International Journal of Digital Crime and Forensics
Volume 12 • Issue 1 • January-March 2020

A Novel Video Forgery Detection Model Based on Triangular Polarity Feature Classification

Chee Cheun Huang, Institute for Infocomm Research, A*STAR, Singapore, Singapore
Chien Eao Lee, Institute for Infocomm Research, A*STAR, Singapore, Singapore
Vrizlynn L. L. Thing, Institute for Infocomm Research, A*STAR, Singapore, Singapore

ABSTRACT

Video forgery has been increasing over the years due to the wide accessibility of sophisticated video editing software. A highly accurate and automated video forgery detection system will therefore be vitally important in ensuring the authenticity of forensic video evidences. This article proposes a novel Triangular Polarity Feature Classification (TPFC) video forgery detection framework for video frame insertion and deletion forgeries. The TPFC framework has high precision and recall rates with a simple and threshold-less algorithm designed for real-world applications. System robustness evaluations based on cross validation and different database recording conditions were also performed and validated. Evaluation on the performance of the TPFC framework demonstrated the efficacy of the proposed framework by achieving a recall rate of up to 98.26% and precision rate of up to 95.76%, as well as high localization accuracy on detected forged videos. The TPFC framework is further demonstrated to be capable of outperforming other modern video forgery detection techniques available today.

KEYWORDS

Frame Deletion, *Frame Insertion*, *Inter-frame Forgery Detection*, *Precision Rate*, *Recall Rate*, *Video Forensic*

INTRODUCTION

Technology advancements in computing and video processing technologies in recent years have enabled the emergence of newer and more sophisticated video editing software tools. With the fact that most of these video editing software tools can be easily accessible online at practically no cost, it is not surprising that criminal acts associated with video forgery such as on surveillance videos are increasingly becoming more prevalent nowadays.

Criminal may be using video forgery as a way to get acquitted on the basis that the video evidences presented in court could not prove that they have performed the crime at a particular time or place. In criminal court cases particularly related to sensitive or high-profile cases, it is likely that a video that is modified or edited even slightly will be deemed as unacceptable to be used as evidence in the court. It is therefore imperative that a highly accurate forgery detection system is developed to ensure the authenticity of the videos used as evidences in court.

Detection of video forgery can be classified into intra-frame or inter-frame forgery detection. In intra-frame forgery detection, the aim is to locate the forged portions or regions within the image associated with a particular video frame whereas in inter-frame forgery detection, the aim is to locate forged frames within the full video sequence (Milani et al., 2012; Kingra, Aggarwal, & Singh, 2016).

DOI: 10.4018/IJDCF.2020010102
This article, originally published under IGI Global's copyright on January 1, 2020 will proceed with publication as an Open Access article starting on January 27, 2021 in the gold Open Access journal, International Journal of Digital Crime and Forensics (converted to gold Open Access January 1, 2021), and will be distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

Criminal video forgery acts are usually associated with inter-frame forgeries especially on frame insertion and deletion forgeries. This is due in part to the relative ease of performing inter-frame forgery with any basic video editing software as compared to intra-frame forgery.

This article proposes a novel Triangular Polarity Feature Classification (TPFC) video forgery detection framework for video frame insertion and deletion forgeries. The proposed algorithm has high precision and recall rates, and is considerably less complex in the system architecture design compared to other more common forgery detection systems based on optical flow consistency as detailed in Chao, Jiang, and Sun (2013) or systems based on evaluation of coding standards that usually involve more complex optimization algorithms such as those detailed in Wang and Farid (2007) and Aghamaleki and Behrad (2016). The lower complexity associated with the proposed algorithm will naturally lead to the advantage of faster processing time in authenticating video evidences in real-world court case scenarios where it is common that a large number of videos may need to be authenticated under a limited time constraint. Other important criteria such as robustness, localization capability and threshold-less system design were also considered in the proposed framework. The good performance of the system together with practical considerations of having a robust, threshold-less and computational efficient algorithm design makes the proposed approach a welcoming addition to the arsenal of algorithms available today for real-world inter-frame forgery detection.

RELATED WORK

In the following subsections, a review of the existing techniques for inter-frame video forgery detection will be presented. These techniques are broadly aggregated into three main categories; (1) camera-based detection techniques, (2) coding-based detection techniques and (3) content inconsistencies detection techniques (Milani et al., 2012; Kingra et al., 2016). Limitations and challenges associated with these techniques will be discussed in the last subsection.

Camera-based Detection Techniques

Camcorders will usually leave a trace or footprint in the recorded videos that could be used for video forgery detection. In particular, Kurosawa, Kuroki, and Saitoh (1999) demonstrated Charge Coupled Device (CCD) Fingerprint method for camera identification based on the usage of fixed pattern noise generated from dark currents on CCD chips. The video sequences however must be recorded in dark places for the method to work. Hsu, Hung, Lin, and Hsu (2008) performed correlation analysis on temporal noise residue based on Gaussian Mixture Model (GMM) technique. The approach however was designed to detect temporal copy-paste inpainting and may not be applicable for inter-frame forgery detection. Kobayashi, Okabe, and Sato (2009, 2010) evaluated video authenticity by detecting inconsistencies in Noise Level Functions (NLFs). Limitations are that they were developed to detect forgery in static scene, and any alteration of brightness of forged region in the video may affect fitting of NLF.

Coding-based Detection Techniques

Camera coding standard can introduce self-generated artifacts that can be used to aid in the video forgery detection (Milani et al., 2012; Kingra et al., 2016). Particularly, Wang and Farid (2006) performed forgery detection on doubly compressed Moving Picture Experts Group (MPEG) video sequence. As MPEG videos perform compression by partitioning video frames into Group Of Pictures (GOP) structure, the resulting motion error from inter-frame forgery will be periodic in nature, occurring in all subsequent GOPs after frame insertion or deletion point. The technique however can only be used for fixed GOP encoding and is unable to detect deleted frames with multiple GOP length. Wang and Farid (2007) performed forgery detection in interlaced and deinterlaced video by utilizing disturbance detection techniques. The approach however is computationally costly, developed mainly for intra-frame forgery detection and may not work well for low quality video. Aghamaleki

and Behrad (2016) performed inter-frame forgery detection by utilizing quantization residual errors in MPEG videos based on an evaluation of spatially constrained residual errors (SCREs) of P frames. The algorithm however requires higher computational time due to threshold optimization algorithms and system performance can suffer slightly for videos with low compression rates.

Jaiswal and Dhavale (2013) utilized Support Vector Machine (SVM) (Vapnik, 1995) for MPEG video forgery classification. Feature vector generated in Discrete Wavelet Transform (DWT) domain that characterizes Prediction Error Sequence (PES) is used for SVM classification. However, there were no localization analysis and robustness evaluation on classifier performance and a small dataset was used. Gironi, Fontani, Bianchi, Piva, and Barni (2014) proposed a method based on detection of misalignment in the video frame structure after double encoding operation. Similar to the limitation in Wang and Farid (2006), this method was investigated only for fixed GOP encoding with constant GOP size and is unable to detect inter-frame forgery associated with addition or removal of a whole GOP. Further, the technique cannot precisely localize the exact forged frame location.

Shanableh (2013) proposed several machine learning approaches to detect frame deletion forgery, namely SVM, K nearest neighbour (KNN), and logistic regression. Eight different features were proposed and they were computed mainly from both P and B frames of the MPEG coding standard based on prediction residual energy, percentage of intra-coded macroblocks, Peak Signal to Nosie Ratio (PSNR) and quantization scales. Limitations with this technique however are the inability to specify exact locations of deleted frames and unable to detect deleted frames with multiple GOP length.

Stamm, Lin, and Liu (2012) performed video forgery detection by detecting increases in P frames PES associated with frame insertion or deletion. Similar to Aghamaleki and Behrad (2016), this method is capable of detecting both frame insertion and deletion with constant or variable group of pictures length. However, a comparative evaluation performed in Aghamaleki and Behrad (2016) demonstrated that this method is less effective for videos with high compression rates.

A clear advantage of using coding-based detection techniques is that these techniques will have the ability to detect the utmost perfectly constructed forgery that is visually imperceptible, as these techniques rely on coding format rather than the actual content of video. A common limitation however in most of these coding based detection techniques, such as those detailed in Wang and Farid (2006), Aghamaleki and Behrad (2016), Jaiswal and Dhavale (2013), Shanableh (2013) and Stamm et al. (2012) is that the application of these techniques tends to be restricted only to videos coded by a specific MPEG compression standard and therefore may not be applicable to different or newer video coding standards.

