

A Novel Deep Learning Framework for Forged Scene Detection in Advanced Video

P.Karthikeyan¹, R.Bhavani², D.Rajiniginath³, R. Priya⁴

¹Research Scholar, Dept. of CSE, Annamalai University

^{2,4}Professor, Dept. of CSE, Annamalai University

³Professor, Dept. of CSE, Sri Muthukumaran Institute of Technology

Article Info

Volume 82

Page Number: 16441 - 16447

Publication Issue:

January-February 2020

Abstract

Video forensics is one of the most prominent areas of digital forensics which helps the real world in finding the originality and authenticity of the videos. Existing techniques which focuses on the video forensics relies on the video data stream for the analysis and verification. In recent times there is a new line of research which encountered by monitoring and analyzing the video containers to verify the authenticity of the videos. Furthermore, existing work on the video containers works based on the manual compression and feature extraction which is a time consuming task and require a huge processing power. To address this issue, this paper proposes a novel deep learning framework for detecting the video forgery. The key idea of the proposed deep learning framework is to utilize the dissimilarity between the frames on the video and to extract the meta-information from the video containers. The proposed framework is experimentally tested and validated using high end workstation with Intel Xeon processor and 64 GB RAM and Titan X Pascal GPU card. The experimental validation is carried out on an average of 100 videos of 360p video format and it is proven that the proposed framework works fine with better accuracy of nearly 98% for the tested samples.

Article History

Article Received: 18 May 2019

Revised: 14 July 2019

Accepted: 22 December 2019

Publication: 28 February 2020

Keywords: Video forensics, image forensics, Deep learning model, Digital forensics, Video forgery.

I. INTRODUCTION

With the rise in popularity of deep learning and machine learning, we can observe their benefits in our day to day life. While this exceptional technology brings comfort and help the human race advance, there have been several misuses reported. As with every other technology, people find a way to misuse it for several ill intentions [1-2]. One of the major misuses of deep learning was to forge faces, fingerprints etc. Such forgeries are extreme security vulnerability [41] which is capable of compromising a huge potential of data. One such example was the “Deep fake” which caused huge mayhem in the world. Deep fake essentially was a deep learning algorithm trained very well to replicate the faces of any person. Later on, it also introduced voice cloning [3-4].

This can be considered a potential identity theft which is punishable by the penal code of any judicial system in

the world. Not just this, such identity thefts can be misused to harm the

reputation of the people on the internet spread false propaganda etc. And if this continues unchecked, it wouldn't be long before people lose trust in the videos they watch. There have been several methods and trusted forensic ways to detect video forgeries. But there still needs upward improvements to combat such forgeries which have a great deal of precision [5-8]. To combat this, researchers at Technical University of Munich have developed a deep learning algorithm which can potentially identify face swaps in videos on the internet. Furthermore, the forgery-classification task aims at identifying duplicate images. This puts the binary classification problem on the basis of each frame of handled videos [9]. Because there are no specific approaches in the current literature for detecting Face 2 Face manipulations, in the forensic community, we decided to consider learning-based

methods used for general manipulation detection, computer generated, and natural image detection, and face damage detection. In addition, we have added a sophisticated deep network [10-11].

Each of these methods is trained in the target reconstruction dataset from a single source consisting of 10 frames from each of the 704 simulators and 704 beautiful videos. Similarly, both the validation and test set have 10 frames extracted from 150 (original) and 150 (duplicate) videos [42]. For each frame, we carve all of the face-centered images, where we use the mask provided by Face 2 Face. When requested, faces are changed to the amount of input on the network; otherwise, a clip of 128×128 pixels centered on the face was extracted as input [11].

Basically, the researchers began by preparing a dataset of over 1000 original videos as well as the forged ones (the ones with face swaps). The basic criteria were to classify the forged frames and the original frames. Thus, this dataset of original and forged videos provides a huge array of data to learn the key difference between them. Moreover, they also added additional computer-generated images to improve the input quality of the data in recognizing such forgery [12-15].

