Chapter

Evolution of Attacks on Intelligent Surveillance Systems and Effective Detection Techniques

Deeraj Nagothu, Nihal Poredi and Yu Chen

Abstract

Intelligent surveillance systems play an essential role in modern smart cities to enable situational awareness. As part of the critical infrastructure, surveillance systems are often targeted by attackers aiming to compromise the security and safety of smart cities. Manipulating the audio or video channels could create a false perception of captured events and bypass detection. This chapter presents an overview of the attack vectors designed to compromise intelligent surveillance systems and discusses existing detection techniques. With advanced machine learning (ML) models and computing resources, both attack vectors and detection techniques have evolved to use ML-based techniques more effectively, resulting in non-equilibrium dynamics. The current detection techniques vary from training a neural network to detect forgery artifacts to use the intrinsic and extrinsic environmental fingerprints for any manipulations. Therefore, studying the effectiveness of different detection techniques and their reliability against the defined attack vectors is a priority to secure the system and create a plan of action against potential threats.

Keywords: intelligent surveillance systems, internet of video things (IoVT), multimedia forgery, environmental fingerprints, forgery detection, DeepFake detection

1. Introduction

The modern smart city infrastructure has advanced by integrating multimedia-based information input and the development of an edge computing paradigm [1, 2]. An increase in visual and auditory input from the deployed sensors has enabled multiple network layer-based processing of incoming information. While most of the intelligent infrastructure depends on a cloud computing-based architecture [3], edge computing has been attracting more and more attention to meet the increasing challenges in terms of scalability, availability, and the requirements of instant, on-site decision making [4–6]. Advancements in artificial intelligence (AI) have equipped the edge computers to process the incoming multimedia feed and deploy recognition and detection software. Machine learning (ML)-based models such as object detection, tracking, speech recognition, and people identification are commonly deployed to enhance the security in infrastructure and private properties [7]. With an increase in such technological advancements, the system's reliability has also exponentially increased where the trust factor established on the system is directly depending on the information retrieved by the multimedia sensor nodes [8]. The edge devices are enhanced with multi-node communication and equipped with Internet connections to provide continuous functionality and security services.

Due to their significance in infrastructure security and functionality, edge computing devices are commonly targeted through networked attacks through Wi-Fi and RF links [9]. The devices are compromised through malicious firmware updates [10] and result in creating a backdoor with admin privileges. The perpetrators then control the device Input/Output (IO) and compromise the network and home security. Specifically, visual layer attacks are developed to manipulate the visual sensor in edge devices and create a false perception of live events monitored by the control station. Simple frame manipulation such as frame duplication or shuffling allows the perpetrator to mask the original frame, where the security of the infrastructure can be easily compromised [11]. There is also no evidence of crimes without the surveillance recordings, and it falters the need for such security devices. Along with the visual layer, the audio channel of the edge nodes is equally targeted. Modern home security is enabled with voice commands and a home assistant system that functions based on the voice commands received. The audio devices are equipped with voice-based home assistant computers and Voice Over IP (VoIP) surveillance recorders. The attackers can target the audio channel through hidden voice commands, control the system, or completely mask the audio channel with noise to disable its functionality [12].

As the ML-based models have enhanced the surveillance system's capabilities, it has also resulted in the development of frame manipulation attacks. Beginning with the traditional copy-move style forgery attacks in spatial regions of a frame, modern deep learning (DL) has enabled generative networks capable of creating a frame based on the user's input. Adversarial networks have rendered some ML models useless due to their targeted attack to disable their functionality. General adversarial networks (GANs) have created DeepFakes, which have become one of the most challenging problems in current multimedia forgery attacks [13]. DeepFake is trained to function in low computing systems such as edge devices and result in manipulations such as Face Swaps, Facial Re-enactments, and complete manipulation of the targeted person's movements resulting in a very realistic media output [14]. It is clear that both the visual and auditory channels require robust security measures and reliable authentication schemes to detect such malicious attacks and secure the network [15, 16].

Advancements in forgery attacks have always been countered with detection schemes. Traditional frame forgery attacks were first detected using watermark technology and compression artifacts [17]. However, when the edge device is compromised, the frames are manipulated at the source level, creating watermarks on false frames. Similarly, with DeepFake being developed, its counterpart detection schemes were also trained. The first stages of DeepFakes carry visual artifacts like face recordings without any eye blinking or face warping artifacts [18]. Still, with more training data and better networks, DeepFakes have evolved to a point where it is almost not distinguishable from real images [19]. Although the technology itself has its own merits when ethically used in the field of medical and entertainment, perpetrators can always use the DeepFake technology with malicious intent without a reliable detector. It is an ongoing effort to create a reliable detection scheme to clearly distinguish between real and fake.

This chapter provides an overview of the evolution of multimedia-based attacks to compromise the edge computing nodes such as surveillance systems and their counterpart forgery detection schemes. The essential features required by a reliable detection system are analyzed and a framework using an environmental fingerprint is introduced that has proven to be effective against such attacks.

2. An overview of audio-visual layer attacks

The networked edge devices are commonly deployed through Wi-Fi or RF links in a private network. The primary means of hijacking a secure device is through network layer attacks where the communication between the devices is intercepted and modified [20]. This allows the source and the destination to believe that the information exchange was secure, while a perpetrator alters the intercepted message as required. Malicious firmware is updated through direct physical access to the USB interface or remote web interface, which allows a perpetrator to gain admin privileges to the edge devices. Some devices are sold through legitimate channels with malicious firmware pre-installed [10]. With complete access to the visual and audio sensor nodes, the attacker can manipulate the media capturing module itself, making the network-level security measures compromised.

Surveillance systems are the most targeted edge devices due to their importance and access medium [11]. Network attacks like Denial of Service (DoS) can disable the network connections of the devices and negate their purposes. Common admin mistakes like using the default credentials on the networks and devices login are primary reasons for backdoor entry. Once the device or the network is compromised, the attacker typically encodes the trigger mechanism into the system. This allows the perpetrator to remotely trigger the selected attack based on remote commands without re-accessing the device. Malicious inputs can be encoded into the multimedia encoding scheme of the edge device. Trigger methods like QR-code-based input to the video recording interpret the command differently [21], face detection-based trigger [22], and hidden voice commands through the audio channels [12] are a few examples of how an attack can be remotely controlled. Wearable technologies like Google Glass are also affected through the backdoor firmware, where the QR-code-based input was used to hack the device [23].

With remote trigger mechanisms, a device can be controlled to manipulate the incoming media signal. Face detection software can be re-programmed to blur selected faces and car plate registrations or disable certain functionality like detecting prohibited items like guns [9]. Popular Xerox scanners and photocopiers were hacked to manipulate the contents of the documents that are scanned and insert random numbers instead of actual data [24]. Surveillance cameras with Pan-Tilt-Zoom (PTZ) capabilities can be controlled to re-position the cameras so that the number of blind spots is increased in a surveillance area [25]. Audio Event Detectors (AED) are commonly deployed in surveillance devices to raise the alarm based on suspicious audio activity or in-home assistant devices to detect the wake commands. Still, the AED system can be directly targeted using the hidden voice commands to interpret its input falsely [12, 26]. Using the adversarial networks, popular ML models on edge devices are targeted so that the input itself can be modified [27]. Frame-level pixel manipulations are made to confuse the ML models and result in the false categorization of object recognition models [28]. A wearable patch is trained to target the person identification ML model, which can be worn by a perpetrator in the form of a t-shirt and escape the identification module [29].

Access to the multimedia sensor nodes can result in many variants of visual and audio layer attacks. To study the effective detection methods, we first narrow the video frame manipulation and audio overlay attacks commonly designed to target the edge-based media input such as surveillance devices and online conferencing technologies.