Content Inconsistencies Detection Techniques

Video forgery will inevitably introduce subtle changes in the physical content of the video such as pixel intensity level changes at frame insertion or deletion boundaries. Recently, there have been an increasing number of studies utilizing such video content inconsistencies to detect video forgery.

Conotter, O'Brien, and Farid (2012) performed video forgery detection via analysis of object ballistic motion in the video. An advantage of this technique is that forgery detection is based on geometry rather than pixel intensity of video content and therefore, it is less sensitive to video resolution and compression. The video however will need to have recorded a full projectile motion for the proper usage of this technique.

Chao et al. (2013) performed inter-frame video forgery detection using optical flow technique (Lucas and Kanade, 1981) on a frame by frame basis. Limitations associated with this approach are the unrealistically created forged databases, reliance on hard threshold settings for classification, no localization analysis on forgery and slightly higher computational cost due to the usage of iterative search methodology in the algorithm. Another optical flow study that was performed by Wang, Jiang, Wang, Wan, and Sun (2013) depends on an anomaly detection scheme to perform inter-frame forgery detection.

Wang, Li, Zhang, and Ma (2014a) employed consistency of correlation coefficients of gray values (CCCoGV) as SVM features for inter-frame forgery detection. Another study by Wang, Li, Zhang, and Ma (2014b) utilized the optical flow approach in Chao et al. (2013) to perform inter-frame forgery classification based on SVM classifier. Limitations of both studies are that the forged databases were unrealistically created, and both studies focused purely on classification without localization analysis. Li, Zhang, Guo, and Wang (2016) utilized consistency of quotient of mean structural similarity (QoMSSIM) that is based on the structural similarity (SSIM) image quality metric (Lowe, 2004) to perform inter-frame forgery detection.

Zheng, Sun, and Shi (2014) demonstrated a computationally efficient, threshold-less Block-wise Brightness Variance Descriptor (BBVD) algorithm that relies on a brightness variance descriptor feature denoted as R_{BVD} to detect inter-frame forgery. Huang, Zhang, and Thing (2017) demonstrated a novel Multi-Level Subtraction (MLS) framework for inter-frame video forgery detection.

Limitations and Challenges

Having described a multitude of forgery detection techniques in previous subsections, it is clear that every technique has its own advantages and limitations, and it is difficult to speculate which may be the best technique for forgery detection without referencing to some forms of system evaluation criteria. This subsection therefore aims to formally define seven system evaluation criteria that can be used to more effectively and systematically compare the different techniques, as below.

1. **System Performance:** Good performance can be measured by high recall and precision rates or high classification accuracies of forged and original videos.
2. **Computational Efficiency:** A system that is easily implementable, algorithmically simple and computationally efficient will typically have a much shorter processing time, thus leading to substantial time savings.
3. **Threshold-less Design:** A threshold-less system design will be beneficial by eliminating human cognitive biases and ambiguities on the usage of correct thresholds. A common limitation with studies such as in Wang and Farid (2006), Gironi et al. (2014), Stamm et al. (2012), Conotter et al. (2012), Chao et al. (2013), Wang et al. (2013), Li et al. (2016) and Huang et al. (2017) is that they do rely on hard threshold settings in their detection algorithms. It is therefore unclear if these algorithms with the same hard threshold settings can generalize and perform well if other databases with different recording conditions were used.
4. **Localization Capability:** Localization refers to the ability to determine the precise frame insertion or deletion boundary in forged video sequences. A common limitation with studies such as in Wang and Farid (2006), Jaiswal and Dhavale (2013), Shanableh (2013), Stamm et al. (2012), Conotter et al. (2012) and Chao et al. (2013) is that they tend to focus purely on classification and lack a detailed localization analysis for both frame insertion and deletion forgeries.
5. **Robustness Evaluation:** A proper robustness evaluation of the forgery detection system will help in ensuring consistency of system performance under various test scenarios. For camera, coding and content inconsistencies-based detection techniques, robustness evaluation may refer to testing on different types of camera models, coding standards and video contents or recording conditions. For machine learning based detection technique, robustness evaluation may refer to usage of different compositions of training and testing databases to ensure consistency of classifier performance.
6. **Open Source Database Usage:** A standardized open-source video database should ideally be used to ensure transparency on database usage and to facilitate a fair performance comparison among different techniques.
7. **Realistic Forgery Evaluation:** For content inconsistencies detection techniques, the location of where frames are inserted or deleted from the original video sequence will have a substantial

impact on the system performance. A notable limitation with most content inconsistencies detection experiments as detailed in Chao et al. (2013), Wang et al. (2014a, 2014b), Li et al. (2016) and Zheng et al. (2014) is that the forged frames being inserted to or deleted from the original video were having too drastically different properties (such as different background environment, object alignment and brightness conditions, etc.) in comparison with neighboring frames of the original video at insertion or deletion point. In real-world forensic caseworks, frame insertion or deletion boundaries are likely to be indistinguishable visually by a lay person if the crime perpetrator has performed a careful editing of the video.

To the best of the authors' knowledge, currently there is not a single existing video forgery detection system that is able to fulfill all the seven system evaluation criteria. While recent study detailed in Huang et al. (2017) was able to achieve good performance, there are certain unfulfilled criteria such as usage of hard thresholds and lack of detailed localization and robustness analysis. This article addresses these particular limitations by proposing a novel Triangular Polarity Feature Classification (TPFC) video forgery detection framework that will have fulfilled all seven evaluation criteria, by incorporating a number of system architectural modifications on the MLS framework detailed in Huang et al. (2017).

The authors compare the performance of the proposed framework with two of the better performing techniques available, i.e. the MLS system detailed in Huang et al. (2017) as discussed above and the BBVD system as detailed in Zheng et al. (2014). The BBVD system was able to satisfy all but the seventh aforementioned system evaluation criteria, where the forged databases were unrealistically created by random insertion of forged frames. In addition, the BBVD system also similarly belongs to the content inconsistencies detection technique category and was experimented on the same open source online database that will be employed in the proposed system. These multitudes of similarities therefore make the BBVD system a reasonable choice to be an additional baseline system for comparison. Due to space constraint, a complete overview of the MLS system will not be included here and readers may refer to Huang et al. (2017) for more detailed descriptions.

SYSTEM OVERVIEW

Baseline System – Block-Wise Brightness Variance Descriptor (BBVD)

The overview of the baseline BBVD system (Zheng et al., 2014) that will be employed in this article is depicted in Figure 1. This section provides a brief summary of the important steps required for BBVD feature extraction and forgery classification. The feature extraction steps are outlined below.

1. Partition the video into sub-sequence groups of 15 frames in length with an overlapping of 5 frames.
2. Perform block partition for each frame within the sub-sequence group into 4×4 blocks as denoted by $B=\{b_1,b_2,b_3,\ldots,b_i,b_{16}\}$, with b_i represents the i^{th} block of the frame.

Calculate the ratio of BBVD for each sub-sequence group based on (1) as defined below

$$R_{BBVD}=\frac{\Delta B_{sblock}}{B_{ave}}=\frac{1}{M\times N}\sum_{i=1}^{M}\sum_{j=1}^{N}\frac{Bf_{ij}-Bl_{ij}}{B_{ave}} \tag{1}$$

where $\mathrm{R}_{\mathrm{BBVD}}$ denotes ratio of BBVD evaluated for each corresponding block between first and last frames in each sub-sequence group. M and N denote total number of row and column of pixel

values within each block. ΔB_{sblock}, Bf_{ij} and Bl_{ij} denote pixel gray value difference and pixel gray value associated with the corresponding block from first and last frames of sub-sequence group respectively. B_{ave} is defined in (2) and denotes average pixel gray value associated with pixels in current block from the first frame of each sub-sequence group.

$$B_{ave} = \frac{1}{M \times N} \sum_{i=1}^{M} \sum_{j=1}^{N} Bf_{ij} \tag{2}$$

4. Calculate the average value R_{BVD} for the R_{BBVD} values for all the 4×4 blocks accordingly using (3).

$$R_{BVD} = \frac{1}{16} \sum_{i=1}^{16} R_{BBVD} \tag{3}$$

For the classification of video forgery, the 3σ rule used for gross error detection from probability theory was employed. This 3σ rule is mathematically expressed in (4) below

$$P\left(-3\sigma < z - \mu < 3\sigma\right) = 0.9974 \tag{4}$$

where $z \sim N\left(¼, Ã^2\right)$ with ¼ and A representing the mean and standard deviation of z respectively.