In total, there were more than half a million images of faces that had been manipulated by software. This curated dataset is then fed to a state-of-the-art deep learning neural network. It had been trained over multiple epochs with several optimization techniques handling the bias, variance and other potential problems. One tremendous leap this CNN has produced is how it produces results even when the input video is highly compressed. Since compressed videos pose a potentially higher challenge [16].

Contribution in this paper

- This paper proposed deep learning framework is to utilize the dissimilarity between the frames on the video and to extract the meta-information from the video containers.
- This paper utilized automated model for extraction of parameters such as motion features and Optical Flow features maintains consistency in finding the tamper points.
- Deep learning model utilized mobile net architecture for automating the stuffs stated in the above points

The organization of the paper is as follows: Section II clearly details the state of the art techniques available in the literature. Further this section elaborates the motivation for this paper. Section III discusses about the

proposed deep learning framework. Section IV showcases the experimental test bed used for experimenting the proposed deep learning framework. Section V exhibits the results obtained during the experimentation and finally the paper is conclude in the section VI.

II. LITERATURE SURVEY

Due to the lack of a lot of research in the video forgery detection area as compared to the image forgery detection, majority of the forgery detection approaches for video are simply the algorithms used on an image applied to each and every frame of the video sequence [17-20]. However, a few types of forgery in video cannot be detected with this approach due to the lack of considerable relationship between the frames [21]. For example, simple duplication of frames cannot be detected since each frame portrays to be authentic and original when evaluated individually. Although with the lack of efficient forgery detection algorithms for videos, there have been several advancements by the years and the research is certainly growing fruitfully [22]. Very common methods available in the literature for the video forgery detection are listed below and reviewed.

Inconsistency based detection

This method effectively tries to detect forgery which actively includes interlacing of multiple videos to tamper with the original one. Several detection algorithms cannot detect such forgery since they operate on individual frames most of the time. This method proposed by [23] checks the consistency of de-interlacing parameters i.e. the fundamental parameters required to convert an interlaced (forged) video into a non-interlaced output. This consistency of such parameters lets us evaluate the authenticity of the video. This algorithm is based strongly on the predicament that interlaced videos have half the vertical resolution to that of the original video and thus the de-interlacing process encircles evaluation of insertion, deletion, duplication and interpolation of frames to create a full-resolution video output. The fundamental de-interlacing parameters are extracted simultaneously by using the EM method. The inter-frame motion is firmly related across several parameters in interlaced videos. Assessing the interference to this relationship brought about by tampering permits this framework i.e. this detection method to recognize forgery in interlaced videos.

Forgery detection with the correlation of noise

This method proves strongly efficient in detecting video tampering where similar regions of the video have been manipulated or if various synthesized objects were to be introduced to the original video. It has been explored in

detail by [24]. This suggests that the volume-level correspondence of the noise residue serves as the defining characteristic for detecting video fraud. On the off chance that a region is implanted by another region in the same video, it is observed that the correlation factor amongst those regions jumps up to an unusually high value, and interestingly, the noise residuals of the orchestrated (synthesized) textured region taken from another video display low consistency with noise outstanding as compared to the other regions. So, by comparing and carefully evaluating the noise residual values, such forgery can be detected. The drawback to this method is that, during noise intensity authentic and the forged videos are extremely different from one another, it fails to accurately reduce the noise and thus can miss a few regions of forgery. There are several noise reduction methods which help to extract the residual values, but because of the vast varied availability of different techniques, even the slightest change in these values can bring about vivid outputs. Thus, specialized equipment is a necessity too [25-27].

Probabilistic intensity similarity: CRF's (Camera response functions) are a key parameter to intelligently detect if a video has been tampered with. These values can be obtained by applying the temporal noise intensity distribution at each pixel, which has been studied in detail by [28]. It has been observed that CRF's are very helpful in detecting the regions of forgery as they exhibit characteristic values for different distribution of noise. To aid this, another method to estimate CRF's was proposed on the basis of probabilistic intensity match [31]. This is the measurement of the observed pixel values and the probability that any two pixel values originate from the same display brightness [43]. This measure gives a clear output on whether the neighboring pixels originate from the same source or not. Using this measure over a distributed are in the video sequence (frames) CRF's are calculated and thus detecting the forged regions of the video [29-32][37].

