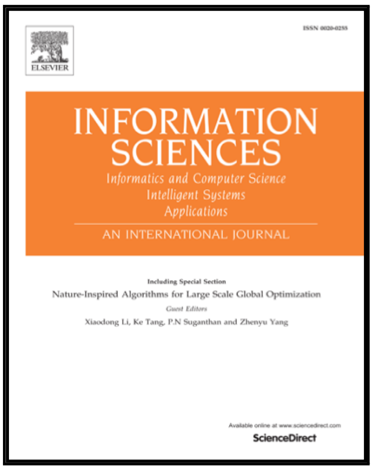


ISSN 0020-0255

ELSEVIER

INFORMATION SCIENCES

Informatics and Computer Science
Intelligent Systems
Applications

AN INTERNATIONAL JOURNAL

Including Special Section
Nature-Inspired Algorithms for Large Scale Global Optimization
Guest Editors
Xiaodong Li, Ke Tang, P.N Suganthan and Zhenyu Yang

Available online at www.sciencedirect.com
ScienceDirect

A Three-stage Framework for Smoky Vehicle Detection in Traffic Surveillance Videos

Huanjie Tao , Peng Zheng , Chao Xie , Xiaobo Lu

PII: S0020-0255(20)30136-5
DOI: https://doi.org/10.1016/j.ins.2020.02.053
Reference: INS 15237

To appear in: *Information Sciences*

Received date: 12 December 2018
Revised date: 9 February 2020
Accepted date: 18 February 2020

Please cite this article as: Huanjie Tao , Peng Zheng , Chao Xie , Xiaobo Lu , A Three-stage Framework for Smoky Vehicle Detection in Traffic Surveillance Videos, *Information Sciences* (2020), doi: https://doi.org/10.1016/j.ins.2020.02.053

A Three-stage Framework for Smoky Vehicle Detection in Traffic Surveillance Videos

Huanjie Tao [1], Peng Zheng [2*], Chao Xie[3], Xiaobo Lu [4,5*]

[1] College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou, 310023, China.
[2] School of Mechanical Engineering, Zhengzhou University, Zhengzhou, 450001, China
[3] College of Mechanical and Electronic Engineering, Nanjing Forestry University, Nanjing 210037, China.
[4] School of Automation, Southeast University, Nanjing 210096 China.
[5] Key Laboratory of Measurement and Control of Complex Systems of Engineering, Ministry of Education, Southeast University, Nanjing 210096, China.

ARTICLE INFO

Key words
Smoky vehicle detection
Pixel Based Adaptive Segmenter
Local Binary Pattern
Co-occurrence Histograms of Oriented Gradients
Motion Boundary Histograms

ABSTRACT

The smoky vehicle exhaust pollutes the air and endangers human health. Existing methods detecting smoky vehicles in traffic surveillance videos are with high false alarm rates due to the diversity of smoke characteristics and continuous interferences of common vehicles. To solve this issue, this paper presents a three-stage framework for smoky vehicle detection. In the first stage, a Robust Pixel Based Adaptive Segmenter (R-PBAS) algorithm, which adapts to cameras shaking, is proposed to extract moving objects. The Cumulative Color Histogram (CCH) is adopted to extract smoke-colored blocks from moving objects. In the second stage, three groups of features, including Non-Redundant Robust Local Binary Pattern (NR-RLBP), Weighted Co-occurrence Histograms of Oriented Gradients (W-CoHOG), and Motion Boundary Histograms (MBH) are proposed to extract texture, gradient, and motion information from smoke-colored blocks, respectively. In the third stage, we fuse smoke blocks to obtain Region of Interest (ROI) and extract frequency domain features based on Discrete Wavelet Transform (DWT). To further improve robustness, the Auto-Regressive and Moving Average (ARMA) Model and Hidden Markov Model (HMM) are adopted to model ROIs in consecutive frames. Extensive experiments show that our method performs better than existing methods.

1. introduction

Smoky vehicle generally refers to the vehicle with black smoke (diesel exhaust particles) exhausted from vehicle exhaust hole. Such vehicles usually use turbocharged diesel engines and have a high proportion in exhaust particles pollution (PM2.5, PM10) [1] among all the motor vehicles. Fig.1 shows a typical smoky vehicle, and the black smoke in the vehicle rear pollutes the air and endangers human health seriously. This paper focuses on automatic smoky vehicle detection in traffic surveillance videos to enforce environmental laws and regulations.

The tradition smoky vehicle detection methods mainly include: the public reporting, regular road inspection and night inspection by the law enforcement workers, installing vehicle exhaust analysis devices, sensor detection, and manual video monitoring that watching traffic surveillance videos to select smoky vehicles, etc. These methods reduce the pollution of smoky vehicles to a certain extent, but due to the rapid growth of vehicle ownerships and the heavy traffic, lots of workers need to be employed, and the purchase and maintenance of vehicle exhaust analysis devices are also costly.

∗Corresponding authors.
E-mail addresses: huanjie_tao@126.com (Huanjie Tao), zpzzut@163.com (Peng Zheng), xblu2013@126.com (Xiaobo Lu), chaoxie@njfu.edu.cn (Chao Xie).

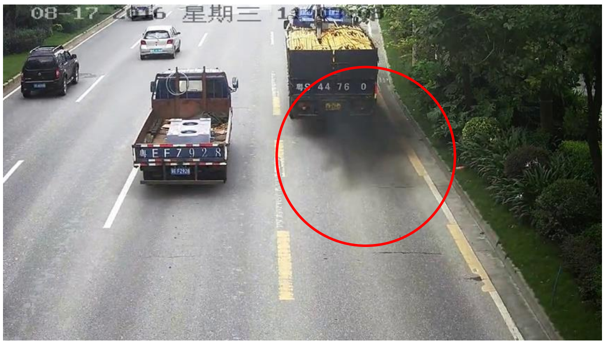


Fig. 1. A smoky vehicle with visible black smoke from its vehicle exhaust pipe. The region marked by the red circle is the smoke position.

The aim of this paper is to develop an automatic smoky vehicle detection system to select and identify vehicles that produce visible smoke. When the system detects a smoky vehicle, it sends a stationary alarm to workers for more detailed measurement. With advanced object detection technology in computer vision, automatic smoky vehicle detection has become possible. However, it is still at its infancy stage, and many existing methods are still with high false alarm rates due to the diversity of smoke characteristics and the continuous interferences of common vehicles.

The features used to characterize smoke is the core of a smoky vehicles detection system. Flora et al. [2] first proposed to estimate carbon emissions caused by vehicular traffic on highway systems based on morphological properties and histogram of oriented gradient (HOG). An adaptive Gaussian mixture model (AGMM) is used to segment vehicles. Pyykonen [3] proposed a smoke detection and traffic pollution analysis system using a far infrared camera and a high-resolution visible wavelength camera. The image graininess, intensity values and histogram are used as features. Banerjee [4] and Zhao [5] gave surveys on the IoT-based (Internet of Things) traffic management systems and the technologies that involves in developing time critical cloud applications. Related methods can be embedded in smoky vehicle detection system by scaling computer vision in the cloud and validating using IoT smoke sensors. In [6-13], a series of smoky vehicle detection methods based on various features are proposed. However, due to the complex road environment, various smoke characteristics, and continuous interferences, etc., the above methods are still with high false alarm rates.

The essence of smoky vehicle detection is smoke detection. Therefore, we discuss it from the perspective of related smoke detection works.

To date, there are lots of forest smoke detection literatures. More specifically, Labati et al [14] proposed a wildfire smoke detection method using computational intelligence techniques enhanced with synthetic smoke plume generation. Gunay et al. [15] proposed an online adaptive decision fusion framework with application to video smoke detection. Decision values of several sub-algorithms are linearly combined with weights updated online. Tian et al [16, 17] proposed to detect and separate smoke from a single frame to obtain the quasi-smoke and quasi-background components based on the atmospheric scattering models and sparse representation. Dimitropoulos et al [18] proposed a video smoke detection method based on higher order linear dynamical systems (h-LDS) to enable the dynamic textures analysis by using the information from various image elements. Yuan et al [19] proposed a method of smoke detection and image classification based on high-order local ternary patterns (HLTP) with locality preserving projection. Yuan [20] proposed a double mapping framework for video smoke detection based on various features including edge orientation, edge magnitude, and local binary pattern (LBP) bit, etc. Chen et al [21] proposed a fast video flame detection method based on the temporal and spatial characteristics of flames, such as ordinary flame movement and color clues, etc. Appana et al. [22] proposed to detect smoke based on optical smoke flow pattern analysis and spatial-temporal energy analysis.

Yin et al [23] proposed a smoke detection method based on deep convolutional recurrent motion-space network (RMSN) to capture the space and motion context information. Hu et al. [24] proposed a multi-task CNN architecture for video smoke recognition using motion information between neighbor frames.

In addition, Yu et al. [25] proposed a video smoke detection method using color and motion features. Yuan [26] extracted an effective feature vector based on the histogram sequences of LBP pyramids and LBP variance (LBPV) pyramids. The shape-variant features may reduce generalization performances. The smoke color is always used as a cue in various color spaces by converting the RGB space into the HSV [27], YUV [28] or YCbCr [29] color spaces. However, it has trouble in differentiating moving objects that are similar in color to the smoke. Barmpoutis et al. [27] proposed to detect smoke based on spatial-temporal wavelet analysis and dynamic texture analysis by means of mathematical models. Filonenko et al [30] proposed a smoke detection method based on shape features and color information. Calderara et al. [31] proposed a smoke detection method based on image energy and color information. Ko et al. [32] proposed an early wildfire smoke detection method based on spatiotemporal bag-of-features.

