Received April 1, 2019, accepted April 16, 2019, date of publication April 26, 2019, date of current version May 7, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2913648

Video-Based Evidence Analysis and Extraction in Digital Forensic Investigation

JIANYU XIAO[1,2], SHANCANG LI[3], (Member, IEEE), AND QINGLIANG XU[1]

[1]School of Computer Science and Engineering, Central South University, Changsha, Hunan 410083, China
[2]Central South Ultrapower Infotech Co. Ltd., Changsha 410205, China
[3]Department of Computer Science and Creative Technologies, University of the West of England, Bristol BS16 1QY, U.K.

Corresponding author: Shancang Li (shancang.li@ieee.org)

This work was partially funded by the Planned Science and Technology Projects of Hunan Province, China (2016JC2010, 2017NK2394); the Planned Science and Technology Project of Changsha, China (kq1801071); and Postgraduate Independent Exploration and Innovation Project of Central South University (2017zzts717).

ABSTRACT As a result of the popularity of smart mobile devices and the low cost of surveillance systems, visual data are increasingly being used in digital forensic investigation. Digital videos have been widely used as key evidence sources in evidence identification, analysis, presentation, and report. The main goal of this paper is to develop advanced forensic video analysis techniques to assist the forensic investigation. We first propose a forensic video analysis framework that employs an efficient video/image enhancing algorithm for the low quality of footage analysis. An adaptive video enhancement algorithm based on contrast limited adaptive histogram equalization (CLAHE) is introduced to improve the closed-circuit television (CCTV) footage quality for the use of digital forensic investigation. To assist the video-based forensic analysis, a deep-learning-based object detection and tracking algorithm are proposed that can detect and identify potential suspects and tools from footages.

INDEX TERMS Forensics investigation, forensic video analysis, video/image enhancement, object detection, deep learning.

I. INTRODUCTION

In forensic investigation, digital cameras and mobile devices are routinely seized as evidence sources. Video and images retrieved from these devices are widely used in crime evidence investigation, which can provide key forensic evidence items, piece together existing evidence items, or establish links between evidence items in particular case. The Closed-circuit television (CCTV) systems are widely used for malls, banks, traffic intersections to park, stores, or even home, where video evidences are retrieved from these systems can be used as evidences much more than ever before [1]. Along with the use of smart devices, such as *mobile phone, smart watch, fitkit, etc.*, audio and video evidences can be easily available in investigation.

In the past few years, the 'image enhancement' techniques have been proposed [2]–[6], most of them can be grouped into *spatial domain methods* and *frequency domain methods*. These techniques shows good potential to improve the quality of images, but only a few of them can be used for low quality of footage, such as cctv footage, mobile video clips, *etc.* [7], [8]. Many cctv surveillance systems export footage in their own formats, which need to be re-format or converted to a suitable format that easier for investigation. However, this can often cause the lower of quality and information loss, which makes the examination process difficult.

The footage in digital forensics is often used for comparative analysis, including forensic analysis, comparison of images of questioned about know objects such as subjects, vehicles, clothing, and weapons, with expert opinion being providing on the findings [9]–[11].

In many modern CCTV systems, *facial recognition* services are embedded to identify online criminals or suspects [26]. Other services such as motion detection, body and face recognition, cross-pose recognition, gait recognition, are widely researched in the past few years. In some hard cases (poor viewing conditions), it is very difficult to identify humans take advantage of face, body, still, *etc.* Although many image processing techniques have been developed in the past few decades, most of them do not take advantage of face, body, *etc.* In video based forensic investigation, following challenges still need to be addressed.

1) Forensic identification, in digital forensics investigation, probe image often "different" from gallery

The associate editor coordinating the review of this manuscript and approving it for publication was Kim-Kwang Raymond Choo.

images due to heterogeneous face recognition due to the low quality, angle of camera, color, *etc.*, new techniques that to improve the quality of footage need to be developed;

2) Establish the links between objects in investigation scenarios and related available evidence resources, such as cctv footage, online image records, or the history trace of these objects;
3) Due to the emerging technologies, such as social networks, internet of things (IoT), mobile devices, *etc.*, the new investigation techniques are not only 'end all' solutions for law enforcement but tools make use of whatever data is available;
4) Intelligent techniques, such as deep learning, artificial intelligence, can assist the digital forensics to quickly identify potential evidences from existing sources collected;
5) Robust evidence extraction methods, such as robust recognition, subject detection, *etc.* , are still need to be developed for forensic investigation, for example, in video-based facial recognition, to detect unfamiliar faces involves facial ageing, marks, forensic sketch recognition, and near-infrared face recognition.

It is clear that the digital forensics investigation over video-related resource highly depends on the quality of footage recordings, poor quality would significantly reduce the confidence level of the investigation process and thus would not make a strong evidence to be presented in a court. In this work, we will address the above challenges and propose new methods aim at assisting effective video-based digital forensics investigation. The main contribution are three-fold:

1) An video-based digital investigation framework is proposed that addresses the concerns in low quality of footage, establishment of links between objects and available digital evidences, detection techniques over video footage, and intelligent techniques that can be used in modern digital forensics;
2) Detailed quality improvement method is provided for low quality footage, include the adaptive histogram equalization (AHE), contrast limited AHE (CLAHE), *etc.*; and a test scenario is provided to compare the different algorithms;
3) A deep learning based object identification scheme in footage is proposed that can be used to establish the links between the objects, subjects, and their behaviors in the available footages.

The main aim of forensic video analysis is to identify strong evidence items at different level. In this paper, we focus on the contents of the video to develop efficient video analysis techniques from the view point of forensics.

II. RELATED WORKS

Traditionally, images and videos related to a case are widely used to retrieve key evidence in forensic investigation [11]. In recent, the new technologies make it much easier to create, collect, and analyze these image materials. The advances of emerging techniques such as mobile devices, low cost image/video capturing devices together with information processing (such as artificial intelligence, machine learning, *etc.*) have significantly improved the forensic analysis level. As a result, many research works have been done on image and video validation of the integrity of image and video by detecting any attempt of forgery [12], [13]. Aimed at determining the image/video sources in criminal investigation, Kamenicky *et al.* introduced a number of video methods and tools for analyzing digital images and videos [14].