From the literature it is observed that the existing method focuses only on the parameters related to the images and frames. Extracting these parameters over consumes the CPU and takes lot of time for processing [33-35]. Further, existing approaches parameters such as motion features and Optical Flow features maintains consistency in finding the tamper points. However the processing time needs to be optimized. Further most of the optimal technique in the literature relies on the correlation part for feature comparison. Correlating all the features extracted in each frames again leads the process complex and over burden the CPU tasks [36][38]. For these problems to be solved, this paper proposes a novel deep learning framework which utilizes the dissimilarity

between the frames on the video and to extract the meta-information from the video containers. The proposed framework is high optimized to work with non-GPU environments also which in turn helps to run the model in any machine.

III. PROPOSED METHOD – DEEP LEARNING FRAMEWORK

Disturbances encode intermediate input and output as they reveal the model's ability to convert from lower-level concepts such as inner layer pixels to higher-level descriptions such as image types. [39-40]. MobileNet consists of two blocks. A Residual block with a stride of 1 and another block with stride of 2 for downsizing. Moreover, there are 3 layers for both the blocks.

Figure 1 shows the Mobilenet architecture. First Layer is 1×1 Convolution with ReLU6. The second layer is equipped with depth wise convolution. The third layer is a 1×1 convolution without any non-linearity. Every input has attached with an expansion factor (t) and $t = 6$ (default expansion factor) in all major operations.

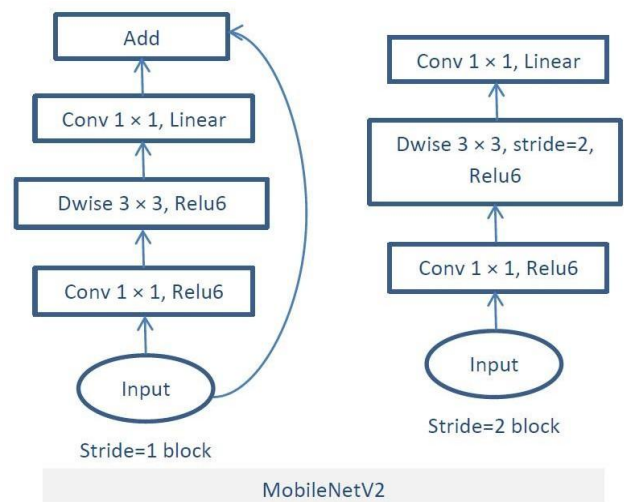


Figure 1 - Mobilenet architecture

If the input has 32 channels, then the internet output would get $32 \times 6 = 192$ channels. In the primary network (width multiplier 1, 224×224), has a 300 million multiply adds as its computational cost and uses 3.4 million parameters.

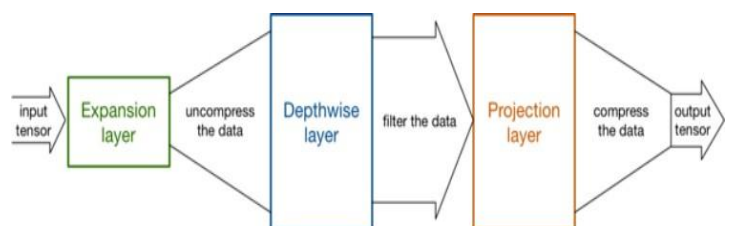


Figure2 - Mobilenet layered architecture.

Figure 2 shows the layered architecture of Mobilenet. The resolutions of the input image can go from 96 to 224 and width multipliers of 0.3 to 1.4. In total, the model consists of 3xconv2d, 7xbottleneck, 1 average pooling system layers. The point wise convolution makes the number of channels smaller. Also known as the projection layer, the data with the largest number of channels are plotted with the smallest number of dimensions in the tensor.