Although the above methods have achieved great successes in forest smoke detection, it is still challenging to detect smoky vehicles from visual scenes due to the large variations in color, texture, shapes of smoke, and continuous interferences of common vehicles. The forest smoke detection and smoky vehicle detection have many differences. First, forest smoke color is usually creamy white, which has obvious differences with the surrounding green trees. But the smoke color of the vehicle exhaust is usually black, which has non-obvious differences with the surrounding black road surface and black shadows. Second, in forest smoke detection, the objects easily with false alarms include: white clouds at the skyline, swaying leaves, moving shadows, small white house and moving vehicles in the mountainside, fog and vapor, etc. But these interferences objects not always occur. In smoky vehicle detection, the objects easily with false alarms include: black shadows of vehicles or roadside trees, black road surfaces, and moving smoke-colored vehicles, etc. These interference objects occur all the time and also connect with smoke.

To sum up, the main challenges in smoky vehicle detection include: 1) Large variations in color, texture, shapes of smoke, especially when the smoke is light and small; 2) Continuous interferences of black shadows of vehicles, black road surfaces, and moving vehicles etc. 3) Some complex situations, such as the smoke covered by vehicle objects or blew by wind, night time and bad weather, etc.

To further reduce false alarm rates (FAR) while maintaining relative high detection rates (DR), we propose a three-stage framework for smoky vehicle detection in traffic surveillance video.

The main contributions of this work are as follows:

(1) We propose a Robust Pixel Based Adaptive Segmenter (R-PBAS) algorithm, which is robust to cameras shaking and slowly moved objects.

(2) We propose the Non-Redundant Robust Local Binary Pattern (NR-RLBP) descriptor to characterize the texture information of smoke-colored blocks. It is insensitive to noise and also robust to relative changes between foreground and background.

(3) We propose the Weighted Co-occurrence Histograms of Oriented Gradients (W-CoHOG) descriptor to characterize gradient information. It not only encodes gradient orientations of neighbouring pixel pairs to capture spatial and contextual information, but also adds gradient magnitude information.

(4) To further improve robustness, the Motion Boundary Histograms (MBH) descriptor is employed to characterize motion information while maintaining resistances to camera shaking. The auto-regressive and moving average (ARMA) model and hidden Markov model (HMM) are employed to model dynamic information of consecutive frames.

The remainder of this paper is organized as follows: The first stage, smoke-colored blocks detection, is provided in section 2. The second stage, smoke blocks detection using three groups of features, is introduced in section 3. The third stage, smoke frames detection, is introduced in section 4. The whole algorithm framework is provided in section 5. Experiments and analysis are provided in section 6. The part of discussion and conclusion is provided in section 7.

2. First stage: smoke-colored blocks detection

In this section, we introduce R-PBAS algorithm and use it to extract moving objects. Then, we introduce the cumulative color histogram (CCH) features to extract smoke-colored blocks.

2.1. Camera installation position

To detect vehicle objects and recognize license plates simultaneously for automatic smoky vehicle alarm, the high-resolution cameras can be installed on a lamp post, beneath an overhead bridge or on other similar structures. Fig.2 shows a schematic diagram of the camera location.

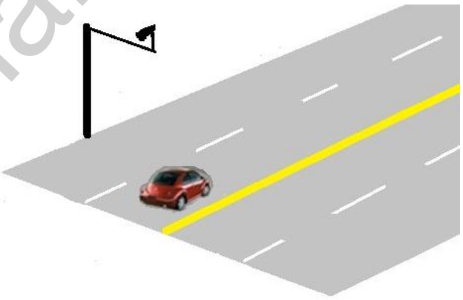


Fig.2. A schematic diagram of the camera location. A recorded high-resolution frame from traffic surveillance videos can be seen in Fig.1.

2.2. Moving object detection based on robust pixel-based adaptive segmenter (R-PBAS)

To narrow search range for smoke recognition in the whole frame, we improve the Pixel-Based Adaptive Segmenter (PBAS) [33] algorithm and use it to extract moving objects.

The PBAS algorithm is not robust to dynamic scenes. It usually leads a low detection accuracy under camera shaking, and the slow-moving objects may be updated to the background. In smoky vehicle detection, the cameras installed around the road shake inevitably when vehicles passed by or in a windy weather, which leads to instantaneous changes in the scene. In addition, the smoke and vehicle far away seem to be moved slowly in the whole frame, which may be updated as the background.

To solve the above problems, we propose a Robust PBAS (R-PBAS) algorithm. The main contributions include: 1) We improve original background complexity by combining regional features; 2) We introduce counting mechanism to prevent slowly moved objects being updated as background; 3) We improve the decision thresholds to suppress the interferences caused by the cameras shaking and leaves wobble of road trees.

The steps of original PBAS algorithm [33] are summarized below:

(1) Moving object detection

The background model $B(x_i)$ is defined by an array of N recently observed pixel values:

$$B(x_i) = \{B_1(x_i), B_2(x_i), ..., B_N(x_i)\} \tag{1}$$

where x_i is the ith pixel. $B(x_i)$ is the background model. $B_k(x_i)$ are the kth pixel values from the background model of pixel x_i.

To classify a new pixel x_i, the following formula is used:

$$F(x_i) = \begin{cases} 1, & N\{dist(I(x_i), B_k(x_i)) < R(x_i)\} < N_{\min} \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

$$dist(I(x_i), B_k(x_i)) = \frac{\alpha}{\overline{I^m}} \left| I^m(x_i) - B_k^m(x_i) \right| + \left| I^v(x_i) - B_k^v(x_i) \right| \tag{3}$$

where $I(x_i)$ denotes the intensity value of pixel x_i in frame I. $dist(I(x_i), B_k(x_i))$ denotes the pixel distance between pixel $I(x_i)$ and pixel $B_k(x_i)$. $R(x_i)$ is a distance threshold of pixel x_i. $N\{*\}$ denotes the number of samples that satisfy the rules marked by symbol *. $N_{\min}$ is a threshold of the sample number. $F(x_i) = 1$ denotes that pixel x_i is a foreground pixel. $\overline{I^m}$ is average gradient magnitude over the last observed frame. $I^m(x_i)$ and $I^v(x_i)$ are the pixel value and gradient magnitude, respectively.

(2) Update of background model

We choose a certain index $k \in 1, 2, ..., N$ uniformly at random, and replace corresponding background model value $B_k(x_i)$ by the current pixel value $I(x_i)$. The above update is only performed with probability $p = 1/T(x_i)$ .Otherwise no update is carried out at all. The parameter $T(x_i)$ is called learning rate. We also randomly update a chosen neighboring pixel with the probability p. The background model $B_k(y_i)$ of this neighboring pixel is replaced by its current pixel value.

The decision threshold and learning rate are updated as follows:

$$R(x_i) = \begin{cases} R(x_i)(1 - R_{inc/dec}), & \text{if } R(x_i) > \bar{d}_{\min}(x_i) R_{scale} \\ R(x_i)(1 + R_{inc/dec}), & \text{otherwise} \end{cases} \tag{4}$$

$$T(x_i) = \begin{cases} T(x_i) + T_{inc} / \bar{d}_{\min}(x_i), & \text{if } F(x_i) = 1 \\ T(x_i) - T_{dec} / \bar{d}_{\min}(x_i), & \text{if } F(x_i) = 0 \end{cases} \tag{5}$$

$$\bar{d}_{\min}(x_i) = \frac{1}{N} \sum_k D_k(x_i) \tag{6}$$

$$d_{\min}(x_i) = \min_k dist[I(x_i), B(x_i)] \tag{7}$$

where $R_{inc/dec}$, R_{scale}, T_{inc} and T_{dec} are fixed parameters. T_{lower} and T_{upper} are the lower and upper bound of T. $d_{\min}(x_i)$ is the minimum pixel distance between x_i and background pixels. $D(x_i) = \{D_1(x_i), ..., D_N(x_i)\}$ is an array to save the minimal decision distances $d_{\min}(x_i)$.

In the above steps, if the object area is large or moving slowly, an incomplete foreground will be obtained. In addition, it is easily disturbed by the camera and leaves shaking.

To improve these problems, we propose a new parameter update mode as follows:

$$R(x_i) = \begin{cases} R(x_i) - 1/S(x_i), & \text{if } R(x_i) > [1 + 2C_{com}(x_i)]^2 \\ R(x_i) + S(x_i), & \text{otherwise} \end{cases} \tag{8}$$

$$T(x_i) = \begin{cases} T(x_i) + T_{inc}/\bar{d}_{\min}(x_i) + T_{inc}/\beta G(x_i) + T_{inc}/\gamma C(x_i), & \text{if } F(x_i) = 1 \\ T(x_i) - T_{dec}/\bar{d}_{\min}(x_i) + T_{dec}/\beta G(x_i) + T_{dec}/\gamma C(x_i), & \text{if } F(x_i) = 0 \end{cases} \tag{9}$$

where $G(x_i)$, $C(x_i)$ and $C_{com}(x_i)$ are background complexities of the regional structure, regional color, and fused information, respectively. They are given by:

$$G(x_i) = \sqrt{E_v^2(x_i) + E_h^2(x_i)} \tag{10}$$

$$C(x_i) = \frac{1}{N_{local} \times N_{local} - 1} \sum_{x_n \in Z(x_i)} \left| I^c(x_i) - I^c(x_n) \right| \tag{11}$$

$$C_{com}(x_i) = \bar{d}_{\min}(x_i) + \beta G(x_i) + \gamma C(x_i) \tag{12}$$

where $Z(x_i)$ denotes a local region with $N_{local} \times N_{local}$ size. $E_v(x_i)$ and $E_h(x_i)$ are two images after convolution using the Sobel operators in horizontal directions and vertical directions, respectively.

A foreground counting mechanism is also proposed. Only when a pixel is judged as foreground pixel many times, it can be gradually upgraded to background. Let $NCO_t(x_i)$ denotes the counting parameter, the updating mechanism is defined as follows:

$$NCO_t(x_i) = \begin{cases} NCO_{t-1}(x_i) + 1, & \text{if } F(x_i) = 1 \\ 0, & \text{otherwise} \end{cases} \tag{13}$$

$$NCO_t(x_i) = NOC_{\max} \Rightarrow F(x_i) = 0 \tag{14}$$

where $NOC_{\max}$ denotes the longest life cycle of a slowly moved foreground object.