Following the validation and legal process, identifying the source of image and video is a challenging task in forensic investigation. Many research efforts focused on the image and video sources identification [15], [16]. In [16], Li *et al.* proposed an algorithm for extracting photo response non-uniformity (PRNU) noise from video files taken by mobile phone camera. Amerini *et al.* introduced a source identification method for video files shared on social networks, such as Twitter, Facebook, Wechat, *etc.*, the proposed method can extract a composite fingerprint for the mobile phone camera using PRNU [17].

With the increasing availability of recording devices, an effective video analysis method is necessary. In [18] and [19], method for forensic analysis of video file formats and anti-forensic techniques are introduced for video shot editing. The proposed methods are able to forensically evaluate the deblocking filtering and quantization parameters that are key for anti-forensics of video editing operations. With the increasing availability of video footage, to automatically analyze a huge volume of video files is very challenging. In [20], an effective forensic procedure is proposed to identify and reconstruct cached online video stream data, such as `YouTube, Twitter, Facebook, WeChat,` *etc.*

The CCTV systems have been widely used in surveillance allow continuous recording image and events happened in specific areas and storages as video footages. These footages can be used in forensic investigation when criminal or other activities occurred. However, the strength of evidenced extracted from these footages significantly relies on the contents recorded [11]. Therefore, the quality of the footage is very important for the contents based evidences recorded.

In [21], Jenkins and Kerr proposed an images analysis by combining the emerging facial recognition techniques to identify hidden facial based information. New high quality recording devices, such as smart mobile phones, *etc.* make this method potentially can be used in forensic image video analysis for criminal investigation. Maksymowicz *et al.* introduced a crime scene and event reconstruction using 3D analysis of image and video in [22].

In the past few years, the automated image analysis from video for public security or detection of dangerous situations has been well explored [23], [24]. A number of histogram equalization (HE) have been widely used for image enhancement [25], in which the histogram of an image is defined as the statistic probabilistic distribution of gray levels [25]. For many cctv footage, the HE-based methods can cause over-enhancement with large smooth-area, which can create unaturalness and wash-out appearance [1]. In low light vision video (such as night vision video), it always contains large smooth area with a narrow dynamic range and can cause over-enhancement problem. The past decades, a number of variants of HE have been proposed, Adaptive histogram equalization (AHE) [26], contrast limited adaptive histogram equalization (CLAHE) [15], [27], dualistic sub-image histogram equalization (DSIHE) [28], include bi-histogram equalization (BBHE) [29], *etc.*

In forensic video analysis, the quality of video footage is always a key challenge. To enhance the low quality of video is very important for assisting to conduct efficient forensic investigation. Jasmine and Annadurai proposed an effective real-time video quality enhance method using AHE [30]. In [31], the authors proposed a color video enhancement using discrete wavelet transform and CLAHE named 'DWT E-CLANE', which can improve the recognition of the front view facial images.

CLAHE is a widely used technique for improving the quality of video files [15], [31], [32], which overcomes the over-enhancement problem in HE and a number of variants of CLAHE have been proposed and demonstrated effective for some specific scenarios.

In [23], an automated image detection and recognition algorithm was proposed that can identify firearms and knives, which can enable CCTV system alert the operator when specific sensitivity objects are recognized.

III. FORENSIC VIDEO ANALYSIS FRAMEWORK

A. FORENSIC VIDEO ANALYSIS FRAMEWORK

As discussed in above Section II, forensic video analysis and multimedia evidence processing are still relatively new compare to tradition photography-based analysis in the courts. The increasingly usage of forensic video analysis is becoming acceptable in court, which has been certified by a number of legitimacy authority as important objective standards, such as the Certified Forensic Video Analyst (CFVA) [33], *etc*. The emerging forensic video analysis techniques mainly focus on following aspects: (1) Law enforced forensic video analysis; (2) Forensic analysis for video and multimedia; (3) Image/video comparison; and (4) Enhanced forensic video analysis. In this work, we focus on the 'improved forensic video analysis' by using the most recent video and data analysis techniques. Figure 1 illustrates an enhanced forensic video analysis framework, which includes three basic components: *crime scene analysis*, *data acquisition*, *video enhancement and analysis*, and *finding presentation and refinement.*

1) FORENSIC VIDEO CAPTURE AND FORMAT CONVERSION

This stage focuses on video capture, screen, and analog video import with legacy format, synchronization of multiple videos to a common time based for playing and processing. Due to the diversity of capture devices, many digital video evidences are not presented in standard formats, such as *avi, mp4, etc*. It is important to convert these videos to standard format without reducing features.

2) TAG AND HASH RELEVANT IMAGES AND CLIPS

Since the original video usually contains large volumes of redundant frames, extracting case-related video clips $\mathbf{v}$ and group them into organized groups $\mathbf{v}_g$ are two key tasks, which will be helpful for combing variety of multimedia evidences, including combination of video sources from multiple channels, detection suspects, and extract relevant images. The organized images and clips will be automatically hashed using $MD5$ and $SHA1$ [34].

3) ENHANCEMENT

The grouped video frames or clips need to be enhanced to provide additional features, uncover hidden information. Meanwhile, evidences can be acquired from different cameras placed in different areas, times but available for the same incident.

4) COMPREHENSIVE FORENSIC VIDEO ANALYSIS

Comprehensive video analysis, including noisy audio recordings, source identification, enhancement, this stage should be capable of reducing many different types of noises in video recording, video editing, extraction, filtering products, *etc.* Combine multiple filters can be previewed and applied by the editing software simultaneously.

5) ADVANCED COMPREHENSIVE EVIDENCE EXTRACTION

Scene reconstruction and objects detection by combing with existing evidence in social network, 3D scene geometries, AI, machine learning technologies to extract further potential evidence.

6) DEMONSTRATIVE EVIDENCE PRESENTATION

In this stage, investigator can select image/video regions of interest and product a demonstrative evidences for court presentations. This stage should highlight both the original untouched footage and the labeled results.

Through the proposed framework, the investigation should be repeatable. An investigator can intuitively apply advanced enhance techniques to their evidence, including existing video/image processing technologies, such as enhancement algorithms, machine learning techniques, producing video with better quality for investigation. In Section IV we will detail the key techniques used in the framework.

IV. METHODOLOGY FOR FORENSIC VIDEO ANALYSIS

It is clear that in above proposed framework the forensic video analysis can be classified into following two main categories: *video type analysis* and *video contents analysis*.