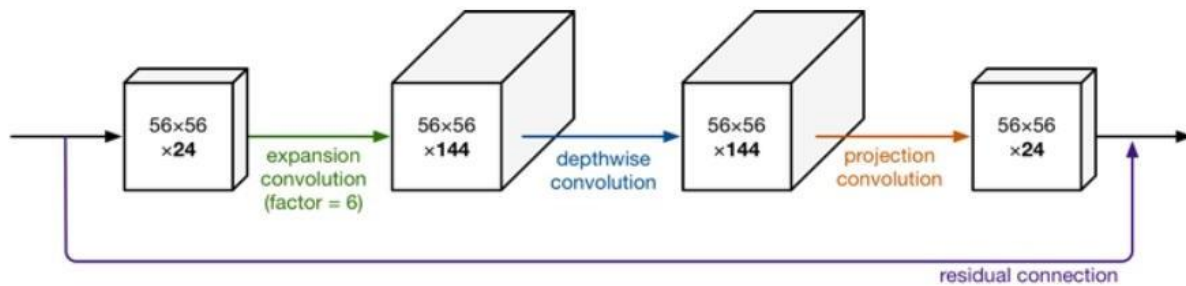


Figure3 - Image block processing

Mean Intersection over Union (mIOU) score is used to assess the semantic segmentation algorithms. Technically speaking this is a very basic evaluation carried out. mIOU is computed pixel wise, where true-positives belong to that class and are correctly predicted as class; false- negatives belong to that class, but are falsely predicted as different class and false. Positives belong to a different class but are predicted as class.

MobileNetV2 uses PASCAL VOC 2012 Semantic Segmentation as a feature extraction to obtain classes from test and train images. MobileNetV2 disables Atrous Spatial Pyramid Pooling (ASPP) as well as Multi-scaling And Flipping (MP), giving a mIOU of 75.32% with the terrific low model size and computational cost. Figure 3 shows the quick explanations of the later image size and channels.

The remaining connection facilitates the flow of gradients through the network. This helps shortcuts enable faster training and better accuracy. Each layer has module normalization and activation functionality - ReLU6, as mentioned earlier. However, the projection layer does not have an activation function unlike other models in the industry.

IV. EXPERIMENTAL SETUP

The experimental setup of the proposed framework is given in this section. All the experimentation is performed in a secure lab system running in windows 10, 64 GB RAM, Intel Xeon processor with Titan X Pascal GPU card. Adobe After effects is used to forge the videos. A few alterations are performed on the videos of set 100. Each set can hold upto 42frames/second. The experimental validation is carried out in two different

The expansion layer has more output channels than the input channels. It acts as a decompression to restore the data to its full potential, then performs layer filtering and scheduling and then converts the data back into smaller ones. Input and output modules are low- dimensional tensors, while the filtration step inside the in-block is done in high-dimensional tensors.

datasets. The experimental validation is carried out on an average of 100 videos of 360p video format. We used our own dataset (self-created), which we collected from the internet. The dimension of the videos in both the datasets are of size 320 × 240.

V. RESULTS AND DISCUSSION

Figure 4 shows the video processing on the two different dataset. Figure 4 also shows the two different variant of the extracted in which left side frames are original and whereas right side frames are tampered. From the Figure 5 it is inferred that the apparent motion of the individual pixel can give good approximation on true moments projected onto the image plane. So, the difference in motion variation between the first and the consecutive frames will be large and effective in traffic video-based applications. Furthermore from the Figure 5, it is also proven that the exact pixel position could be easily identified in the forged video.