2.3. Smoke-colored blocks detection

To facilitate temporal information analysis, we evenly divide the whole frame into blocks, and each block is $b_{block} \times b_{block}$ pixels ($b_{block} = 32$ is used). We first extract blocks that covered by moving objects, and then remove the blocks that with lower proportion of foreground objects, i.e.,

$$\text{Rule 1: } S_{fore\_block} / (b_{block} \times b_{block}) > T_{fore} \tag{15}$$

where S_{fore_block} is the area of foreground object in current block. T_{fore} is a threshold with a recommended range [0.1, 0.4].

The color information is a useful cue in smoke recognition. We take it as a sieve to remove the obvious non-smoke blocks to reduce computation amount and improve algorithm speed.

Color histogram is commonly used in image retrieval and recognition with low computation complexity. Because smoke blocks tend to be low gray value levels and cannot contain all grayscale levels. Therefore, the cumulative color histogram (CCH) is introduced as color features. The support vector machine (SVM) is used as classifier.

$$F_{CCH} = \{CH_1, CH_2, CH_3\} \tag{16}$$

$$CH_i(j) = \sum_{k=1}^{j} \frac{n_i(k)}{3N}, i = 1, 2, 3. j = 1, 2, ..., L. \tag{17}$$

where N denotes the pixel number of the block. $n_i(k)$ denotes the pixels number in k th gray level and i th color channel $i(i = 1: R, 2: G, 3: B)$. L denotes total color bins. F_{CCH} denotes the cumulative color histogram.

Fig. 3 shows an example of smoke-colored blocks detection. The blue blocks mean that it is covered by moving objects. If a blue block is then detected as smoke-colored block, the blue block will be replaced by a green block. We can see that about 30% of the blocks covered by moving object are removed, but some vehicle-body blocks still cannot be removed.

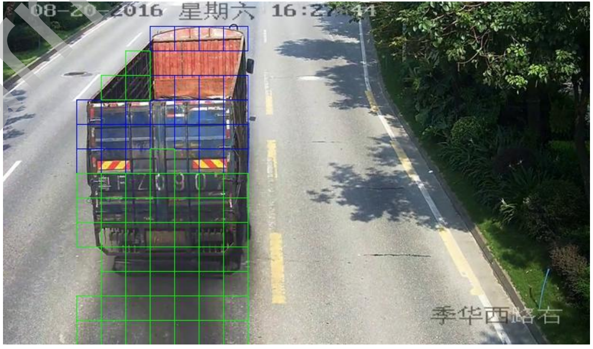


Fig. 3. An example to show the smoke-colored blocks detection. The blue blocks denote the blocks covered by moving objects. The green blocks denote the detected smoke-colored blocks.

3. Second stage: smoke blocks detection using three groups of features

In this section, we introduce three groups of features, including Non-Redundant Robust Local Binary Pattern (NR-RLBP), Weighted Co-occurrence Histograms of Oriented Gradients (W-CoHOG), and Motion Boundary Histograms (MBH), which used to characterize texture, gradient, and motion information, respectively. For each smoke-colored block, three groups of features are extracted to characterize the block and classify it to smoke block or non-smoke block.

3.1. Texture features: non-redundant robust local binary pattern (NR-RLBP)

The original LBP descriptor was first proposed by Ojala [34] and used for texture classification. It captures local appearance information with high discriminative power, low computational complexity and high robustness under illumination changes. Lots works take LBP-based descriptor as features to recognize smoke and have achieved great success, such as LBP [9, 16, 17, 20, 26, 35], LBPV [26], HLTP [19], volume LBP (VLBP) [36], spatio-temporal LBP (STLBP) [37], etc.

However, the above descriptors still have drawbacks: 1) LBP is sensitive to relative changes between foreground and background. As shown in Fig. 4, the LBP codes of local regions marked by blue arrows in two images with the same smoke spatial characteristics are different. 2) The original LBP is sensitive to noise. To overcome the above drawbacks, we propose a new LBP variant called Non-Redundant Robust Local Binary Pattern (NR-RLBP).

First, we briefly introduce the original LBP descriptor [30]. Given a pixel $c=(x_c,\ y_c)$, its LBP code can be obtained by comparing its intensity with those of its neighbors, i.e.,

$$LBP_{P,R}(x_c,y_c)=\sum_{p=0}^{P-1}s(g_p-g_c)2^P \tag{18}$$

$$s(x)=\begin{cases}1, & \text{if } x\geq 0\\ 0, & \text{if } x<0\end{cases} \tag{19}$$

where g_c is the gray value of center pixel. $g_p(p=0,1,...,P-1)$ are gray values of P equally spaced pixels on a circle with radius R.

To make LBP descriptor insensitive to relative changes between foreground and background and also make feature dimension reduced, a simple idea can be easily obtained to define NR1-LBP descriptor [38], as follows:

$$NR1-LBP_{P,R}(x_c,y_c)=\min\left\{LBP_{P,R}(x_c,y_c),2^P-1-LBP_{P,R}(x_c,y_c)\right\} \tag{20}$$

The above idea is based on the fact that the sum of the complementary LBP codes is 2^P-1. As shown in Fig.4, $LBP_{8,1}$ of local (b) in the two smoke images are 00011111 and 11100000, respectively. The two codes are complementary and their sum is 2^8-1. This verifies that $NR1-LBP_{P,R}$ seems to be effective.

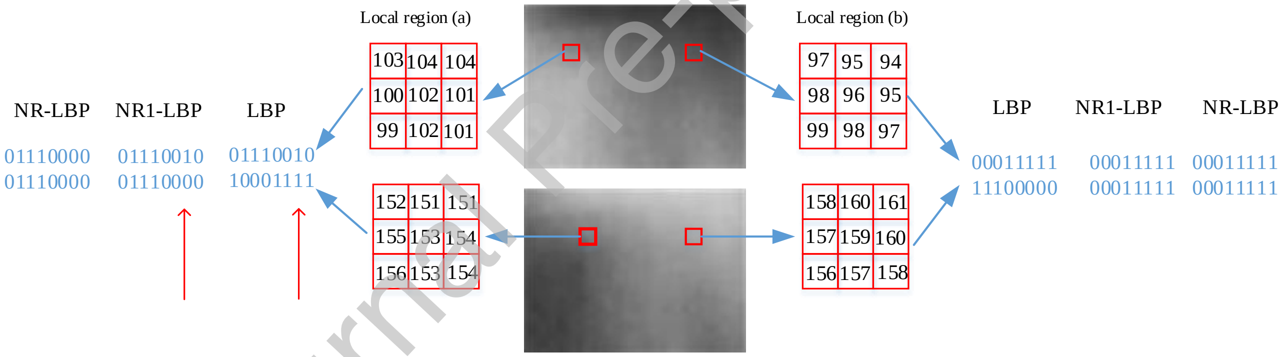


Fig. 4. Two images represent the same smoke structure with an inverted background and foreground. The original LBP codes of local regions in the two images with the same smoke spatial characteristics are different, and the NR-LBP of them are the same.

However, $LBP_{8,1}$ of local (a) in the two smoke images are 01110010 and 10001111, respectively. The two codes are not complementary, since there is one intensity value of neighbors are equal to the intensity value of center pixel. In our observations, this case always occurs, since smoke has a smooth effect which makes the adjacent pixels more similar.

Based on the above discussions, we propose another idea to define NR-LBP, as follows:

$$NR-LBP_{P,R}(x_c,y_c)=\min\left\{OLBP_{P,R}(x_c,y_c),2^P-1-OLBP_{P,R}(x_c,y_c)\right\} \tag{21}$$

$$OLBP_{P,R}(x_c,y_c)=\sum_{p=0}^{P-1}h(g_p,g_c)2^P \tag{22}$$

$$h(g_p,g_c)=\begin{cases}1, & \text{if } g_p>g_c\\ h(g_{p-1},g_c), & \text{if } g_p=g_c\\ 0, & \text{if } g_p<g_c\end{cases} \tag{23}$$

where g_{-1} is equal to g_P. If all the intensity values of neighbors are equal to the intensity of center pixel, all the elements of NR-LBP code are set to 0.

In Fig.4, $LBP_{8,1}$ codes of local (b) in the two smoke images are 01110010 and 10001111, respectively. $NR1-LBP_{8,1}$ codes of local region (b) in the two smoke images are 01110010 and 01110000, respectively. $NR-LBP_{8,1}$ codes of local region (b) in the two smoke images are 01110000 and 01110000, respectively. The idea of NR-LBP is more effective. It is insensitive to relative changes between foreground and background and also reduce the final feature dimension.

We can also set NR-LBP to $\min\left\{LBP_{P,R}(x_c, y_c), LBP_{P,R}(\overline{x}_c, \overline{y}_c)\right\}$ where $LBP_{P,R}(\overline{x}_c, \overline{y}_c)$ is the common LBP code at position (x_c, y_c) of complementary image. But this method should calculate the LBP codes of the same position twice.

In NR-LBP, we consider LBP code and its complement as the same case to reduce intra-class distances. This makes feature dimension reduced by half. In addition, NR-LBP characterizes relative contrast between foreground and background. Therefore, it is more discriminative compared with the original LBP.

To make NR-LBP more robust, the key is to find a threshold which is insensitive to noise and invariant to monotonic gray scale transformation. It means that the parameter g_c in equation (22) should be carefully selected. In our strategy, we replace parameter g_c by parameter ξ_c, which given by

$$\xi_c = \left(\alpha g_c + \sum_{i=1}^{P} g_{ci}\right) \Big/ \left(\alpha + P\right) \tag{24}$$

where ξ_c is the weighted local gray level used to make a balance between the information and noise of the individual pixel. g_c is the intensity value of center pixel. $g_{ci}(i = 1, 2..., P)$ is the intensity value of i th neighbor pixel of g_c. α is a regulatory factor.