A. FORENSIC VIDEO TYPE ANALYSIS

In forensic video type analysis, one of main aims is to examine whether the video is illegally re-producted, or if

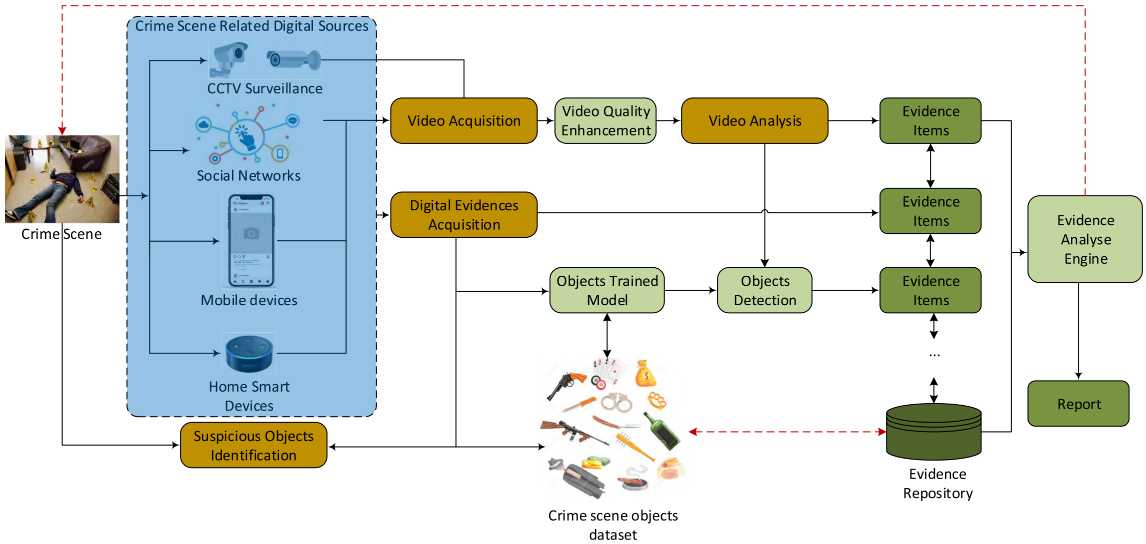


FIGURE 1. Enhanced forensic video analysis framework.

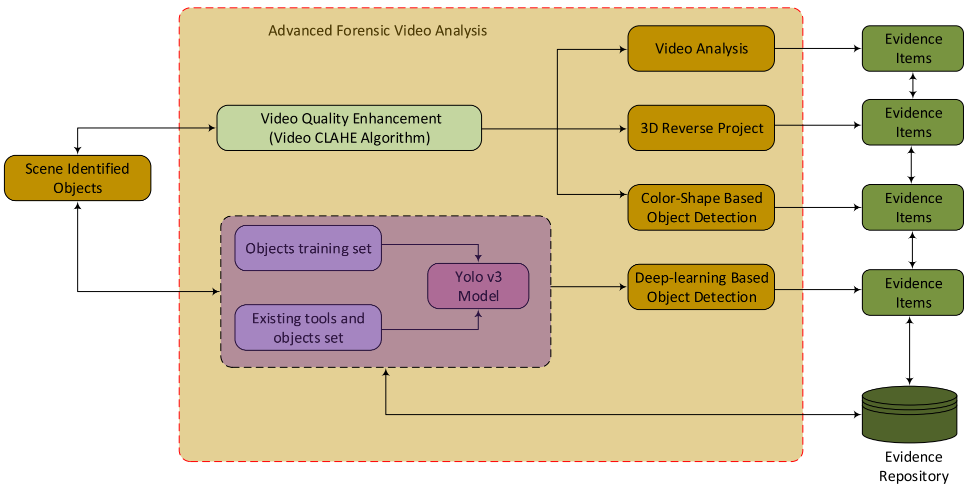


FIGURE 2. Procedures for advanced forensics video analysis.

it is a tampered video. This analysis also conducts video source identification and video steganography analysis to uncover hidden information. Specifically, the video source identification is an important evidence source to identify the sources camera or devices that token this video or image [26], [34].

With the widely use of smart mobile devices such as smart mobile phone, tablet, wifi-camera, *etc.*, the video source identification is an important issue in video forensic analysis (VFC) [13], [34]. In the past decade a number of sensor noise pattern (SNP) based video/image source identification have been developed, which are based on the digital camera sensor imperfections. In general, it involves following three stages: (1) video/image features extraction; (2) devices sensor pattern noise extraction; (3) match the video/image features to device features. The source identification video source identification can be very useful in real scenarios, can identify the video belongs to the source.

$$g_f = u_f + u_f \cdot \Gamma + \Theta, \quad (f = 1 \ldots N) \tag{1}$$

in which g_f is the frame $f \in \{1, \ldots, N\}$ in a video, u denotes the noise-free frame, Γ is the noise (PRNU), and γ denotes additive noise. For a specific device, the PRNU Γ can be estimated using the average of multiple images/frames (between 50 and 300) captured by the device. The camera sensor noise can be described by

$$\Gamma_d = \frac{1}{N}\sum_{i=1}^{N} g_i = \frac{1}{N}\sum_{i=1}^{N}(g_i - \Theta(g_i)) \tag{2}$$

in which $g_i - \Theta(g_i)$ is the denoising filter and averaging the noise residuals [35]. To evaluate the possibility of an image/frame took by a specific devices, we use a normalized correlation

$$corr(\Gamma_{f_i}, \Gamma_d) = \frac{< \Gamma_{f_i}, \Gamma_d >}{\|\Gamma_{f_i}, \Gamma_d\|} \tag{3}$$

where Γ_d represents the mean of f, the operate $<, >$ represents inner product, and $\| \cdot \|$ represents l_2-norm, respectively. In our previous works [26], [34], we have proposed an efficient video/image source identification framework using the PRNU.

B. VIDEO CONTENTS FORENSIC ANALYSIS

This mainly focus on the contents the video or image presents. In traditional forensic video analysis, investigators mainly focus on manually check the contents by play the video and identify the evidence items by going through the contents of video. Actually, the traditionally methods is very time consuming and inefficient when huge volumes of video footage are available. In many case, the analysis significantly relies on the investigator. With the advances of emerging techniques, such as *facial recognition, objects detection, deep learning, etc.*, it makes the automated forensic video analysis possible, however there are still many challenges need to be addressed.

