Video smoke detection with block DNCNN and visual change image

Tong Liu[1], Jianghua Cheng[1*], Zhimin Yuan[2], Honghu Hua[1], and Kangcheng Zhao[1]
[1] College of Electronic Science, National University of Defense Technology
Changsha, CO 410073 – China
[2] Naval University of Engineering
Wuhan, CO 430033 – China
[e-mail: hankson@126.com]
*Corresponding author: Jianghua Cheng

Abstract

Smoke detection is helpful for early fire detection. With its large coverage area and low cost, vision-based smoke detection technology is the main research direction of outdoor smoke detection. We propose a two-stage smoke detection method combined with block Deep Normalization and Convolutional Neural Network (DNCNN) and visual change image. In the first stage, each suspected smoke region is detected from each frame of the images by using block DNCNN. According to the physical characteristics of smoke diffusion, a concept of visual change image is put forward in this paper, which is constructed by the video motion change state of the suspected smoke regions, and can describe the physical diffusion characteristics of smoke in the time and space domains. In the second stage, the Support Vector Machine (SVM) classifier is used to classify the Histogram of Oriented Gradients (HOG) features of visual change images of the suspected smoke regions, in this way to reduce the false alarm caused by the smoke-like objects such as cloud and fog. Simulation experiments are carried out on two public datasets of smoke. Results show that the accuracy and recall rate of smoke detection are high, and the false alarm rate is much lower than that of other comparison methods.

Keywords: smoke detection, deep learning, object detection, convolutional neural networks

---

1. Introduction

$\mathbf{O}$f all disasters, fire is one of the most frequent and widespread threats to public safety and social development. Fire not only destroys material property and causes social order chaos but also directly or indirectly endangers life. In addition to various fire prevention measures, fire detection technology is needed to timely detect and deal with fire and ultimately reduce fire hazards. Smoke usually appears when a fire occurs, so rapid smoke detection is conducive to early fire detection and fire hazard reduction. Existing physical smoke sensors are usually used in indoor environments, and the alarm is required only after smoke enters the sensor and reaches a certain concentration. With its large coverage area and low cost, vision-based smoke detection technology is not limited by space, which is the main direction of outdoor smoke detection research [1].

Vision-based smoke detection technology often uses tranditional image features such as color, texture and shape to detect smoke objects [2-7], and the motion features are useful for video smoke detection [8-12]. With the development of deep learning technology, deep features have also been widely used in the field of smoke detection in recent years [13-19]. Although deep learning methods greatly improve the accuracy of smoke detection, detections of smoke-like objects such as cloud and fog remain erroneous.

This paper proposes a video smoke detection method combined with block DNCNN(Deep Normalized CNN) and visual change image. The main contributions are listed below.

(1) A two-stage smoke detection approach is used. First, DNCNN [16] is used to initially detect suspected smoke regions on each image block. Second, these suspected smoke regions are sub-classified according to visual change image and SVM classifier, to reduce error detections for smoke-like objects such as cloud and fog.

(2) The concept of visual change image is put forward. Visual change image can effectively describe the physical diffusion characteristics of smoke in time and space domains. It is effective for distinguishing smoke objects and smoke-like objects such as cloud and fog.

The remainder of the paper is organized as follows. Related work is reviewed in Section 2. The proposed method is described in detail in Section 3. The experimental results and analysis are presented in Section 4. Conclusions are drawn in Section 5.

2. Related Work

Color, texture, shape, and motion are commonly used as tranditional features for smoke detection. Chen et al [2] pointed out that the gray values of smoke color in the RGB model on the three channels are very close, and mainly distributed within the range of 80 to 220. Krstinić et al [3] used HS'I model to describe the color features of smoke, where S' is a a transformation of S for increase separability between smoke and nonsmoke classes. While S<0.5, S'=2S; otherwise S'=1. This color model is better than RGB, YCbCr, CIELab, and HIS for smoke detection. The color features of smoke are insignificant, and many objects in the natural world are similar in color to smoke. So there are more false detections. Gubbi et al [4] put forward a video smoke detection method based on wavelet transformation and SVM. This method extracted 60 characteristics, such as arithmetic mean, geometric mean, deviation, gradient, peak, and entropy, for describing smoke on all subband images of three-level wavelet decomposition. These features were put into a SVM classifier for outputing the smoke detection result. Russo et al [5] used Local Binary Pattern (LBP) and SVM to detect smoke in

image. LBP values and histograms were calculated from the pixels of motion regions to form a feature vector, to describe the texture of smoke. The features classifier was also SVM. Yuan et al [6] introduced a smoke detection algorithm based on the multi-scale features of the LBP and LBPV (LBP based on variance) pyramids. A three-level pyramid was used in the process of feature extraction. Each level pyramid included a Gaussian low-pass filter and a downsampling operation with a step size of 2. However, the interior of smoke is relatively uniform and irregular, and the texture features are insignificant. Yuan et al [7] proposed a double mapping framework for extraction of shape-invariant features based on multi-scale partitions with Adaboost. They eliminated the shape dependency generated by using rules of partitions of detection windows, thus presenting a robust video smoke feature including strengths of edges, LBP and color. The problem is that the smoke is erratic, so the shape feature is weak. Guillemant et al [8] generated a link table and extracted motion features according to the table to detect smoke objects. Toreyin et al [9] extracted the fuzzy and fluctuating features of smoke by using wavelet transformation. Yuan et al [10] proposed an improved fast Horn-Schunck optical flow algorithm to obtain optical flow field, by which the suspected smoke motion regions were detected. The smoke and other interference objects were then distinguished by features of the mean direction and the average velocity of the optical flow vector. Kopilovic et al [11] extracted the distributed entropy of the direction of motion optical flow and identified the irregular features of smoke movement to detect smoke. Tung et al [12] presented a four-stage video smoke detection algorithm. In Stage 1, the motion regions were extracted by using an approximate median method. In Stage 2, the motion regions were clustered to obtain the candidate smoke regions by using the fuzzy c-means method. In Stage 3, the space–time features of the candidate regions were extracted. In Stage 4, the SVM classifier was used to make a judgment. Relying solely on motion features to detect smoke is insufficient because many objects in the scene may have motion.