We extract discrete histogram features from NR-RLBP codes, and nominalize it to form $F_{NR-RLBP}$:

$$F_{NR-RLBP}(i) = H_{NR-RLBP_{P,R}}(i) \Big/ \sum_{k=1}^{2^{P-1}} H_{NR-RLBP_{P,R}}(k) \tag{25}$$

where $H_{NR-RLBP_{P,R}}$ is the discrete histogram from NR-RLBP codes, and the feature dimension is equal to 2^{P-1}.

The NR-RLBP descriptor characterize texture information of smoke-colored blocks. It is insensitive to noise and also robust to relative changes between foreground and background.

3.2. Gradient features: Weighted Co-occurrence histograms of oriented gradients (W-CoHOG)

The well-known HOG descriptor is first proposed by Dalal [39] for human detection. It is widely used to describe statistics information of oriented gradients of pixels in an image. It is robust to illumination variation and invariance to local geometric and photometric transformations, but it ignores spatial information with respect to neighbouring pixels.

The Co-occurrence HOG (CoHOG), an extension of HOG, is first proposed by Watanabe [40] and used for pedestrian/human detection [40] and scene character recognition [41]. The CoHOG encodes gradient orientations of neighbouring pixel pairs and accordingly captures more spatial and contextual information. It has showed good performances since the co-occurrences at each pixel position represent various spatial information of object shape. However, it ignores gradient magnitude information.

To add gradient magnitude information, we propose a Weighted Co-occurrence Histograms of Oriented Gradients (W-CoHOG), which is more powerful to describe an object precisely and effectively.

Few works take CoHOG-based descriptors as features to recognize smoke, and so we make a brief review about original CoHOG descriptor [40]. CoHOG descriptor captures rich spatial information by counting frequency of the co-occurrence of oriented gradients between pixel pairs. The frequency is captured at each offset via a CoHOG matrix. The CoHOG at a specific offset (dx, dy) is given by

$$H_{dx,dy}(i, j) = \sum_{(p,q)\in B} c_{dx,dy}(p, q, i, j) \tag{26}$$

$$c_{dx,dy}(p, q, i, j) = \begin{cases} 1, & \text{if } O(p, q) = i \text{ and } O(p + dx, q + dy) = j \\ 0, & \text{otherwise} \end{cases} \tag{27}$$

where $H_{dx,dy}$ denotes the element in position (i, j) of CoHOG matrix (a square matrix with the dimension equalling to orientation bins) at offset (dx, dy). $O(p, q)$ is gradient orientation at location (p, q) of the input image which is quantized into N_{bin} orientation bins. B is a block image.

To extract CoHOG descriptor, we can vectorize and concatenate the CoHOG matrix of all blocks in the image. Obviously, CoHOG descriptor is the same as HOG descriptor when the offset is $(0, 0)$. In this paper, 8 orientation bins are used to evenly divide gradient orientation, which ranges between 0° and 360°. However, original CoHOG ignores the difference between strong gradient and weak gradient pixels. Therefore, we propose W-CoHOG descriptor to add gradient magnitude information.

The gradient magnitude is determined by the L2 norm of horizontal and vertical magnitude computed by Sobel operation. We propose the concept of Co-occurrence Histograms of Gradient Magnitudes (CoHGM). The CoHGM at a specific offset (dx, dy) is given by

$$M_{dx,dy}(i, j) = \sum_{(p,q)\in B} b_{dx,dy}(p, q, i, j) \tag{28}$$

$$b_{dx,dy}(p,q,i,j)=\begin{cases} m_{(p,q)}^{(p+dx,q+dy)}, \text{ if } O(p,q)=i \text{ and } O(p+dx,q+dy)=j \\ 0, \qquad \text{otherwise} \end{cases} \tag{29}$$

where $M_{dx,dy}$ denotes the element in position (i, j) of CoHGM matrix at offset (dx, dy). $m_{(p,q)}^{(p+dx,q+dy)}$ denotes gradient magnitude between pixel at position (p,q) and pixel at position $(p+dx,q+dy)$. $O(p,q)$, N_{bin} and B have the same meaning in Equation (27).

To extract W-CoHOG descriptor of an image, we vectorize and nominalize CoHOG matrix and CoHGM matrix of each block of the image. Then we concatenate them to form final feature vector $F_{W-CoHOG}$ to characterize smoke-colored blocks.

3.3. Motion features: motion boundary histograms (MBH)

Besides texture descriptor and gradient descriptor, we also employed a motion descriptor to describe motion information. The motivation is based on the follow observations that different objects on the road have different motion information. The successive smoke blocks describe the motion of smoke that gradually extends around. The successive road surface blocks are with small motion information. The successive vehicle blocks describe the motion of a vehicle in one direction.

Optical flow information can characterize the absolute motion between two adjacent frames. However, in the realistic road surveillance videos, the extracted optical flow information may contain motion from the camera motion (e.g., zooming, tilting, rotation, etc.), which may corrupt smoke-colored blocks classification. To characterize smoke or vehicle motion well while maintaining resistances to camera shaking, the Motion Boundary Histograms (MBH) descriptor is introduced into smoky vehicle detection.

The MBH descriptor was first proposed by Dalal [42] and used for human detection by computing derivatives separately for horizontal and vertical components of optical flow to encode relative motion between pixels. MBH descriptor removes locally constant camera motion and keeps information about changes in the flow field, since it is the gradient of optical flow. Therefore, MBH descriptor is more robust to camera motion than optical flow, and so more discriminative for smoke recognition.

In extracting MBH descriptor, we separate optical flow $\omega=(u,v)$ into vertical and horizontal components. And then calculate spatial derivatives for each of them. For each component (i.e., MBH_x and MBH_y), we quantize the orientation information into 8-bin histograms with magnitude as their weights. Both histogram vectors are normalized separately with the L2 norm. Then we concatenate them as final feature vector F_{MBH} to characterize smoke-colored blocks.

3.4. Recognition and classification of smoke-colored blocks

After extracting the above groups of features from each smoke-colored block and forming final features vector, SVM is adopted as classifier to train and classify new smoke-colored blocks. Though there are many classification techniques, we chose SVM classifier because that fewer hyper parameters are enough in SVM classifier, and it requires less grid searching to get a reasonably accurate model.

4. Third stage: frame classification

In this section, we introduce the method of smoke frame detection. We first fuse smoke blocks into a Region of Interest (ROI). Then we extract frequency domain information from the ROI of continuous frames, i.e. the same position in different time. Finally, we analyze frequency domain information and three groups of features using autoregressive moving-average model (AMAM) and Hidden Markov Model (HMM) to classify the current frame to smoke or non-smoke.

4.1. Fusion of smoke blocks

Unlike most methods that make frame classifications by analyzing smoke blocks directly, we look smoke from a wide field of view and propose a strategy of smoke block fusion.

The strategy of smoke blocks fusion can be seen in Fig.5. We first obtain a rectangular region that covers smoke blocks. Then, we check the four edges of the rectangular region, and remove the edge on which the number of non-smoke blocks is higher than the number of smoke blocks. For the up edge in Fig.5, the number of non-smoke blocks is higher than that of smoke blocks, and so the up edge is removed. Otherwise, for the down edge in Fig.5, the down edge is retained.

4.2. Feature extraction from frequency domain

Frequency domain features can detect the patterns that are not visible in the images and have been used in literature [41] for smoke detection based on Discrete Wavelet Transforms (DWTs). The DWT decomposes original image into four sub-bands: approximation and three detail sub-bands in horizontal, vertical and diagonal directions.

The main motivation of using DWT comes from the following fact: smoke usually has smooth effects in the scene and leads to more low frequency components. The DWT can capture frequency information of an image at different scales to maintain spatial information and find smoke presence.

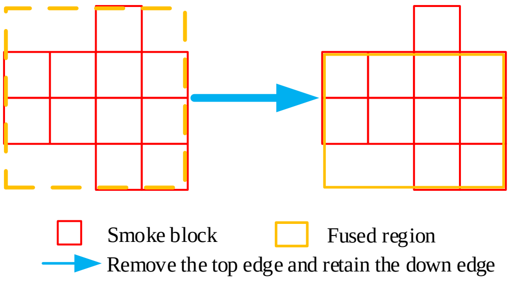


Fig.5 The strategy to obtain the fused region.

In our implementation, we first divide the ROI into 2×2 sub-blocks and calculate the Wavelet energy from each sub-block, 3 directions and 2 levels are used in this paper. Then we concatenate the Wavelet energy information into a feature vector and normalize them to form frequency domain features denoted by F_{DWT} .

4.3. Dynamic features extraction using auto-regressive moving average (ARMA) model and hidden Markov model (HMM)

Dynamic features are sequences of images that exhibit a certain stationary property in time, such as the shaking leaves. The motivation using dynamic features is based on the fact that different objects have different dynamic features. For a smoke block sequence, i.e. the blocks at the same position in different time, dynamic features describe the gradual diffusion and motion process of smoke; For a vehicle-body-covered block sequence, dynamic features describe the process from the vehicle entry to the vehicle departure; For a roadside leaves-covered block sequence, the dynamic features describe the process of the leaves moving back and forth; For a shadow-covered block sequence, the dynamic features describe the process of the shadow moving.

To characterize dynamic features, some models has been introduced into smoke detection, such as h-LDS in [18], Hidden Markov Tree (HMT) in [44]. In this paper, we borrow tools from the system identification for modeling and learning the essence of dynamic textures. A method based on Auto-Regressive and Moving Average Model (ARMA) and Hidden Markov Model (HMM) is employed.

