Smoke recognition network based on dynamic characteristics

Dahan Wang¹, Sheng Luo², Li Zhao², Xiaoming Pan², Muchou Wang² and Shunzhi Zhu¹

Abstract
Fire is a fierce disaster, and smoke is the early signal of fire. Since such features as chrominance, texture, and shape of smoke are very special, a lot of methods based on these features have been developed. But these static characteristics vary widely, so there are some exceptions leading to low detection accuracy. On the other side, the motion of smoke is much more discriminating than the aforementioned features, so a time-domain neural network is proposed to extract its dynamic characteristics. This smoke recognition network has these advantages: (1) extract the spatiotemporal with the 3D filters which work on dynamic and static characteristics synchronously; (2) high accuracy, 87.31% samples being classified rightly, which is the state of the art even in a chaotic environments, and the fuzzy objects for other methods, such as haze, fog, and climbing cars, are distinguished distinctly; (3) high sensitiveness, smoke being detected averagely at the 23rd frame, which is also the state of the art, which is meaningful to alarm early fire as soon as possible; and (4) it is not been based on any hypothesis, which guarantee the method compatible. Finally, a new metric, the difference between the first frame in which smoke is detected and the first frame in which smoke happens, is proposed to compare the algorithms sensitivity in videos. The experiments confirm that the dynamic characteristics are more discriminating than the aforementioned static characteristics, and smoke recognition network is a good tool to extract compound feature.

Keywords
Smoke detection, fire alarming, neural network, dynamic characteristics, convolution 3D network, motion information

Date received: 29 February 2020; accepted: 21 April 2020

Introduction
About 6–7 million times of fires happen every year, which cause huge economic losses. About 0.2% of gross domestic product of the Global are destroyed and 1 billion people are killed by fire.¹ This is only the direct loss, and the indirect loss is about five times of the direct loss. So, it is very important to detect fire and provide an alarm as early as possible. It is well-known that smoke always happens before flame and smoke is difficult to be sheltered. Therefore, it is a good idea to detect smoke rather than flame.

Many methods are developed to detect smoke based on the chrominance, texture, shape, frequency, transparency, and so on.¹⁰ But smoke chrominance is widely varied according to fire material, and its varying range is too wide to cover all samples. The common smoke chrominance are black, white, gray, and pink. Smoke also has many types of textures, which is related to frequency. The

¹ School of Computer and Information Engineering, Xiamen University of Technology, Xiamen, China
² College of Computer Science and Artificial Intelligence, Wenzhou University, Wenzhou, China

Corresponding author:
Sheng Luo and Li Zhao, College of Computer Science and Artificial Intelligence, Wenzhou University, Wenzhou 325000, China.
Emails: ls2008@wzu.edu.cn, lizhao@wzu.edu.cn

Creative Commons CC BY: This article is distributed under the terms of the Creative Commons Attribution 4.0 License (https://creativecommons.org/licenses/by/4.0/) which permits any use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (https://us.sagepub.com/en-us/nam/open-access-at-sage).
frequency of heaven smoke is easy to obtain, but this computing is difficult for the one of thin smoke because the frequency of the latter is flooded by chaotic environments. So, the methods based on chrominance, texture, frequency, and transparency are difficult to differentiate smoke from other similar objects. If smoke is thin or less, which usually happens at the early stage of fire, the classifying accuracies would be more worse. Haze and fog are usually mistaken as smoke. That is also the reason why these traditional methods are not enough sensitive.

Smoke is always rising, so it is a good idea to make use of motion. Some methods have been developed based on this characteristic,\textsuperscript{7,10–14} and these works are concluded by Luo et al.\textsuperscript{15} Smoke motion is very different from other objects, but there is still some confusing objects, such as waving flags, climbing vehicles on mountain, moving light, and so on.\textsuperscript{11} Though the appearances of flag, vehicle, and light are completely different from the one of smoke, their motion characteristics are similar. These exceptions usually happen when smoke is assumed to move vertically up. If this hypothesis is not assumed and the moving direction is random, there will be more exceptions. Though it is true that smoke always goes upwardly, there is a implicit assumption that an axis of the camera frame should be perpendicular to the horizon. In most cases it is met, but it is a hypothesis with some exceptions.