Compared with traditional features of smoke such as color, texture and shape, deep features are more salient and robust. Xu et al [13] proposed a novel video smoke detection method based on deep saliency network, which highlights the most important object regions in an image by using visual saliency detection method. For extracting the informative smoke saliency map, they combined the pixel-level and object-level salient CNNs. Frizzi et al [14] used CNNs to automatically extract the features of the smoke with strong differentiation, which are more generalized than the features of artificially selected LBP and wavelets. Zhang et al [15] used faster R-CNN to detect smoke, and produced synthetic smoke images by inserting real or simulative smoke into forest background to solve the lack of training data. Yin et al [16] proposed a DNCNN for image smoke detection. The network improved the convolution layer in traditional CNN to the batch standardized convolution layer, which effectively solved gradient dispersion and overfitting in the course of network training. Solving these problems can speed up the training process and improve the detection effect. In addition, the data of the training sample were enhanced to address the imbalance between positive and negative samples and lack of training samples. Although deep learning methods greatly improve the accuracy of smoke detection, false detections for smoke-like objects such as cloud and fog remain.

Combining the deep features with motion features helps further reduce false detections. Yin et al [17] proposed a RCN (Recurrent Convolutional Network) for video-based smoke detection. This method first captured the space and motion context information by using deep convolutional motion-space networks, and then trained the smoke model by using a temporal pooling layer and RNNs. Hu et al [18] proposed a multitask learning strategy to jointly recognize smoke and estimate optical flow, for capturing intra-frame appearance features and

inter-frame motion features simultaneously. However, the features of motion and appearance (color, texture, shape) largely vary. When using the same network, they often affect each other, thus possibly reducing the ability to distinguish each feature. The two-stage method can be used to deal with the apparent and motion features of smoke. Luo et al [19] proposed a fire smoke detection algorithm based on motion characteristic and convolutional neural networks. This method first used a moving object detection algorithm based on background dynamic update and dark channel priori to extract suspected smoke regions, and then used high-capacity convolutional neural networks to extract the feature vectors of the suspected smoke regions and recognize real smoke regions. In this two-stage detection framework, the deep features and motion features work independently and do not affect each other.

3. The Proposed Method

In this paper, we also use two-stage method to deal with the apparent and motion features of smoke. The main defference between our method and methods such as reference [19] is that, we first use deep features to detect the suspected smoke regions, and then use apparent and motion features to recognize real smoke regions. That is because the moving object detection algorithm such as reference [19] will detect many motion regions. However,the objects in these moving regions are not necessarily all moving, and the color and other characteristics of some non-moving objects may be very close to the smoke object, such as clouds and fog, which makes it difficult to distinguish these objects and smoke objects by using subsequent deep features. The framework of our method is shown in **Fig. 1**. We use two-stage detection to detect smoke objects in videos. In the first stage of space domain, we use block DNCNN to detect suspected smoke regions from each sub-block of image. In the second stage of time domain, we compute visual change image of suspected smoke regions, and use SVM to detect any real smoke regions, in this way to reduce the false alarms of smoke detection caused by smoke-like objects such as cloud and fog. In our framework, the reliability of the suspected smoke regions detected by using deep features is significantly higher than that detected by using moving object detection algorithm. Further, the second judgment of the suspicious smoke regions makes full use of the apparent and motion features, which is effective for reducing false alarms of smoke-like objects such as clouds and fog.
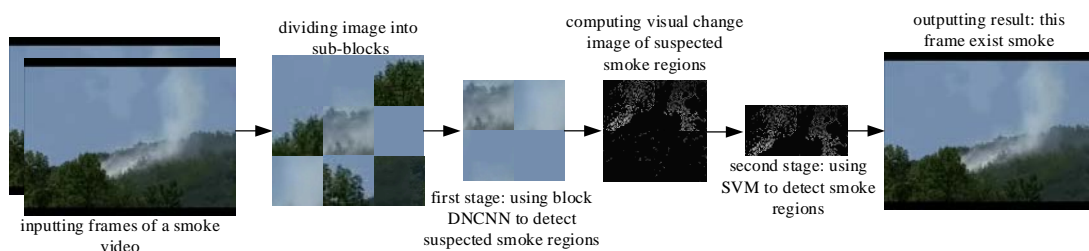

Fig. 1. The framework of our method

3.1 Block DNCNN

DNCNN replaces traditional convolutional layers with normalization and convolutional layers to accelerate the convergence speed of training and improve performance simultaneously. The architecture of the DNCNN classification system for smoke detection is shown in **Fig. 2**. In this figure, each cuboid drawn in dotted lines denotes a stack of feature maps. For example, F^i

stands for a stack of feature maps in the *i*th layer. The widths, heights and numbers of each feature map are listed in **Table 1**, and the types and hyper-parameters of each layer are shown in **Table 2**. DNCNN has fourteen layers. The first eight layers are normalization and convolutional layers alternated with three pooling layers for feature extraction, and the remaining three layers are fully connected layers for classification.
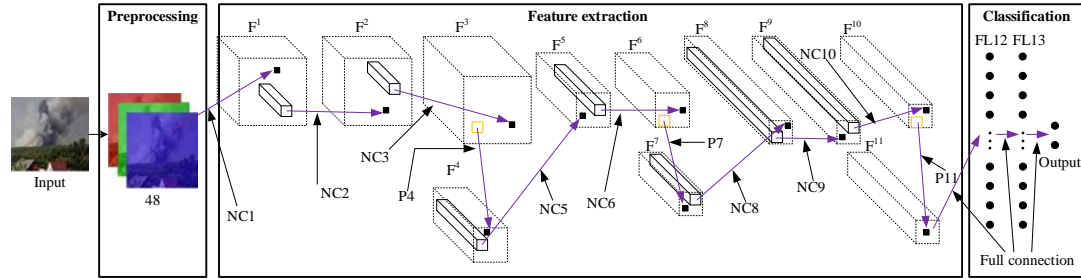