The model can be specified by a 4-tuple and defined as follows [44],

$$M^{(p,b,q)} = \{\pi, A, \mu(X), \sigma(X), \varphi, \omega, \theta\} \tag{30}$$

where π denotes the initial state probabilities, A denotes the transition matrix of hidden states, X denotes the hidden state variable. $\mu(*)$ and $\sigma(*)$ parameterize the emission distributions for each state. φ , ω and θ are filter coefficients. $\varphi=(\varphi_1,\varphi_2,...,\varphi_p)$ denote the autoregressive coefficients and p is the autoregressive order. $\omega=(\omega_1,\omega_2,...,\omega_b)$ denote the moving average coefficients and b is the moving average order. $\theta=(\theta_1,\theta_2,...,\theta_q)$ are filter coefficients with number q .

The process related to above model is observation $V=(V_t)_{0\le t\le N}$ as follows [44]:

$$V_t = \sum_{i=0}^{p} \varphi_i \mu(X_{t-i}) + \zeta_t \tag{31}$$

$$\zeta_t = \sum_{i=0}^{b} \omega_i \sigma(X_{t-i}) \varepsilon_{t-i} + \sum_{i=0}^{q} \theta_i \eta_{t-i} \tag{32}$$

This model includes ARMA filtered output noise as well as additional ARMA filtering of the resulting HMM signals. So we call it as ARMA-HMM.

The likelihood of observation sequence V is estimated by the approximation method presented in [44]. In the testing, before solving recognition task as an optimal-state sequence problem for the trained model, the ARMA model is used to enrich testing input sequences. Specifically, each feature vector dimension is treated as a univariate time series, and so ARMA models can be trained respectively for each dimension based on the observed sequence $V_1,V_2,...,V_t$. Next we adopt all m ARMAs to do k steps prediction to enrich the testing input observation sequence V by incorporating k predicted features $V_{t+1},...,V_{t+k}$. Lastly, we calculate $\Pr(V_{1:t}V_{t+1:t+k} \mid M_i)$ for each trained ARMA-HMM M_i and select M_{c^*} :

$$c^* = \arg\max_i \{\Pr(V_{1:t}V_{t+1:t+k} \mid M_i)\} \tag{33}$$

4.4. Smoke frame recognition

In this part, we analyze frequency domain features and the three groups of features mentioned in section 3 to classify current frame to a smoke frame or a non-smoke frame. The frame with at least one smoke ROI is defined as a smoke frame.

For one ROI in the current frame, let B_i^j denotes the $i(i=1,2,...,N_{ROI})$ th block of ROI region at $j(j=1,2,...,T)$ th frame. Let $B_i=\left\{B_i^1,B_i^2,...,B_i^T\right\}$ denotes the set of continuous blocks, which with different time at the same i th blocks. Let F_i denote the three groups of features extracted from block sequence B_i. Let $p_{block}(i)$ and $\hat{p}_{block}(i)$ denote the smoke block possibility and non-smoke block possibility of features F_i by using ARMA-HMM model. The classification of ROI is given by:

$$\begin{cases}\text{smoke}, & \text{if } P_{smoke}>P_{non-smoke}\\ \text{non-smoke}, & \text{if } P_{smoke}\le P_{non-smoke}\end{cases} \tag{34}$$

$$P_{smoke}=\alpha p_{ROI}+(1-\alpha)\sum_{i}^{N_{block}}\frac{w(i)}{\sum_k w(k)}p_{block}(i) \tag{35}$$

$$P_{non-smoke}=\alpha \hat{p}_{ROI}+(1-\alpha)\sum_{i}^{N_{block}}\frac{w(i)}{\sum_k w(k)}\hat{p}_{block}(i) \tag{36}$$

where P_{smoke} and $P_{non-smoke}$ denotes the smoke possibility and non-smoke possibility of current ROI, respectively. N_{blocks} denotes the number of blocks in current ROI. $w(i)$ is a vector with weight coefficients. It satisfies Gaussian distribution. The farther away from ROI center, the bigger the weight value.

If there are one or more ROIs been detected as smoke ROIs in one frame, current frame is classified as a smoke frame.

5. The whole algorithm framework

The whole algorithm framework is shown in Fig.6. It includes three stages: smoke-colored blocks detection, smoke blocks detection, frames classification.

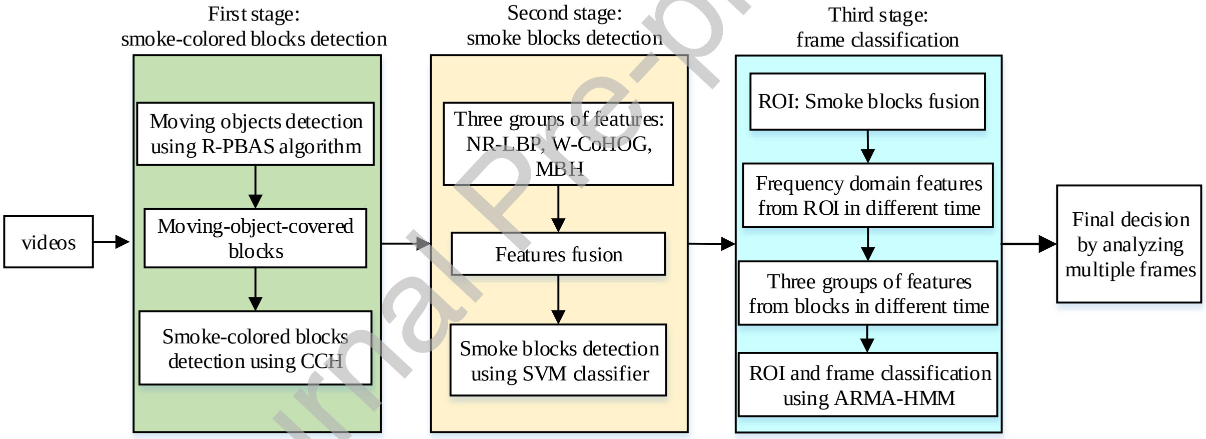

First stage:
smoke-colored blocks detection
Second stage:
smoke blocks detection
Third stage:
frame classification
videos
Moving objects detection using R-PBAS algorithm
Moving-object-covered blocks
Smoke-colored blocks detection using CCH
Three groups of features: NR-LBP, W-CoHOG, MBH
Features fusion
Smoke blocks detection using SVM classifier
ROI: Smoke blocks fusion
Frequency domain features from ROI in different time
Three groups of features from blocks in different time
ROI and frame classification using ARMA-HMM
Final decision by analyzing multiple frames

Fig.6. The whole algorithm framework.

In the first stage, we use R-PBAS algorithm to detect moving objects. Then we divide the whole frame into blocks to search moving-object-covered blocks. The CCH features and SVM classifier are used to further detect smoke-colored blocks.

In the second stage, we extract three groups of features (NR-RLBP, W-CoHOG and MBH) to characterize smoke-colored blocks. Then we classify smoke-colored blocks into smoke blocks or non-smoke blocks using SVM classifier.

In the third stage, we extract frequency domain features from the ROIs and three groups of features from the blocks in ROI with different time. The aim is to obtain features of the same position (blocks or ROIs) in different time to form time series data. The ARMA-HMM model is used to analyze the sequence features and classify the ROI into a smoke or non-smoke region. The current frame is classified as a smoke frame if there are one or more ROIs been detected as smoke ROIs in it.

Different from existing methods, the above algorithm framework has three characteristics. Firstly, the color, texture, gradient, motion and frequency domain features are well integrated to characterize smoke. Using ARMA-HMM model to describe dynamic information makes it more robust to reduce false alarm rates while keeping high detection rates. Secondly, various improvements and new strategies are proposed to improve algorithm performances and increase robustness to various interferences from camera shaking and other smoke-liked objects. Thirdly, we analyze smoke from different scales with a block fusion strategy to makes it more robust in recognizing small smoke.

6. Experiments and Analysis

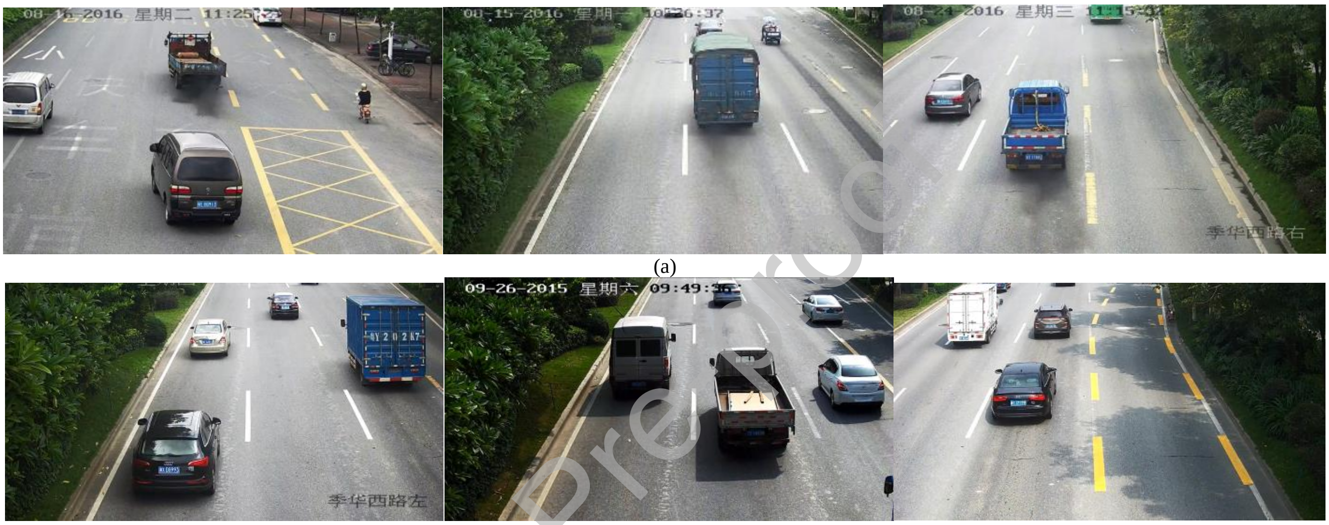
In this section, we verify the robustness and effectiveness of the proposed method through extensive experiments on challenging datasets.