2.1 Frame manipulation attacks

Video recordings used for temporal correlation of the live events are primarily targeted using frame shuffling or duplication attacks [30]. The perception of live events is affected, which disables the effectiveness of live monitoring [31]. Adaptive replay attacks are designed such that the frame duplication attack can adapt to the changes in the environments such as light intensity variations, object displacement, and camera alignments. With adjusting frame masking, the operator in the monitoring station cannot distinguish between the real and fake images since the duplicated frames are originally copied from the same source camera [22]. The effect of source device identification and watermarking technique is negated since the frames originated in the same camera. **Figure 1** represents a frame replay attack where the attack is triggered remotely by either a QR-code or face detection module, and the resulting frame is masked with a static background [21, 32].

Spatial manipulation of a frame includes changes to the pixels like object addition or deletion, while the static frame is maintained. Frame-level manipulations are commonly made to deceive the viewer with the presence of a subject [33, 34]. The figure shows the spatial manipulation of the video frame.

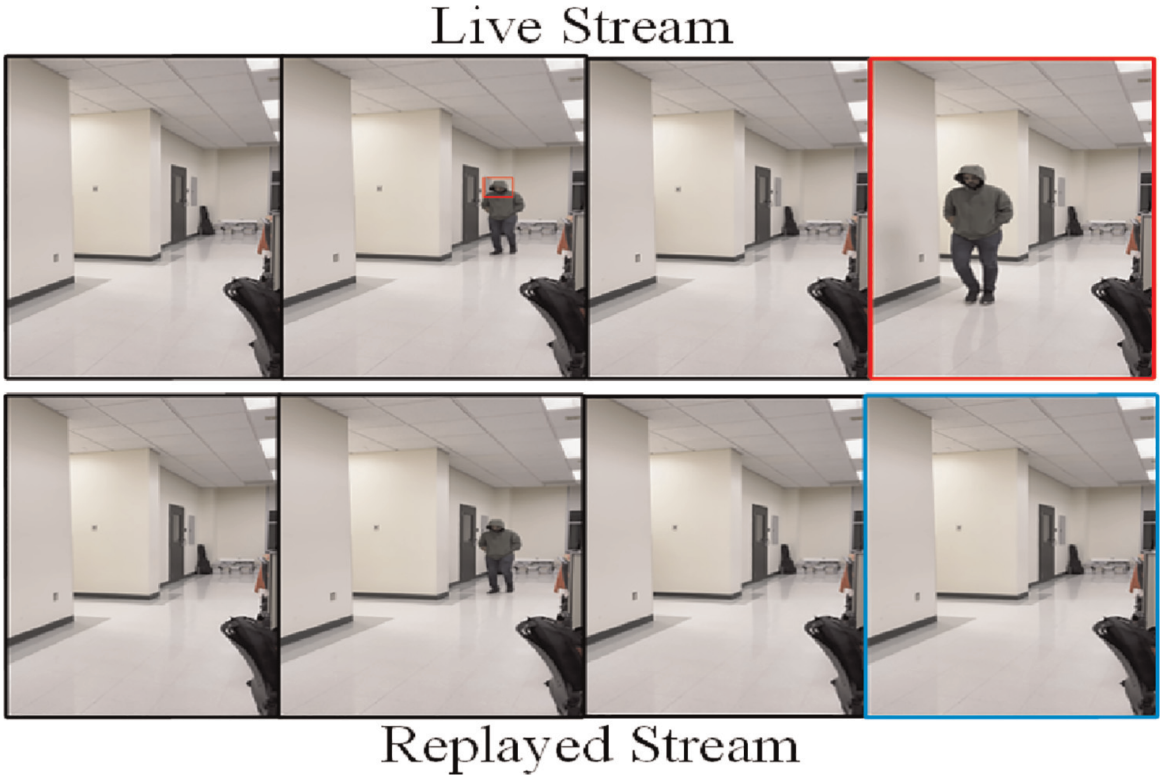

Figure 1.
Frame duplication attack to manipulate the perception of live events triggered by the perpetrator's face detection.

2.2 Audio masking and overlay

Most edge nodes are equipped with audio recording capabilities making them a target for forgery attacks [3]. Every household is equipped with surveillance cameras, home assistants, and edge devices capable of two-way communications. The AED module is responsible for wake command detection or event detection based on audio like gunshot sounds. The input audio sensor nodes are disabled by compromising the AED module by replacing the actual event with the quiet static noise. The input is also affected by adding additional white noise to disrupt the AED module [26].

2.3 DeepFake attack

DeepFake attacks developed using GAN architecture [13] have resulted in a large quantity of fake media generation. With enough training data available and the computation resources, the quality of the generated media keeps improving to a point where a person cannot distinguish between the real and fake media [35]. Although DeepFake technology has its application merits, any technology can cause more harm than good in the wrong hands. The developing software technologies have made it easier and more convenient for the generation of DeepFake media using their mobile phone.

A simple face manipulation software where two people can swap their facial landmarks originated in the form of mobile applications. Soon, advanced technologies were made to make the swap more realistic [36]. Many organizations and institutions rely on online conferencing solutions for their daily communications. Face-swapping technologies allow perpetrators to mimic a source facial landmark and duplicate their online personality [37]. However, with the capability to extract facial landmarks and skeletal features from a source subject, a new form of DeepFake emerged to project source movement on a targeted subject (**Figure 2**).

The facial re-enactment software [38] allows the model to extract the face landmark movements from a source subject. These landmarks are projected on a targeted

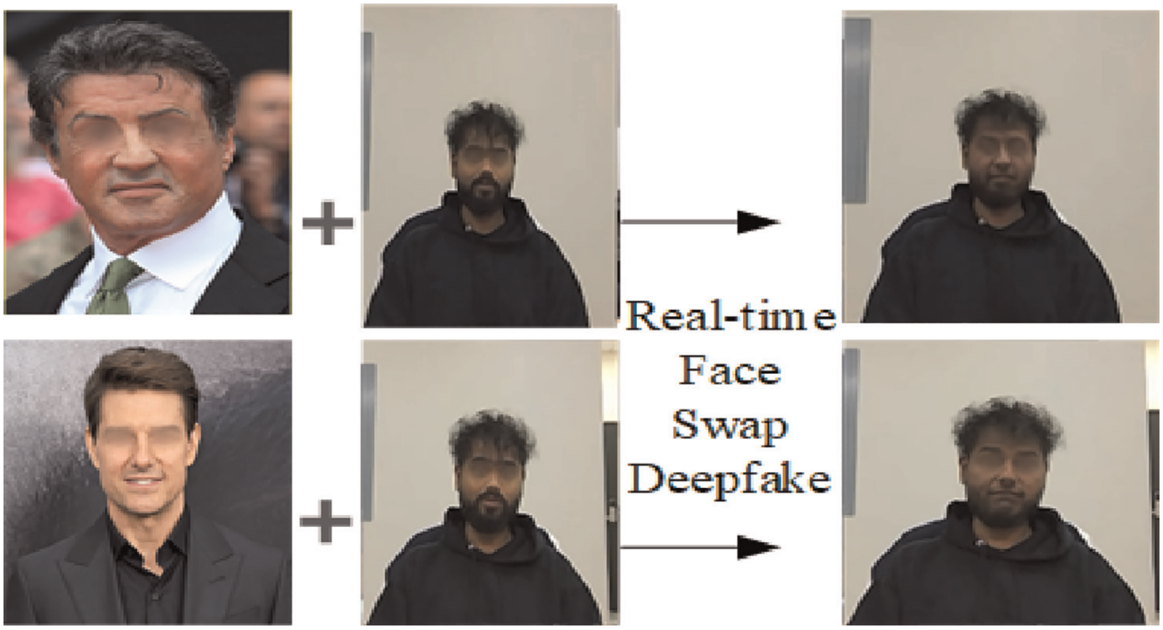

Figure 2.
DeepFake Face Swap Attack to project a source face on a target.

victim resulting in a media where the victim is projected to act out however the perpetrator wishes. Although the model was created to demonstrate the capabilities of deep learning models, it resulted in targeting politicians and celebrities to develop fake media. A GAN model is created where the source body actions are projected on a targeted person [39]. The model introduced resulted in creating an entertainment application, and it could also be alternatively used to frame a victim by forging their actions in surveillance media. The style-based transfer learning has enabled the GAN technology to create more realistic and indistinguishable output [19].

