WJCMS

Wasit Journal of Computer and Mathematics Science

Journal Homepage: https://wjcm.uowasit.edu.iq/index.php/WJCM

e-ISSN: 2788-5879 p-ISSN: 2788-5879

Automatic Detection of Object-Based Video Forgery Using Various Groups of Pictures (GOP)

Alex Khang[1,*] and Neyara Radwan[2]

[1]Global Research Institute of Technology and Engineering, USA

[2]Industrial Engineering Dept., College of Applied Sciences, Al Maarefa University, Saudi Arabia

*Corresponding Author: Alex Khan

DOI: https://doi.org/10.31185/wjcms.234

Received: September 2023; Accepted: November 2023; Available online: December 2023

ABSTRACT: In recent years, there has been a lot of interest in detecting object-based video forgeries. There has been a lack of satisfactory performance with object-based forgery detectors until recently since a majority of them are still based on handcrafted features. There has been a great deal of interest in passive video forensics in recent years. Forgery of video encoded with advanced codec frameworks remains one of the biggest challenges in object-based forgery research. An object-based forgery detection approach is presented in this paper. To evaluate the proposed method, a derived test dataset of variable video lengths and frame sizes is also used in addition to the SYSU-OBJFORG dataset. This process's efficacy is verified by comparing its results with other methods. When tested on datasets with degraded-quality videos, the proposed framework performed better in real-life scenarios.

1. INTRODUCTION

The camera generates a video file with a specific extension based on the container, which contains metadata about the file's structure. Quality is determined mainly by the codec, which encodes the video and is the main component of the content. Many codecs can be stored in a popular container, for example. H.264 video file in MOV format. A video is a sequence of frames organized by a group of pictures (GOP). In a three-dimensional plane, it is a series of consecutive frames or GOPs with a temporal dependency. An action composed of multiple frames is referred to as a shot. A scene refers to one or more shots forming a coherent whole. The sequence leaves a unique fingerprint when encoded, such as by quantization. Video forgeries fall into two subcategories: intraframe and interframe. Adobe Premiere Pro, Adobe Photoshop, and other video editors are effective tools for performing these forgeries. A digital video forgery is shown in Figure 1.

As passive techniques are not dependent on pre-embedded data, such as watermarks or signatures, it is still possible to check a video's integrity and authenticity without using pre-embedded data. The researcher faces a challenge when he is unaware of the embedded information in the video when using passive techniques [1]. In recent years, the scientific community has increasingly focused on passive techniques to detect video forgeries. The passive detection of digital video forgery creates a visual comparison of the original video with the tampered video after the forgery was detected. Passive techniques can also be called blind techniques since they rely on static and temporal artifacts that can be detected in a video to identify manipulated material. According to the features/artifacts used to detect passive video forgery, techniques for detecting it are categorized in Figure 2.

Various technologies are used to detect video forgeries.

- **Pristine frames:** Video frames that are not modified during compression..
- **Double compressed frames:** Stream frames that have been recompressed.