1) VIDEO QUALITY ISSUES

In data acquisition stage, all potential data are collected to make maximum use of the information in the scene. The collected images, video clips, such as cctv footage, video captured by mobile devices, are not suitable for forensics analysis because it cannot meet the legal and quality requirements of the criminal justice system (CJS) [36].

In traditional forensic video examination, a number of processing, such as enlarging, re-scale, *etc.*, are widely used to identify potential evidences from distorted contents in low quality of image and video. However, rescaling an image cannot help to identify additional details. The potential for enhancement of low resolution imagery is generally very limited.

2) BRIGHTNESS

The brightness can significantly affect the examination in low quality video and image due to it can cause mistakes for different contents. In practical, many cctv surveillance system works in different environment and it can cause the captured video over or under-exposed. In these situations, adjusting the brightness of the video/image can reveal more details.

3) COMPRESSION

In forensic video examination, it is usually noted that to maximize the use of storage space, many digital camera systems heavily compress video/images to reduce files size. This will cause in a loss of details that cannot be recovered and will introduce visible artefacts into the image [37].

4) OTHER CONCERNS

In video analysis, other concerns, such as camera angle, number of visible features, the rotation of a subject or object may have significant effect upon its appearance [34].

V. ADVANCED FORENSIC VIDEO ANALYSIS AND USE CASES

A. FORENSIC OBJECT DETECTION/IDENTIFICATION IN VIDEO/IMAGE

Object detection, identification, and tracking are a key techniques in many video surveillance systems [26]. As discussed above, forensic video analysis plays an increasingly important role in identify key evidences in forensic investigation. The video analysis can allows investigator recreate a scene to the presence of a large volume of data. The specific detection of object related to the case can significantly improve the investigation analysis efficiency and provide high quality of evidences items. In many existing works, the footage analysis mainly based on the human retrieve and itâĂŹs is very inefficient, and the evidence identification significantly relies on the experiences of the investigators.

As a key technology, the object detection aimed at detecting objects of interest in digital sources such as videos, audio, and images in an investigation. Given objects identified in a crime case, it is possible to detect and identify addition information that related with these object in other related evidence sources, such as footage, image, *etc.*, and further analysis uncover more details or key crucial information (i.e. relating objects of various crime scenes) and establish links between key evidences.

The objection detection have been well researched many areas, such as medical diagnosis, radar signal processing, using machine learning, AI, etc. The object detection system can also be joined with other techniques to extract useful evidence items in digital investigation. For example, facial recognition have been widely used in China to detect the suspects or it is reported there many suspects have been caught up at pop music concert after being detected using the facial recognition technology [38].

In forensic video analysis, the emerging new technologies, such as artificial intelligence, machine learning, can significantly push the automatization of forensic investigation. Specifically, the object detection can play key roles in two aspects: (1) to identify objects in video surveillance that can be a threat to society and nation; (2) to identify the abnormal behaviors or malicious behaviors; (3) to detect and identify crucial information based on the contents of video/images.

In this work, we address the problem of analyzing the visual data acquired in a crime scene using the object detection. The suspicious or specific subjects can be automatically detected and presented in the evidences. This can extract some intelligence and establish relationships that can help an investigation to relate different evidence items or even crime scenes.

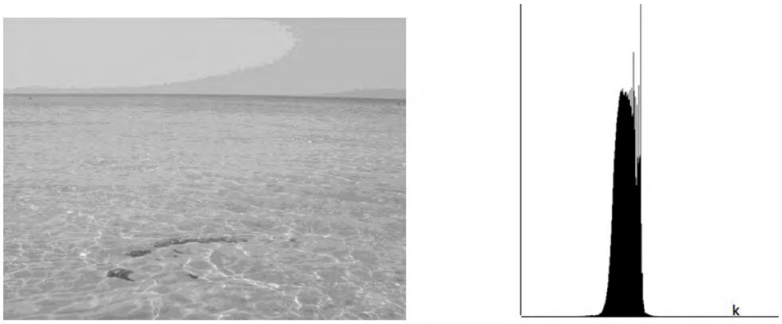


FIGURE 3. CLANE example (left: Original; right: Histogram).

B. VIDEO ENHANCEMENT TECHNIQUES

In the past decade, a number of video enhancement techniques have been developed for video monitoring system, intelligent-highway system, safety monitoring system, *etc.* in [8], [39]. For example, Tzanidou *et al.* proposed a baggage detection for low quality video footage using color information in [8]. This analysis make it possible to identify the moving direction, construction human-like temporal templates, and their align with the best matched examplars. Ghanem and Davis proposed an approach to detect baggage [40]. Similarly, in [41], Chuang *et al.* developed a ratio histogram based individual tracking model, which is an variable of color histograms to detect the missing colors. From the view point of forensics, the basic aim is to retrieve as much information as possible from acquired low quality videos. In this section, we focus on techniques that can be helpful for retrieving addition information from enhanced videos. The Histogram Equalization (HE)-based techniques can significantly improve the probability to uncover addition information from low quality video/images.

1) AHE AND CLAHE

The histogram of an image can be described using

$$h(k) = n_k/(M \times N), \quad k \in [0, \ldots, L-1] \tag{4}$$

in which L denotes the number of values of each pixel ranged between 0 and 255, n_k denotes the number of pixels with intensity k, and $M \times N$ is the pixels in the image.

We use Eq.(5) to describe the estimation of the probability distribution of the pixel intensity

$$\sum_k h(k) = 1 \tag{5}$$

Let $F(k)$ denote the cumulative function, then we have

$$F(k) = \sum_{i=0}^{k} h(i) \tag{6}$$

HE-based methods have been widely used to equalize a gray-scale images [42]. The AHE algorithm does not work effectively for many cctv footage and CLAHE is more useful for general frame/images. The CLAHE algorithm is very simple and can be easily implemented in over digital camera.

Figure. 3 is an example of CLAHE equalization, where low contrast image have a narrow histogram, which means the algorithm is more effective for homogeneous image. In forensic investigation, this is very useful in low quality cctv footage examination. Figure. 3 (a) shows the original image and its histogram, and Figure. 3 (b) shows the image and histogram after equalization using CLAHE. Obviously, from the enhancement, more details can be retrieved.

2) VIDEO ENHANCEMENT USING AHE AND CLAHE

Different with ordinary HE, the AHE is based on histograms of each corresponding to a distinct region, which is suitable for local contrasting of an image. Algorithm 1 address the video enhancement using CLANE.