6.1. Datasets

The testing datasets contains 102 traffic surveillance videos, including 4 long videos and 98 short videos filmed using a static camera with 25fps. The frames are with the spatial resolution of 1920×1080 pixels (downsampled to 768×448 pixels using bilinear interpolation). These surveillance videos are taken at different locations in the daytime with sunny weather. The training dataset contains 1000 smoke and 1000 non-smoke frames.

According to the intensity values of black smoke and if there are shadows, all frame images (Dataset 5) are divided into four categories, which correspond to four datasets, namely Dataset 1, Dataset 2, Dataset 3 and Dataset 4 (as shown in Table 1). One vehicle always has many frames in surveillance video, and we collect all the vehicles to form dataset 6.

Some frame images from the datasets are shown in Fig.7. Fig.7(a) show three smoke frames including a heavy smoke frame, a light smoke frame, and a smoke frame with shaking leaves. Fig.7(b) show three non-smoke frames including a no-smoke frame without shadows, a non-smoke frame with vehicle shadows and a non-smoke frame with tree shadows. In the testing all the frames are analyzed in corresponding frame sequences, not just on a single frame.


(a)

(b)

Fig.7. Some smoke frames and non-smoke frames from the testing datasets. (a)Three smoke frames from traffic surveillance videos. (b) Three non-smoke frames from traffic surveillance videos.

The surveillance videos are provided by a company that specializes in smoky vehicle detection. Each vehicle in our datasets is labeled by an engineer in this company and he has rich experiences in judging whether the current vehicle is a smoky vehicle or not.

Table 1. All the datasets used in this paper.

Dataset name	Brief description
Dataset 1	3714 smoke frames with heavy smoky vehicles
Dataset 2	2223 smoke frames with light smoky vehicles
Dataset 3	89417 non-smoke frames without shadows
Dataset 4	62196 non-smoke frames with shadows
Dataset 5	5937 smoke frames and 151613 non-smoke frames
Dataset 6	104 smoky vehicles and 3417 non-smoky vehicles

6.2. Comparison experiments of different methods on frames classification

The possibility of correct classification (Pcc) is used as the evaluation criteria. The definition of Pcc is given by:

$$Pcc = N_{correct} / N_{all} \tag{37}$$

where N_{all} denotes the number of all the samples in the used dataset, and $N_{correct}$ denotes the number of correctly classified samples. Based on this concept, if Pcc is the result of recognizing non-smoke frames from non-smoke frames dataset, 1-Pcc will be the false alarm rate that recognize non-smoke frames as smoke frames.

To validate algorithm performances, some advanced methods [12, 21, 31, 32, 45, 46] are selected as baseline methods to make comparisons. It should be noted that in method [46], 4000 smoke images and 8000 non-smoke images are resized to $227 \times 227 \times 3$ to fine-tune the pre-trained CNN. The other settings are the same with method [46].

Table 2 reports the experimental results of different methods on frames classification. Our method performs better than all the baseline methods, especially with lower false alarm rates. In Table 2, all the methods achieve the best results in Dataset 1. This is reasonable since Dataset 1 contains all the smoke frames with heavy smoky vehicles. The performances of all the methods in Dataset 2 perform worse than that in Dataset 1, and this illustrates that the light and small smoke is easily missed. In addition, most methods perform badly in Dataset 4. This verifies that the shadows of vehicles and trees lead to false alarms.

Table 2. Experimental results of different methods on frames classification in Dataset 5.

Different methods	Dataset 1	Dataset 2	Dataset 3	Dataset 4
Method in [12]	0.9094	0.8226	0.8806	0.6678
Method in [21]	0.8603	0.7897	0.8445	0.5897
Method in [31]	0.8505	0.7904	0.8458	0.5993
Method in [32]	0.8822	0.8206	0.8702	**0.6686**
Method in [45]	0.8866	**0.8328**	0.8754	0.6371
Method in [46]	**0.9299**	0.8109	**0.9025**	0.6646
Our method	**0.9309**	**0.8818**	**0.9238**	**0.7156**

From Dataset 1, the advantages of our method are not obvious, but in the more challenging Dataset 2 and Dataset 4, our method has obvious advantages than the baseline methods and achieves a higher Pcc. This illustrates that our method is more robust to resist the shadow interferences and more sensitive to recognize small light smoke. It is exciting that our method performs better than the CNN-based method in [46], especially on challenging Dataset 2 and Dataset 4. Table 3 shows the average time consumptions of different methods in processing one frame, we can see that the CNN-based method in [46] is the most time-consuming.

Table 3. The average time consumptions of different methods in processing one frame.

Different methods	Average time consumptions/ms
Method in [12]	91.34
Method in [21]	114.98
Method in [31]	128.56
Method in [32]	136.56
Method in [45]	113.45
Method in [46]	231.33
Our method	140.34

6.3. Comparison experiments using different κ on vehicle classification

A vehicle always has many frames in the traffic surveillance video. If we produce a smoky vehicle alarm only based on the fact that the current frame is recognized as a smoke frame in the current short sequence, it will lead to high false alarms. To make it more robust, we add the following strategy. For each continuous 100 frames, if Rule 2 is satisfied, we make the decision that there are smoky vehicles in current short sequence,

$$\text{Rule 2}: k_{frame} > \kappa \tag{38}$$

where κ is a threshold. k_{frame} is the total number of frames been detected as smoke frames in continuous 100 frames.

The experiment of this part is on Dataset 6, which contains 104 smoky vehicles and 3417 non-smoky vehicles. We adopt detection rate (DR) and false alarm rate (FAR) to evaluate the algorithm performances.

Table 4 shows the experiment results by setting κ to 2, 4, 6, 8, 10 and 15 to obtain the DRs and FARs. With the increase of κ, the FARs are reduced while the DRs are also reduced.

Table 4. Experimental results of different κ on vehicles classification of Dataset 6.

Different parameters	Dataset 6	
	DR	FAR
κ =2	0.9327	0.2809
κ =4	0.9231	0.1978
κ =6	0.9135	0.1630
κ =8	0.8942	0.1504
κ =10	0.8365	0.1405
κ =15	0.8077	0.1367

6.4. Comparison experiments of the original PBAS algorithm and R-PBAS algorithm

To make it more robust to the camera shaking and slowly moved objects, R-PBAS algorithm is proposed in section 2.2. In this section, we verify the robustness of R-PBAS algorithm.

Table 5 shows the experiment results of using original PBAS algorithm and the R-PBAS algorithm on frames classification in Dataset 5. Here we make brief introductions about the parameters setting of N, β and γ in R-PBAS algorithm. The other parameters setting can be seen in [33].

The parameter N in equation (1) is the number of components of background model. A large N increases the memory space and computational time complexity. In our application, we set N to 30. The parameter β in equation (9) represents the weight coefficient of the regional gradient information in background complexity. The locations with abundant gradient information are always edge or corner points that belong to complicated regional structure. We can appropriately change β to prevent them being misjudged. Of course, if β is too large, it may miss parts of objects edges. In traffic surveillance videos, increasing β can eliminate the error detections caused by trees leaves and roadside trees shadows. So as to further reduce false alarms. In our application, we set β to 4. Table 5 shows the algorithm performances using different β with other parameters unchanged. With the increase of parameter β from 0 to 4, the false alarms rates reduced. The parameter γ in equation (9) represents the weight coefficient of regional color information in background complexity. When the object color is single, and interference factors are complex, parameter γ should be decreased. The smoke is black, and so a relatively small value for γ is better. In our application, $\gamma = 0.3$ is suggested. Table 5 shows the algorithm performances using different γ with other parameters unchanged. With the increase of parameter γ from 0.1 to 0.4, the detection rates increased.

Table 5. The experimental results of using original PBAS algorithm and R-PBAS algorithm on frames classification of Dataset 5

Different methods		DR	FAR
Original PBAS algorithm		0.9014	0.1804
R-PBAS algorithm with different (β, γ)	(0, 0.3)	0.9013	0.1811
	(2, 0.3)	0.9102	0.1734
	(4, 0.3)	0.9125	0.1610
	(6, 0.3)	0.9112	0.1623
	(4, 0.1)	0.9017	0.1808
	(4, 0.2)	0.9110	0.1702
	(4, 0.3)	0.9125	0.1610
	(4, 0.4)	0.9201	0.1621

6.5. Experimental comparison to verify the three groups of features in Dataset 5

In section 3.1, we propose the NR-RLBP descriptor to characterize texture information. It is insensitive to noise and relative changes between foreground and background.

To verify the advantages, we keep other features unchanged, and replace NR-RLBP by original LBP, LBPV, respectively. Table 6 shows the experiment results. We can see that NR-RLBP descriptor performs better than other descriptors, which verifies that our improvements are necessary. To analyze the samples of miss detections and false alarms, we find that the shadows and light smoke are correctly recognized by NR-RLBP descriptor while wrongly recognized by other descriptors.

Table 6. Experiment results of replacing NR-RLBP descriptor by other descriptors on frames classification in Dataset 5.

Different methods	DR	FAR
LBP	0.8674	0.3103
LBPV	0.8322	0.2307
NR-RLBP	0.9125	0.1610

In section 3.2, we propose W-CoHOG descriptor to characterize gradients information. It not only encodes gradient orientation of neighbouring pixel pairs and captures more spatial and contextual information, but also adds gradient magnitude information.

To verify the advantages, we keep other features unchanged, and replace W-CoHOG by original HOG, CoHOG, respectively. Table 7 shows the experiment results. We can see that W-CoHOG descriptor performs better than other descriptors.

Table 7. Experiment results of replacing W-CoHOG descriptor by other descriptors on frames classification in Dataset 5.