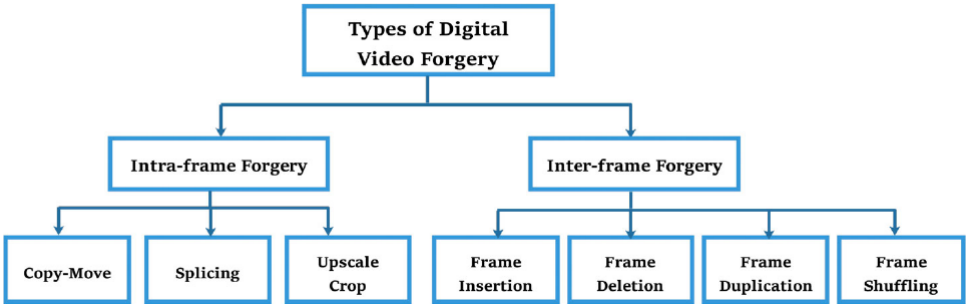

FIGURE 1. **Various types of digital video forgery.**

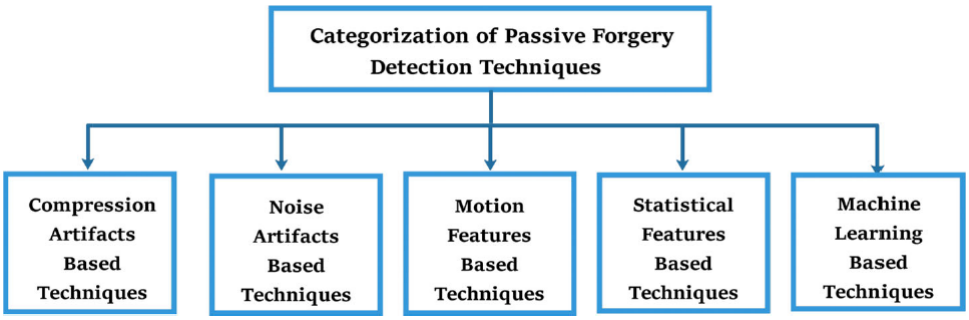

FIGURE 2. **Detection of passive video forgeries categorized**

- **Innocent double- compressed frames:** There is no forged content in those frames..
- **Forged frames:** A tampering operation has been performed on those frames.
- **I-frames (intra-coded frames):** GOPs begin with I-frames. Unlike video, it's an independently encoded still image that includes the entire image..
- **P-frames (predictive-coded frames):** A P-frame is a frame in which the difference between the preceding frames is motion-compensated.

The trustworthiness of digital media is waning due to its easy alteration and manipulation. A digital authenticator is necessary to restore original content. A method for categorizing an object based on its attributes is developed for identifying forged content. A method for automatically detecting object-based forgeries is developed using its GOP structure.

2. LITERATURE REVIEW

Author [2] uses AWOB to account for variable widths of neighbouring areas to detect video forgery. An image edge is exploited locally using the Gaussian distribution model with non-subsampled contourlet coefficients. Rayleigh distributions are used to analyze edge intensities. In Libsvm, the SVM classifier [3] is based on the library's RBF and distinguishes natural objects from forged objects. It is necessary to use trained sample databases as a drawback. There is still a need for standard video forensic datasets. An author has proposed a spatiotemporal scheme to extract moving objects from video sequences [4]. A block-based motion detector is used to identify changed regions by localizing moving objects during temporal segmentation [5]. It is taken into account how objects move slowly and quickly. Once the temporal segmentation is complete, the watershed algorithm is used for spatial segmentation. It is then possible to detect moving objects by combining these segmentations with a fixed threshold. When the video sequence is short, the results are not as impressive. In the future, it will be necessary to address the problem of identifying moving objects with global motion estimation.

The MPEG-1, MPEG-2, MPEG-3, MPEG-4, and H-264 coding standards are generally used to compress digital videos for improved storage and transmission [6]. Video forgeries are detected using compression artifacts or coding clues acquired during compression. Figure 3 illustrates how video forgeries are detected based on compression artifacts.