The steps required to perform video forgery classification for the BBVD system are outlined as below.

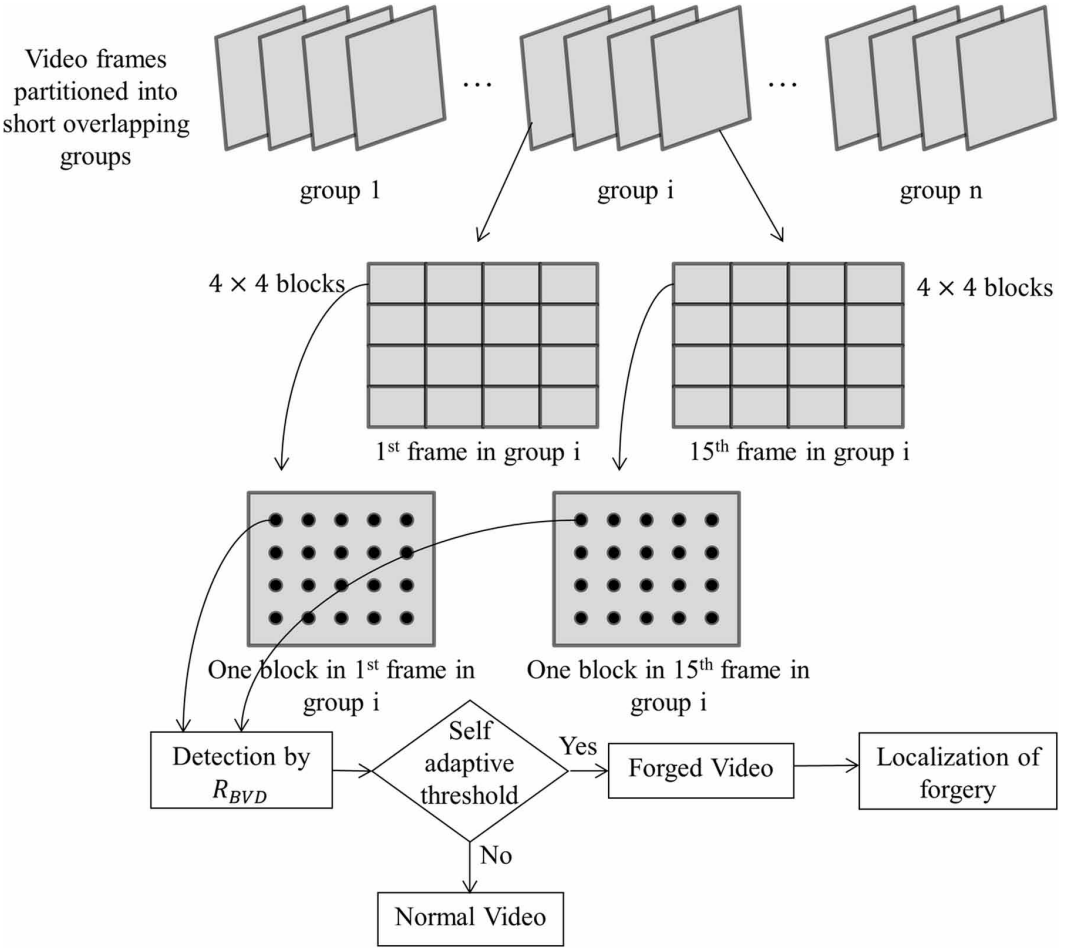
Generate the series of R_{BVD} values, $\left\{R_{BVD1}, R_{BVD2}, R_{BVD3} \ldots R_{BVDN}\right\}$ that represent the whole video sequence and determine mean μ_{BVD} and standard deviation σ_{BVD} of the R_{BVD} series as in (5) and (6):

$$\mu_{BVD} = \frac{1}{N} \sum_{i=1}^{N} R_{BVDi} \tag{5}$$

$$\sigma_{BVD} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \left(R_{BVDi} - \mu_{BVD}\right)^2} \tag{6}$$

where N denotes the total number of R_{BVD} values in the video.

Figure 1. Baseline block-wise brightness variance descriptor (BBVD) video forgery detection framework (Zheng et al., 2014)



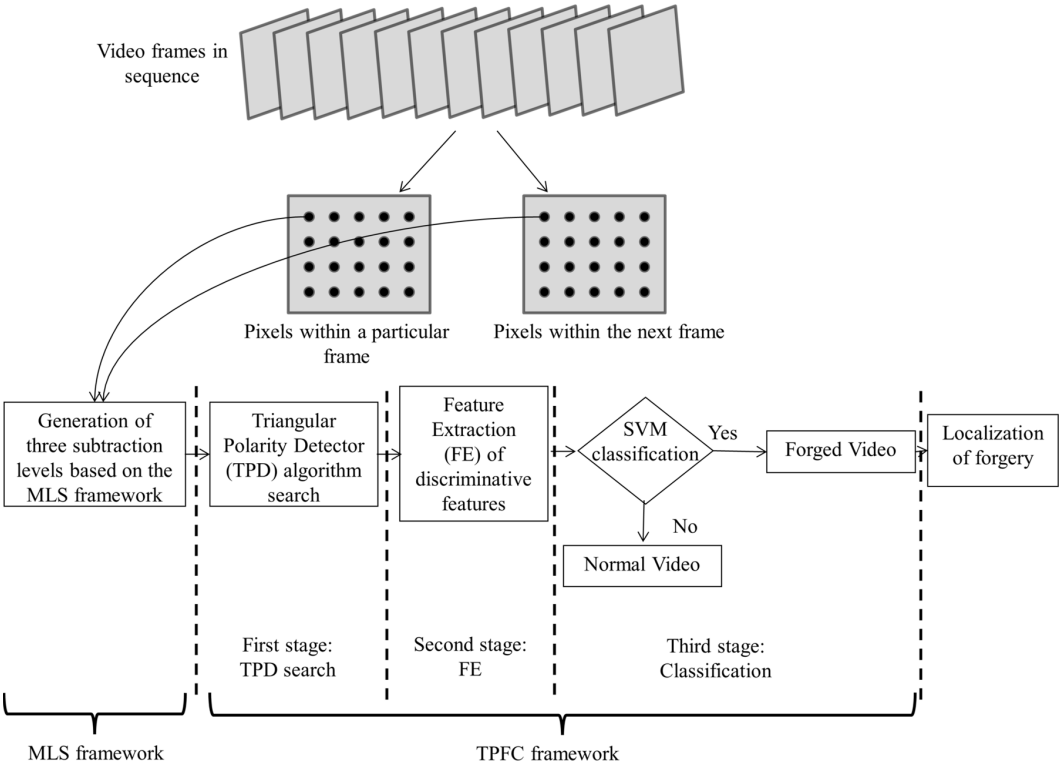
2. If there is a value greater than $3\sigma_{BVD}$ in the R_{BVD} series, then the video is classified as forged video; otherwise the video is classified as normal video.

Proposed System – Triangular Polarity Feature Classification (TPFC)

A system overview of the proposed TPFC framework is illustrated in Figure 2. As the TPFC framework is an enhancement built on the MLS framework (Huang et al., 2017), a review of MLS framework will first be outlined and followed by discussions on the three system stages in the TPFC framework.