Different methods	DR	FAR
HOG	0.8445	0.2182
CoHOG	0.8961	0.1883
W-CoHOG	0.9125	0.1610

In section 3.3, MBH descriptor is introduced to characterize motion information. It can describe motion of different objects well while maintaining resistances to camera shaking.

To verify the advantages, we keep other features unchanged, and replace MBH by original optical flow features. Table 8 shows the experiment results. We can see that the MBH descriptor performs better than original optical flow features.

Table 8. Experimental results to verify the advantages of MBH descriptor on frames classification of Dataset 5.

Different methods	DR	FAR
Optical flow	0.8945	0.1842
MBH	0.9125	0.1610

6.6. The memory space and computation time

The algorithm is implemented using VS2012 with OpenCV 2.4.9. All experiments are executed on the PC with 2.4 GHz Intel(R) Core(TM) i7 CPU and 4GB memory. In this part, the memory space complexity and computation time complexity are analyzed.

Most computation is spent on performing the R-PBAS algorithm and extracting features of NR-RLBP, W-CoHOG, MBH and Wavelets.

Let T_1 and T_2 denote the frame width and frame height, respectively. Let N denote the saved pixels number in background model of a pixel. The memory space complexity in the step of R-PBAS algorithm is $O(T_1T_2N)$. In processing a region of current frame, we save the regions in some front frames and back frames of the video to analyze dynamic information. For a region with the size of $R_1 \times R_2$ pixels. Let T denote the regions number. The memory space complexities of extracting features of the NR-RLBP, W-CoHOG, MBH and Wavelets are all $O(R_1R_2T)$.

For time complexity, Table 9 shows the time consumptions of extracting different features on a region with size 64×64 . The Wavelets energy is the most time-consuming.

It should be noted that the algorithm is highly parallel because each block or each group of features can be processed independently.

Table 9. Time consumptions in extracting different features.

Features	Time / ms
NR-RLBP	80±5
W-CoHOG	83±5
MBH	40±5
Wavelets	125±10

6.7. Some smoke frames captured from surveillance video

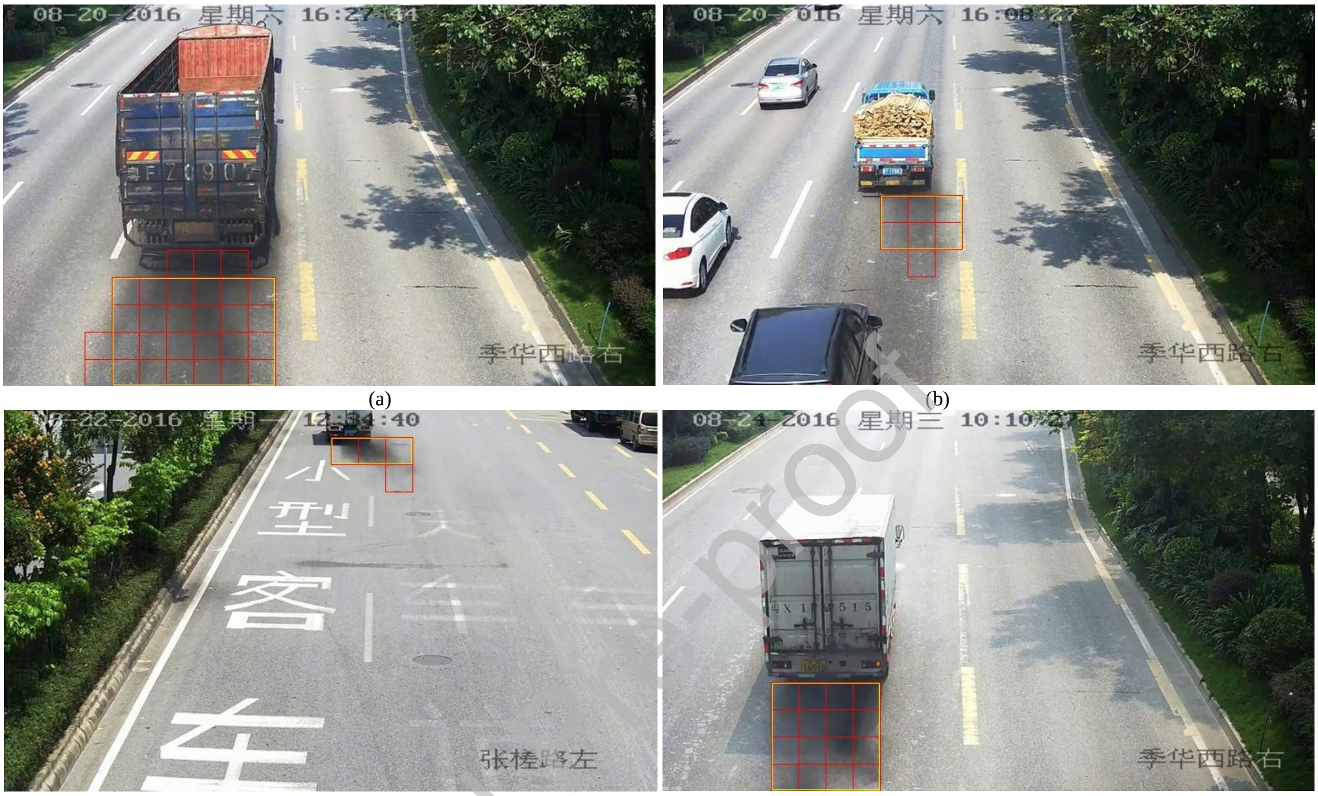
Some smoke frames captured from traffic surveillance videos can be seen in Fig.8. The test results show the robustness and effectiveness of our algorithm framework under different conditions.

Fig.8(a) shows a heavy smoky vehicle. This kind of smoky vehicles can be easily detected. Fig.8(b) shows a smoky vehicle with light and small smoke. Most of existing methods are not robust to this kind of smoky vehicles. Fig.8(c) shows a smoky vehicle with smoke at the edge of the camera's field of view. The vehicle far away is easily updated to background since they move slowly, but our method can detect the smoke successfully. Fig.8(d) shows a smoky vehicle with shadows. It can be seen that our method is robust to shadows.

7. Discussion and Conclusion

Existing smoky vehicle detection methods are still with high false alarm rates due to many factors: 1) Large variations in color, texture, shapes features of smoke, especially when the smoke is light and small; 2) Continuous interferences from black vehicles shadows, black road surfaces, and moving black vehicles, etc. 3) Some complex situations, such as the smoke covered by vehicle objects, and the smoke at night time or bad weather, etc. Most of existing methods are designed for forest smoke detection, and they are vulnerable to false alarms in smoky vehicle detection task.

This paper presents a three-stage framework for smoky vehicle detection in traffic surveillance videos. For the first stage, R-PBAS algorithm is proposed to extract moving objects and adapt to cameras shaking, CCH features and some rules are adopted to extract smoke-colored blocks. For the second stage, three groups of features, including NR-RLBP, W-CoHOG and MBH are proposed to characterize texture, gradient, and motion information of smoke-colored blocks, respectively. NR-RLBP is insensitive to noise and also robust to relative changes between foreground and background. W-CoHOG not only encodes gradient orientations of neighboring pixel pairs to capture spatial and contextual information, but also adds gradient magnitude information. MBH describes the motion of different objects well while maintaining resistances to camera shaking. For the third stage, we fuse smoke blocks to obtain ROI and extract frequency domain information based on DWT. To further improve robustness, the ARMA-HMM model is adopted to model the ROIs in consecutive frames.


08-20-2016 星期六 16:27:44
季华西路右
08-20-2016 星期六 16:08:27
季华西路右
12:54:40
小型客车
张槎路左
08-24-2016 星期三 10:10:27
季华西路右

(a) (b)

(c) (d)

Fig.8. Some smoke frames captured from traffic surveillance videos. The red box represents smoke blocks. The yellow box represents the smoke region (ROI) fused by smoke blocks. (a) A smoke frame containing a heavy smoky vehicle. (b) A smoke frame containing a smoky vehicle with light and small smoke. (c) A smoke frame containing a smoky vehicle with smoke at the edge of the camera's field of view. (d) A smoke frame containing a smoky vehicle with shadows.

Our method can be easily embedded into a smoky vehicle alarm system to automatically detect smoky vehicles from traffic surveillance videos. We can also take into consideration additional sustainability technology optimization applications [47]. When a camera in a new road position is installed, an experienced engineer should stay there for a few days and debug the system to get the best parameters for the new road scene.

However, our method is not robust to detect smoky vehicles at night time. A visible smoke in surveillance videos at daytime may be difficult to distinguish at night, especially for the faraway smoky vehicle with light and small smoke. In addition, the reflections of road will further increase the invisibility of smoke, and the vehicle shadows caused by street lights can also lead to false alarms. Analyzing videos filmed by the infrared imaging cameras or thermal imaging cameras may be a solution in the future.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (No.51775515, No.61871123, No.61901221), Key Research and Development Program in Jiangsu Province (No.BE2016739), a Project Funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions, the Postgraduate Research and Practice Innovation Program of Jiangsu Province (No. KYCX18_0101), the Scientific Research Foundation of Graduate School of Southeast University (No. YBPY1871).

Credit Author Statement

Huanjie Tao: **Conceptualization, Methodology, Writing- Original draft preparation.** Peng Zheng**: Writing- Review & Editing, Visualization, Investigation, Funding acquisition.** Chao Xie**: Software, Data curation.** Xiaobo Lu**: Supervision, Project administration, Funding acquisition, Resources.**

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Liu, Y. H., et al. "Vehicle emission trends in China's Guangdong Province from 1994 to 2014", Science of the Total Environment, Volume 586, 15 May 2017, pp. 512–521.

[2] Flora, J. B., Alam, M., & Iftekharuddin, K. M. Improvements to vehicular traffic segmentation and classification for emissions estimation using networked traffic surveillance cameras. Optics and Photonics for Information Processing VIII. Vol.9216, pp.92160K,2014

[3] Pyykonen, P., Peussa, P., Kutila, M., et al.: 'Multi-camera-based smoke detection and traffic pollution analysis system', Proc. Int. Conf. Intelligent Computer Communication and Processing (ICCP), Cluj-Napoca, 2016, pp. 233-238.