Introducing perturbations in real objects or images can cause edge layer object classifiers to make incorrect predictions, which could have serious repercussions. A study showed that making small changes in a stop sign could cause an object detector to wrongly classify it as a different object as depicted in 3(a) [40]. This phenomenon has been analyzed and the Fast Gradient Sign Method attack was proposed, which uses the gradient of the loss function of the classifier to construct the perturbations necessary to carry out the attack [41]. The attack begins by targeting an image and observing the confidence of the classifier in its predictions of the class. Next, the minimum perturbation that maximizes the loss function of the classifier is found iteratively. Using this method, the image can be manipulated such that incorrect classification is achieved without producing any discernible difference to the human eye as shown in 3(b). The Jacobian-based Saliency Map Attack [42] algorithm computes the Jacobian matrix of the CNN being used for object classification and produces a salient map. The map denotes the scale of influence each pixel of the image has on the prediction of the CNN-based classifier. The original image is manipulated in every iteration, such that the two most influential pixels, which are chosen from the saliency map, are changed. The salient map is updated in each iteration, and each pixel is changed only once. This stops when the adversarial image is successfully classified to the target label (**Figure 3**).

Table 1 summarizes the multimedia attack techniques and their respective targeted systems. Along with video manipulation, audio is also equally targeted when creating realistic fake media. Paired with technology like facial re-enactment, DeepFake audio can create an illusion of a targeted person with manipulated actions. Software like Descript [43] can recreate source audio with training data for few as 10 minutes. Emerging technologies like DeepFake need a reliable detector that can distinguish between real and fake media to preserve security and privacy in the modern digital era. Due to the inconsistencies in earlier stages of DeepFake media, many detector modules were created to identify the artifacts introduced during media generation. However, with more training data and advanced computing, the output benefited and rendered the previous detection scheme useless. In the following section, we study the key parameters required for a reliable detector to establish an authentication system for digital media.

3. Detection techniques against multimedia attacks

Countering forgery attacks led to the development of detection techniques relying on artifacts related to the in-camera processing module or the post-processing methods. The prior knowledge of the source of the media recordings has been an advantage in detecting forgery; however, without that knowledge, some techniques depend on the artifacts introduced by forgery itself. Techniques based on blind techniques, prior knowledge, and forgery artifacts using the conventional methods are first discussed, followed by neural networks trained to identify the forgery.

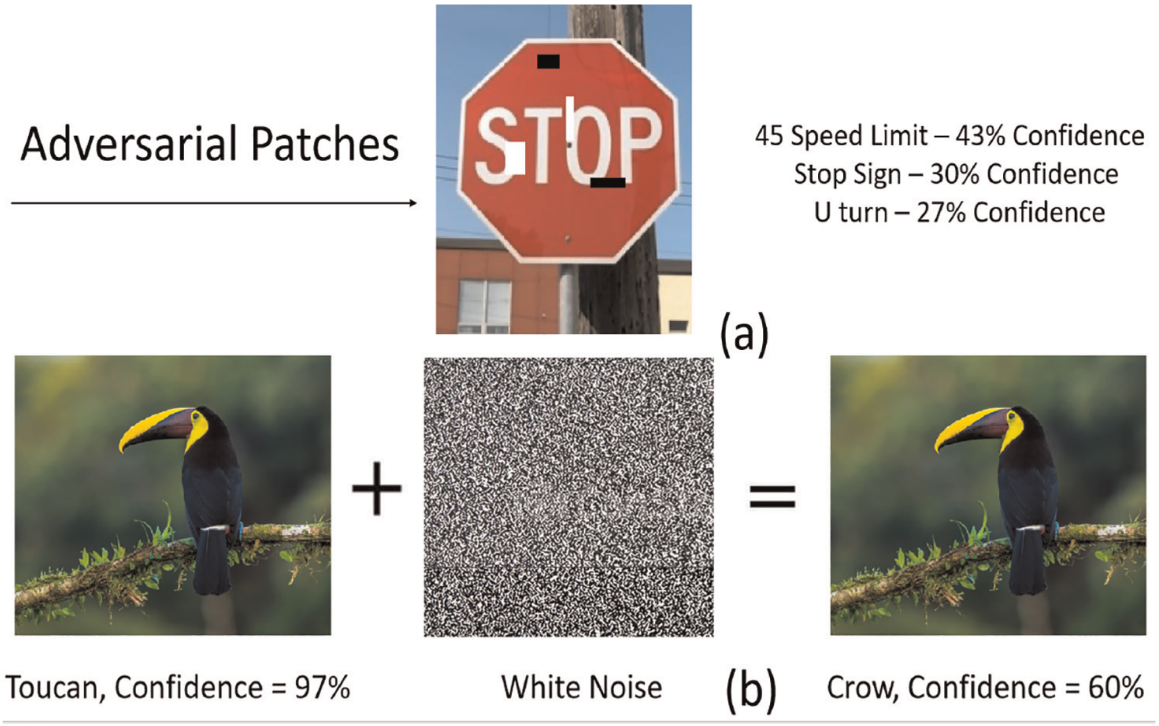

Figure 3.
a) Adversarial patches cause the classifier to wrongly classify the stop sign. b) FGSM attack based on introducing pixel-based perturbations.

3.1 Conventional detection methods

The processing modules present in-camera and post-processing of the media captured result in generating unique features and artifacts, which are exploited to identify frame forgeries. Each image capturing device is equipped with wide or telescopic lenses, where the unique interaction between the lens and the imaging sensor creates chromatic aberrations. A profile of unique chromatic aberrations is created to identify foreign frames inserted from a different lens and sensor [44, 45]. Along with lens distortion artifacts, another module present in in-camera processing after image acquisition is the Color Filter Array (CFA). The CFA is used to record light at a certain wavelength, and the demosaicing algorithm is used to interpolate the missing colors. A periodic pattern emerges due to the in-built CFA module, and whenever a frame is forged, it disrupts the periodic pattern. For frame region splicing attacks, the interrupted periodic pattern from CFA is analyzed to detect the forgery and localize the attack [46, 47].

Each camera sensor manufactured has a unique interaction with the light capturing mechanism due to its sensitivity and photodiode. A unique Sensor Pattern Noise (SPN) is generated for every source camera [48]. It can identify the image acquisition device based on prior knowledge of the camera's sensor noise fingerprint. The SPN noise is similar for RGB and Infrared video; however, it is weaker in Infrared due to low light [49]. Since SPN is used for source device identification, frames moved from an external camera can be identified with any localized in-frame manipulation. The frame and audio acquisition process introduce noise level to the media recordings based on the sensor light sensitivity and localized room reverberations. Using the Error Level Analysis, rich features can be extracted from the noise level present and reveal possible anomalies from image splicing [50].