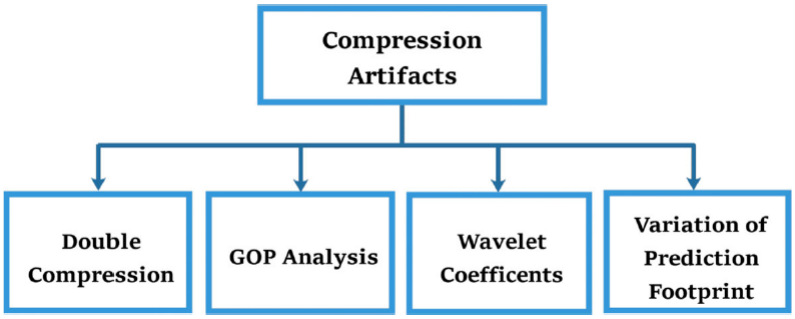

FIGURE 3. Video forgery detection artifacts caused by compression

Using the histogram of oriented gradients (HOG) feature, [7] devises an algorithm that combines oriented gradients (OG) and histograms (HOG) [8]. Feature extraction from blocks and comparison with other blocks is used to perform temporal and spatial tempering analysis. Copy-paste tempering is the only thing detected by the algorithm. They detect temper by looking for statistical or mathematical correlations that tend to be peculiar to certain devise manufacturers. Detection techniques within this group are very similar to source identification techniques.

Several methods have been proposed for detecting image-forged documents; however, very few have been proposed for detecting video-forged documents. The following algorithms are the most promising in detecting image copy-paste forgeries [9–11]. SIFTS (scale-invariant feature transforms) are the techniques proposed in [9], [10]. The author of [5] achieves robustness against geometric transformations using the Fourier-Mellin Transform (FMT). In addition to these techniques, you can find others in [12]. It is possible to detect spatial copy-paste forgeries using techniques [9]– [11], but extending them to detect temporal copy-paste forgeries is very difficult. Although the copies of the tampered region come from different sources, the algorithms [9]– [11] assume that the tampered region is at a different spatial location. Timoral tampering, however, places copy-paste regions close to one another.

Spatial and temporal tampering are the two types of copy-paste video forgeries. Copying and pasting regions from one frame to another may be considered spatial tampering. For example, an enlarged version of Figure 4a has been pasted into Figure 4d. It is possible to duplicate an entire frame while temporal tampering takes place. The same spatial location can be duplicated across frames as well. Since objects and actions may occur in multiple frames simultaneously, temporal tampering becomes necessary for convincing forgeries. Many frames have been pasted across the enlarged trash bin in Figures 4-d to 4-f. The assumption makes sense since a frame or region must be duplicated within a GOP (Group of Pictures) to detect temporal tampering. As well as video regions, forged regions may come from other sources. Pasting other objects or backgrounds in a scene can hide undesirable events and fabricate evidence in both spatial and temporal tampering. However, in Figure 4-c, it appears the person keeps the bag, but in Figure 3-f, it is unclear whether they are keeping or taking it.

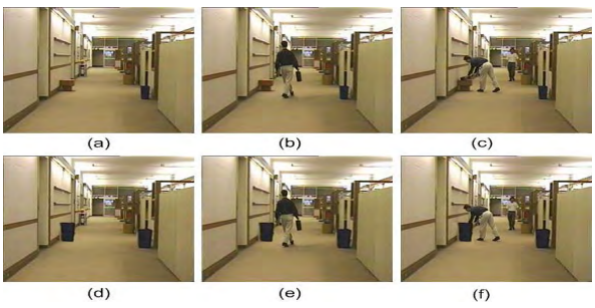

FIGURE 4. (a - c) Original video sequence (d - f) Forged video sequence.

3. PROPOSED METHODOLOGY

An object-based forgery is a technique for altering video scenes by adding or removing objects. Object-based forgery is a common way to tamper with videos due to its critical role in conveying content. Compressed videos are generally

available. Forgers first decompress video frames into individual still images before performing object-based forgeries. After the segment has been tampered with, the rest of the sequence is left intact. The manipulations are recompressed after they are complete to create forged versions of the frame sequence.

