}(t)}{\partial t^n}, for\ n > 0 \tag{8}$$

Substituting formula (7) into formula (8), we obtain:

$$D^n[P_{mcfi}(t)] = \sum_{t-1}^{t+1} P_{origin}(\chi)D^n\varpi(\frac{t}{\Delta} - \chi) \tag{9}$$

Suppose the original video follows a variance $\tilde{A}^2$ (Jung,& Lee,2018; Dar,& Bruckstein,2015; Mahdian,& Saic,2008), it can be further expressed as

$$\mathrm{var}\{D^n[P_{mcfi}(t)]\} = \sigma^2 \cdot \sum_{t-1}^{t+1} D^n\varpi(\frac{t}{\Delta} - \chi)^2 \tag{10}$$

Further, for $\epsilon \in \mathrm{Z}$, this equation can be derived as

$$\mathrm{var}\{D^n[P_{mcfi}(t + \varepsilon \cdot \Delta)]\}$$
$$= \sigma^2 \cdot \sum_{t-1+\varepsilon}^{t+1+\varepsilon} D^n\varpi(\frac{t + \varepsilon \cdot \Delta}{\Delta} - \chi)^2 = \sigma^2 \cdot \sum_{t-1+\varepsilon}^{t+1+\varepsilon} D^n\varpi(\frac{t}{\Delta} - (\chi - \varepsilon))^2 \tag{11}$$
$$= \sigma^2 \cdot \sum_{t-1+\varepsilon-\varepsilon}^{t+1+\varepsilon-\varepsilon} D^n\varpi(\frac{t}{\Delta} - \chi)^2 = \sigma^2 \cdot \sum_{t-1}^{t+1} D^n\varpi(\frac{t}{\Delta} - \chi)^2$$

Hence, the variance of position is periodic with Δ

$$\mathrm{var}\{D^n[P_{mcfi}(t)]\} = \mathrm{var}\{D^n[P_{mcfi}(t + \varepsilon \cdot \Delta)]\} \tag{12}$$

This imply that the position of synthesized block occur period change. Here, the magnitude values of OF can beutilized to reflect the change of position. Moreover, due to different ˛ for distinct MCFI, the values of OF will be different for various MCFI methods.

Xiangling Ding is currently an associate professor with Hunan University of Science and Technology, xiangtan, China. He received the Ph.D. degree from Hunan University in June 2018. His current research interests include optical image encryption, authentication, digital media forensics, and image/video processing.

Yanming Huang is currently a librarian with Hunan University of Science and Technology, xiangtan, China. He received the Bachelor degree from Huaihua University in June 2016. His current research interests include optical image encryption, and image/video processing.

Dengyong Zhang is an associate professor in the School of Computer and Communication Engineering, Changsha University of Science and Technology, Changsha, China. He received the B.S. degree and M.S. degree from Changsha University of Science and Technology, Changsha, China, in 2003 and 2006 respectively, and the Ph.D. degree from Hunan University, China in 2018. His research interests include digital media forensics and video passive forensics.

Junlin Ouyang received the B.S. degree in computer application from the central south University, China, in 2007, and a Ph.D. degree in computer science from the Southeast University, China, in 2015. He is now a Lecturer in the School of Computer Science and Engineering, Hunan University of Science and Technology, China. His research interests include color image processing, information security and image retrieve, deep learning, image forensics, etc.