Figure 3 presents an overview of the MLS framework where light-colored frames represent the original video frames, and dark–colored frames represent the inserted forged frames for frame insertion forgery or the remaining frames after deletion for frame deletion forgery. The frame insertion or deletion boundary is located between the light-colored frames and dark-colored frames accordingly. To effectively characterize and detect this forged boundary location, three levels of subtraction of pixel gray values are implemented as illustrated in Figure 3. In the first level subtraction of pixel gray values, any pair of adjacent light-colored frames (i.e. frames within the original video) or any pair of adjacent dark-colored frames (i.e. frames within the forged insertion video or within the remaining frames of a forged deletion video) will exhibit a higher degree of similarity than those exact two

Figure 2. Proposed triangular polarity feature classification (TPFC) video forgery detection framework



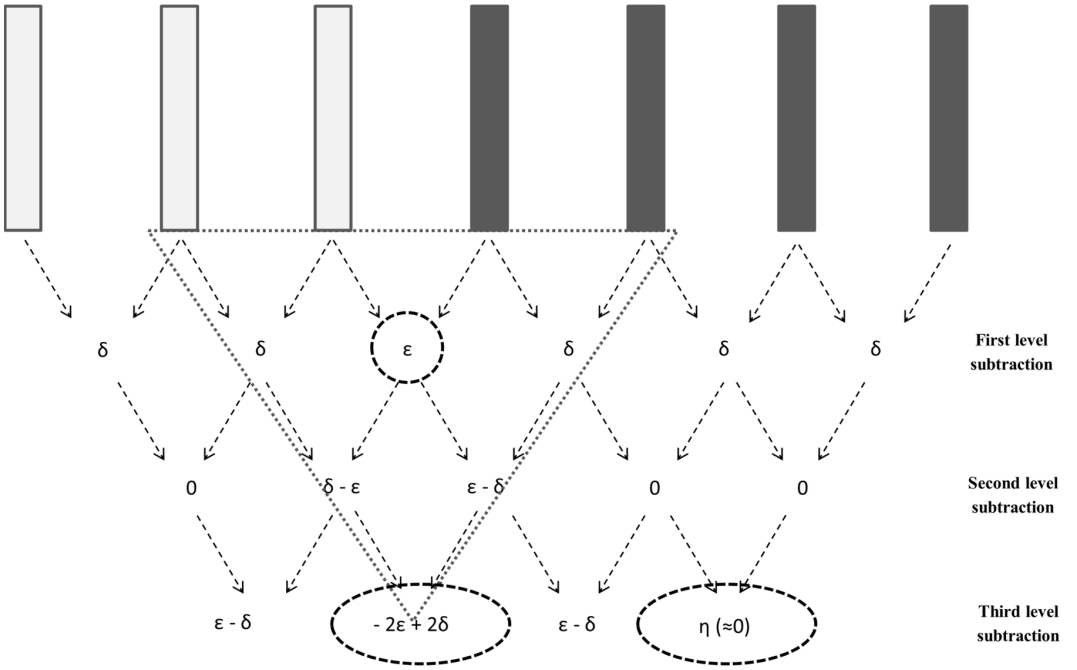
frames (i.e. one light-colored frame from original video, and one dark-colored frame from forged video) at the frame insertion or deletion boundary. This effectively implies that the first level of subtraction will return a small value (i.e. as represented by the symbol δ in Figure 3) for those adjacent pairs of purely light-colored or dark-colored frames but will return a much larger value (i.e. as represented by the symbol ε and circled in first level subtraction of Figure 3) for those exact two frames (one light-colored frame and one dark-colored frame) at the frame insertion or deletion boundary. Note that the symbols ε and δ in Figure 3 refers to the average pixel difference calculated between two video frames.

In the second level subtraction of pixel gray values, two prominent subtracted values (i.e. $\delta - \varepsilon$ and $\varepsilon - \delta$ as illustrated in Figure 3) will be obtained near the forged boundary by simply subtracting adjacent values of the first level subtraction.

In the third level subtraction of pixel gray values, one particularly prominent subtracted value of $-2\varepsilon + 2\delta$ will be obtained at the forged boundary by simply subtracting adjacent values obtained from second level subtraction. In contrast, subtraction of pixel gray values associated with frames beyond the forged boundary (i.e. using all light-colored frames or all dark-colored frames) will result in a very small value η that is close to zero. The absolute value of $-2\varepsilon + 2\delta$ will be close to the value of 2ε as ε is much larger than δ. This implies that third subtraction level of MLS framework is capable of creating a stark contrast in pixel difference by amplifying pixel difference ε. caused by forged boundary by roughly twice in magnitude while diminishing pixel difference δ at other locations away from the forged boundary to a very small value that is close to zero.

Having described the inner workings of MLS framework, the subsequent three stages of the TPFC framework will be described in detail below.

Figure 3. Multi-Level Subtraction (MLS) forgery detection framework (Huang et al., 2017)



Stage 1: Triangular Polarity Detector (TPD)

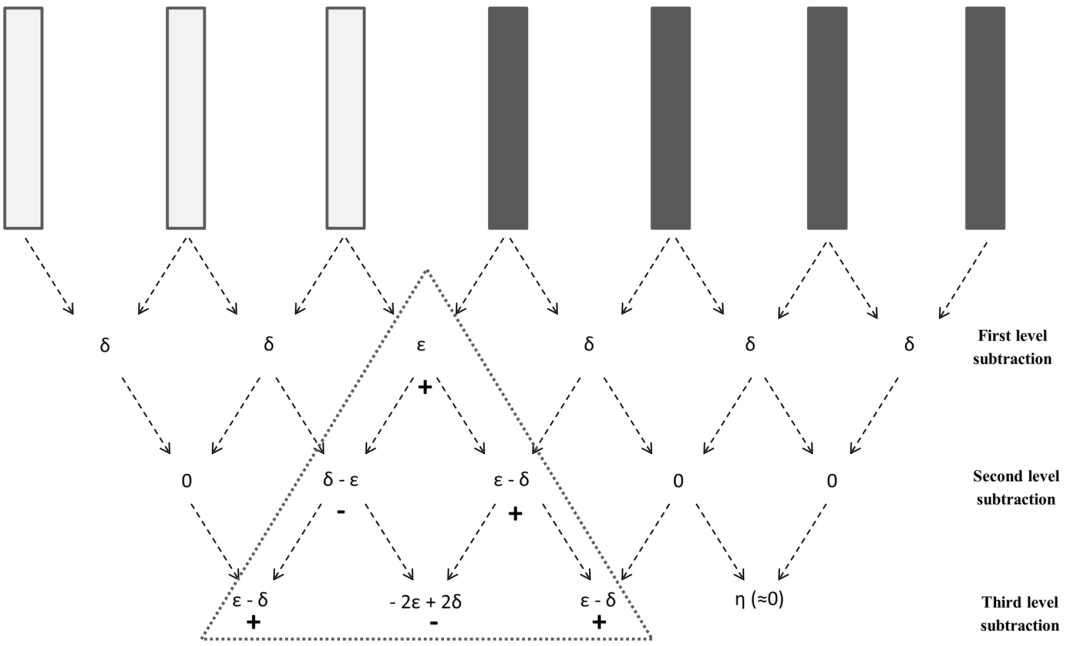
In the first stage of the TPFC framework, a novel forgery boundary search mechanism named Triangular Polarity Detector (TPD) is employed. The searching mechanism relies on detection of a triangular shaped polarity signs of the pixel difference gray values in each of the three subtraction levels as illustrated in Figure 4. Considering that the ε value at the forged boundary is having a positive polarity sign, the values $\delta - \varepsilon$ and $\varepsilon - \delta$ at the second level of subtraction will have negative and positive polarity signs respectively as ε is a larger number than δ in terms of magnitude. Similarly, the values of $-2\varepsilon + 2\delta$ and $\varepsilon - \delta$ in the third subtraction level will have negative and positive polarity signs respectively. The polarity signs for all three subtraction levels associated with the forged boundary are captured in the triangle as illustrated in Figure 4.

Stage 2: Extraction of Discriminative Features

The second stage of detection process involves extraction of discriminative features from MLS framework at forged boundary location obtained via TPD search mechanism as discussed in the first stage of detection. Three discriminative features were extracted as detailed in Figure 5. The steps required to extract the first feature x_1 is detailed below.

1. Locate the two circled terms $\delta - \varepsilon$ and $\varepsilon - \delta$ from the second subtraction level in Figure 5.
2. Denoting the absolute value of these two terms as A_1 and A_2 where $A_1 = \left|\delta - \varepsilon\right|$ and $A_2 = \left|\varepsilon - \delta\right|$, calculate the mean of A_1 and A_2 and denote this as m as defined in (7).