Attack type	Attack surface[a]	Trigger	Attack vector[a]	Complexity[b]
Frame Manipulation	Visual layer attack (duplication/ shuffling)	• QR code scan • Face/Object Detection • Remote Trigger	• Denial of Service • Malicious Firmware Injection • Live Event Monitoring	• Access: Easy • Low Computation
Audio Masking	Auditory layer attack (noise addition/ audio suppression)	• Voice Command • Programmable noise input	• Compromised AED system • Malicious Firmware Injection	• Access: Easy • Low Computation
DeepFake Manipulation	• Visual layer attack • Auditory layer attack	• Target Face Detection • Remote Trigger	• Face/Object Detection • Live Event Monitoring • Identity Spoofing	• Access: Medium • High Computation
Adversarial Perturbations	• Visual layer attack • Auditory layer attack	• Target Object Detection • Pretrained Noise broadcast	• Object Detection/ Classification • AED Systems	• Access: High (Reconnaissance required) • High Computation

[a] *Targeted/Compromised systems and attack technique.*
[b] *Attack launching complexity—varied based on ease of access and computational requirements.*

Table 1.
Summary of attack vectors and affected modules.

In the media capturing post-processing, each compression algorithm uses unique encoding. Therefore, multiple processing of the same media and multiple compression can result in some artifacts identifying prior changes. Analyzing the compression algorithms used by H.264 coding, the presence of any recompression artifacts is used to identify frame manipulations [51]. The spatial and temporal correlation is used to create motion vector features [30, 52]. The de-synchronization caused by removing a group of frames introduces spikes in the Fourier transform of the motion vectors. However, these techniques are sensitive to resolution and noise in the recordings.

The frame manipulations have also inadvertently introduced their unique artifact, and attacks can be identified with prior knowledge of attack nature. Many types of research were developed using custom hand-crafted features. The scale-invariant feature transform key points are used as features for the comparison of duplicated frames in a video recording [53]. The features comprise illumination, noise, rotation, scaling, and small changes in viewpoint. For a continuous frame capture, the standard deviation of residual frames can result in inter-frame duplication detection [54, 55]. Histograms of Oriented Gradients (HOGs) are a unique presentation of pixel value fluctuations, which can be used to identify copy-move forgery based on the HOG feature fluctuation [56]. The optical flow represents the pattern of apparent motion of an image between consecutive frames and its displacement. Using the feature vector designed from the optical flow, copy-move forgery can be identified [31]. Features are generated for each frame and then lexicographically sorted [57]. The Root Mean Square Error (RMSE) is calculated for the frames, and any frame that crosses the threshold is identified as the duplicated frame. However, the technique takes higher processing time due to the sorting and RMSE algorithm and is not applicable in real-time applications.

3.2 Machine learning-based detection methods

The development of AI in computer vision has efficiently enabled media processing for forgery detection using trained neural networks. The anomalies introduced in the media recordings result in the forgery-specific artifact, which many research approaches exploit.

3.2.1 Artifacts and feature-based detection

Convolutional neural network (CNN) is the most commonly used frame processing feed-forwarding neural network model, enabling pixel data processing. Forgery attacks such as frame manipulation in the temporal and spatial domain and the DeepFake create an underlying artifact extracted to identify the forgery [58]. In the initial stages of DeepFake development, the resulting media generated visible frame-level artifacts such as inconsistent eye-blinking, face warping, and head-poses. Later, a CNN model is trained to identify the abnormalities introduced by DeepFakes by observing for face warping artifacts [59]. The synthesized face region is spliced into the original image, and a 3D head pose estimation model is created to identify the pose inconsistencies [18]. With the help of pixel information obtained from videos, filters can be designed to identify any tampering. Filters based on discrete cosine transform and video re-quantization errors combined with Deep CNN are used [60].

The DeepFake generation tools are integrated with online conferencing tools to create a fake virtual presence by mimicking a targeted person. The video chat liveness detection in [61] can identify the fake personality due to its fake behavior. The model is trained on behavioral expression in online presence, and any abnormality is marked as fake. For offline media, the audio and video are manipulated to create a video statement; however, the underlying synchronization error for the video lip sync and its corresponding audio are used to identify fake media [62]. To counter DeepFake videos in edge-based computers and online social media, lightweight machine learning models are trained based on the facial presence and its respective spatial and temporal features [63]. Video conferencing solutions are also protected by analyzing the live video stream and passing it through a 3D convolution neural network to predict video segment-wise fakeness scores. The fake online person is identified by the CNN trained on large DeepFake datasets such as Deeperforensics, DFDC, and VoxCeleb.

Along with video forgeries, audio forgeries targeting the AED system in IoT devices like Echo dot by Amazon and Nest Hub by Google are designed. Using the audio perturbations, the AED system misclassifies the incoming voice commands or completely ignores the commands [64]. Training a CNN and recurrent neural network (RNN) [26] has secured the AED system from white noise to disrupt the commands.

3.2.2 Fingerprint-based detection

Modern DeepFake videos are almost perfect without any visual inconsistencies. However, the underlying pixel information is modified due to the project of foreign information on existing media. With advancing DeepFake technology, the current research has developed techniques to identify the underlying pixel fluctuations and use unique fingerprints due to GAN models and in-camera processing. Authors in [65, 66] have identified that GAN leaves unique fingerprints in the media generated from its network. By creating a profile of these unique fingerprints, the forgery can be detected, and the source GAN model used to create the forgery can also be identified. The DeepFake models introduce pixel-level frequency fluctuations, which result in spectral inconsistencies. Inspecting the spectral inconsistencies in a fake image shows that due to the up-sampling convolution of a CNN model in GAN, the frequency artifact is introduced [67, 67]. A filter-based design is used in [68] to highlight the frequency component artifacts introduced by GAN. The two filters used are used in the high-frequency region of an image and the pixel level to observe the changes in the pixels in the background of the image. A biological signature is created from the portrait videos by collecting the signals from different image sections such as facial regions and under image distortions [69].

3.2.3 Adversarial training-based detection

Deep neural networks have been proven to be effective tools in extracting features exclusive to DeepFaked images and can thus detect DeepFake-based image forgery. The traditional approach uses a dataset containing real and fake images to train a CNN model, and to identify artifacts that point to forgery. However, this could lead to the problem of generalization as the validation dataset is often a subset of the training dataset. To avoid this, the images can be preprocessed by using Gaussian Blur and Gaussian noise [70]. Doing so suppresses noise due to pixel-level-high-frequency artifacts. Hybrid models have also been proposed that use multiple streams in parallel to detect fake images [71]. It uses one branch to prepare a model trained on the GoogleNet dataset to differentiate between benign and faked images, and another branch that uses a steganalysis feature extractor to capture low-level details. Results from both the branches are then fused together to formulate the ultimate decision on whether a particular image has been tampered with or not.

There are various approaches to detecting fake or tampered videos using machine learning techniques and can be broadly categorized into those that use biological features for detection, and those that observe spatial and temporal relationships to achieve the same objective. A study proposed a novel approach based on eye blinking to detect tampered videos [72]. It is common knowledge that forgery techniques such as DeepFakes produce little-to-no eye blinking in the fake videos that they produce. Using a combination of CNNs and RNNs that were trained on an eye blinking-based dataset, a binary classifier can be produced, which in turn can be used to detect fake videos with reasonable accuracy. Facial regions of interest were used to train models to differentiate between real and DeepFaked videos [73]. Specifically, photoplethysmography (PPG), which uses color intensities to detect heartbeat variations, was used to train a GAN to distinguish between real and fake face videos. However, the drawback lies in the fact that this method is limited to high-resolution videos containing faces only.