Fig. 2. Architecture of the DNCNN classification system for smoke detection

Table 1. Widths, heights and numbers of each feature map

Feature maps	Widths	Heights	Numbers
F^1	48	48	32
F^2	48	48	32
F^3	48	48	64
F^4	24	24	64
F^5	24	24	64
F^6	24	24	64
F^7	12	12	64
F^8	12	12	384
F^9	12	12	384
F^{10}	12	12	256
F^{11}	6	6	256

Table 2. Types and hyper-parameters of each layer

Layer name	Layer type	Hyper-parameters
input	input	Image size: 48×48
NC1	Normalization and convolution	Filer size k1×k1: 3×3
		Filer number n1: 32
		Stride:s1×s1: 1×1
		Activation function: ReLU
NC2	Normalization and convolution	Filer size k2×k2: 3×3
		Filer number n2: 32
		Stride:s2×s2: 1×1
		Activation function: ReLU
NC3	Normalization and convolution	Filer size k3×k3: 3×3
		Filer number n3: 64
		Stride:s3×s3: 1×1
		Activation function: ReLU
P4	Pooling	Pooling region size k4×k4: 3×3
		Stride:s4×s4: 2×2
		Pooling method: max-pooling

		Filer size k5×k5:3×3
NC5	Normalization and convolution	Filer number n5:64
		Stride:s5×s5: 1×1
		Activation function: ReLU
NC6	Normalization and convolution	Filer size k6×k6: 3×3
		Filer number n6: 64
		Stride:s6×s6: 1×1
		Activation function: ReLU
P7	Pooling	Pooling region size k4×k4: 3×3
		Stride: s7×s7: 2×2
		Pooling method: max-pooling
NC8	Normalization and convolution	Filer size k8×k8: 3×3
		Filer number n8: 384
		Stride:s8×s8: 1×1
		Activation function: ReLU
NC9	Normalization and convolution	Filer size k9×k9: 3×3
		Filer number n9: 384
		Stride:s9×s9: 1×1
		Activation function: ReLU
NC10	Normalization and convolution	Filer size k10×k10: 3×3
		Filer number n10: 256
		Stride:s10×s10: 1×1
		Activation function: ReLU
P11	Pooling	Pooling region size k11×k11: 2×2
		Stride: s11×s11: 2×2
		Pooling method: max-pooling
FL12	Fully-connected	Neurons number n12: 2048
		Dropout probability p: 0.5
		Activation function: ReLU
FL13	Fully-connected	Neuron number n13: 2048
		Dropout probability p: 0.5
		Activation function: ReLU
Output	Output	Neuron number n14: 2

DNCNN is used to classify smoke and non-smoke images. The size of the input image is usually compressed to 48×48. The conclusion whether the image is a smoke image is the output result. The experimental results show that the smoke classification effect of DNCNN is better than that of classic deep learning networks such as AlexNet, ZF-Net, and VGG16 [16].

In outdoor scenes, the smoke region typically takes up a small region throughout the scene. If the image of the entire scene is compressed to 48×48, the smoke region is insignificant in the small resolution image and is prone to misdetection. In this case, a multi-scale sliding window traversal is commonly used to search for smoke in each region of the scene. This method requires classifying many window images, so its efficiency is low. In this paper, the image block [4] approach is used as basis to solve this problem. The input image is divided into non-overlapping image sub-blocks, and then each image sub-block is classified with DNCNN to judge which image sub-block is a suspected smoke region. This step lays the foundation for subsequent fine classification of smoke. Specifically, the size of the input video image is 720×576, and it is divided into non-overlapping 48×48 image sub-blocks from left to right and from top to bottom. As illustrated in **Fig. 3**, the 720×576 image can be divided into 180 image sub-blocks with size of 48×48. Each 48 × 48 image sub-block is input in a trained DNCNN

model, and the image sub-blocks with an output of 0 are marked as suspected smoke image region. The smoke classification model trained by DNCNN has a low rate of missed detection on smoke images, but it is prone to false detection of images with texture and color similar to smoke, such as cloud and fog.

Fig. 3. Block partition of an image

3.2 Visual change image

Suspected smoke regions in the image can be detected with block DNCNN. However, because objects such as cloud and fog are similar in color spatial distribution and shape to the smoke, the use of a block DNCNN can still incorrectly detect some objects such as cloud and fog. We analyze the occurrence and development of smoke and find that it has clear upward diffusion characteristics. Smoke starts with a fire point and continues to spread up and around. The motion and diffusion of cloud and fog have no starting point and no fixed direction, which are markedly different from the diffusion of smoke. In this paper, the concept of visual change image is proposed to describe the diffusion characteristics of smoke as the basis for distinguishing between smoke objects and smoke-like ones such as cloud and fog.

The visual change image described in this paper is obtained cumulatively by the variation image of the suspected smoke object in the time domain. The variation image of the suspected smoke object is used to reflect changes between adjacent frames of the suspected smoke object and is constructed by the following steps.