Figure 4. Triangular Polarity Detector (TPD) for the case of subtracting values of the second and third subtraction levels from left to right



$$m = \frac{A_1 + A_2}{2} \tag{7}$$

3. The feature x_1 can be computed by taking ratio of the absolute difference of A_1 and A_2 to the mean m as in (8).

$$x_1 = \frac{\left|A_1 - A_2\right|}{m} \tag{8}$$

Note that the by substituting (7) into (8) the feature x_1 can also be expressed in (9) below.

$$x_1 = 2\frac{\left|A_1 - A_2\right|}{A_1 + A_2} \tag{9}$$

The discriminative property of this feature can be observed by evaluating numerator and denominator terms in (9) for forged and normal videos. For frame insertion or deletion videos at forged boundary location, the numerator term $\left|A_1 - A_2\right|$ will have a small value as both A_1 and A_2 will have similar magnitudes while denominator term $A_1 + A_2$ will have a large value as ε is much larger than δ. The resulting ratio of a small-valued numerator to a large-valued denominator will result in an even smaller value close to 0. This implies that feature x_1 will be small if the video has

inter-frame forgery. For the case of normal videos, as there is no forged boundary, the location determined by TPD search mechanism will have a much smaller ε value relative to the ε value obtained from a real insertion or deletion boundary in a forged video. Since ε is much smaller in this case, this means that the individual terms A_1 and A_2 will also be small. This will result in numerator term $\left|A_1 - A_2\right|$ to be small as well, however the denominator term $A_1 + A_2$ will be reduced to a much smaller value relative to the denominator value obtained for the case of forged video. Since the denominator is having a much smaller value, overall this will have the effect of increasing the feature x_1 value to a larger value compared with the case of forged video.

Similarly, the steps required to extract the second feature x_2 is detailed below.

1. Locate the two $\varepsilon - \delta$ terms from the third subtraction level in Figure 5.
2. Denote the absolute value of these two terms as A_1 and A_2 where $A_1 = \left|\varepsilon - \delta\right|$ and $A_2 = \left|\varepsilon - \delta\right|$. By similar reasoning, feature x_2 can also be expressed as in (10) and will have similar behavior to feature x_1 by having small value for forged videos and large value for normal videos.

$$x_2 = \frac{\left|A_1 - A_2\right|}{m} = 2\frac{\left|A_1 - A_2\right|}{A_1 + A_2} \tag{10}$$

Lastly, the steps required to extract the third feature x_3 is detailed below.

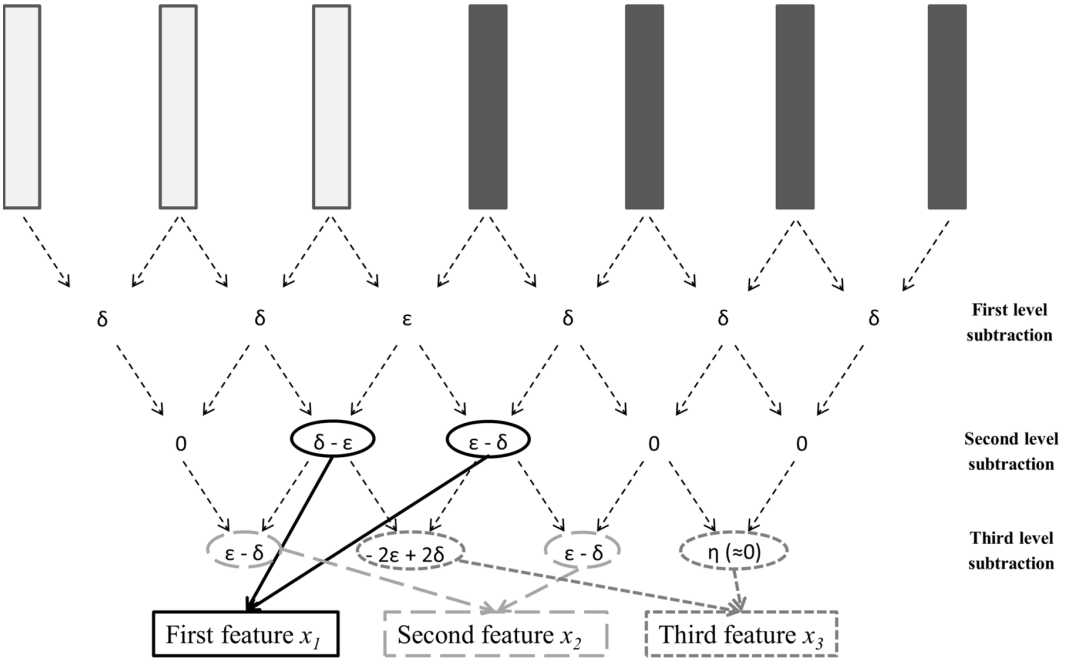
1. Locate the two terms $-2\varepsilon + 2\delta$ and η from the third subtraction level in Figure 5.
2. The feature x_3 can be computed by taking ratio of absolute value of η to the absolute value of $-2\varepsilon + 2\delta$ as in (11).

$$x_3 = \frac{\left|\cdot\right|}{\left|-2\varepsilon + 2\delta\right|} \tag{11}$$

At the forged boundary location in a forged video, the numerator term $\left|\eta\right|$ will have a small value while denominator term $\left|-2\varepsilon + 2\delta\right|$ will have a large value as ε is much larger than δ. The resulting ratio of a small-valued numerator to a large-valued denominator will result in an even smaller value close to 0. This implies that feature x_3 will be small if the video has inter-frame forgery. For the case of normal videos, the location determined by TPD search mechanism will have a much smaller ε value in comparison to the ε value obtained from a real insertion or deletion boundary in a forged video as discussed previously. As ε is much smaller in this case, the denominator term $\left|-2\varepsilon + 2\delta\right|$ will also be smaller. This will then have the effect of increasing the feature x_3 value to a larger value compared with the case of forged video.

Stage 3: Support Vector Machine (SVM) Classification

Figure 5. Feature extraction process where three discriminative features are calculated at forged boundary locationi



Finally, the third stage will utilize the extracted features to perform video forgery classification on the basis of SVM (Vapnik, 1995). The concept of SVM is to devise a separation hyperplane that optimally separates two classes of training samples with the largest margin of separation and it can expressed as in (12) below:

$$f\left(\mathrm{x}\right) = \sum_{i=1}^{L} \alpha_i t_i K\left(\mathrm{x}, \mathrm{x}_i\right) + d \tag{12}$$

where represent the class labels, and are the learned constant with and . represents the total number of support vectors with and denoting the input test vector and the support vector respectively. Support vectors are those training data points that lie closest to the hyperplane. The SVM performs classification of any input test vector by determining if the value of is above or below a threshold. The kernel function can be used to map data in the input space to a higher dimensional kernel feature space such that the data becomes linearly separable in the kernel feature space. There are numerous kernel functions associated with SVM; one of the more popular choices is the Radial Basis Function (RBF) kernel as expressed mathematically in (13) below:

$$K\left(\mathrm{x}, \mathrm{x}_i\right) = \exp\left(-\gamma \left\|\mathrm{x} - \mathrm{x}_i\right\|^2\right) \tag{13}$$

where $^{3} > 0$ is the adjustable RBF parameter (Chang & Lin, 2011). In this study, SVM with RBF kernel was employed via a publicly available toolkit named LIBSVM (Chang and Lin, 2011) and the RBF parameter empirically defined at the value of 0.001.

To summarize, the steps required to perform video forgery detection based on the TPFC framework are as below:

1. Generate the required three levels of subtraction based on the MLS framework for the whole video sequence.
2. Sort the values generated at the third subtraction level from the highest to the lowest in terms of magnitude.
3. Perform TPD check for each of the sorted third subtraction level values starting from the highest magnitude.
4. Once TPD check has returned a frame boundary location, determine the corresponding location in the video and perform extraction of the three discriminative features.
5. Perform SVM training and classification accordingly on the basis of the three extracted discriminative features.

EXPERIMENTAL PROCEDURES AND DATABASES

The video database employed is the Recognition of Human Actions Database (Schuldt, 2004). This database is open-source and was used by numerous past video forgery detection studies such as those in Chao et al. (2013), Wang et al. (2014a, 2014b), Li et al. (2016), Zheng et al. (2014) and Huang et al. (2017). It contains six classes of actions (i.e. boxing, handclapping, hand waving, jogging, running and walking) performed by 25 human subjects under four different recording conditions, S1-S4 with S1: outdoor, S2: scaled version of S1, S3: outdoor with different clothing and S4: indoor. Videos were recorded over static homogenous background in AVI format with 25 frames per second.