Though dynamic characteristics of smoke is different from many others, it is difficult to describe and extract this motion information by hand. That is why there is not many researches to detect smoke by dynamic characteristics. Deep learning is a good tool to extract features, so this article proposes a deep neural network to detect smoke. This network focuses on time-domain features of consecutive images. Though there is no such benchmark as ImageNet, COCO to evaluate the proposed method, a contrasting experiment is still conducted based on several small data sets, such as visor,\textsuperscript{16} fesb,\textsuperscript{17} yfn,\textsuperscript{18} and VisiFire.\textsuperscript{19} The experiments validate that: (1) dynamic characteristics is extraordinarily discriminating with better scores even based on these extern data sets and (2) smoke recognition network is a good tool to extract compound feature. The pipeline of the proposed network is illustrated in Figure 1. The probabilities of smoke, cloud, fire, flag, fog and haze, vehicle, and others of a typical sample are $[99.88\ 0\ 0\ 0\ 0.12\ 0\ 0]$, respectively, and the maximum, 99.88%, shows that the image includes smoke.

The rest of this article is organized as four sections. In the second section, we describe a brief state of the art in smoke dynamic characteristics. The third section presents the details of our proposed method. The fourth section highlights some experiments with comparison with other methods. Finally, the fifth section provides a conclusion.

**Relevant work**

Such static characteristics as chrominance,\textsuperscript{2,3} texture,\textsuperscript{4–6} shape,\textsuperscript{7} frequency,\textsuperscript{9} and transparency\textsuperscript{19} are adopted to detect smoke, but the wide-ranging varieties of these characteristics make the recognition difficult.\textsuperscript{15} Even some of these characteristics are synthesized with machine learning methods.\textsuperscript{4,8,19–23,24} Based on surface wave and Markov tree model to extract smoke textures, Ye et al. adopted support vector machine (SVM) to judge whether there is any smoke.\textsuperscript{4} After synthesizing the edge blurring, the gradual changing of smoke energy and the gradual changing of smoke color, Lee et al. also adopted SVM to make decisions.\textsuperscript{20} After spectral features are extracted, Li et al.\textsuperscript{25} classified smoke, cloud, and underlying surface by a back-propagation neural network. Yuan et al.\textsuperscript{8} extracted...
Haar-like features and statistical features, and made decisions by AdaBoost.

Dynamic characteristics, such as moving speed, moving direction, moving trail, and the changing of smoke contour during moving, which usually are accompanied with tracking algorithms, have been already adopted to detect smoke. Yu et al. adopted optical flow to trace smoke. Lu et al. adopted the spatiotemporal histogram to calculate the speed and acceleration information of smoke and judged whether it is smoke. Yuan adopted a new frame to collect the moving areas and created an accumulative motion model to estimate the motion orientation of smoke. Calderara et al. detected scope of moving smoke through energy variation. Tian et al. adopted NRLBP (no repeated LBP algorithm) to extract smoke features, recorded the continual changes in time domain, and made the final decision by SVM. These works have been proved that dynamic characteristics is discriminating, but the distinguishing ability is not very outstanding because the dynamic characteristics are extracted into some hand crafted features. And there are some hypotheses underlying. The commonly used hypothesis is that camera should be fixed with one axis being vertical, so smoke rises upward. It is helpful to improve the accuracy but the improvement is limited. It is still difficult to be applied in surveillance scenario.

Since both static and dynamic characteristic are helpful, some synthesized method are proposed. Calderara et al. used Bayesian model combined texture and color features of moving objects to detect smoke. But the result is still worrying. The confusing objects, such as cloud, fog, and haze, cannot be rightly distinguished. The reason might be that smoke are too various, and hand crafted features are not robust enough to include all diversification. So, there is still not an automatic product which could recognize smoke for other objects.