Two types of video attributes can be evaluated based on their inherent statistical properties: intra-frame attributes describing how the video looks spatially and inter-frame attributes describing how it looks temporally. Video frames in a local temporal window correlate strongly with neighbouring frames based on their motion and static nature. Static anchor frames are used for local temporal windows, whereas motion anchor frames represent movement residuals above them. Motion residuals, the main source of visual information in video frames, contain a significant portion of their intra-frame properties. There are two kinds of motion residuals: those that maintain the intrinsic characteristics of the corresponding frame and those that describe the temporal change from the fundamental anchor frame to the corresponding frame. Since motion residuals contain intraframe and interframe inherent properties, we use them as our primary analysis object. The GOPs in a video stream have a local window associated with them or a temporal window. Pframes/Bframes into a GOP correspond to motion residuals from corresponding I-frames in that GOP. Our collision operator in advanced video frameworks relies on a fixed local temporal window structure rather than GOPs, which are flexible and cannot guarantee a fixed local temporal window. Sequences of video frames that have not been compressed are denoted by

$$V \triangleq \left(F^1,\ F^2, \ldots.,\ F^N\right\} N \in Z \tag{1}$$

An 8-bit grayscale still image of $n_1 \times n_2$ is the k^{th} decompressed video frame $F^k = \left(F^k_{i,j}\right) \in (0,\ \ldots,\ 255\}^{n_1 \times n_2}$. In a video F^k frame sequence, it is defined as follows: a window with a size of$L = 2\times L_h+1(L_h\ is\ the\ number\ of\ the \frac{left}{right} neighbors\ of\ F^k)$, is performed:

$$C^k = C^k_{i,j} = H\left(\left(F^{k-L_h}_{i,j}\right), \ldots\ldots,\ \left(F^k_{i,j}\right), \ldots..,\ \left(F^{k+L_h}_{i,j}\right)\right] \tag{2}$$

When F^k and H are collided, C^k is the result. As a result of the collision operator H, each frame in the temporal window is grouped in the corresponding coordinates by a convolution function that generates the $C^k_{i,j}$ by grouping the pixels according to each frame's coordinates. As a result of F^k, a motion residual is defined as follows:

$$R^k = \left(F^k - C^k\right| = (R^k_{i,j}) = \left(\left|F^k_{i,j} - C^k_{i,j}\right|\right) \tag{3}$$

Amount in absolute terms is (·| denoted by. H_{MIN} represents minimum collusion in Eq (4), where collusion is defined as a function of the collusion operator.

$$H_{MIN} \triangleq \underset{l\in(-L_h,L_h]}{min} \left(F^{k+l}_{i,j}\right) \tag{4}$$

Consequently, the result $R^k$is a grayscale still image that has 8 bits.

It is evident from Eq. (3) that object-based video forgery becomes a roach when motion residual is introduced. Motion residuals are modified by changing the pixel values. The tampering with motion residuals of a video object can thus be considered object-based video forgery. Modeling an untouched video's intra-frame and inter-frame properties is key to detecting object-based forgeries in advanced video systems. There are several ways to represent a grayscale still image ($R^k_{i,j}$). In image forensics, the goal is to detect manipulations and processing of images; therefore, motion residuals are a useful tool for detecting tampering. According to recent research, state-of-the-art image tampering/processing features can detect data hiding/steganography as part of a state-of-the-art image steganalysis feature [13]. We can model object-based forgery using motion residuals by borrowing powerful statistical features from image steganalysis. As the I-frames themselves or motion vectors in P-frames and B-frames are compressed using frequency-domain lossy compression schemes, it is possible to use frequency-domain oriented feature sets to model the inherent properties of intra-frame and inter-frame frames in video streams. A frequency-domain image steganalytic feature set incorporating the CC-PEV frequency-domain image steganalytic feature set and extracting it from motion residuals has been developed [14], incorporating the intra-frame and inter-frame inherent properties contained in the motion residuals. We use the ensemble classifier described to construct the frame manipulation detector [15].

As P-frames and B-frames strongly correlate with I-frames, CC-PEV features do not have to be extracted from every frame in each GOP. The leading I frame is selected from each GOP, and several subframes are extracted following the I frame at equal intervals. All the frames in the GOP structure are classified by their classification results. Frames that represent a GOP structure are called represent frames. Motion residuals are calculated for each GOP structure based on a target video clip. The following steps extract CC-PEV feature vectors from motion residuals using the CC-PEV feature set.