First, for each suspected smoke region detected by block DNCNN, a rough smoke mask extraction is implemented on the basis of color model. From the analysis of physical characteristics, smoke is mostly gray-white due to relatively low temperature. The color values (R, G, and B) of the pixel points in the RGB space are very close, and the value of saturation of S in HSV space is relatively small. Therefore, when the color value of the pixel

point (x, y) satisfies Formula (1), the pixel may be a smoke pixel point.

$$\begin{cases} \max(x, y) - \min(x, y) < T_1 \\ \dfrac{\max(x, y) - \min(x, y)}{\max(x, y) + 0.001} < T_2 \end{cases} \tag{1}$$

where, $max(x, y)$ and $min(x, y)$ represent the maximum and minimum values of R, G, and B at the pixel point (x, y) on the RGB color model, respectively. T_1 and T_2 are two fixed thresholds. In this paper, $T_1=10$ and $T_2=0.1$. In this way, a mask image corresponding to the image of the kth frame can be constructed and marked as I_k, which is expressed as follows:

$$I_k(x, y) = \begin{cases} 255, \text{pixel in the suspected area and the pixel values satisfy formula(1)} \\ 0, \text{otherwise} \end{cases} \tag{2}$$

That is, the point with a pixel value of 255 in the image is a pixel of a suspected smoke region.

On this basis, the inter-frame difference method is used to calculate the variation image of the suspected smoke region. Specifically, for the kth frame image f_k, the variation image d_k of the suspected smoke region can be expressed as

$$d_k(x, y) = \begin{cases} |f_k(x, y) - f_{k-1}(x, y)| & , I_k(x, y) > 0 \\ 0 & , I_k(x, y) = 0 \end{cases} \tag{3}$$

The variation image of the suspected smoke region described in this paper can reflect changes between adjacent frames of the suspected smoke object.

The visual change image is the sum of adjacent N variation images, marked as F_k, and represented as

$$F_k(x, y) = \sum_{i=k-N+1}^{k} d_i(x, y) \tag{4}$$

To partially display and extract features, the value of each pixel point in F_k is compressed to 255, specifically,

$$F_k^*(x, y) = \frac{255}{\max\limits_{x, y}\left(F_k(x, y)\right)} F_k(x, y) \tag{5}$$

where, $\max\limits_{x, y}\left(F_k(x, y)\right)$ is the maximum value of each pixel point in F_k. F_k^* then can be regarded as a grayscale image.

For the characteristic analysis of smoke diffusion, the smoke occurs from the point of ignition, and the area changes most of the time. While the smoke spreads upward and outward, the surrounding area begins to change. The longer the area is affected by the spread of smoke, the longer the time lags. These characteristics are represented in the visual change image. Overall, the brightness level in the visual change image of smoke object is larger. Moreover, the upper the pixel position is, the darker the brightness is, the more the number of non-zero pixels is, as shown in **Fig. 4(a)**. This characteristic is obviously different from the smoke-like objects such as cloud and fog. For cloud and fog, their position movement or diffusion is relatively consistent, and the change time of each position is relatively close, so the brightness level in the visual change image is small, as shown in **Fig. 4(b)**. With this characteristic, the visual change image is used to distinguish smoke from smoke-like objects such as cloud and fog.
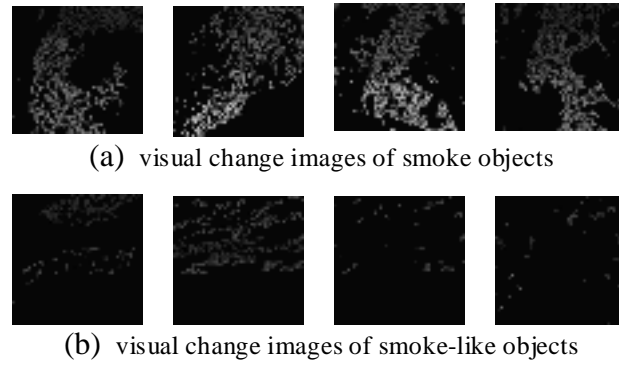
(a) visual change images of smoke objects

(b) visual change images of smoke-like objects

Fig. 4. Comparison of visual change images of smoke and smoke–like objects

3.3 Implementation method

The step-by-step implementation is depicted in **Fig. 5**. The steps are as follows:

Step 1: The image of kth frame is inputted, and the image size is unified to 720×576. If the image size does not meet the requirements, then the image size is scaled to 720×576 by bilinear interpolation method.

Step 2: If $k>1$, then the current image is not the first frame image, so the next step is taken. Otherwise, the current frame image is cached, $k=k+1$, and Step 1 is repeated.

Step 3: Block DNCNN is used to detect the suspected smoke regions, called ROI, and the number of ROI is named N_{ROI}.

Step 4: If $N_{ROI}>0$, then the suspected smoke regions are detected, and the next step is taken. Otherwise, the current frame image is cached, $k=k+1$, and Step 1 is repeated.

Step 5: The mask image I_k corresponding to the kth frame image is constructed using the color models of Formulas (1) and (2).

Step 6: Formula (3) is used to extract the variation image d_k of the kth frame.

Step 7: The variation image d_k is cached. N represents the maximum value of the number of variation images in the cache. When the number of variation images in the cache is greater than N, the oldest image stored in the cache space is first deleted, that is, d_{k-N}, and the current variation image d_k is cached. In this paper, $N=25$.

Step 8: If $k>N$, then the next step is taken. Otherwise, the current frame image is cached, $k=k+1$, and Step 1 is repeated.

Step 9: Formulas (4) and (5) are used to extract the visual change image F_k^*.