Deep learning is a new branch of machine learning with the capability to automatically extract features by a series of convolutional layers. Convolutional neural network (CNN) has been proved to be good enough to extract static features from images with an outstanding discriminatory power, equaling or exceeding human level. Sheng et al. extracted smoke static features by CNN and dynamic characteristic by recurrent neural network to recognize smoke. But the 2D convolution kernel could not collect the temporal information between the consecutive images. Fortunately, Tran et al. has proposed a 3D convolution (C3D) network that included 3D convolution kernel and 3D pooling layer for action recognition, which could automatically extract the temporal and spatial features and achieve an accuracy of 85.2% on UCF-101 data set. Inspired by this network, a simplified network is proposed to extract dynamic characteristics, which achieved an accuracy of 98.07% in visor data set, exceeding the best score of today, 85.20%.

The proposed method

The network structure

The 3D ConvNet is very suitable for the spatiotemporal signals. Compared with 2D ConvNet, 3D ConvNet can better model temporal information through 3D convolution and 3D pooling. In 3D ConvNets, convolution and pooling operations are performed spatially and spatially. While in 2D ConvNets, they are only performed spatially. The input of our network is 16 consecutive images, and the kernels of all convolution and pooling layers are 3D, defined by depth, length, and width. For examples, the size of all convolutional kernels are $3 \times 3 \times 3$, which has been proved to be better than $7 \times 7 \times 7$ and $5 \times 5 \times 5$ in the smoke task, and the one of all pooling kernels is $2 \times 2 \times 2$. The convolutional layers can find various localized features, and output different features. The pool layers condense the large amount features from the convolutional layers into small vectors.

Our network has 13 layers, as shown in Figure 2. The network includes five convolution layers, and the channels of the first to fifth convolutional layers are 64, 128, 256, 256, and 256, respectively. The strides of all convolution layers are: (1) 5 max pool layers are included and their strides are 2.
Table 1. The sample data set.

| Data sets         | Number of videos (duration time) | Number of sub-video clips | Train    | Test     | Label |
|-------------------|----------------------------------|---------------------------|----------|----------|-------|
| Smoke             | 35 (168 min)                     | 6412                      | 5130     | 1282     | 0     |
| Cloud             | 18 (91 min)                      | 5302                      | 4242     | 1060     | 1     |
| Fire              | 34 (86 min)                      | 4171                      | 3336     | 835      | 2     |
| Flag              | 5 (15 min)                       | 848                       | 678      | 170      | 3     |
| Fog and haze      | 10 (46 min)                      | 1913                      | 1530     | 383      | 4     |
| Vehicle           | 8 (17 min)                       | 1489                      | 1191     | 298      | 5     |
| Others            | 42 (12 min)                      | 2705                      | 2164     | 541      | 6     |

(2) After these layers, there are two fully connected layers, each of which has 1024 hidden neural units. A softmax layer follows the last fully connected layers, which converts video clip with size of $16 \times 3 \times 112 \times 112$ into a $7 \times 1$ output during a transitional $512 \times 1 \times 4 \times 4$ result. The output is a probability array with seven elements, which represent the probabilities of smoke, cloud, fire, flag, fog and haze, vehicle, and others, respectively. The position with the maximum value in the array is the most possible category. For example, the result $[P_1 P_2 P_3 P_4 P_5 P_6 P_7] (P_1 \geq P_i, i = 2:7)$ is that the scene includes smoke with a little probability accompanied with other objects.

The loss function is defined to drive a backpropagation function to update the network parameters. It has two parts, one being the difference between the probability of the right classification $s_j$ and the maximum probability of all wrong classification $s_{yi}$, and the other one being a regularization term to prevent over-fitting during training

$$L = \frac{1}{N} \sum_{i=1}^{N} \sum_{j \neq y_i} \max(0, s_j - s_{y_i} + 1) + 0.21 \times ||w||^2$$

(1)

where $N$ is the sample number, $j$ is the wrong label, $y_i$ is the ground truth label, $s_j$ is the maximum probability of all wrong classification $s_{yi}$, and the weight is the weights of the last fully-connected layer, $||w||^2$ is a regularization term, and the coefficient 0.21 is an empiric value.