Frame Insertion Forgery Dataset

For frame insertion experiment, a total of 100 videos associated with the 25 human subjects from each of the walking, jogging and running action classes in all 4 conditions (S1, S2, S3, and S4) were used. Each video in the database has 4 sets of enter-exit frame number markings that indicate when the person enters and exits the camera viewing range for a total of 4 times. To create an appropriate frame insertion database, the second or third set of enter-exit markings from each human subject video was replaced with the corresponding second or third set of enter-exit markings from another human subject video in the same action class and condition.

From the generated insertion database, a total of 600 forged videos were randomly selected for the experiments. The replacement of frames was performed to be forensically realistic such that there is no abrupt appearance or disappearance of human subject in the forged video. In addition, excluding one missing video in the database, a total of 599 normal videos from all 25 human subjects in all action classes and conditions were utilized to form a database of normal videos.

Frame Deletion Forgery Dataset

For frame deletion experiment, video associated with all six action classes under all conditions were used. To create a frame deletion video database, the second or third set of frame sequence according to the provided enter-exit markings for each video was removed. From the generated deletion database, a total of 600 forged videos were randomly selected for the experiments. Similar to the insertion forgery experimental setup, the 599 normal videos were used to form a database of normal videos.

Experiment Setup

For the TPFC framework, three-fold cross-validation was performed. The videos in forged and normal databases were divided into three sets. One set of forged and normal videos were used for testing while the remaining sets were used for training during cross-validation to examine the robustness of the SVM classification performance.

Localization accuracy of the TPFC framework can be calculated for each test video by determining if the frame location where the three discriminative test features $\mathrm{x} = \left[\mathrm{x}_1, \mathrm{x}_2, \mathrm{x}_3\right]$ were extracted matches exactly to the ground truth label of the actual frame insertion or deletion boundary location. Furthermore, localization accuracy can similarly be calculated for the baseline BBVD system by evaluating if sub-sequence group number that contains the maximum RBVD value also contains the ground truth frame insertion boundary location.

To comparatively evaluate the TPFC system performance, results for BBVD framework (Zheng et al., 2014) and MLS framework (Huang et al., 2017) were both evaluated under the same experimental database setting to facilitate a fair comparison between the three different forgery detection frameworks. While the BBVD system relies on an adaptive threshold-based algorithm for forgery detection, the MLS framework detailed in Huang et al., 2017 is a hard threshold based detection system with threshold percentages α and β to characterize the forged frame boundary. An additional evaluation on the effect of varying these threshold percentages α and β on MLS system performance was also performed.

EVALUATION METRICS

Recall and precision rates were used as performance metrics, as expressed in (14) and (15) below:

$$R_r = \frac{N_c}{N_c + N_m} \times 100\% \tag{14}$$

$$R_p = \frac{N_c}{N_c + N_f} \times 100\% \tag{15}$$

where N_c denotes the number of correctly detected normal and forged videos, N_m denotes the number of missed video forgeries and N_f is the number of falsely detected video forgeries.

EXPERIMENTAL RESULTS AND DISCUSSIONS

Comparison of Experimental Results Between the BBVD and TPFC Frameworks

Experimental results for frame insertion and deletion forgeries evaluated for the BBVD and TPFC frameworks are tabulated in Table 1 and 2 respectively.

For the BBVD system evaluated in the frame insertion experiment, 419 frame insertion videos were correctly detected out of a total of 600 frame insertion videos. A total of 482 normal videos were correctly detected out of a total of 599 normal videos. The number of correctly detected video forgeries N_c, number of missed video forgeries Nm and number of falsely detected video forgeries N_f were 901, 181 and 117 respectively. For localization accuracy, 359 videos out of a total of 419 detected frame insertion videos were correctly localized at the precise frame insertion boundary. For the BBVD system evaluated in the frame deletion experiment, 129 frame deletion videos were correctly detected out of a total of 600 frame deletion videos. N_c, N_m and N_f were 611, 471 and 117 respectively. For localization accuracy, only 22 videos out of a total of 129 detected frame deletion videos were correctly localized at the precise frame deletion boundary.

For the proposed TPFC system, SVM classification and localization accuracies are presented in Table 1 for frame insertion experiment and Table 2 for frame deletion experiment.

Results outlined in Table 1 and 2 demonstrated the TPFC framework is able to outperform the BBVD framework with improvement in classification accuracy of up to 26.84% and 43.00% for frame insertion and deletion forgeries respectively. High recall and precision rates at 98.26% and 95.76% respectively for frame insertion experiment and 80.60% and 89.76% respectively for frame deletion experiment were also exhibited by TPFC framework. Moreover, substantial improvement in localization accuracy of TPFC framework of up to 14.32% and 74.16% was also observed for frame insertion and deletion forgeries respectively.

One of the contributing factors in good TPFC system performance is associated with the highly discriminative features as demonstrated in feature distribution plots in Figure 6. Images illustrated in first column of Figure 6 shows features distribution extracted from normal database (i.e. containing 599 normal videos), whereas images in the second and third columns plot the corresponding features distribution extracted from frame insertion database (i.e. containing 600 frame insertion videos) and frame deletion database (i.e. containing 600 frame deletion videos), respectively. These plots indicate that feature values were relatively smaller for tampered videos than the normal videos. This high feature separation between the two classes therefore translates to good classification performance.

Comparing the features distribution between frame insertion and frame deletion videos, it can be noticed that extracted features from frame insertion videos have consistent smaller values. This can be attributed to the fact that frame deletion videos generally have higher correlation (i.e. lower pixel difference) between neighboring frames at the frame deletion boundary as compared with neighboring frames at the frame insertion boundary from frame insertion videos. Thus, it is generally more challenging to detect forgery in frame deletion videos compared to frame insertion videos

Table 1. Results for frame insertion forgery detection

	BBVD (%)	TPFC (%)	Improvement (%)
Classification accuracy on forged videos	69.83	96.67	26.84
Classification accuracy on normal videos	80.47	91.66	11.19
Recall rate R_r	83.27	98.26	14.99
Precision rate R_p	88.51	95.76	7.25
Localization accuracy on forged videos	85.68	100.00	14.32

Table 2. Results for frame deletion forgery detection

	BBVD (%)	TPFC (%)	Improvement (%)
Classification accuracy on forged videos	21.50	64.50	43.00
Classification accuracy on normal videos	80.47	83.14	2.67
Recall rate R_r	56.47	80.60	24.13
Precision rate R_p	83.93	89.76	5.83
Localization accuracy on forged videos	17.05	91.21	74.16

through comparison derived from pixels intensity as can be observed in Table 1 and 2 where higher recall and precision rates were exhibited in frame insertion detection experiments.

Another key contribution to good TPFC system performance is the effectiveness of MLS framework in detecting sudden inconsistency of pixel intensity at forged boundary. Figure 7 and 8 show comparisons between the R_{BVD} plots and MLS third level subtraction plots of two frame insertion videos. Both tampered videos in Figure 7 and 8 utilized the same original video, but were inserted with different video segments at the same start and end boundaries, i.e. frame numbers 152 and 225 respectively.

Figure 7 shows that while BBVD system failed to detect any forgery in the video, the MLS framework was able to identify the forgery and frame insertion boundaries. In Figure 8, both BBVD system and MLS framework were able to classify the video as forged however with distinctive characteristics. As illustrated in both Figure 7 and 8, the third subtraction level of MLS framework has the beneficial effect of boosting the signal amplitudes at forged boundaries while at the same time diminishing signal amplitudes at all other normal frame locations. This is in contrast to the behaviour of R_{BVD} feature values illustrated in the corresponding plots where although the BBVD framework was able to correctly classify the frame inserted video as a forged video, there were other undesirable R_{BVD} signal peaks at other locations where frame insertion has not occurred as well. It should be noted that BBVD algorithm runs on a block-based approach, and instead of the frame number, sub-sequence group numbering which is a multiple of 10 frames is labelled on the x-axis in Figure 7 and 8 (left images).

The effectiveness of MLS framework in modeling of the forged boundaries significantly helps in ensuring an accurate TPD forged boundary search in the first stage and highly discriminative feature extraction in the second stage of the TPFC framework.