Spatiotemporal analysis-based methods treat videos as a collection of frames related to time. Here, in addition to CNNs, Long-Short Term Memory (LSTM) models are used due to their ability to learn temporal characteristics. One such combination that used a CNN to extract frame level features and an LSTM for temporal sequence analysis was proposed [74]. Simply put, the input to the LSTM is a concatenation of features extracted per frame by the CNN. The final output is a binary prediction as to whether the video is genuine or not. GANs have also been proposed as means of analyzing spatiotemporal relationships of videos. An information theory-based approach was used to study the statistical distribution of fake and real frames, and the differential between them was used to make a decision [75].

4. Measure of effective detection techniques

Evaluating the state of the current media authentication system, the existing state-of-the-art technique relies on a fundamental forgery-related artifact or training a deep neural network to identify specific forgery. However, the same deep learning technology has allowed the perpetrator to hijack the existing detection scheme and counteract its purpose. A source device identification methodology used to locate the device used to capture a certain media recording by leveraging the Sensor Pattern Noise fingerprint can be spoofed. The counter method uses a GAN-based approach to inject camera traces into synthetic images, deceiving the detectors into realizing that the synthetic images are real [76]. Development in GAN technology and abundantly available computing resources have generated many fake media that are indistinguishable. A style transfer technique can project facial features into a targeted person and re-create a realistic image [19].

Modern infrastructure relying on machine learning algorithms for seamless people detection and tracking are targeted by adversarial training. A wearable patch can be trained and used to escape the detection or fool the detector into misclassifying the object [29]. The remote trigger mechanism for frame-level attacks is triggered using visual cues and avoids detection by face blur or frame duplication [22]. Tools with simple instructions are designed to allow users to create DeepFake in online video conferences by portraying a targeted person [77].

The need for secure media authentication that spans multiple media categories becomes more and more compelling because of an increase in counterattacks on existing detection techniques. Based on our analysis of the current state-of-the-art detection methods and their counterattacks, here we highlight the key ingredients of the most successful and reliable approaches:

- *Spatial and temporal correlation*: Forgeries involve manipulating spatial frame regions or shuffling the frame itself, which affects the temporal region. A reliable detector should exploit both spatial and temporal correlations to identify forgeries in both layers.

- *Unique Fingerprint*: Deep learning has enabled architectures that are capable of replicating unique device-related fingerprints given sufficient training data. The detector should utilize a fingerprint that is independent of external factors and the device to avoid predictions and re-creation of a unique fingerprint. Inability to control the source of fingerprint generation correlates with difficulty in recreating its unique nature.

- *Multimedia Applicability*: Detectors target specific attacks, which allows a perpetrator to adjust the artifacts and bypass the detection. Both audio and video recordings are the primary input sources for edge devices, and it is equally important to secure both media channels against attacks. A detector should equally account for changes and manipulations in both channels, thereby creating a redundant system capable of dual authentication.

- *Heterogeneous Platform*: Modern smart infrastructure consists of many different types of edge-based IoT smart devices. Each device has its designated functionality relying on either video or audio sensors. Each edge device is also limited in its computational capability due to its power source preservation.

The forgery detection technique should account for enabling its authentication measures across all devices capable of capturing any multimedia.

- *Online Detection*: Attacks are focused on interrupting the active state of the detection system, and most existing techniques are offline systems. Given the state of infrastructure security, it is crucial to immediately raise the alarm upon forgery detection. Enabling instant, online detection can actively observe the media capture and process for any manipulations.

- *Attack Localization*: Lastly, it is important to localize the forgery for further inspection along with attack detection. A detection method that is capable of tracking spatial and temporal changes to the media can locate changes made to the collected samples.

Analyzing the critical traits of a reliable detection system, we propose an environmental fingerprint capable of justifying the qualities aforesaid using the power system frequency. The following section discusses the rationale behind our fingerprint-based authentication system for edge-based IoT devices.

5. Environmental fingerprint-based detection

Electrical Network Frequency (ENF) is a power system frequency with a nominal value of 60 Hz in the United States and 50 Hz in most European and Asian countries. The power system frequency fluctuates around its nominal values, making it a time variant, and the resulting signal is referred to as the ENF signal. The ENF-based media authentication was first introduced for audio forgery detection in law enforcement systems [78]. The fluctuations in ENF are similar to a power grid interconnect and originate from the power supply demand, making the fluctuations unique, random, and unpredictable. For audio recordings, ENF is induced in the recordings through electromagnetic induction from being connected to the power grid [78]. Later, it was discovered that battery-operated devices could also capture ENF fluctuations due to the background hum generated by grid-powered devices [79]. In the case of video recordings, ENF is captured in the form of illumination frequency from artificially powered light sources [80]. The capturing of ENF signal through photos depends on the type of imaging sensor used in the camera. For a CCD sensor with a global shutter mechanism, one sample is captured per frame since the whole sensor is exposed at one time instant. However, for a CMOS sensor with a rolling shutter mechanism, each row in the sensor is exposed sequentially, resulting in collecting the ENF samples from spatial and temporal regions of a frame [81, 82].

ENF estimation from media recordings allows many applications due to its time-varying unique nature. For geographical tagging of media recordings, the ENF signal estimated is compared with the global reference database, and its recording location can be identified [83]. Similar fluctuations in ENF signal throughout the power grid are used to synchronize the multimedia recordings in audio and video channels [84]. The fluctuations in ENF and the standard deviations of the signal from its nominal value are observed to study the load effects on the grid and predict blackouts [85].

The estimation of ENF from media recordings is thoroughly studied for a reliable signal estimation [86, 87] and the factors that affect its embedding process [82, 88]. An ENF-based authentication system is integrated for false frame forgery detection in both spatial and temporal regions due to the nature of the ENF signal. In DeFake [77, 89], the distributed nature of ENF is exploited by utilizing ENF as a consensus mechanism for distributed authentication among the edge-based system. The media collected from online systems are processed, and the ENF signal is estimated along with the consensus ground truth signal. With the help of the correlation coefficient, any mismatch in the signal is located, and an alarm is raised. For detailed system implementation and ENF integration techniques, interested readers are referred to papers on the ENF-based authentication system [90, 91].

6. State of multimedia authentication

The state of the detection system and forgery attacks never reach an equilibrium where the presented detection scheme can function as a solution for all types of attacks. This chapter discussed the evolution of forgery attacks from subtle frame-level modifications to advanced generated images with fake people, along with its parallel development in detection methods. Based on the critical observations discussed in Section 4, **Table 2** presents a comparison of several current forgery detection techniques.

ENF is a reliable detection method given the signal embedded in the media recordings. The current limitation of this approach involves the recording environment where the ENF-inducing equipment is not present. Due to the absence of artificial lights for outdoor recording, the ENF is not captured in the video recordings. However, in the case of outdoor surveillance recordings, the device is connected to the power grid directly, and the ENF signal is induced in the audio recordings.

Most of the DeepFake detection techniques presented utilize higher computational resources for each frame analysis, and in general, edge devices are not equipped with such power. A different approach would be to design lightweight algorithms utilizing the artifacts or fingerprints for its detection. However, the DeFake approach avoids any training step, and the ENF estimation can be performed in low-computing hardware like Raspberry Pi [91]. Although computer vision has advanced with the emergence of deep learning architecture, DeFake is an environmental fingerprint-based approach relying on signal processing technologies and with encouraging results.

System	FakeCatcher [69]	FakeBuster [92]	Noiseprint [66]	UpConv [93]	MesoNet [94]	DeFake[a] [77]
Spatial	✓	✓	✓	✓	✓	✓
Temporal	✓					✓
Unique	✓		✓			✓
Multimedia						✓
Heterogeneous		✓		✓		✓
Online		✓		✓	✓	✓
Localization	✓		✓			✓

[a] *ENF-based authentication System.*

Table 2.
A comparison of recently proposed forgery detection techniques.