The sample data set

The public data sets, such as UCF-101 and Sports-1M, lack smoke samples. So, 108 videos is selected from YouTube and four videos from VisiFire as a smoke data set. Each video is randomly cut into video clips at different parts with size of $3 \times 112 \times 112$ and 16 consecutive frames and be assigned with the corresponding labels. The positive subset includes samples from various scenes, such as indoor, outdoor, urban, forest, grassland, open areas, and so on, and from various perspectives. The negative subset includes samples without smoke, and the samples with some confusing objects are specifically appended, such as cloud, fire, waving flags, waving leaves, fog and haze, vehicle lighting at night, climbing vehicles, and others. Both the positive subset and the negative subset include samples of distant view and close view. The samples of the data set is shown in Table 1 and six typical clips are shown in Figure 3.

The training process

The hardware platform is configured with Nvidia GTX1080 and Intel i7-6700k. The relevant parameters of the training process are listed in Table 2.

The dynamic characteristics revealed by deep neural network

Our network costs about 18 min 16 s every epoch, and achieves the best accuracy 90.10%. The loss values and the test accuracies are evolving along the training epochs, as shown in Figure 4.

To demonstrate the dynamic characteristics revealed by deep neural network, four typical channels of the 1st, the 2nd, the 3rd, and the 4th convolution layers are shown in Figure 5, corresponding to the 3rd column, the 4th column, the 5th column, and the 6th column. Only the strongest responses of every convolutional layer are listed. For example, the 8th channel is the strongest response of the 1st convolution layer, which is shown in the 3rd column. The 3rd column, the response of the 1st convolutional layer look like optical flow, and the ones of the 4th convolution layer look like motion vectors of smoke. The higher the level of the convolution layer, the more abstract the meaning, and the more difficult to understand its physical meaning. So, only some responses of the first four convolutional layer are shown. It is clear that the network has learned to find the changing areas and provides dynamic characteristics, which means that the network could effectively extract temporal and spatial feature information. Of course, the sizes of the 3rd column, the 4th column, the 5th column, and the 6th column should be a half of the preceding one because the sizes of the pooling kernels are $2 \times 2 \times 2$, but be enlarged as the same size in Figure 5.

Experiments

Two experiments are conducted to validate the network. The first one is to measure how good the network judges whether there is smoke by classifying accuracy. Two tests are executed to demonstrate the accuracy of the proposed network, one based on our data set and the other one based
on four external data sets. The second experiment is to measure how sensitive the network could be, and a new metric is proposed. The proposed network finds smoke averagely at the 23rd frame, which is more sensitive than other algorithms.

The classifying accuracy

The proposed network recognizes smoke, fire, fog and haze, cloud, vehicle, flag, and others on our data set with accuracy 78.00%, 84.06%, 99.43%, 100.00%, 95.00%, 95.00%, and 91.00%, respectively. Besides, the average classifying accuracy is 87.31%. The classifying accuracy is shown in Figure 6.

To compare the proposed method with other methods, a contrasting experiment is executed. To be fair, our data set for training and preceding testing is abandoned and four external famous data sets, which are visor, fesb, yfn, and VisiFire, are selected to conduct the experiment. As shown in Table 3, our smoke detecting rates achieve 98.07%, 94.30%, 100%, and 92.61% on visor, fesb, yfn, and VisiFire, respectively, and the corresponding ones are 85.20% contributed by:

\[\text{Table 2. The hyper parameters.}\]

| Image size | Crop size | Batch size | Initial learning rate | Learning policy | Step size | Gamma |
|------------|-----------|------------|-----------------------|----------------|-----------|-------|
| 240 × 320  | 112 × 112 | 25         | 0.003                 | Step           | 2400      | 0.5   |

\[\text{Table 3. The contrasting accuracies between the proposed method and the others.}\]

| Data sets | Smoke video numbers (total time) | Detection accuracy | Nonsmoke video number (total time) | False alarm rate | *The scores of our method.

| visor16   | 14 (8 min) | 85.20%33  | 98.07%* | 31(6 min) | 1.75%33  | 2.70%* |
| fesb17    | 9 (58 min) | 84.80%20  | 94.30%* | 3(3 min)  | 2.08%36  | 0.00%* |
| yfn18     | 3 (3 min)  | 75.00%36  | 100.00%* | 3(3 min)  | 5.21%32  | 1.30%* |
| VisiFire19 | 20 (35 min)| 74.05%32  | 92.61%* | 3(2 min)  |          |        |

*: no data are provided for these indexes on this data set.