Figure 4 - Normal image extracted from video frame vs. the tampered frame

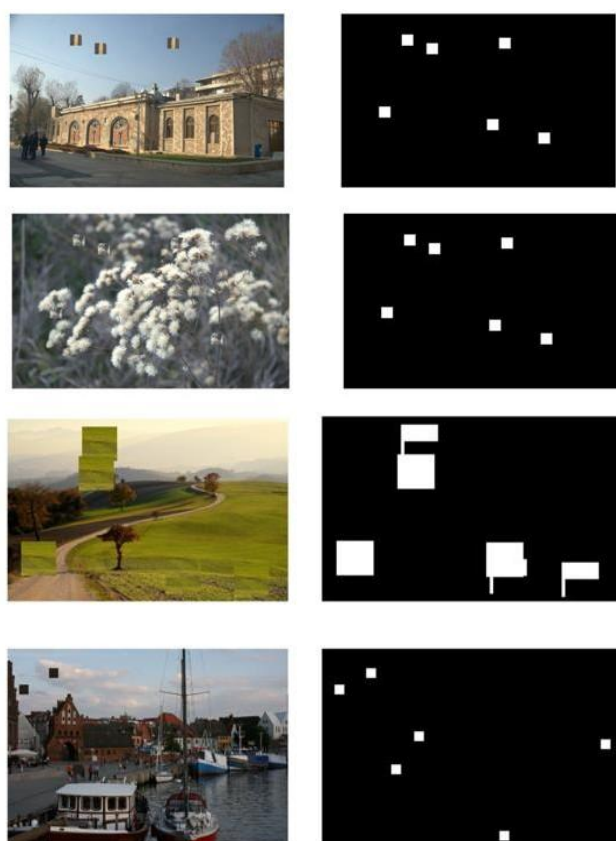


Figure 5 - Tampered frame detected by the proposed deep learning framework

VI. CONCLUSIONS

This paper is concluded by proposing a novel deep learning framework. The proposed deep learning algorithm achieves better results when compared to the state of the art methods. The key idea of the proposed deep learning model is to effectively use video meta-information feature video container. Further the training characteristic was also an important feature which validates the entire system to avoid over fitting problem. The challenge involved in using the proposed method is that fixation of total epochs for training and validation.

Further monitoring of flow level in pixels leads the pathway to detect the tampered contents from the video at ease.

REFERENCES

1. S. Vatansever, A. E. Dirik and N. Memon, "Analysis of Rolling Shutter Effect on ENF-Based Video Forensics," in IEEE Transactions on Information Forensics and Security, vol. 14, no. 9, pp. 2262- 2275, Sept. 2019.
2. S. Chen, A. Pande, K. Zeng and P. Mohapatra, "Live Video Forensics: Source Identification in Lossy Wireless Networks," in IEEE Transactions on Information Forensics and Security, vol. 10, no. 1, pp. 28-39, Jan. 2015.
3. M. Iuliani, D. Shullani, M. Fontani, S. Meucci and A. Piva, "A Video Forensic Framework for the Unsupervised Analysis of MP4-Like File Container," in IEEE Transactions on Information Forensics and Security, vol. 14, no. 3, pp. 635-645, March 2019.
4. V Mohanraj, S. SibiChakkaravarthy , I Gogul, V Sathiesh Kumar, Ranajit Kumar, V Vaidehi ; "Hybrid Feature Descriptors to Detect face Spoof Attacks", Journal of Intelligent & Fuzzy Systems, vol. 34, no. 3, pp. 1411-1419, 2018, IOS press.
5. M. C. Stamm, W. S. Lin and K. J. R. Liu, "Temporal Forensics and Anti-Forensics for Motion Compensated Video," in IEEE Transactions on Information Forensics and Security, vol. 7, no. 4, pp. 1315-1329, Aug. 2012.
6. S. Chen, S. Tan, B. Li and J. Huang, "Automatic Detection of Object-Based Forgery in Advanced Video," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 26, no. 11, pp. 2138- 2151, Nov. 2016.
7. V. Amanipour and S. Ghaemmaghami, "Median Filtering Forensics in Compressed Video," in IEEE Signal Processing Letters, vol. 26, no. 2, pp. 287-291, Feb. 2019.
8. U. Budhia, D. Kundur and T. Zourntos, "Digital Video Steganalysis Exploiting Statistical Visibility in the Temporal Domain," in IEEE Transactions on Information Forensics and Security, vol. 1, no. 4, pp. 502-516, Dec. 2006.
9. A. Hajj-Ahmad, A. Berkovich and M. Wu, "Exploiting Power Signatures for Camera Forensics," in
10. IEEE Signal Processing Letters, vol. 23, no. 5, pp. 713-717, May 2016.
11. S. SibiChakkaravarthy , D. Sangeetha, M.VenkataRathnam, K.Srinithi, V. Vaidehi;