7. Conclusions

The development of forgery attacks has exponentially accelerated with growing computer vision technologies, and the need for a reliable and secure authentication system becomes more compelling. Most detection systems are exploited for their weakness, and attackers frequently launch attacks targeting the system and its security system. This chapter studied the evolution of multimedia attacks using traditional frame-level modification and advanced machine learning-based techniques like DeepFakes. Countering each forgery, we analyzed the detection techniques proposed over time and their progress with the attacks. For a reliable detection and authentication system, we constitute vital ingredients that a system should possess to counter forgery attacks. A thorough analysis and comparison of existing detection techniques are performed to understand the current state of multimedia authentication. Based on the key qualities introduced for a reliable system, we highlight DeFake, an environmental fingerprint-based authentication system, and describe its applications for frame forgeries like a DeepFake attack. Given the state of current edge computing technologies and the constant attacks targeted to disable the system, DeFake is the potential to provide a unique approach for detecting such forgery attacks and protecting the information integrity.

Acknowledgements

This work is supported by the U.S. National Science Foundation (NSF) via grant CNS-2039342 and the U.S. Air Force Office of Scientific Research (AFOSR) Dynamic Data and Information Processing Program (DDIP) via grant FA9550-21-1-0229. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the U.S. Air Force.

Abbreviations

ML	Machine Learning
AI	Artificial Intelligence
IoT	Internet of Things
VoIP	Voice over Internet Protocol
DL	Deep Learning
GAN	General Adversarial Networks
AED	Audio Event Detector
PTZ	Pan-Tilt-Zoom
CFA	Color Filter Array
SPN	Sensor Pattern Noise
HOG	Histogram of Oriented Gradients
RMSE	Root Mean Square Error
CNN	Convolutional Neural Network
RNN	Recurrent Neural Network
LSTM	Long Short Term Memory
PPG	Photoplethysmography

ENF	Electrical Network Frequency
CMOS	Complementary Metal Oxide Semiconductor
CCD	Charge-Coupled Device

Author details

Deeraj Nagothu, Nihal Poredi and Yu Chen*
Department of Electrical and Computer Engineering, Binghamton Univerisity, Binghamton, New York, USA

*Address all correspondence to: ychen@binghamton.edu

References

[1] Chen J, Li K, Deng Q, Li K, Philip SY. Distributed Deep Learning Model for Intelligent Video Surveillance Systems with Edge Computing. NY, United States: IEEE Transactions on Industrial Informatics; 2019

[2] Nikouei SY, Chen Y, Song S, Choi BY, Faughnan TR. Toward intelligent surveillance as an edge network service (isense) using lightweight detection and tracking algorithms. IEEE Transactions on Services Computing. 2019;**14**(6): 1624-1637

[3] Obermaier J, Hutle M. Analyzing the security and privacy of cloud-based video surveillance systems. In: Proc. 2nd ACM Int. Work. IoT Privacy, Trust. Secur. NY, United States: ACM; 2016. pp. 22-28

[4] Shi W, Cao J, Zhang Q, Li Y, Xu L. Edge computing: Vision and challenges. IEEE Internet of Things Journal. 2016; **3**(5):637-646

[5] Chen N, Chen Y, Blasch E, Ling H, You Y, Ye X. Enabling smart urban surveillance at the edge. In: 2017 IEEE International Conference on Smart Cloud (Smart Cloud). NY, United States: IEEE; 2017. pp. 109-119

[6] Chen N, Chen Y, Ye X, Ling H, Song S, Huang CT. Smart city surveillance in fog computing. In: Advances in Mobile Cloud Computing and Big Data in the 5G Era. Cham, Switzerland: Springer; 2017. pp. 203-226

[7] Nikouei SY, Xu R, Nagothu D, Chen Y, Aved A, Blasch E. Real-time index authentication for event-oriented surveillance video query using blockchain. In: 2018 IEEE Int. Smart Cities Conf. ISC2 2018. 2019

[8] Xu R, Nagothu D, Chen Y. Decentralized video input authentication as an edge Service for Smart Cities. IEEE Consumer Electronics Magazine. 2021; **10**(6):76-82

[9] Mowery K, Wustrow E, Wypych T, Singleton C, Comfort C, Rescorla E, et al. Security analysis of a full-body scanner. In: 23rd USENIX Security Symposium (USENIX Security 14). San Diego, CA, United States; 2014. pp. 369-384

[10] Olsen M. Beware, Even Things on Amazon Come with Embedded Malware. 2016. Available from: https://artfulhacker.com/post/142519805054/beware-even-things-on-amazon-come

[11] Costin A. Security of Cctv and video surveillance systems: Threats, vulnerabilities, attacks, and mitigations. In: Proc. 6th Int. Work. Trust. Embed. Devices. NY, United States: ACM; 2016. pp. 45-54

[12] Carlini N, Mishra P, Vaidya T, Zhang Y, Sherr M, Shields C, et al. Hidden voice commands. In: 25th USENIX Security Symposium (USENIX Security 16). Austin, TX, United States; 2016. pp. 513-530

[13] Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative Adversarial Nets. In: Advances in Neural Information Processing Systems. Montreal, Canada; 2014

[14] Verdoliva L. Media forensics and deepfakes: an overview. IEEE Journal of Selected Topics in Signal Processing. NY, United States. 2020;**14**(5):910-32

[15] Westerlund M. The emergence of Deepfake technology: A review. Technology Innovation and Management Review. 2019;**9**(11):40-53

[16] Nagothu D, Chen YY, Blasch E, Aved A, Zhu S. Detecting malicious false frame injection attacks on surveillance Systems at the Edge Using Electrical Network Frequency Signals. Sensors (Basel). 2019;**19**(11):1-19

[17] Wolfgang RB, Delp EJ. A watermark for digital images. In: Proceedings of 3rd IEEE International Conference on Image Processing. Lausanne, Switzerland: IEEE; 1996. pp. 219-222

[18] Yang X, Li Y, Lyu S. Exposing deepfakes using inconsistent head poses. In: ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Brighton, United Kingdom: IEEE; 2019. pp. 8261-8265

[19] Karras T, Laine S, Aila T. A style-based generator architecture for generative adversarial networks. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. Long beach, CA, United States; 2019. pp. 4401-4410

[20] Mena DM, Papapanagiotou I, Yang B. Internet of things: Survey on security. Information Security Journal: A Global Perspective. Philadelphia, PA, United Stated. 2018;**27**(3):162-82. DOI: 10.1080/19393555.2018.1458258

[21] Kharraz A, Kirda E, Robertson W, Balzarotti D, Francillon A. Optical delusions: A study of malicious QR codes in the wild. In: 2014 44th Annual IEEE/IFIP International Conference on Dependable Systems and Networks. Atlanta, GA, United States; 2014. pp. 192-203

[22] Nagothu D, Schwell J, Chen Y, Blasch E, Zhu S. A study on smart online frame forging attacks against video surveillance system. In: Proc. SPIE - Int. Soc. Opt. Eng. Bellingham, Washington, United States; 2019

[23] Zhang C, Shahriar H, Riad ABMK. Security and privacy analysis of wearable health device. In: 2020 IEEE 44th Annual Computers, Software, and Applications Conference (COMPSAC). Los Alamitos, CA, United States; 2020. pp. 1767-1772

[24] Kriesel D. Xerox Scanners/Photocopiers Randomly Alter Numbers in Scanned Documents. 2017. https://www.dkriesel.com/en/blog/ 2013/0802_xerox-workcentres_are_switching_written_numbers_when_scanning