Algorithm 1 Video Enhancement Algorithm Using Contrast Limited AHE

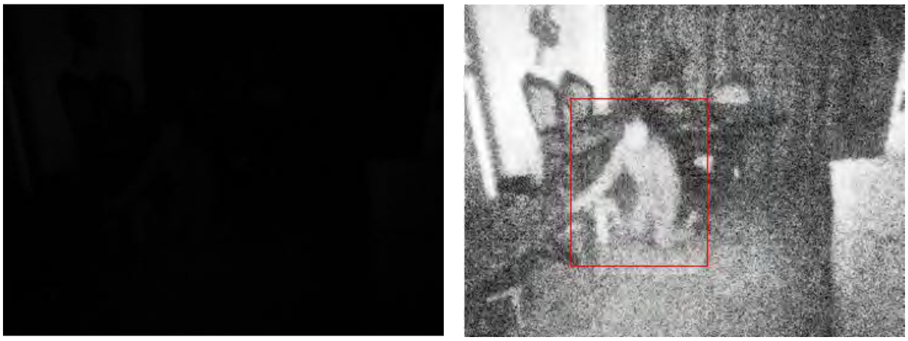
Input: Video or camera input V

Output: Labeled $\overline{V}$

```
v ← video_capture(V)
v_g ← video_group(v)
for each frame f ∈ v_g do
    Initialize array Hist to zero;
    for every contextual pixel j do
        Hist[g(j)] = Hist_l[g[j] + 1
    end for
    CHist_l = Σ_{k=0}^{l} Hist(k)
    l' = CHist_l × L/W^2
    v̄_g ← update(f, l')
end for
V̄ ← update(v, v̄_g)
```

It is noted that in Algorithm 1, we have to calculate the $Hist[g(j)]$ for each pixel is computationally expensive, for an image with size $M \times N$, it cost $O(M \times N \times W^2)$. To further improve the performance of Algorithm 1 for real-time application, it can be improved by reducing the computation.

Actually, the AHE, CLAHE, RETINEX, *etc.*, can be used to provide a services as "automatic color



(a) (b)

FIGURE 4. **Enhanced cctv footage frame in a night visual. (a) Original frame. (b) Enhanced frame.**

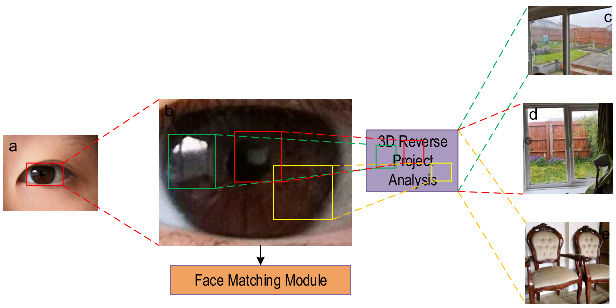


FIGURE 5. **Reveal hidden information from eye reflection.**

enhancement (ACE)'' in forensic video analysis, which is expected to provide more efficient, intelligent digital forensic investigation but a few challenges such as computation, *etc.* need to be addressed. In Figure 6, an example to enhance a mobile camera night vision clip using proposed algorithm. It can be seen that in Figure 6 (a) shows the night vision without infrared, and Figure 6 (b) shows the enhanced frame and a 'suspect' can be identified as shown in the red block. Since it is very difficult using traditional forensic video analysis to identify the hidden information as shown in above example, with the help of proposed framework and algorithms, in forensic video analysis can maximize the use of available visual data.

C. HIDDEN INFORMATION EXTRACTION FORM CORNEAL REFLECTIONS IN VIDEO AND IMAGES

In criminal investigation, high resolution image evidence can be used to identify suspects. As reported in [21], investigator can recover images of unseen bystanders from reflections in the subject's eye and further conduct facial recognition to identify hidden information. Actually, in forensic video analysis, it is possible to mine hidden formation, with which it is possible to apply 3D reverse projection analysis results in highly accurate crime scenes reconstruction, suspect height, subject perspective, and vehicle identification demonstratives, *etc*.

In the example in Figure 5, it is possible to extract the hidden scene information from the eye reflection. Zooming the view of the region of interest with contrast enhanced, introducing the ''3D reverse project analysis'' module can make it possible to reconstruct the scene. In the example, zooming in the interested region in Figure 5(b) can extract hidden scene information and can further reconstruct the scene that token the video/image as shown in Figure 5 (c), (d), and (e). It is also possible to conduct a ''face matching'' to identify faces of subjects if there are. However, the ''face matching'' requires very high-resolution according to Figure 5.

D. AUTOMATED SUSPICIOUS OBJECT DETECTION IN VIDEO

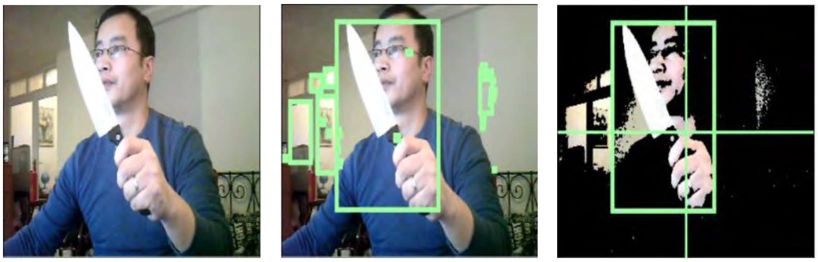
In forensic investigation, it is very challenging to identify specific suspicious objects already identified in some evidences from large volume of footage. This section aimed at: (1) to identify specific suspicious objects that already identified from evidence but hope to find more information or behaviors related with these objects. For example, in a crime scene, firearms and knives were found but the investigator need identify how the crime were conducted from video files, the automated methods will be very helpful to extract related clips. (2) In forensic investigation, it is possible that frequently used and deadly weapons, such as knife, firearms, etc., can be categorized as suspicious objects that may be used in crime behaviors. It will be very useful in investigation to quickly identify these objects, e.g., to identify objects held in a hand might be an example of a sign of danger or crime activity. Specifically, the automated weapons detection over real-time video (such as CCTV system) can be very useful in crime alerts or quick response.