A frame manipulation detector incorporates two ensemble classifiers as inputs. Double-compressed or pristine ensemble classifiers. Incoming frames are reviewed to determine whether they are "pristine" or "double compressed". Comparison of "pristine" and "dirty" outputs. "Double compressed" classifiers use simple majority algorithms to select "pristine" video clips. The term "pristine" refers to a video clip that contains more than fifty percent of the frames that do not appear to have been tampered with or if it appears to have been tampered with already. According to each feature vector, this ensemble classifier compares the "innocent double compressed" video clips to the suspicious ones. It determines whether or not an input frame is "innocent" or "forged" by using the "innocent double compressed" or "forged" frame classifier. An innocent double-compressed classifier defines a GOP structure as "forged" if all the I frames and P/B frames are deemed "forged" instead of "innocent double-compressed". It is suspicious if at least one "forged" GOP structure appears in an innocent double-compressed video clip; otherwise, it is considered forged. A GOP structure labelled "forged" can identify segments in those videos that have been forged." Two I frames and two P frames are derived from each GOP structure.

To generate a forged video:

1. Assume that each video frame is a still image by decompressing the video into individual frames.
2. The rest of the sequence remains unchanged while frames are manipulated in certain segments.
3. Frames are recompressed to create forged frames.

A GOP structure that is labelled "forged" by the "innocent double compressed" vs "forged" classifier is deemed "forged". Without a forged GOP structure, the suspicious video clip is considered innocent and double-compressed; otherwise, it is considered forged.

4. RESULT ANALYSIS & DISCUSSION

Currently, the largest dataset for video tampering is SYSU-OBJFORG, which we tested our algorithm and tool on. All 100 pairs of video sequences can be marked in 3.5 hours. Manually marking 5% or less of the bounding box is required. The computer-aided annotation algorithm generates all of the other bounding boxes. To fast annotate forged video sequences, researchers use this algorithm.

4.1 PERFORMANCE METRIC

Here, we use three types of video databases: basic, low resolution, and low bitrate. The "pristine" video clips comprise half of the selected video clips. Together with their forged counterparts, they make up the training set. Test video clips make up the remaining 50%. In all experiments, the average result is reported after 10 repetitions. Experiments are conducted using the following criteria, where Σ stands for the number of the set elements:

$$Pristine\ frame\ accuracy\ (PFACC): \frac{\sum Correctly\ classified\ pristine\ frames}{\sum Pristine\ frames}$$

$$Forged\ frame\ accuracy\ (FFACC): \frac{\sum Correctly\ classified\ forged\ frames}{\sum forged\ frames}$$

$$Double-compressed\ frame\ accuracy\ (DFACC): \frac{\sum Correctly\ classified\ Double-compressed\ \ frames}{\sum Double-compressed\ \ frames}$$

$$Frame\ accuracy\ (FACC): \frac{\sum Correctly\ classified\ frames}{\sum\ frames}$$

$$Video\ accuracy\ (VACC): \frac{\sum Correctly\ classified\ video\ clips}{\sum\ All\ the\ video\ clips}$$

$$Precision\ of\ forged\ segment\ localization\ (Precision): \frac{\sum Forged\ frames\ with\ correct\ labels}{\sum\ All\ the\ frames\ labeled\ as\ "forged"}$$

$$Recall\ of\ forged\ segment\ localization\ (Recall): \frac{\sum Forged\ frames\ with\ correct\ labels}{\sum\ All\ the\ forged}$$

$$Balanced\ F-score\ segment\ localization\ (F-score):\ 2\times\frac{precision\times Recall}{precision+Recall}$$

Except for VACC, all other criteria are applied to frames. The proposed frame manipulation detector is evaluated using PFACC, FFACC, DFACC and FACC; in contrast, Precision, Recall, and F1 scores are used to evaluate the proposed algorithm for localizing forged segments. Despite the similar formulas, one significant difference exists between Recall and FFACC. Several frames may be relabelled thanks to the segment localization process in suspiciously forged videos. Frame manipulation detectors calculate FFACC differently than Recall and Precision when they classify frames as "forged".