Figure 6. Distribution plots of extracted TPFC features, x_1, x_2, x_3 from normal database (column 1), frame insertion database (column 2) and frame deletion database (column 3)

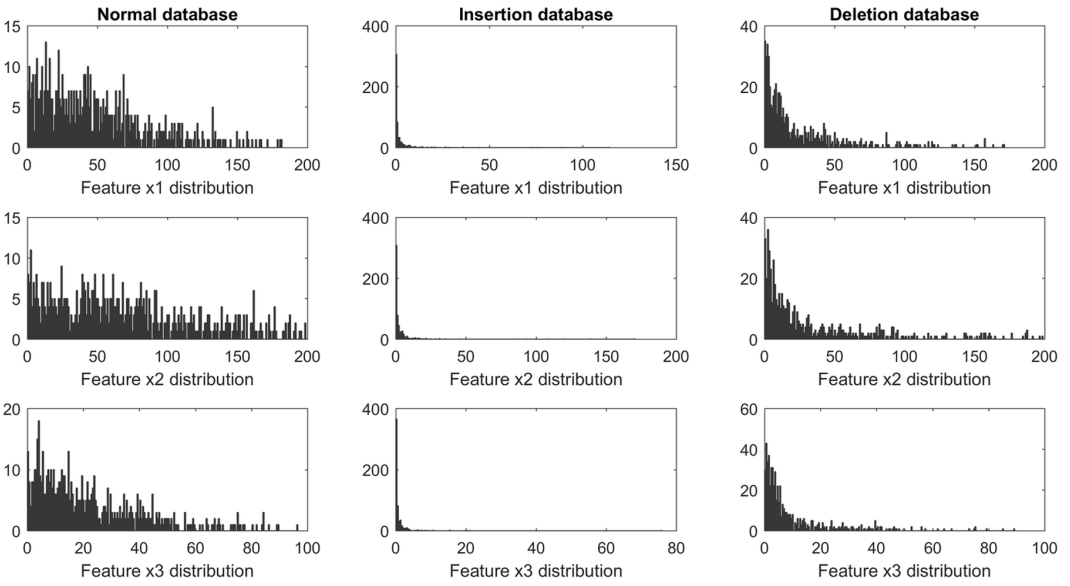


Figure 7. R_{BVD} **feature values from BBVD system that failed forgery detection using the 3σ rule (left image) and the corresponding third level subtraction feature values from MLS framework (right image)**

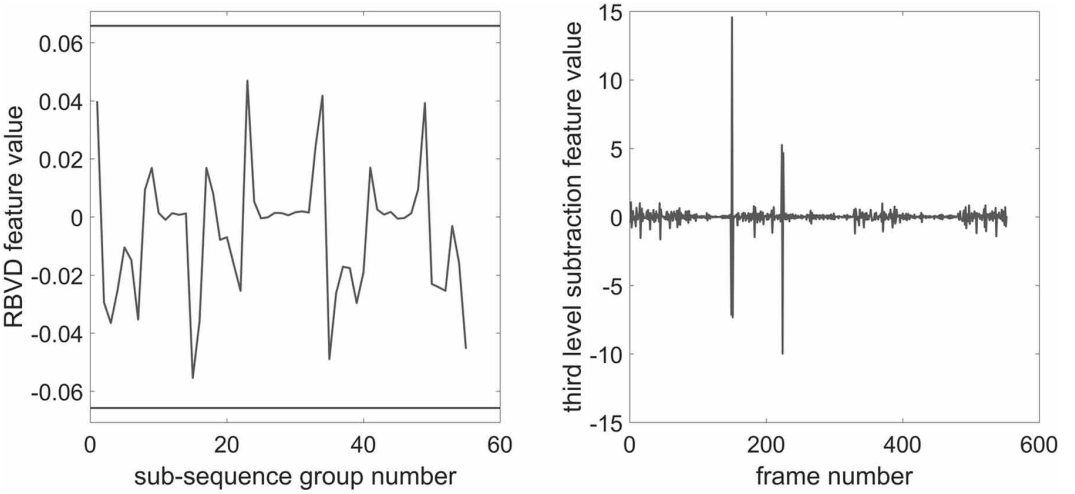
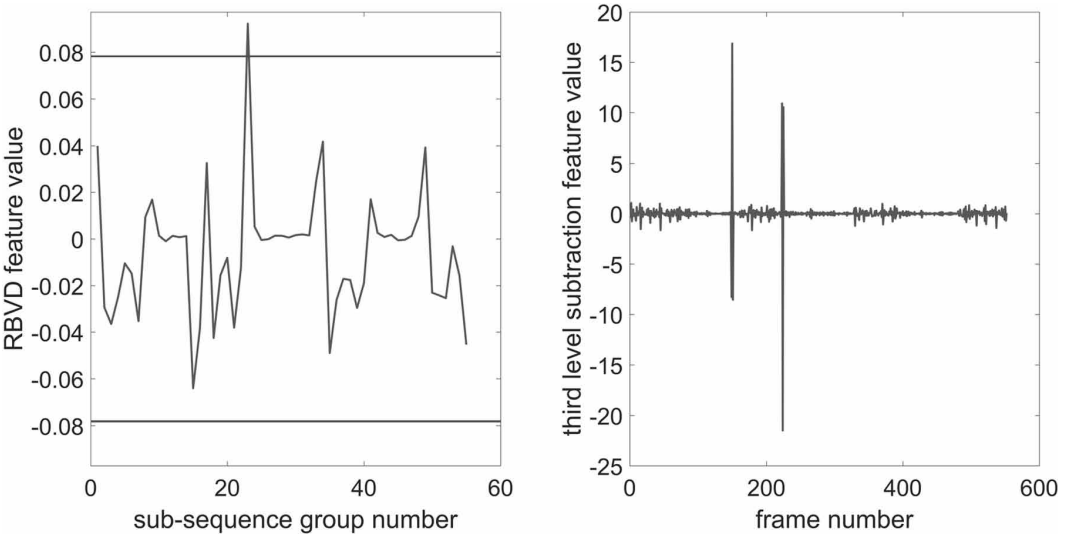


Figure 8. R_{BVD} **feature values from BBVD system that successfully detect forgery (left image) and the corresponding third level subtraction feature values from MLS framework (right image)**



Comparison of Experimental Results Between the MLS and TPFC Frameworks

Experimental results for frame insertion and deletion forgeries evaluated for the MLS and TPFC frameworks are tabulated in Table 3 and 4 respectively. Various settings for both threshold percentages α and β at 1%, 5% (i.e. setting used in Huang et al., 2017) and 50% were evaluated.

With the increase of threshold percentages, classification accuracy results from Table 3 and 4 exhibited a tendency to increase on forged videos and a tendency to decrease on normal videos.

Correspondingly, a similar trend of increasing recall rate with a decreasing precision rate was also observed with the increase of threshold percentages in Table 3 and 4. Furthermore, the MLS results from Table 3 and 4 indicated a strong dependency on threshold settings with the possibility of classification accuracies on forged and normal videos reaching as low as 9.83% and 80.63% respectively, and for recall and precision rates reaching as low as 54.84% and 88% respectively. These sharp increasing/decreasing trends suggested a high dependency of MLS performance to the threshold settings. Without an automatic approach in determining the optimal threshold settings, manual selection of threshold settings based purely on human judgement will be prone to error due to human bias and ambiguity. Further, optimal threshold setting may differ according to different database content or recording conditions, making the task of manually determining the optimal setting an inconvenient and challenging one. This therefore presents a clear advantage of TPFC framework over the MLS framework in that the TPFC framework is a threshold-less system with an overall better system performance in comparison with MLS system as shown in Table 3 and 4.

CONCLUSION

This article has investigated a novel TPFC video forgery detection framework on a specific set of forged video databases constructed to be closely representative to real-world forensic casework applications. The TPFC framework was demonstrated to achieve high recall and precision rates as

Table 3. Results for frame insertion forgery detection

	MLS (%)			TPFC (%)
Threshold percentages α and β	**1**	**5**	**50**	
Classification accuracy on forged videos	71.00	91.17	96.83	96.67
Classification accuracy on normal videos	99.83	97.66	80.63	91.66
Recall rate R_r	85.48	95.53	98.25	98.26
Precision rate R_p	99.90	98.78	90.17	95.76
Localization accuracy on forged videos	100.00	100.00	100.00	100.00

Table 4. Results for frame deletion forgery detection

	MLS (%)			TPFC (%)
Threshold percentages α and β	**1**	**5**	**50**	
Classification accuracy on forged videos	9.83	39.17	61.33	64.50
Classification accuracy on normal videos	99.83	97.66	80.63	83.14
Recall rate R_r	54.84	69.20	78.58	80.60
Precision rate R_p	99.85	98.32	88.00	89.76
Localization accuracy on forged videos	98.31	95.32	90.22	91.21

well as high localization accuracy. Practical aspects such as threshold-less design, system robustness and efficiency were also considered.