Step 10: Feature extraction is conducted in a suspected smoke regions. For each suspected smoke region detected in the kth frame image, the sub-image in the corresponding area is cropped from the visual change image F_k^*. Next, the HOG features of the image is extracted. The details of HOG features extraction can be seen in reference [20]. In this paper, the value of block size is set to 8×8, the value of block stride is set to 4×4, the value of cell size is set to 4×4, and the value of $n$bins is set to 9.

Step 11: SVM classification entails a secondary judgment on suspected smoke regions. The main goal is to distinguish smoke images from smoke-like images such as cloud and fog. The specific method involves using SVM classifier constructed in the training phase to classify the HOG features of each suspected smoke region of the kth frame. As long as one area is classified as smoke, then smoke alarm is the output. After the classification, the current frame

image is cached, regardless of whether the alert is output, $k=k+1$, and Step 1 is repeated.
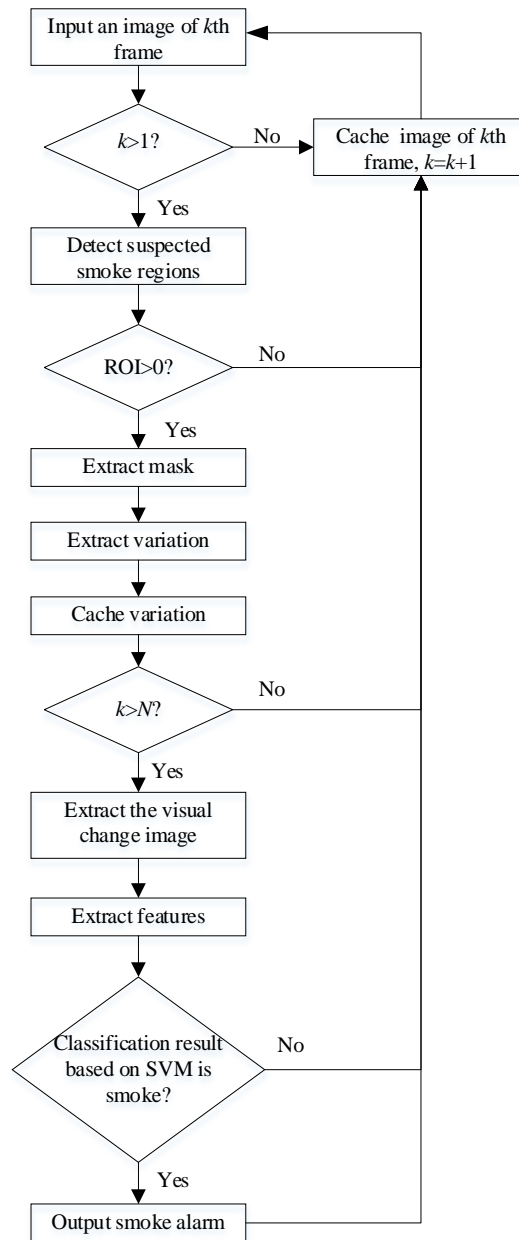

Fig. 5. Smoke detection flow chart of the proposed method

4. Experiment and Analysis

4.1 Experiment description

(1) Experimental dataset

In this paper, the performance of the algorithms are tested on two public smoke datasets. One

is the dataset provided by Prof. Yuan, which includes 4 sets of image datasets and a video dataset for 6 scenes [21]. Image set 1 includes 552 smoke images and 831 non-smoke images; set 2 includes 668 smoke images and 871 non-smoke images; set 3 includes 2,201 smoke images and 8,511 non-smoke images; and set 4 includes 2,254 smoke images and 8,363 non-smoke images. The video dataset has 3 smoke scenes (Dry leaf smoke 02, Cotton rope smoke 04, and Black smoke 05) and 3 non-smoke scenes (Traffic 1000, Basketball yard, and Waving leaves). The other is CVPR Lab's video dataset, which includes smoke videos of 2 indoor scenes and 4 outdoor scenes as well as 10 non-smoke videos [22]. The description of the dataset used in this paper is shown in **Table 3**. Some samples are shown in **Fig. 6**.

Table 3. Description of the dataset used in this paper

Name	Purpose	Description
Yuan-imageset-train	training	This dataset includes 4 sets of image data of Prof. Yuan. The sizes of all images in the dataset are resized to 48×48 by bilinear interpolation method for DNCNN model training.
Yuan-videoset-train	training	This dataset includes one segment of smoke video and one segment of non-smoke video from the video datasets of Prof. Yuan.
Yuan-videoset-test	testing	This dataset includes two segments of smoke video and two segments of non-smoke video from the video datasets of Prof. Yuan.
CVPR-videoset-test	testing	This dataset includes two segments of indoor smoke scene and four segments of outdoor smoke scene from the video datasets of CVPR Lab.

(a) some smoke frames   (b) some non-smoke frames

Fig. 6. Some samples in dataset

(2) Evaluation metrics

Accuracy Rate (AR), Recall Rate (RR) and False Alarm Rate (FAR) are three commonly used metrics for quantitatively comparing different smoke detection algorithms. They are denoted as follows:

$$AR = \frac{TT+FF}{T+F} \tag{6}$$

$$RR = \frac{TT}{TT+FF} \tag{7}$$

$$FAR = \frac{TF}{TT+TF} \qquad (8)$$

TT represents the number of samples with a category of True predicted to be True; FT represents the number of samples with a category of True predicted to be False, and T=TT+FT. FF represents the number of samples with a category of False predicted to be False; TF represents the number of samples with a category of False predicted to be True, and F=FF+TF. The category of smoke images is True, and the category of non-smoke images is False.