Basic steps: (1) extract frames from video; (2) back ground detection; (3) Canny edge detection; (4) Sliding window, Scaling, PCA, and NN; (5) Candidate regions; (6) MPEG classification; (7) Filter and decision making.

$$e_i = log_{10}[1 + p_i] \tag{7}$$

1) COLOR AND SHAPE-BASED OBJECT DETECTION AND TRACKING

The advances in computational capability and graphic processing using OpenCV and GPU make it possible to conduct



(a) (b) (c)

FIGURE 6. Color and shape based object detection over webcam. (a) Original frame. (b) Enhanced multiple objects recognition. (c) Enhanced single objects recognition.

specific object detection and tracking on video or even live camera. In this subsection, we will focus on the color and shape-based object detection using the opencv and cuda. Since the red-green-blue (RGB) color space cannot well separate color information from intensity information, so we use the Hue Saturation Value (HSV) and YCrCB for the object detection and tracking, in which each color has a specific range in the hue channel that can be utilized for its detection. Algorithm 2 details the procedures of color and shape-based objects detection and tracking in forensic video analysis.

Algorithm 2 Color and Shape-based Object Detection and Tracking Algorithm

Input: Video or camera input V

Output: Labeled $\overline{V}$

```
v ← video_capture(V)
v_g ← video_group(v)
for each frame f_i ∈ v_g do
    f_hsv ← covColor(f_i, COLOR_RGB2HSV)
    f_shsv ← split(f_hsv)
    for each channel c ∈ {hue, saturatin, value} do
        threshld(f_shsv[c], d_th[c], low[c], upper[c])
    end for
    f_result ← bitwise_and(d_th[0], d_th[1], d_th[2])
end for
V̄ ← update(v, v̄_g)
```

In Algorithm 2, we apply features provided by the `OpenCV` and `CUDA` to make it can work with real-time video streaming from camera. Figure 2 shows an example to use proposed algorithm to identify objects over a webcam. It can be seen that in Figure 2 (b), multiple objects can be identified, however, in Figure 2 (c) specific object can be recognized.

2) DEEP LEARNING BASED KEY OBJECTS DETECTION IN FOOTAGE

In many case, the color of a specific object in video is not stable due to multiple factors, such as the change of light, quality of video, other objects with similar color, *etc.*, can significantly affect the identification. Fortunately, the evolutionary of machine learning can offer much better identification performance.

In recent, the *deep learning* shows great advantages in image-based object detection [35]. In forensic investigation, the traditional methods usually manually review and extract related frames and features from available footage, which requires experiences through years of accumulation. The emerging deep learning techniques can automatically extract and learn features through a large amount of existing data. A well designed and trained deep learning model can accurately analyze features of video and image. The current popular object detection models include: convolution neural network (CNN) [43], region - CNN (RCNN), single shot multibox detector, fast-RCNN, You Only Look Once (Yolo), resNet, *etc.* In this works, we introduce (Yolo) network model in our real-time video analysis to detect suspicious activities, events, image, or frames in footage.

In [44], a trained Yolo model can predict bounding boxes using dimension clusters as anchor boxes, in which each bounding box can be coordinated using its centre coordinates, weight, and height, as (t_x, t_y, t_w, t_h). With the bounding box prior width and height p_w, p_h, the predictions can be derived by Eq.8

$$\begin{aligned} b_x &= \sigma(t_x) + c_x \\ b_y &= \sigma(t_y) + c_y \\ b_w &= p_w e^{t_w} \\ b_h &= p_h e^{t_h} \end{aligned} \tag{8}$$

With Yolov3, for each bounding box we can predict its ''objectness score'' using logistic regression [44]. Each box predicts the classes the bounding box may contain using multi-label classification. In this work, we apply the Yolo v3 to model the detection of suspicious objects in crime scene. Basically, the detection procedure includes following four phases:

1) **Frame picker**, this phase picks out objects from existing objects image set, which means in forensic investigation we first needs to identify suspicious objects in crime scene and take as much labeled pictures as possible, and picks up images of objects within the class and crops the annotated regions; this can be done automatically using scripts;
2) **Scale modifier**, the main aim of this phase to reduces the size of objects cropped in first phase to an

(a) (b) (c)

FIGURE 7. Object detection and tracking using deep learning algorithms (Yolo v3). (a) Image objects detection. (b) CCTV footage object detection. (c) Specific object detection image.

TABLE 1. Yolo v3 based object recognition.

Classification	Avg. Reg. Accurate	Avg Time(ms)
Person	97%	16.32
Knife	86.4%	16.29
Truck	89.6%	16.06
Bicycle	95.4%	16.71
Horse	92.1%	21.50
Car	96.3%	15.48
Blood	86%	17.99
Fire	84%	16.81
Gun, trifle, machine gun	92.1%	16.58

appropriate size, to make the features learning more efficient;

3) **Object marker**, in this phase the modified image is pasted into base images with similar backgrounds and size;
4) **Annotation creator**, this phase mainly focuses on annotating the position and size of newly placed objects over base image. The new created annotation annotations can be calculated based on the tasks applied in image marker phase.

3) USE CASE

Figure 7 shows examples to detect and recognize person, objects, cars, from good quality image, cctv footage, respectively. Specific, we trained a model that can recognize dangerous tools, such as knives, guns, firearms, *etc.*

In this work, based on the Yolo v3 model, a knife, gun, and firearms model are trained which can be used to identify 'knife, gun, and other firearms' in video. Figure 7 shows an example of identify and track the knife, it is clear that the model can be used to identify knife based on a pre-trained model. Table 1 shows the object detection accurate and time consuming for the detection. Since the Yolov3 is a well-designed framework which can be easily used in video detection using: `./darknet detector demo cfg/coco.data cfg/yolov3.cfg yolov3.weights <video file>`, and real-time detection on a web-online camera or even powerful mobile devices `./darknet detector demo cfg/coco.data cfg/yolov3.cfg yolov3.weights <video file>`.

In our experiments, an important issue is when we change the direction of camera, the detect result will be different, as shown in Figure 8(a), only a person was detected when we turned the camera -90 degree. In Figure 8(b), it is noted that a remote 'person', a 'bottle', and a 'bird' were detected, however, the model fails to detect the 'person'.

(a) (b)

FIGURE 8. Object detection in video with different angle. (a) Labeled Image. (b) Origin Image.

(a) (b)

FIGURE 9. Low quality of video can significantly affect the detection performance. (a) Labeled Image. (b) Origin Image.