As future works, a higher number of MLS subtraction levels of pixel gray values beyond the three levels of subtraction that were used currently could be explored. This will lead to the possibility of creating more complex and discriminative features and thus improving the detection of forged frame boundary. Another research avenue could involve the partitioning each video frame into smaller blocks of pixels as multiple individual inputs to the TPFC framework. A weighted combination of the various TPFC system outputs could then be used as the final system result. Moreover, different modeling approaches apart from SVM could also be evaluated within the TPFC framework.

REFERENCES

Aghamaleki, J. A., & Behrad, A. (2016). Inter-frame video forgery detection and localization using intrinsic effects of double compression on quantization errors of video coding. *Signal Processing Image Communication*, *47*, 289–302. doi:10.1016/j.image.2016.07.001

Chang, C. C., & Lin, C. J. (2011). LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, *2*(3), 27. doi:10.1145/1961189.1961199

Chao, J., Jiang, X., & Sun, T. (2013). A novel video inter-frame forgery model detection scheme based on optical flow consistency. In *Proceedings of the International Workshop on Digital Forensics and Watermarking 2012* (pp. 267-281). Springer. doi:10.1007/978-3-642-40099-5_22

Conotter, V., O'Brien, J. F., & Farid, H. (2012). Exposing digital forgeries in ballistic motion. *IEEE Transactions on Information Forensics and Security*, *7*(1), 283–296. doi:10.1109/TIFS.2011.2165843

Gironi, A., Fontani, M., Bianchi, T., Piva, A., & Barni, M. (2014, May). A video forensic technique for detecting frame deletion and insertion. In ICASSP (pp. 6226-6230). Academic Press. doi:10.1109/ICASSP.2014.6854801

Hsu, C. C., Hung, T. Y., Lin, C. W., & Hsu, C. T. (2008, October). Video forgery detection using correlation of noise residue. In *Proceedings of the 2008 IEEE 10th Workshop on Multimedia Signal Processing* (pp. 170-174). IEEE.

Huang, C. C., Zhang, Y., & Thing, V. L. (2017, August). Inter-frame video forgery detection based on multi-level subtraction approach for realistic video forensic applications. In *Proceedings of the 2017 IEEE 2nd International Conference on Signal and Image Processing (ICSIP)* (pp. 20-24). IEEE. doi:10.1109/SIPROCESS.2017.8124498

Jaiswal, S., & Dhavale, S. (2013). Video Forensics in Temporal Domain using Machine Learning Techniques. *International Journal of Computer Network and Information Security*, *5*(9), 58–67. doi:10.5815/ijcnis.2013.09.08

Kingra, S., Aggarwal, N., & Singh, R. D. (2016). Video Inter-frame Forgery Detection: A Survey. *Indian Journal of Science and Technology*, *9*(44). doi:10.17485/ijst/2016/v9i44/105142

Kobayashi, M., Okabe, T., & Sato, Y. (2009, January). Detecting video forgeries based on noise characteristics. In *Proceedings of the Pacific-Rim Symposium on Image and Video Technology* (pp. 306-317). Springer. doi:10.1007/978-3-540-92957-4_27

Kobayashi, M., Okabe, T., & Sato, Y. (2010). Detecting forgery from static-scene video based on inconsistency in noise level functions. *IEEE Transactions on Information Forensics and Security*, *5*(4), 883–892. doi:10.1109/TIFS.2010.2074194

Kurosawa, K., Kuroki, K., & Saitoh, N. (1999). CCD fingerprint method-identification of a video camera from videotaped images. In *Proceedings. 1999 International Conference on Image Processing ICIP 99* (Vol. 3, pp. 537-540). IEEE. doi:10.1109/ICIP.1999.817172

Li, Z., Zhang, Z., Guo, S., & Wang, J. (2016). Video inter-frame forgery identification based on the consistency of quotient of MSSIM. *Security and Communication Networks*, *9*(17), 4548–4556. doi:10.1002/sec.1648

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, *60*(2), 91–110. doi:10.1023/B:VISI.0000029664.99615.94

Lucas, B. D., & Kanade, T. (1981). An iterative image registration technique with an application to stereo vision.

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*.

Shanableh, T. (2013). Detection of frame deletion for digital video forensics. *Digital Investigation*, *10*(4), 350–360. doi:10.1016/j.diin.2013.10.004

Stamm, M. C., Lin, W. S., & Liu, K. R. (2012). Temporal forensics and anti-forensics for motion compensated video. *IEEE Transactions on Information Forensics and Security*, *7*(4), 1315–1329. doi:10.1109/TIFS.2012.2205568

Vapnik, V. (2013). *The nature of statistical learning theory*. Springer science & business media.

Wang, Q., Li, Z., Zhang, Z., & Ma, Q. (2014a). Video inter-frame forgery identification based on consistency of correlation coefficients of gray values. *Journal of Computer and Communications*, *2*(04), 51–57. doi:10.4236/jcc.2014.24008

Wang, Q., Li, Z., Zhang, Z., & Ma, Q. (2014b). Video inter-frame forgery identification based on optical flow consistency. *Sensors & Transducers*, *166*(3), 229.

Wang, W., & Farid, H. (2006, September). Exposing digital forgeries in video by detecting double MPEG compression. In *Proceedings of the 8th workshop on Multimedia and security* (pp. 37-47). ACM. doi:10.1145/1161366.1161375

Wang, W., & Farid, H. (2007). Exposing digital forgeries in interlaced and deinterlaced video. *IEEE Transactions on Information Forensics and Security*, *2*(3), 438–449. doi:10.1109/TIFS.2007.902661

Wang, W., Jiang, X., Wang, S., Wan, M., & Sun, T. (2013, October). Identifying video forgery process using optical flow. In *Proceedings of the International Workshop on Digital Watermarking* (pp. 244-257). Springer.

Zheng, L., Sun, T., & Shi, Y. Q. (2014, October). Inter-frame video forgery detection based on block-wise brightness variance descriptor. In *Proceedings of the International Workshop on Digital Watermarking* (pp. 18-30). Springer.

Chee Cheun Huang is a data science professional with a broad spectrum of domain expertise and technical knowledge in the areas of machine learning algorithm development, digital signal processing and electronics engineering. Chee Cheun holds a PhD in Electrical Engineering from The University of New South Wales, Australia and his research interests include image and video analytics, automatic phonetic segmentation and speech recognition, and statistical modelling with applications in forensic voice comparison and automatic speaker recognition.

*Chien Eao Lee is currently a research engineer in the Data Security unit of Cyber-Security and Intelligence department at Institute for Infocomm Research (I2R), Agency for Science, Technology and Research (A*STAR). Chien Eao holds a B.Eng (Hons) degree in electronics engineering, M.Eng in telecommunications engineering, and M.IT in digital communications. Her research interests include areas in image processing and multimedia forensics.*

*Vrizlynn Thing is the Head of Cyber Security and Intelligence at the Institute for Infocomm Research (I2R), Agency for Science, Technology and Research (A*STAR). She is also an Adjunct Associate Professor at the National University of Singapore (School of Computing), and holds the appointment of Honorary Assistant Superintendent of Police (Specialist V) at the Singapore Police Force, Ministry of Home Affairs. Her research draws on her multidisciplinary background in computer science (Ph.D. from Imperial College London, United Kingdom), and electrical, electronics, computer and communications engineering (M.Eng. by Research and B.Eng (Hons) from Nanyang Technological University, Singapore, and Diploma from Singapore Polytechnic). During her career, she has taken on various roles with the key focus to lead and conduct cyber security R&D that benefit our economy and society. She also participates actively as the Lead Scientist of collaborative projects with industry partners and government agencies, and takes on advisory roles at national and international level on cyber security initiatives.*