[25] Stamm MC, Lin WS, Liu KR. Temporal forensics and anti-forensics for motion compensated video. IEEE Transactions on Information Forensics and Security. 2012;**7**(4):1315-1329

[26] dos Santos R, Kassetty A, Nilizadeh S. Disrupting audio event Detection deep neural networks with white noise. Technologies. 2021;**9**(3):64

[27] Akhtar N, Mian A. Threat of adversarial attacks on deep learning in computer vision: A survey. IEEE Access. 2018;**6**:14410-14430

[28] Quan W, Nagothu D, Poredi N, Chen Y. Cri PI: An efficient critical pixels identification algorithm for fast one-pixel attacks. In: Sensors and Systems for Space Applications. Bellingham, Washington, United States: SPIE; 2021. pp. 83-99

[29] Thys S, Van Ranst W, Goedeme T. Fooling automated surveillance cameras: Adversarial patches to attack person Detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Long Beach, CA, United States; 2019

[30] Wang W, Farid H. Exposing digital forgeries in video by detecting

duplication. In: Proc. 9th Work. Multimed. Secur. Dallas, Texas, United States: ACM; 2007. pp. 35-42

[31] Al-Sanjary OI, Abdullah Ahmed A, Ahmad HB, Ali MAM, Mohammed MN, Irsyad Abdullah M, et al. Deleting object in video copy-move forgery Detection based on optical flow concept. In: 2018 IEEE Conference on Systems, Process and Control (ICSPC). Melaka, Malaysia; 2018. pp. 33-38

[32] Ulutas G, Ustubioglu B, Ulutas M, Nabiyev V. Frame duplication/mirroring Detection method with binary features. IET Image Processing. 2017;**11**(5): 333-342

[33] Su L, Li C. A novel passive forgery Detection algorithm for video region duplication. Multidimensional Systems and Signal Processing. 2018;**29**(3): 1173-1190

[34] Wahab AWA, Bagiwa MA, Idris MYI, Khan S, Razak Z, Ariffin MRK. Passive video forgery detection techniques: A survey. In: 2014 10th Int. Conf. Inf. Assur. Al Ain, United Arab Emirates; 2014. pp. 29-34

[35] Korshunov P, Marcel S. Deep fakes: a new threat to face recognition? Assessment and detection. In: 2018 Computer Vision and Pattern Recognition. Salt Lake City, Utah, United States; 2018. DOI: 10.48550/arXiv.1812.08685

[36] Bitouk D, Kumar N, Dhillon S, Belhumeur P, Nayar SK. Face swapping: Automatically replacing faces in photographs. In: ACM SIGGRAPH 2008 Papers. SIGGRAPH '08. New York, NY, USA: Association for Computing Machinery; 2008. pp. 1-8

[37] Perov I, Gao D, Chervoniy N, Liu K, Marangonda S, Umé C, et al. Deep face lab: Integrated, flexible and extensible face-swapping framework. In: 2021 Computer Vision and Pattern Recognition. NY, United States; 2021

[38] Thies J, Zollhofer M, Stamminger M, Theobalt C, Niessner M. Face 2Face: Real-time face capture and Reenactment of RGB videos. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, Nevada; 2016. pp. 2387-2395

[39] Chan C, Ginosar S, Zhou T, Efros AA. Everybody dance now. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. Seoul, South Korea; 2019. pp. 5933-5942

[40] Eykholt K, Evtimov I, Fernandes E, Li B, Rahmati A, Xiao C, et al. Robust physical-world attacks on deep learning visual classification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, Utah; 2018. pp. 1625-1634

[41] Goodfellow IJ, Shlens J, Szegedy C. Explaining and harnessing adversarial examples. In: 6th International Conference on Learning Representations (ICLR). San Diego, CA, United States; 2015. DOI: 10.48550/arXiv.1412.6572

[42] Papernot N, McDaniel P, Jha S, Fredrikson M, Celik ZB, Swami A. The limitations of deep learning in adversarial settings. In: IEEE European Symposium on Security and Privacy (Euro S and P). Saarbrücken, Germany: IEEE; 2016. pp. 372-387

[43] Descript | Create Podcasts, Videos, and Transcripts. 2021. Available from: https://www.descript.com/

[44] Yerushalmy I, Hel-Or H. Digital image forgery detection based on lens and sensor aberration. International

Journal of Computer Vision. 2011;**92**(1): 71-91

[45] Fu H, Cao X. Forgery authentication in extreme wide-angle Lens using distortion Cue and fake saliency map. IEEE Transactions on Information Forensics and Security. 2012;**7**(4): 1301-1314

[46] Bayram S, Sencar H, Memon N, Avcibas I. Source camera identification based on CFA interpolation. In: IEEE International Conference on Image Processing. Genoa, Italy; 2005

[47] Cao H, Kot AC. Accurate Detection of Demosaicing regularity for digital image forensics. IEEE Transactions on Information Forensics and Security. 2009;**4**(4):899-910

[48] Lukas J, Fridrich J, Goljan M. Digital camera identification from sensor pattern noise. IEEE Transactions on Information Forensics and Security. 2006;**1**(2):205-214

[49] Hyun DK, Lee MJ, Ryu SJ, Lee HY, Lee HK. Forgery detection for surveillance video. In: Jin JS, Xu C, Xu M, editors. The Era of Interactive Media. New York, NY: Springer; 2013. pp. 25-36

[50] Cozzolino D, Poggi G, Verdoliva L. Splicebuster: A new blind image splicing detector. In: 2015 IEEE International Workshop on Information Forensics and Security (WIFS). Rome, Italy; 2015. pp. 1-6

[51] González Fernández E, Sandoval Orozco AL, García Villalba LJ. Digital video manipulation Detection technique based on compression algorithms. IEEE Transactions on Intelligent Transportation Systems. 2022;**23**(3): 2596-2605

[52] Wang W, Farid H. Exposing digital forgeries in interlaced and deinterlaced video. IEEE Transactions on Information Forensics and Security. 2007;**2**(3): 438-449

[53] Kharat J, Chougule S. A passive blind forgery Detection technique to identify frame duplication attack. Multimedia Tools and Applications. 2020;**79**(11): 8107-8123

[54] Fadl SM, Han Q, Li Q. Authentication of surveillance videos: Detecting frame duplication based on residual frame. Journal of Forensic Sciences. 2018;**63**(4): 1099-1109

[55] Bestagini P, Milani S, Tagliasacchi M, Tubaro S. Local tampering Detection in video sequences. In: 2013 IEEE 15th International Workshop on Multimedia Signal Processing (MMSP). Pula, Sardinia, Italy; 2013. pp. 488-493

[56] Subramanyam AV, Emmanuel S. Video forgery Detection using HOG features and compression properties. In: 2012 IEEE 14th International Workshop on Multimedia Signal Processing (MMSP). 2012. pp. 89-94

[57] Singh VK, Pant P, Tripathi RC. Detection of frame duplication type of forgery in digital video using sub-block based features. Int. Conf. Digit. Forensics Cyber Crime. Seoul, South Korea: Springer; 2015. pp. 29–38

[58] Güera D, Delp EJ. Deepfake video detection using recurrent neural networks. In: 2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS). Auckland, New Zealand; 2018

[59] Li Y, Lyu S. Exposing deepfake videos by detecting face warping

artifacts. 2018. In Computer Vision and Pattern Recognition. Salt Lake City, Utah, United Stated; 2018