- "Futuristic cyber- attacks", International Journal of Knowledge based and Intelligent System Engineering, Vol.22, no.3, pp. 105- 204, 2018. IOS press.
12. A. L. Sandoval Orozco, C. QuintoHuamán, J. A. Cifuentes Quintero and L. J. GarcíaVillalba, "Digital Video Source Acquisition Forgery Technique Based on Pattern Sensor Noise Extraction," in IEEE Access, vol. 7, pp. 157363-157373, 2019.
 13. W. Chuang, R. Garg and M. Wu, "Anti-Forensics and Countermeasures of Electrical Network Frequency Analysis," in IEEE Transactions on Information Forensics and Security, vol. 8, no. 12, pp. 2073-2088, Dec. 2013.
 14. F. Ma, X. Jing, X. Zhu, Z. Tang and Z. Peng, "True-Color and Grayscale Video Person Re-Identification," in IEEE Transactions on Information Forensics and Security, vol. 15, pp. 115-129, 2020.
 15. X. Liang, Z. Li, Y. Yang, Z. Zhang and Y. Zhang, "Detection of Double Compression for HEVC Videos With Fake Bitrate," in IEEE Access, vol. 6, pp. 53243-53253, 2018.
 16. M. M. Esmaili, M. Fatourehchi and R. K. Ward, "A Robust and Fast Video Copy Detection System Using Content-Based Fingerprinting," in IEEE Transactions on Information Forensics and Security, vol. 6, no. 1, pp. 213-226, March 2011.
 17. X. Jiang, P. He, T. Sun, F. Xie and S. Wang, "Detection of Double Compression With the Same Coding Parameters Based on Quality Degradation Mechanism Analysis," in IEEE Transactions on Information Forensics and Security, vol. 13, no. 1, pp. 170-185, Jan. 2018.
 18. Y. Zheng, Y. Cao and C. Chang, "A PUF-Based Data-Device Hash for Tampered Image Detection and Source Camera Identification," in IEEE Transactions on Information Forensics and Security, vol. 15, pp. 620-634, 2020.
 19. X. Zhu and C. W. Chen, "A Joint Source-Channel Adaptive Scheme for Wireless H.264/AVC Video Authentication," in IEEE Transactions on Information Forensics and Security, vol. 11, no. 1, pp. 141- 153, Jan. 2016.
 20. A. Hajj-Ahmad, S. Baudry, B. Chupeau, G. Doërr and M. Wu, "Flicker Forensics for Camcorder Piracy," in IEEE Transactions on Information Forensics and Security, vol. 12, no. 1, pp. 89-100, Jan. 2017.
 21. X. Ma, W. K. Zeng, L. T. Yang, D. Zou and H. Jin, "Lossless ROI Privacy Protection of H.264/AVC Compressed Surveillance Videos," in IEEE Transactions on Emerging Topics in Computing, vol. 4, no. 3, pp. 349-362, July-Sept. 2016.
 22. X. Ma, W. K. Zeng, L. T. Yang, D. Zou and H. Jin, "Lossless ROI Privacy Protection of H.264/AVC Compressed Surveillance Videos," in IEEE Transactions on Emerging Topics in Computing, vol. 4, no. 3, pp. 349-362, July-Sept. 2016.
 23. JoshanAthaneious, S. SibiChakkaravarthy , S. Vasuhi and V. Vaidehi, "Trajectory based Abnormal Event Detection in Video Traffic Surveillance using General Potential Data field with Spectral clustering", Multimedia Tools and Applications, Vol.78, Issue. 14, July 2019, pp. 19877 - 19903, Springer.
 24. B. C. Hosler, X. Zhao, O. Mayer, C. Chen, J. A. Shackelford and M. C. Stamm, "The Video Authentication and Camera Identification Database: A New Database for Video Forensics," in IEEE Access, vol. 7, pp. 76937-76948, 2019.
 25. L. M. Mathew, Suma R and J. J. Kizhakkethottam, "A survey on different video restoration techniques," 2015 International Conference on Soft-Computing and Networks Security (ICSNS), Coimbatore, 2015, pp. 1-3.
 26. M. C. Stamm and K. J. R. Liu, "Anti-forensics for frame deletion/addition in MPEG video," 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Prague, 2011, pp. 1876-1879.
 27. S. Verde, L. Bondi, P. Bestagini, S. Milani, G. Calvagno and S. Tubaro, "Video Codec Forensics Based on Convolutional Neural Networks," 2018 25th IEEE International Conference on Image Processing (ICIP), Athens, 2018, pp. 530-534.
 28. Yan Zhou, Fan-ZhiZeng and Guang-Fa Yang, "The research for tamper forensics on MPEG-2 video based on compressed sensing," 2012 International Conference on Machine Learning and Cybernetics, Xian, 2012, pp. 1080-1084.
 29. Yu-Ming Liang, "Video condensation for video forensics," 2012 IEEE International Conference on Computational Intelligence and Cybernetics (CyberneticsCom), Bali, 2012, pp. 180-184.
 30. S. SibiChakkaravarthy, D. Sangeetha and V. Vaidehi, "A Survey on malware analysis and mitigation techniques", Computer Science Review, Vol. 32, pp 1 - 23, May 2019, Elsevier.
 31. M. F. E. M. Senan, S. N. H. S. Abdullah, W. M. Kharudin and N. A. M. Saupi, "CCTV quality assessment for forensics facial recognition analysis," 2017 7th International Conference on Cloud Computing, Data Science & Engineering - Confluence, Noida, 2017, pp. 649-655.