In many cases, it is very difficult to identify related suspects, objects, or scenarios that related with a specific case using low quality cctv footage. Even if the Yolo3 model is well trained, due to the low quality of the video, the detection rate is pretty low. In Figure 9, we use the same trained model, when it is used to detect video with different quality, it is obvious that in Figure 9(a), the system is failed even to identify the main suspect but in the Figure 9(b), all persons were identified.

It can be obtained by using the Yolo as: `./darknet detect cfg/yolov3-crime.cfg yolov3-crime.weights`.

E. IMPROVE THE YOLO V3

In our test scenarios, it is noted that the average people recognition rate is pretty high, however when using the

(a) (b)

FIGURE 10. Deep learning based object tracking in real-time video. (a) Labeled Image. (b) Origin Image.

yolov3.weights, the recognition for 'knife' is pretty low, which means a new model is necessary for object detection in crime scenes. Another issue is the recognition can be improved when using a high quality images or video. As shown in Fig 10.

It is noted that if the knife matches the one found in crime scenes, then the holder in the video will be a suspect that might have links with the case. By doing this, or knife related subjects, environment, scenarios, and timeline will be identified, which will be very useful for the case investigation.

VI. CONCLUSION

It is noted that in digital forensic investigation, the low quality cctv footages are widely used to extract potential evidence items. In this work, we proposed a framework for video based digital forensics investigation, and further we developed a way to enhance the quality of video to extract as much as evidence items. Specifically, we proposed an method to extract more evidence items in a reverse way. It is also useful for anti-crime or fast response when crime activities or behaviors are detected. In the future works, we will further establish the links between existing evidence items and the detected evidence item.

REFERENCES

[1] 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," in *Proc. 7th Int. Conf. Cloud Comput., Data Sci. Eng.-Confluence*, Jan. 2017, pp. 649–655.

[2] G. Gilboa, N. Sochen, and Y. Y. Zeevi, "Image enhancement and denoising by complex diffusion processes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 8, pp. 1020–1036, Aug. 2004.

[3] S. Park, S. Yu, M. Kim, K. Park, and J. Paik, "Dual autoencoder network for retinex-based low-light image enhancement," *IEEE Access*, vol. 6, pp. 22084–22093, 2018.

[4] W. Fan, K. Wang, C. François, and Z. Xiong, "Median filtered image quality enhancement and anti-forensics via variational deconvolution," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 5, pp. 1076–1091, May 2015.

[5] C.-Y. Li, J.-C. Guo, R.-M. Cong, Y.-W. Pang, and B. Wang, "Underwater image enhancement by dehazing with minimum information loss and histogram distribution prior," *IEEE Trans. Image Process.*, vol. 25, no. 12, pp. 5664–5677, Dec. 2016.

[6] S. Mandal, X. L. Deán-Ben, and D. Razansky, "Visual quality enhancement in optoacoustic tomography using active contour segmentation priors," *IEEE Trans. Med. Imag.*, vol. 35, no. 10, pp. 2209–2217, Oct. 2016.

[7] S. Kim, W. Kang, E. Lee, and J. Paik, "Wavelet-domain color image enhancement using filtered directional bases and frequency-adaptive shrinkage," *IEEE Trans. Consum. Electron.*, vol. 56, no. 2, pp. 1063–1070, May 2010.

[8] G. Tzanidou, I. Zafar, and E. A. Edirisinghe, "Carried object detection in videos using color information," *IEEE Trans. Inf. Forensics Security*, vol. 8, no. 10, pp. 1620–1631, Oct. 2013.

[9] H. Walker and A. Tough, "Facial Comparison from CCTV footage: The competence and confidence of the jury," *Sci. Justice*, vol. 55, no. 6, pp. 487–498, 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1355030615000635

[10] E. Verolme and A. Mieremet, "Application of forensic image analysis in accident investigations," *Forensic Sci. Int.*, vol. 278, pp. 137–147, Sep. 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0379073817302463

[11] D. Seckiner, X. Mallett, C. Roux, D. Meuwly, and P. Maynard, "Forensic image analysis—CCTV distortion and artefacts," *Forensic Sci. Int.*, vol. 285, pp. 77–85, Apr. 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0379073818300380

[12] R. C. Pandey, S. K. Singh, and K. K. Shukla, "Passive forensics in image and video using noise features: A review," *Digit. Invest.*, vol. 19, pp. 1–28, Dec. 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1742287616300809

[13] L. J. G. Villalba, A. L. S. Orozco, R. R. López, and J. H. Castro, "Identification of smartphone brand and model via forensic video analysis," *Expert Syst. Appl.*, vol. 55, pp. 59–69, Aug. 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S095741741600035X

[14] J. Kamenicky *et al.*, "PIZZARO: Forensic analysis and restoration of image and video data," *Forensic Sci. Int.*, vol. 264, pp. 153–166, Jul. 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0379073816301827

[15] L. G. Moré, M. A. Brizuela, H. L. Ayala, D. P. Pinto-Roa, and J. L. V. Noguera, "Parameter tuning of CLAHE based on multi-objective optimization to achieve different contrast levels in medical images," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2015, pp. 4644–4648.

[16] J. Li, B. Ma, and C. Wang, "Extraction of PRNU noise from partly decoded video," *J. Vis. Commun. Image Represent.*, vol. 57, pp. 183–191, Nov. 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1047320318302670

[17] I. Amerini, R. Caldelli, A. D. Mastio, A. Di Fuccia, C. Molinari, and A. P. Rizzo, "Dealing with video source identification in social networks," *Signal Process., Image Commun.*, vol. 57, pp. 1–7, Sep. 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0923596517300759

[18] T. Gloe, A. Fischer, and M. Kirchner, "Forensic analysis of video file formats," *Digit. Invest.*, vol. 11, pp. S68–S76, May 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1742287614000140

[19] P.-C. Su, P.-L. Suei, M.-K. Chang, and J. Lain, "Forensic and anti-forensic techniques for video shot editing in H.264/AVC," *J. Vis. Commun. Image Represent.*, vol. 29, pp. 103–113, May 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1047320315000309

[20] G. Horsman, "Reconstructing streamed video content: A case study on YouTube and Facebook live stream content in the Chrome Web browser cache," *Digit. Invest.*, vol. 26, pp. S30–S37, 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1742287618301932

[21] R. Jenkins and C. Kerr, "Identifiable images of bystanders extracted from corneal reflections," *PLoS ONE*, vol. 8, no. 12, 2013, Art. no. e83325.