4.2 EXPERIMENTAL RESULTS

Our dataset encodes 10 original and 10 forged videos in H.264 format. From these videos, we extracted 154 GOPs for both original and forged videos. We used 84 untampered and 84 forged GOPs as training sets. The remaining 70 GOPs were used as testing sets. Based on the algorithm proposed, we can observe the difference within each frame in Figure 5. To make the data set homogeneous, we collect the forged and pristine sequences corresponding to the forged video and the same number of double-compressed frames. In the next step, processed frames will be used as motion residues derived from the proposed algorithm.

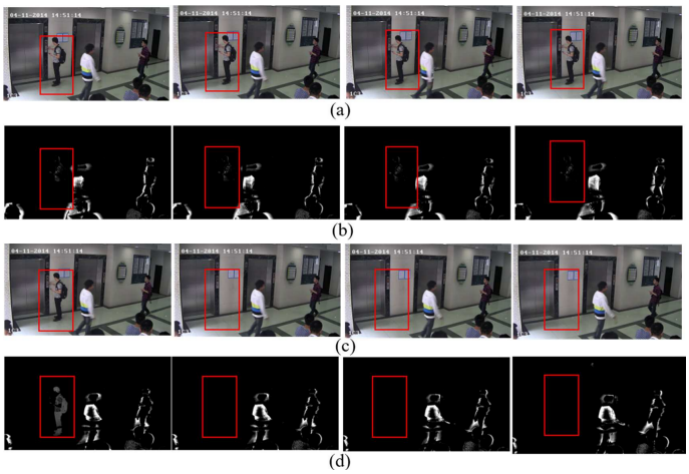

FIGURE 5. (a) Pristine frames, (b) Pristine frames with motion residue extracted, (c) This video has been forged, and these are some of its frames, (d) An extract of the forged video's motion residue.

We model the intraframe and interframe inherent properties of the motion residuals using seven state-of-the-art image steganography features, CCPEV [14], [16], SPAM [17], CDF [18], CF [15], SRM [19] and CC-JRM [20]. Figure 6 shows the graphical representation of identification score with respect to the steganography features. An algorithm for accurate positioning contributes less to the advanced steganalytic feature set. Steganalytic feature sets with higher performance can better classify frames around forged segment boundaries, leaving less promotion space for accurate positioning algorithms. While accurate positioning algorithms are more useful for locating forged segments in high-resolution videos, they also perform well on low-quality videos as show in the Figure 7. Our method also performed better than existing image steganalytic features, as shown in Table 1.

5. CONCLUSION

Digital video authenticity has become a significant concern in recent years due to the availability of video editing gadgets and applications. Digital multimedia technology is rapidly advancing, increasing the importance of multimedia data, such as digital images/videos, in diverse applications. Several civil/criminal trials nowadays generate false/unlawful evidence by using digital video forgeries based on interframes and intraframes. The frames or objects within a video can be reproduced, removed, inserted or replaced. Our approach uses advanced frameworks to detect the automatic forgery of video objects. A similarity analysis was conducted between object-based video forgery and steganography to detect hidden data in motion residuals corresponding to object-based video forgery. Our training and testing datasets of five videos each achieved an accuracy of 99.99%.

Table 1. Comparison of the experimental resultswith various image steganalytic features for 1280 × 720 videos.