32. N. Khanna and E. J. Delp, "Source scanner identification for scanned documents," 2009 First IEEE International Workshop on Information Forensics and Security (WIFS), London, 2009, pp. 166-170.
33. S. K. Yarlagadda et al., "Shadow Removal Detection and Localization for Forensics Analysis," ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, United Kingdom, 2019, pp. 2677-2681.
34. S. D. Lin, C. Chuang, M. Chen and H. Meng, "A Novel Video Watermarking Scheme in H.264/AVC Encoder," 2009 Fourth International Conference on Innovative Computing, Information and Control (ICICIC), Kaohsiung, 2009, pp. 357-360.
35. V. Joshi and S. Jain, "Tampering detection in digital video - a review of temporal fingerprints based techniques," 2015 2nd International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, 2015, pp. 1121-1124.
36. S. Lameri, L. Bondi, P. Bestagin and S. Tubaro, "Near-duplicate video detection exploiting noise residual traces," 2017 IEEE International Conference on Image Processing (ICIP), Beijing, 2017, pp. 1497-1501.
37. C. Hsieh, C. Chiu and P. Su, "Video Forensics for Detecting Shot Manipulation Using the Information of Deblocking Filtering," 2018 IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC), Tokyo, 2018, pp. 353-358.
38. D. Coppi, S. Calderara and R. Cucchiara, "Iterative active querying for surveillance data retrieval in crime detection and forensics," 4th International Conference on Imaging for Crime Detection and Prevention 2011 (ICDP 2011), London, 2011, pp. 1-6.
39. Y. Kakde, P. Gonnade and P. Dahiwal, "Audio-video steganography," 2015 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), Coimbatore, 2015, pp. 1-6.
40. S. Safie, A. A. Samah, G. Sulong, H. A. Majid, R. Muhammad and H. Hasan, "Block matching algorithm for moving object detection in video forensic," 2017 6th ICT International Student Project Conference (ICT-ISPC), Skudai, 2017, pp. 1-5.
41. R. D. Singh and N. Aggarwal, "Detection of re-compression, transcoding and frame-deletion for digital video authentication," 2015 2nd International Conference on Recent Advances in Engineering & Computational Sciences (RAECS), Chandigarh, 2015, pp. 1-6.
42. S. Law and N. Law, "PRNU-based source identification for network video surveillance system," 2018 IEEE International Conference on Consumer Electronics (ICCE), Las Vegas, NV, 2018, pp. 1-2.
43. S. Lameriet et al., "Who is my parent? Reconstructing video sequences from partially matching shots," 2014 IEEE International Conference on Image Processing (ICIP), Paris, 2014, pp. 5342-5346.
44. G. Mathew, "Architectural considerations for highly scalable computing to support on-demand video analytics," 2017 IEEE International Conference on Big Data (Big Data), Boston, MA, 2017, pp. 1646-1649.