[22] K. Maksymowicz, W. Tunikowski, and J. Kościuk, "Crime event 3D reconstruction based on incomplete or fragmentary evidence material—Case report," *Forensic Sci. Int.*, vol. 242, pp. e6–e11, Sep. 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0379073814002886

[23] M. Grega, A. Matiolański, P. Guzik, and M. Leszczuk, "Automated detection of firearms and knives in a CCTV image," *Sensors*, vol. 16, no. 1, p. 47, 2016.

[24] D. M. Jang and M. Turk, "Car-Rec: A real time car recognition system," in *Proc. IEEE Workshop Appl. Comput. Vis. (WACV)*, Jan. 2011, pp. 599–605.

[25] L. Zeng, J. Chen, L. Tong, B. Yan, and X. Ping, "Image contrast enhancement based on histogram similarity," in *Proc. IEEE Int. Conf. Med. Imag. Phys. Eng.*, Oct. 2013, pp. 269–273.

[26] S. Li, S. Li, K.-K. R. Choo, Q. Sun, W. J. Buchanan, and J. Cao, "IoT forensics: Amazon echo as a use case," *IEEE Internet Things J.*, to be published.

[27] S.-D. Chen and A. R. Ramli, "Minimum mean brightness error bi-histogram equalization in contrast enhancement," *IEEE Trans. Consum. Electron.*, vol. 49, no. 4, pp. 1310–1319, Nov. 2003.

[28] Y. Wang, Q. Chen, and B. Zhang, "Image enhancement based on equal area dualistic sub-image histogram equalization method," *IEEE Trans. Consum. Electron.*, vol. 45, no. 1, pp. 68–75, Feb. 1999.

[29] Y.-T. Kim, "Contrast enhancement using brightness preserving bi-histogram equalization," *IEEE Trans. Consum. Electron.*, vol. 43, no. 1, pp. 1–8, Feb. 1997.
[30] J. Jasmine and S. Annadurai, "Real time video image enhancement approach using particle swarm optimisation technique with adaptive cumulative distribution function based histogram equalization," *Measurement*, to be published.
[31] T. Ayyavoo and J. J. Suseela, "Illumination pre-processing method for face recognition using 2D DWT and CLAHE," *IET Biometrics*, vol. 7, no. 4, pp. 380–390, Jul. 2018.
[32] A. Hendrawan and S. Asmiatun, "Identification of picnosis cells using contrast-limited adaptive histogram equalization (CLAHE) and K-means algorithm," in *Proc. 1st Int. Conf. Comput. Appl. Inf. Secur. (ICCAIS)*, Apr. 2018, pp. 1–3.
[33] LEVA. *LEVA Certification Program*. Assessed: May 1, 2019. [Online]. Available: https://leva.org/index.php/certification
[34] S. Li, Q. Sun, and X. Xu, "Forensic analysis of digital images over smart devices and online social networks," in *Proc. IEEE 20th Int. Conf. High Perform. Comput. Commun. (HPCC/SmartCity/DSS)*, Jun. 2018, pp. 1015–1021.
[35] S. Saikia, E. Fidalgo, E. Alegre, and L. Fernández-Robles, "Object detection for crime scene evidence analysis using deep learning," in *Image Analysis and Processing*, S. Battiato, G. Gallo, R. Schettini, and F. Stanco, Eds. Cham, Switzerland: Springer, 2017, pp. 14–24.
[36] NCA. *Forensic Image Comparison and Interpretation Evidence: Guidance for Prosecutors and Investigators Issue 2*. [Online]. Available: https://assets.publishing.service.gov.uk/
[37] X. Liu, W. Lu, Q. Zhang, J. Huang, and Y.-Q. Shi, "Downscaling factor estimation on Pre-JPEG compressed images," *IEEE Trans. Circuits Syst. Video Technol.*, to be published.
[38] BBC. *Chinese Man Caught by Facial Recognition at Pop Concert*. [Online]. Available: https://www.bbc.co.uk/news/world-asia-china-43751276
[39] Y. Chang, C. Jung, P. Ke, H. Song, and J. Hwang, "Automatic contrast-limited adaptive histogram equalization with dual gamma correction," *IEEE Access*, vol. 6, pp. 11782–11792, 2018.
[40] N. M. Ghanem and L. S. Davis, "Human appearance change detection," in *Proc. 14th Int. Conf. Image Anal. Process. (ICIAP)*, Sep. 2007, pp. 536–541.
[41] C. H. Chuang, J. W. Hsieh, L. W. Tsai, S. Y. Chen, and K. C. Fan, "Carried object detection using ratio histogram and its application to suspicious event analysis," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 6, pp. 911–916, Jun. 2009.
[42] Z. Wang and J. Tao, "A fast implementation of adaptive histogram equalization," in *Proc. 8th Int. Conf. Signal Process.*, vol. 2, Nov. 2006, pp. 1–4.
[43] H.-J. Jeong, K.-S. Park, and Y.-G. Ha, "Image preprocessing for efficient training of YOLO deep learning networks," in *Proc. IEEE Int. Conf. Big Data Smart Comput. (BigComp)*, Jan. 2018, pp. 635–637.
[44] J. Redmon and A. Farhadi. (2018). "YOLOv3: An incremental improvement." [Online]. Available: https://arxiv.org/abs/1804.02767

JIANYU XIAO received the Ph.D. degree in computer science from Xi'an Jiaotong University, China, in 2006. He is currently a Professor with the School of Computer Science and Engineering, Central South University, China. He is also with Central South Ultrapower Infotech Co. Ltd., Changsha, China. His current research interests include cyber security, healthcare, online social network, and so on.

SHANCANG LI is currently a Senior Lecturer with the Department of Computer Science and Creative Technologies, University of the West of England, Bristol, U.K. His current research interests include digital forensics for emerging technologies, cyber security, the IoT security, data privacy-preserving, the Internet of Things, Blockchain technology, and the lightweight cryptography in resource constrained devices. He has authored over 60 papers published in high profile journals and conferences. He is an Associate Editor of the IEEE Access and the *Journal of Industrial Information Integration*. He is a member of the British Computer Society.

QINGLIANG XU is currently pursuing the master's degree with Central South University. His research interests include embedded systems, cloud computing, and bigdata.