	Identification					Forged segment localization		
	PFACC	FFACC	DFACC	FACC	VACC	Precision	Recall	F1-Score
CC-PEV	0.999	0.839	0.952	0.957	0.998	0.904	0.918	0.911
SPAM	0.997	0.768	0.890	0.924	0.990	0.789	0.830	0.809
CF	0.995	0.775	0.936	0.941	0.995	0.870	0.858	0.864
CDF	0.999	0.840	0.956	0.958	0.997	0.902	0.910	0.906
SRM	0.999	0.764	0.932	0.937	0.984	0.831	0.826	0.828
CC-JRM	0.999	0.843	0.978	0.965	0.999	0.931	0.915	0.923
J+SRM	0.999	0.849	0.975	0.965	1.00	0.928	0.915	0.921

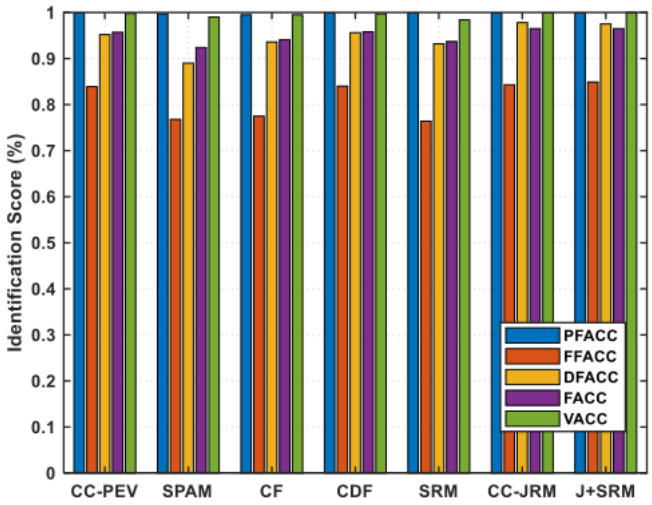

FIGURE 6. Identification Score (%)with various image steganalytic features.

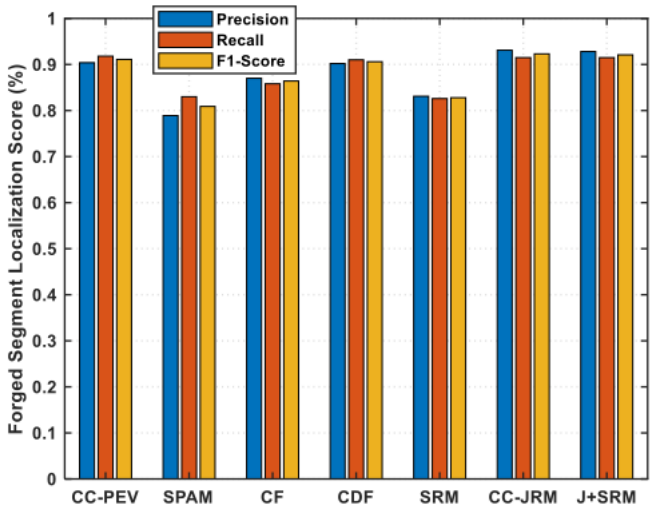

FIGURE 7. Forged Segment Localization Score (%)versus various image steganalytic features in term of the precision, recall and F1-Score.

FUNDING

None

ACKNOWLEDGEMENT

None

CONFLICTS OF INTEREST

The author declares no conflict of interest.