[4] Banerjee, S., Chinmay C., Sumit C., "A Survey on IoT Based Traffic Control and Prediction Mechanism." In Internet of Things and Big Data Analytics for Smart Generation, pp. 53-75. Springer, Cham, 2019.

[5] Zhao, Z., Paul Ma., Wang j., Ari T., Andrew Jones, Ian Taylor, Vlado Stankovski et al. "Developing and operating time critical applications in clouds: the state of the art and the SWITCH approach." Procedia Computer Science 68 (2015): 17-28.

[6] Bhatnagar, A., Vishvabodh S., Gaurav R., "IoT based Car Pollution Detection Using AWS." In 2018 International Conference on Advances in Computing and Communication Engineering (ICACCE), pp. 306-311. IEEE, 2018.

[7] Tao, H., Lu, X., "Smoke vehicle detection based on multi-feature fusion and hidden Markov model", Journal of Real-Time Image Processing, 2019. [Online] https://doi.org/10.1007/s11554-019-00856-z

[8] Tao, H., Lu, X., "Smoky vehicle detection based on multi-scale block Tamura features," Signal, Image and Video Processing, vol. 12, no. 6, pp. 1061-1068, 2018.

[9] H. Tao, X. Lu, "Smoky vehicle detection based on multi-feature fusion and ensemble neural networks," Multimedia Tools and Applications, vol. 77, no. 24, pp. 32153-32177, 2018.

[10] Tao, H., Lu, X., "Smoke vehicle detection in surveillance video based on gray level co-occurrence matrix," in Proc. 10th International Conference on Digital Image Processing, Shanghai, SPIE, vol. 10806, id.1080642, pp.1-7, 2018.

[11] Tao, H., Lu, X., "Contour-based smoky vehicle detection from surveillance video for alarm systems," Signal, Image and Video Processing. vol. 13, no. 2, pp. 217-225 2018.

[12] Tao, H., Lu, X., "Automatic smoky vehicle detection from traffic surveillance video based on vehicle rear detection and multi-feature fusion" IET Intelligent Transport Systems, vol. 13, no. 2, pp. 252-259, 2018.

[13] Tao, H., Lu, X., "Smoky vehicle detection based on range filtering on three orthogonal planes and motion orientation histogram", IEEE Access, vol.6, no.1, pp.57180-57190, 2018.

[14] Labati, R.D., Genovese, A., Piuri, V., "Wildfire smoke detection using computational intelligence techniques enhanced with synthetic smoke plume generation," IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 43, no. 4, pp. 1003–1012, 2013.

[15] Gunay, O., Toreyin, B.U., Kose, K., A. E. Cetin, "Entropy-functional- based online adaptive decision fusion framework with application to wildfire detection in video," IEEE Transactions on Image Processing, vol. 21, no. 5, pp. 2853–2865, 2012.

[16] Tian, H., Li, W., Ogunbona, P. O., et al., "Detection and Separation of Smoke from Single Image Frames". IEEE Transactions on Image Processing, 27(3), 1164-1177. 2018

[17] Tian, H., Li, W., Wang, L., Ogunbona, P., "Smoke detection in video: an image separation approach". International journal of computer vision, 106(2), 192-209. 2014

[18] Dimitropoulos, K., Barmpoutis, P., Grammalidis, N., "Higher order linear dynamical systems for smoke detection in video surveillance Applications". IEEE Trans. Circuits Syst. Video Techno. 27(5), 1143-1154. 2017

[19] Yuan, F., Shi, J., Xia, X., et al. High-order local ternary patterns with locality preserving projection for smoke detection and image classification. Information Sciences, 372(C), 225-240, 2016

[20] Yuan, F., "A double mapping framework for extraction of shape-invariant features based on multi-scale partitions with AdaBoost for video smoke detection," Pattern Recognition, vol. 45, no. 12, pp. 4326–4336, 2012.

[21] Chen J, He Y, Wang J. Multi-feature fusion based fast video flame detection. Building & Environment, 2010, 45(5):1113-1122.

[22] Appana D K, Islam R, Khan S A, et al. A Video-based Smoke Detection using Smoke Flow Pattern and Spatial-Temporal Energy Analyses for Alarm Systems. Information Sciences, 2017, s 418–419:91-101.

[23] Yin, M., Lang, C., Li, Z., Feng, S., & Wang, T. Recurrent convolutional network for video-based smoke detection. Multimedia Tools and Applications (8), 1-20. 2018.

[24] Hu, Y., Lu, X., "Real-time video fire smoke detection by utilizing spatial-temporal ConvNet features", Multimed Tools Appl. vol. 77, no. 22, pp.29283–29301, 2018,

[25] Yu, C., Fang, J., Wang, J., "Video Fire Smoke detection using motion and color features", Fire Technology, 46 (3) (2010) 651–663.

[26] Yuan, F.N., Video-based smoke detection with histogram sequence of LBP and LBPV pyramids, Fire Safety Journal, 46 (3), 132–139, 2011.

[27] Barmpoutis, P., Dimitropoulos, K., Grammalidis, N., "Smoke detection using spatio-temporal analysis, motion modeling and dynamic texture recognition", in: 2014 22nd European Signal Processing Conference (EUSIPCO), IEEE, 2014, pp. 1078–1082.

[28] Morerio, P., Marcenaro, L., Regazzoni, C.S., et al., "Early fire and smoke detection based on colour features and motion analysis", in: 2012 19th IEEE International Conference on Image Processing, IEEE, 2012, pp. 1041–1044.

[29] Seo, J., Kang, M., Kim, C.H., et al., “An optimal many-core model-based supercomputing for accelerating video-equipped fire detection”, The Journal of Supercomputing. 71, 2015, pp.2275–2308.

[30] Filonenko, A., Hernández, D. C., Jo, K.H., “Fast Smoke Detection for Video Surveillance Using CUDA”. IEEE Transactions on Industrial Informatics, 2018, 14(2), 725-733.

[31] Calderara, S., Piccinini, P., Cucchiara, R., “Vision based smoke detection system using image energy and color information,” Machine Vision and Applications, vol. 22, no. 4, pp. 705–719, Jul. 2011.

[32] Ko, B., Park, J., Nam, J.Y., “Spatiotemporal bag-of-features for early wildfire smoke detection,” Image and Vision Computing, vol. 31, no. 10, pp. 786–795, Oct. 2013.

[33] Hofmann, M., Tiefenbacher, P., Rigoll, G.. "Background segmentation with feedback: The pixel-based adaptive segmenter." Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference on. IEEE, 2012.

[34] Ojala, T., Pietikainen, M., Maenpaa, T., “Multiresolution gray-scale and rotation invariant texture classification with local binary patterns,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 7, pp. 971–987, Jul. 2002.

[35] Liu, Z. G., Yang, Y., Ji, X.H. Flame detection algorithm based on a saliency detection technique and the uniform local binary pattern in the YCbCr color space. Signal, Image and Video Processing, 2016, 10(2), 277-284.

[36] Lin, G., Zhang, Y., Zhang, Q., Jia, Y., Xu, G., & Wang, J. Smoke detection in video sequences based on dynamic texture using volume local binary patterns. Ksii Transactions on Internet & Information Systems, 2017, 11(11), 5522-5536.

[37] Favorskaya, M., Pyataeva, A., Popov, A., Verification of smoke detection in video sequences based on spatio-temporal local binary patterns[J]. Procedia Computer Science, 2015, 60: 671-680.

[38] Tian, H., Li, W., Ogunbona, P., et al. “Smoke detection in videos using non-redundant local binary pattern-based features”, In Multimedia Signal Processing (MMSP), 2011, pp. 1-4, IEEE.

[39] Dalal, N., Triggs, B., “Histograms of oriented gradients for human detection”, in: Computer Vision and Pattern Recognition (CVPR), 2005, pp. 886–893.

[40] Watanabe, T., Ito, S., Yokoi, K., “Co-occurrence histograms of oriented gradients for pedestrian detection,” in Proc. 3rd IEEE Pacific-Rim Symp. Image Video Technol., 2009, pp. 37–47.

[41] Tian, S., Bhattacharya, U., Lu, S., Su, B., Wang, Q., Wei, X., Tan, C. L., Multilingual scene character recognition with co-occurrence of histogram of oriented gradients. Pattern Recognition, 2016, 51, 125-134.

[42] Dalal, N., Triggs, B., Schmid, C. Human detection using oriented histograms of flow and appearance. In European conference on computer vision, 2006, pp. 428-441. Springer, Berlin, Heidelberg.

[43] Ye W, Zhao J, Wang S, et al. Dynamic texture-based smoke detection using Surfacelet transform and HMT model. Fire Safety Journal, 2015, 73:91-101.

[44] Michalek, S., Wagner, M., Timmer, J., “A new approximate likelihood estimator for ARMA-filtered hidden Markov models”. IEEE Transactions on Signal Processing, 48(6):1537–1547, 2000.

[45] Wang, S., He, Y., Yang, H., et al. “Video smoke detection using shape, color and dynamic features,” Journal of Intelligent & Fuzzy Systems, vol. 33, no.1, pp. 305-313. 2017.

[46] Luo, Y., Zhao, I., Liu, P., et al. “Fire smoke detection algorithm based on motion characteristic and convolutional neural networks,” Multimedia Tools & Applications, vol. 77, no.12, pp.15075–15092, 2018

[47] Rodger, J.A., George J.A., "Triple bottom line accounting for optimizing natural gas sustainability: A statistical linear programming fuzzy ILOWA optimized sustainment model approach to reducing supply chain global cybersecurity vulnerability through information and communications technology." Journal of cleaner production, 2017, 142: 1931-1949.