[60] Zampoglou M, Markatopoulou F, Mercier G, Touska D, Apostolidis E, Papadopoulos S, et al. Detecting tampered videos with multimedia forensics and deep learning. In: Kompatsiaris I, Huet B, Mezaris V, Gurrin C, Cheng WH, Vrochidis S, editors. Multi Media Modeling. Lecture Notes in Computer Science. Cham: Springer International Publishing; 2019. pp. 374-386

[61] Liu H, Li Z, Xie Y, Jiang R, Wang Y, Guo X, et al. Live Screen: Video Chat Liveness Detection Leveraging Skin Reflection. In: IEEE INFOCOM 2020 - IEEE Conference on Computer Communications. Toronto, ON, Canada: IEEE; 2020. pp. 1083-1092

[62] Zhou Y, Lim SN. Joint audio-visual Deepfake Detection. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. Montreal, Canada; 2021. pp. 14800-14809

[63] Chen HS, Rouhsedaghat M, Ghani H, Hu S, You S, CCJ K. Defake Hop: A Light-Weight High-Performance Deepfake Detector. In: 2021 IEEE International Conference on Multimedia and Expo (ICME). Shenzhen, China; 2021

[64] Santos RD, Nilizadeh S. Audio attacks and defenses against AED systems – A practical study. In: Audio and Speech Processing. Ithaca, NY, United States; 2021. DOI: 10.48550/arXiv.2106.07428

[65] Marra F, Gragnaniello D, Verdoliva L, Poggi G. Do GANs Leave Artificial Fingerprints? In: 2019 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR). San Jose, CA, United States; 2019. pp. 506-511

[66] Cozzolino D, Verdoliva L. Noiseprint: A CNN-based camera model fingerprint. IEEE Transactions on Information Forensics and Security. 2020;**15**:144-159

[67] Durall R, Keuper M, Pfreundt FJ, Keuper J. Unmasking deep fakes with simple features. In: Computer Vision and Pattern Recognition. Seattle, Washington, United States; 2020

[68] Jeong Y, Kim D, Min S, Joe S, Gwon Y, Choi J. BiHPF: Bilateral high pass filters for robust deepfake detection. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. Waikoloa, HI, United States; 2022. pp. 48-57

[69] Ciftci UA, Demir I, Yin L. Fake catcher: Detection of synthetic portrait videos using biological signals. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2020;**2020**:1-1

[70] Xuan X, Peng B, Wang W, Dong J. On the generalization of GAN image forensics. In: Chinese Conference on Biometric Recognition. Zhuzhou, China: Springer; 2019. pp. 134-141

[71] Zhou P, Han X, Morariu VI, Davis LS. Two-stream neural networks for tampered face detection. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Honolulu, Hawaii, United States: IEEE; pp. 1831-1839

[72] Li Y, Chang MC, Lyu S. In ictu oculi: Exposing ai created fake videos by detecting eye blinking. In: 2018 IEEE International workshop on information

forensics and security (WIFS). Hong Kong, China: IEEE; 2018. pp. 1-7

[73] Ciftci UA, Demir I, Yin L. How do the hearts of deep fakes beat? deep fake source detection via interpreting residuals with biological signals. In: 2020 IEEE International Joint Conference on Biometrics (IJCB). Houston, TX, United States: IEEE; 2020. pp. 1-10

[74] Güera D, Delp EJ. Deepfake video detection using recurrent neural networks. In: 2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS). Auckland, New Zealand: IEEE; 2018. pp. 1-6

[75] Agarwal S, Girdhar N, Raghav H. A novel neural model based framework for detection of GAN generated fake images. In: 2021 11th International Conference on Cloud Computing, Data Science Engineering (Confluence). Uttar Pradesh, India; 2021. pp. 46-51

[76] Cozzolino D, Thies J, Rossler A, Niessner M, Verdoliva L. Spo C: Spoofing camera fingerprints. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. NY, United States; 2021. pp. 990-1000

[77] Nagothu D, Xu R, Chen Y, Blasch E, Aved A. DeFake: Decentralized ENF-consensus based deep fake detection in video conferencing. In: IEEE 23rd International Workshop on Multimedia Signal Processing. Tampere, Finland; 2021

[78] Grigoras C. Applications of ENF criterion in forensic audio, video, computer and telecommunication analysis. Forensic Science International. 2007;**167**(2–3):136-145

[79] Chai J, Liu F, Yuan Z, Conners RW, Liu Y. Source of ENF in battery-powered digital recordings. In: Audio Eng. Soc. Conv. Rome, Italy: Audio Engineering Society; 2013

[80] Garg R, Varna AL, Hajj-Ahmad A, Wu M. "Seeing" ENF: Power-signature-based timestamp for digital multimedia via optical sensing and signal processing. IEEE Transactions on Information Forensics and Security. 2013;**8**(9):1417-1432

[81] Vatansever S, Dirik AE, Memon N. Analysis of rolling shutter effect on ENF based video forensics. IEEE Transactions on Information Forensics and Security. 2019;**14**(7):2262-2275

[82] Nagothu D, Chen Y, Aved A, Blasch E. Authenticating video feeds using electric network frequency estimation at the edge. EAI Endorsed Transactions on Security and Safety. 2021;**7**(24):e4-e4

[83] Wong CW, Hajj-Ahmad A, Wu M. Invisible geo-location signature in a single image. In: 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Calgary, Alberta, Canada: 2018. pp. 1987-1991

[84] Vidyamol K, George E, Jo JP. Exploring electric network frequency for joint audio-visual synchronization and multimedia authentication. In: 2017 International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT). Kannur, Kerala, India; 2017. pp. 240-246

[85] Liu Y, You S, Yao W, Cui Y, Wu L, Zhou D, et al. A distribution level wide area monitoring system for the electric power grid–FNET/grid eye. IEEE Access. 2017;**5**:2329-2338

[86] Hua G, Zhang H. ENF signal enhancement in audio recordings. IEEE Transactions on Information Forensics and Security. 2020;**15**:1868-1878

[87] Hajj-Ahmad A, Garg R, Wu M. Spectrum combining for ENF signal estimation. IEEE Signal Processing Letters. 2013;**20**(9):885-888

[88] Hajj-Ahmad A, Wong CW, Gambino S, Zhu Q, Yu M, Wu M. Factors affecting ENF capture in audio. IEEE Transactions on Information Forensics and Security. 2019;**14**(2): 277-288

[89] Xu R, Nagothu D, Chen Y. Econ ledger: A proof-of-ENF consensus based lightweight distributed ledger for IoVT networks. Future Internet. 2021; **13**(10):248

[90] Nagothu D, Xu R, Chen Y, Blasch E, Aved A. Detecting compromised edge smart cameras using lightweight environmental fingerprint consensus. In: Proceedings of the 19th ACM Conference on Embedded Networked Sensor Systems. New York, NY, USA: Association for Computing Machinery; 2021. pp. 505-510

[91] Nagothu D, Xu R, Chen Y, Blasch E, Aved A. Deterring Deepfake attacks with an electrical network frequency fingerprints approach. Future Internet. 2022;**14**(5):125

[92] Mehta V, Gupta P, Subramanian R, Dhall A. FakeBuster: a DeepFakes detection tool for video conferencing scenarios. In: 26th International Conference on Intelligent User Interfaces-Companion. College Station, TX, United States; 2021. pp. 61-63

[93] Durall R, Keuper M, Keuper J. Watch your up-convolution: CNN based generative deep neural networks are failing to reproduce spectral distributions. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. NY, United States; 2020. pp. 7890-7899

[94] Afchar D, Nozick V, Yamagishi J, Echizen I. Meso net: A compact facial video forgery Detection network. In: 2018 IEEE International Workshop on Information Forensics and Security (WIFS). Hong Kong, China; 2018. pp. 1-7