(3) Experimental environment

The experimental environment is Windows10 system, Python 3.6.2, Tensorflow1.11.0, Keras2.2.4,   In the hardware device section, the graphics card is NVIDIA GeForce GTX1080Ti, and the CPU is Intel Core i7-8700K.

4.2. Classifier training

(1) DNCNN model training

The DNCNN model training refers to Yin et al [16]'s method. Training-relevant hyper-parameters is listed in **Table 4**, and the training dataset uses the Yuan-imageset training dataset in **Table 3**. The training curves are illustrated in **Fig. 7**.

Table 4. Training-relevant hyper-parameters of DNCNN

Hyper-parameters	Initial learning rate	Mini-batch size	Momentum coefficient	Learning rate decay coefficient
Value	0.01	96	0.9	0.01

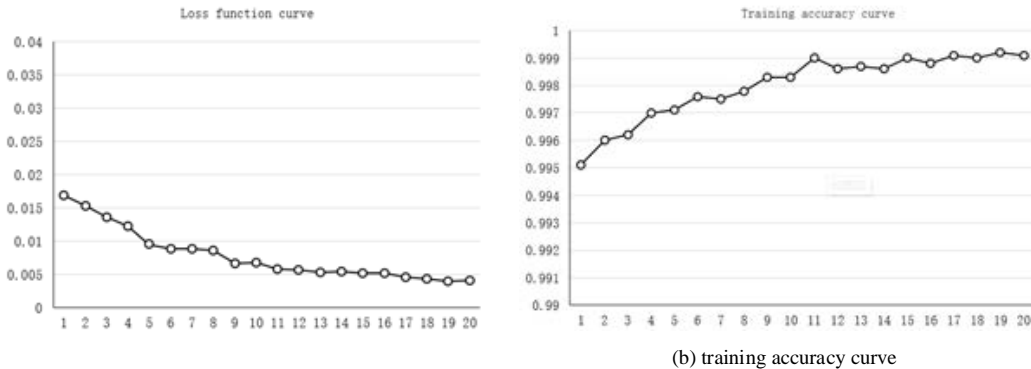(b) training accuracy curve

Fig. 7. Training curves of DNCNN model

(2) SVM classifier training

Yuan-videoset training dataset is selected to conduct SVM classifier training. A segment of video is inputted to generate a visual change image by following Steps 1 to 9 in Section 3.3. Step 10 is followed to crop the sub-images of the corresponding region from the visual change images. Next, two datasets are constructed with these images, wherein the sub-images corresponding to the real smoke region are placed in the positive sample dataset and the sub-images of the non-smoke region are placed in the negative sample dataset. The image size is 48×48. Finally, according to the HOG feature extraction method described in Step 10, the

image features of the training set are extracted. LIBSVM software development package is used for training, and the SVM classifier is constructed. The kernel function selects the radial base function.

4.3. Performance comparison

(1) Two-stage smoke detection performance test

In this paper, two-stage detection is used. In the space domain, each frame of the image of the suspected smoke region is initially detected using block DNCNN. In the time domain, the suspected smoke region is classified in detail using a visual change image. To verify the performance of the two-stage smoke detection method, the following experiments are carried out:

Experiment-1: Using the detection steps in Section 3.3, Step 11 directly outputs the smoke alarm regardless of the SVM classification result, i.e., only block DNCNN is used to detect whether there is smoke.

Experiment-2: Using the detection steps in Section 3.3, the presence of smoke is detected with the two-stage classification of block DNCNN and SVM.

The experiment results are provided in **Table 5**. Some detection results of Experiment-2 are illustrated in **Fig. 8**. Under the two datasets, the AR and FAR indicators of the two-stage detection method are better than those using the block DNCNN method alone. In particular, when the FAR value has been decreased more than twice, the RR index is equal. Using two-stages to detect smoke can obviously reduce the false alarm rate.

Table 5. Comparisons with two-stage detection of our algorithm

Experiments	Yuan-videoset-test			CVPR-videoset-test		
	AR	RR	FAR	AR	RR	FAR
Experiment-1	0.9949	0.9953	0.0015	0.8968	0.9044	0.0037
Experiment-2	**0.9972**	0.9953	**0.0007**	**0.9285**	0.9031	**0.0009**

(a) some correct detection results

(b) some wrong detection results. up: non-smoke frames incorrectly detected as smoke;down: undetected smoke frames

Fig. 8. Some detection results

Fig. 9 illustrates comparison results when the number of training samples is different. Among them, 'Rate' represents the proportion of the actual number of samples participating in

the training to the total number of samples in the training dataset 'Yuan-imageset-train', 'value' represents the average AR value under two datasets including 'Yuan-videoset-test' and 'CVPR-videoset-test'. As 'Rate'increases, the AR value of two experiments all increase, too. However, compared with Experiment-1, the AR value in Experiment-2 is less affected by 'Rate'. This shows that the two-stage detection method in this paper is less affected by the number of training samples and more adaptable.
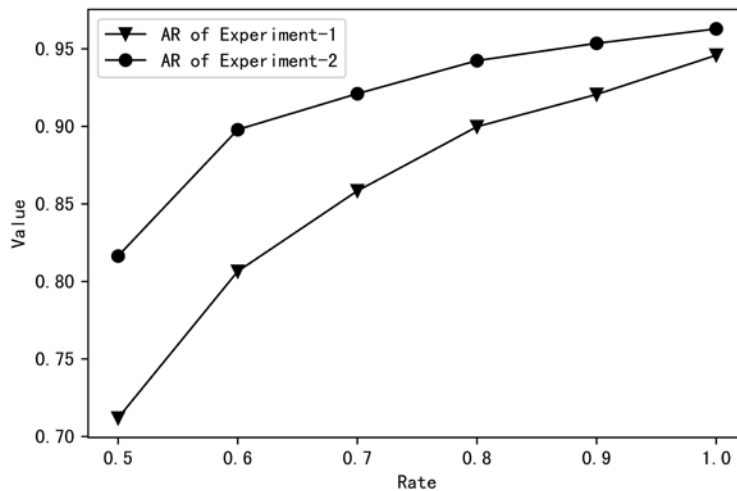