REFERENCES

[1] S. P. Yadav, M. Jindal, P. Rani, V. H. C. D. Albuquerque, C. D. S. Nascimento, and M. Kumar, "An improved deep learning-based optimal object detection system from images," *Multimed. Tools Appl*, 2023.
[2] R. Chen, Q. Dong, H. Ren, and J. Fu, "Video forgery detection based on non-subsampled contourlet transform and gradient information," *Inf. Technol. J*, vol. 11, no. 10, pp. 1456–1456, 2012.
[3] P. Rani and R. Sharma, "Intelligent transportation system for internet of vehicles based vehicular networks for smart cities," *Comput. Electr. Eng*, vol. 105, pp. 108543–108543, 2023.
[4] R. Li, S. Yu, and X. Yang, "Efficient spatio-temporal segmentation for extracting moving objects in video sequences," *IEEE Trans. Consum. Electron*, vol. 53, no. 3, pp. 1161–1167, 2007.
[5] G. Ansari, P. Rani, and V. Kumar, "A Novel Technique of Mixed Gas Identification Based on the Group Method of Data Handling (GMDH) on Time-Dependent MOX Gas Sensor Data," in *Proceedings of International Conference on Recent Trends in Computing* (R. P. Mahapatra, S. K. Peddoju, S. Roy, , and P. Parwekar, eds.), vol. 600, pp. 641–654, Springer Nature, 2023.
[6] P. Rani and R. Sharma, "An Experimental Study of IEEE 802.11n Devices for Vehicular Networks with Various Propagation Loss Models," in *Advanced IoT Sensors, Networks and Systems* (A. K. Dubey, V. Sugumaran, , and P. H. J. Chong, eds.), vol. 1027, pp. 125–135, Springer Nature, 2023.
[7] A. V. Subramanyam and S. Emmanuel, "Video forgery detection using HOG features and compression properties," *2012 IEEE 14th international workshop on multimedia signal processing (MMSP)*, pp. 89–94, 2012.
[8] J. Chao, X. Jiang, and T. Sun, "A novel video inter-frame forgery model detection scheme based on optical flow consistency," in *The International Workshop on Digital Forensics and Watermarking 2012: 11th International Workshop, IWDW 2012*, pp. 267–281, Springer, 2012.
[9] I. Amerini, L. Ballan, R. Caldelli, A. D. Bimbo, and G. Serra, "A sift-based forensic method for copy-move attack detection and transformation recovery," *IEEE Trans. Inf. Forensics Secur*, vol. 6, no. 3, pp. 1099–1110, 2011.
[10] X. Pan and S. Lyu, "Region duplication detection using image feature matching," *IEEE Trans. Inf. Forensics Secur*, vol. 5, no. 4, pp. 857–867, 2010.
[11] W. Li and N. Yu, "Rotation robust detection of copy-move forgery," *2010 IEEE International Conference on Image Processing*, pp. 2113–2116, 2010.
[12] T. V. Lanh, K. S. Chong, S. Emmanuel, and M. S. Kankanhalli, "A survey on digital camera image forensic methods," *2007 IEEE international conference on multimedia and expo*, pp. 16–19, 2007.
[13] X. Qiu, H. Li, W. Luo, and J. Huang, "A universal image forensic strategy based on steganalytic model," *Proceedings of the 2nd ACM workshop on information hiding and multimedia security*, pp. 165–170, 2014.
[14] J. Kodovský and J. Fridrich, "Calibration revisited," *Proceedings of the 11th ACM workshop on Multimedia and security*, pp. 63–74, 2009.
[15] J. Kodovsky, J. Fridrich, and V. Holub, "Ensemble classifiers for steganalysis of digital media," *IEEE Trans. Inf. Forensics Secur*, vol. 7, no. 2, pp. 432–444, 2011.
[16] T. Pevny and J. Fridrich, "Merging Markov and DCT features for multi-class JPEG steganalysis," *Security, steganography, and watermarking of multimedia contents IX, SPIE*, pp. 28–40, 2007.
[17] T. Pevný, P. Bas, and J. Fridrich, "Steganalysis by subtractive pixel adjacency matrix," *Proceedings of the 11th ACM workshop on Multimedia and security*, pp. 75–84, 2009.
[18] J. Kodovský, T. Pevný, and J. Fridrich, "Modern steganalysis can detect YASS," *Media Forensics and Security II, SPIE*, pp. 9–19, 2010.
[19] J. Fridrich and J. Kodovsky, "Rich models for steganalysis of digital images," *IEEE Trans. Inf. Forensics Secur*, vol. 7, no. 3, pp. 868–882, 2012.
[20] J. Kodovský and J. Fridrich, "Steganalysis of JPEG images using rich models," *Media Watermarking, Security, and Forensics*, pp. 81–93, 2012.