Fig. 9. Comparison results when the number of training samples is different

(2) Comparison of smoke detection performance of different methods

Table 6 compares the performance of the proposed method with some popular smoke detection methods on Yuan-videoset test and CVPR-videoset testing datasets. It needs to be explained that the implementation process of these comparison methods in our experiment is designed according to the implementation steps from corresponding references [3, 5, 11, 14, 16-19]. In order to improve the performance, references [3] and [5] also uses image block approach described in section 3.1, and the feature classification uses a SVM classifier. For comparison purposes, the testing and training dataset are the same as shown in **Table 3**, and experimental software and hardware environments are the same, too.

In **Table 6**, the FAR value of HS'I [3] is very large, because this method mainly used color features to detect smoke objects, and many objects similar to smoke color are misdetected as smoke objects. The method of LBP+LBPV[5] mainly uses texture features to detect smoke objects. Although texture features are slightly better than color features, they are not significant enough too. Therefore the FAR is still large. Motion features used in reference [11] can greatly reduce false detection in non-moving regions, leading to a reduction in FAR value. However, there are many moving object in some scenes, and it is difficult to accurately identify the smoke objects due to motion features. In general, the AR and RR values of smoke detection methods based on deep features(such as references [14, 16-19] and our method) are higher than that of traditional features (such as references [3, 5, 11]). That is because the deep features learned through big data can more fully mine the smoke features with stronger discrimination ability. Further, the combination of deep features and motion features can significantly reduce false detection in non-moving regions, so the FAR value of references

[17-19] and our method reduces. References [17, 18] use same network to extract deep features related to apparent and motion features, and may reduce discrimination ability while apparent and motion features affect each other. Reference [19] and our method uses two-stage method to deal with the apparent and motion features of smoke for reducing mutual interference. Specifically, reference [19] first uses a moving object detection algorithm to extract suspected smoke regions, and then uses high-capacity convolutional neural networks to select real smoke regions from the suspected smoke regions. The FAR value of this method reduces, but the AR and RR values are not high because the suspected smoke regions exist some non-moving objects with similar color and texture features like smoke. Our method also uses two-stage method to deal with the apparent and motion features of smoke. The main defference is that, we first use deep features to detect the suspected smoke regions, and then use apparent and motion features to recognize real smoke regions. In our method, block DNCNN is used to detect the suspected smoke region, and secondary judgment is made on the basis of visual change image characteristics. As long as an image sub-block is determined to be smoke, the smoke alarm is output, which further improves the AR and RR values as well as greatly reduces the FAR value by the secondary judgment. Therefore,the AR, RR, and FAR values of our method are better than those of other methods under both Yuan-videoset and CVPR-videoset-test datasets, that can be seen from **Table 6**. In addition, the test results of various methods on the CVPR-videoset testing dataset are worse than those on the Yuan-videoset testing dataset, because the training dataset is more similar to the smoke patterns in the latter.

Table 6. Comparisons with other methods

Methods	Yuan-videoset-test			CVPR-videoset-test		
	AR	RR	FAR	AR	RR	FAR
HS'I [3]	0.7157	0.8933	0.0128	0.6784	0.8623	0.0143
LBP+LBPV[5]	0.8949	0.9182	0.0046	0.8429	0.8810	0.0069
optical flow [11]	0.9194	0.9037	0.0019	0.8298	0.8774	0.0027
CNN [14]	0.9628	0.9411	0.0022	0.8836	0.9009	0.0029
DNCNN [16]	0.9793	0.9616	0.0019	0.8917	0.8903	0.0033
RCN [17]	0.9686	0.9632	0.0013	0.8992	0.9044	0.0019
Spatial-temporal ConvNet [18]	0.9577	0.9368	0.0011	0.8799	0.8906	0.0023
Motion detection+CNN[19]	0.9673	0.9402	0.0012	0.8924	0.8911	0.0020
Our method	**0.9972**	**0.9953**	**0.0007**	**0.9285**	**0.9031**	**0.0009**

5. Conclusions

Video smoke detection has been a hot spot in fire detection in recent years. Detection is difficult because of the uncertainty of the smoke object. Traditional color, texture, and other detection methods are impractical because their high false alarm rate. Deep learning-based smoke detection methods greatly improve the performance of smoke detection, but relying solely on deep features is difficult to effectively distinguish smoke-like objects such as cloud and fog. In this paper, the deep features are combined with the diffusion characteristics of smoke motion. Following the idea of two-stage detection, DNCNN method is first used to extract the deep features of smoke image for preliminary detection of suspected smoke regions. The probability of visual change image is then put forward for secondary classification of the visual change image of suspected smoke regions combined with SVM, significantly reducing

the false alarm rate of smoke detection. With visual change images, the physical diffusion characteristics of smoke can be described effectively in time and space domains, which are crucial for distinguishing smoke and smoke-like objects. The simulation experiment proves that the two-stage detection proposed in this paper improves the performance of video smoke detection and especially reduces greatly the false alarm rate of smoke detection. How to improve the efficiency of the algorithm will be further studied.

References

[1] Shi J, Yuan F, and Xue X, "Video smoke detection: a literature survey," *Image Graph*, Vol. 23, No. 3, pp. 303-322, 2018. Article (CrossRef Link)

[2] Chen T H, Yin Y H, and Huang S F, eds., "The smoke detection for early fire-alarming system base on video processing," in *Proc. of International Conference on Intelligent Information Hiding and Multimedia. Pasadena, CA, USA: IEEE*, pp. 427-430, 2006. Article (CrossRef Link)

[3] Krstinić D, Stipaničev D, and Jakovčević T, "Histogram-based smoke segmentation in forest fire detection system," *Information Technology and Control*, Vol.38, No. 3, pp. 237-244, 2009.

[4] Gubbi J, Marusic S, and Palaniswami M, "Smoke detection in video using wavelets and support vector machines," *Fire Safety Journal*, Vol. 44, No. 8, pp. 1110-1115, 2009. Article (CrossRef Link)

[5] Russo A U, Deb K, and Tista S C, eds., "Smoke detection method based on LBP and SVM from surveillance camera," in *Proc. of 2018 International conference on computer, communication, chemical, material and electronic engineering (IC4ME2). IEEE*, pp. 1-4, 2018. Article (CrossRef Link)

[6] Yuan F, "Video-based smoke detection with histogram sequence of LBP and LBPV pyramids," *Fire Safety Journal*, Vol. 46, No. 3, pp. 132-139, 2011. Article (CrossRef Link)

[7] 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. Article (CrossRef Link)

[8] Vicente J, and Guillemant P, "An image processing technique for automatically detecting forest fire," *International Journal of Thermal Sciences*, Vol. 41, No. 12, pp. 1113-1120, 2002. Article (CrossRef Link)

[9] Toreyin B U, Dedeoglu Y, and Cetin A E, "Contour based smoke detection in video using wavelets," in *Proc. of 14th European Signal Processing Conference. Florence, Italy: IEEE*, pp. 1-5, 2006. Article (CrossRef Link)

[10] Yuan P, Hou X, and Pu L, "A Smoke Recognition Method Combined Dynamic Characteristics and Color Characteristics of Large Displacement Area," in *Proc. of 2018 2nd IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC). IEEE*, pp. 501-507, 2018. Article (CrossRef Link)

[11] Kopilovic I, Vagvolgyi B, and Szirányi T, "Application of panoramic annular lens for motion analysis tasks: surveillance and smoke detection," in *Proc. of 15th International Conference on Pattern Recognition. Barcelona, Spain: IEEE*, pp. 714-717, 2000. Article (CrossRef Link)

[12] Tung T X, and Kim J-M, "An effective four-stage smoke-detection algorithm using video images for early fire-alarm systems," *Fire Safety Journal*, Vol. 46, No. 5, pp. 276-282, 2011. Article (CrossRef Link)

[13] Xu G, Zhang Y, and Zhang Q, eds., "Video Smoke Detection Based on Deep Saliency Network," *Fire Safety Journal*, Vol. 105, pp. 277-285, 2019. Article (CrossRef Link)

[14] Frizzi S, Kaabi R, and Bouchouicha M, eds., "Convolutional neural network for video fire and smoke detection," in *Proc. of IECON 2016 - 42nd Annual Conference of the IEEE Industrial Electronics Society. Florence: IEEE*, pp. 877-882, 2016. Article (CrossRef Link)

[15] Zhang Q, Lin G, and Zhang Y, eds., "Wildland forest fire smoke detection based on faster R-CNN using synthetic smoke images," *Procedia engineering*, Vol. 211, pp. 441-446, 2018. Article (CrossRef Link)

[16] Yin Z , Wan B , and Yuan F, eds., "A deep normalization and convolutional neural network for image smoke detection," *IEEE Access*, Vol. 5, pp. 18429-18438, 2017. Article (CrossRef Link)

[17] Yin M X, Lang C Y, and Li Z, eds., "Recurrent convolutional network for video-based smoke detection," *Multimedia Tools and Applications*, Vol. 78, No. 1, pp. 237-256, 2018. Article (CrossRef Link)

[18] Hu Y C, and Lu X B, "Real-time video fire smoke detection by utilizing spatial-temporal ConvNet features," *Multimedia Tools and Applications*, Vol. 77, No. 22, pp. 29283-29301, 2018. Article (CrossRef Link)

[19] Luo Y , Zhao L , and Liu P , eds., "Fire smoke detection algorithm based on motion characteristic and convolutional neural networks," *Multimedia Tools and Applications*, Vol. 77, pp. 15075-15092, 2018. Article (CrossRef Link)

[20] Dalal N, and Triggs B, "Histograms of oriented gradients for human detection," in *Proc. of international Conference on computer vision & Pattern Recognition (CVPR'05). IEEE Computer Society*, Vol. 1, pp. 886-893, 2005. Article (CrossRef Link)

[21] http://staff.ustc.edu.cn/~yfn/vsd.html.

[22] https://cvpr.kmu.ac.kr/.

Tong Liu received his B.S., M.S. and Ph.D. degrees from the National University of Defense Technology in 2004, 2006 and 2015. His main research interests include computer vision and electrical system.

Jianghua Cheng received his B.S., M.S. and Ph.D. degrees from the National University of Defense Technology in 2002, 2005 and 2013. His main research interests include computer vision and electrical system.

Zhimin Yuan received his B.S., M.S. and Ph.D. degrees from the Nati0onal University of Defense Technology in 2004, 2006 and 2012. His main research interests include computer vision and cyber security.

Honghu Hua received his B.S. and M.S. degrees from the National University of Defense Technology in 2004 and 2006. His main research interests include computer vision and electrical system.

Kangcheng Zhao is a master's student at the National University of Defense Technology. His main research interests include computer vision and electrical system.