The classifying accuracy

The proposed network recognizes smoke, fire, fog and haze, cloud, vehicle, flag, and others on our data set with accuracy 78.00%, 84.06%, 99.43%, 100.00%, 95.00%, 95.00%, and 91.00%, respectively. Besides, the average classifying accuracy is 87.31%. The classifying accuracy is shown in Figure 6.

To compare the proposed method with other methods, a contrasting experiment is executed. To be fair, our data set for training and preceding testing is abandoned and four external famous data sets, which are visor, fesb, yfn, and VisiFire, are selected to conduct the experiment. As shown in Table 3, our smoke detecting rates achieve 98.07%, 94.30%, 100%, and 92.61% on visor, fesb, yfn, and VisiFire, respectively, and the corresponding ones are 85.20% contributed by
Sebastien et al.\textsuperscript{33} on visor, 84.80\% contributed by Lee et al.\textsuperscript{20} on fesb, 75.00\% contributed by Li et al.\textsuperscript{36} on yfn, and 74.05\% contributed by Zhang et al.\textsuperscript{32} on VisiFire. Our false alarm rates achieve 2.70\% on visor, 0.00\% on yfn, and 1.30\% on VisiFire, and the corresponding ones are 1.75\%,\textsuperscript{33} 2.08\%,\textsuperscript{36} and 5.21\%,\textsuperscript{32} respectively. It is obvious that the proposed method is much better than 2D CNN network\textsuperscript{32} and traditional method\textsuperscript{36} with the best scores in all metrics.

Due to the coexistence of smoke and fire, the images whose predicted label are fire with Top-1 hit, usually include smoke also. The reason might be that fire is more obvious than smoke when both are visible. So, this brings down the classifying accuracy of smoke. So, the recall of smoke is provided to show how good the network is in recognizing smoke. The mean recall is 82.40\%, which is shown in Figure 7.

**The sensitiveness**

The sensitiveness is the serial number difference between the frame which are recognized as smoke and the frame...
in which smoke has occurred. The sensitiveness is very important to alarm early fire as soon as possible. The test is executed on the four external data sets. Our network finds smoke at the 25th, 16th, and 32nd frame. The mean sensitiveness of the proposed method is 23, which is based on the four external data sets. The corresponding ones are 150 on visor, 88 on yfn, 99 on VisiFire, and 91 on VisiFire. The sensitiveness is shown in Table 4.

The mean sensitiveness is provided to contrast these methods. The mean sensitiveness of the proposed network are 23, 88, 99, and 150, respectively. This testifies that the proposed network is more sensitive than the AdaBoost method and the compound method, which is shown in Figure 8. This is very meaningful for fire alarm.

### The test speed

The test speed, a round of a forward pass, by the proposed network is 51.8 fps in our hardware platform, an Intel i7-6700k CPU with a Nvidia GTX1080 GPU. It is enough for surveillance applications.

### The test samples

Some typical scenes, such as indoor, outdoor, urban areas, forest, grassland, and open areas, are tested in this experiment. The samples with various perspectives are included, such as from overlook, looking up, aerial view, and looking horizontally. Thin and heavy smoke, even with various colors are included. Twenty five samples are shown in Figure 9. The 1st and the 2nd are outdoor and indoor samples, the 3rd is the smoke in a forest, and the 5th happens in a factory. The smoke of the former four samples are white, and the one of the 5th sample is black. In fact, smoke with different appearances, shown in samples from the 1st to the 10th, and the 11th to the 15th, are recognized rightly. The 12th, the 13th, and the 14th samples are identified as fire with a small probability of smoke. In fact, there is some tiny smoke. The 15th sample is classified as smoke with a small probability of fire. The 18th and the 20th samples are recognized as fog and haze and cloud, respectively with high confidence, though they look like smoke and some human cannot classify rightly. The objects with similar appearance are identified clearly. The 21st and 22nd samples are moving vehicle lights from far to close. The 23rd sample is a waving flag. The objects with similar motions, for examples, waving flags, climbing vehicles on mountain, and moving light, are identified clearly.
This experiment illustrates that the method is generally used without any hypothesis and can accurately distinguish smoke with similar appearance objects and similar motion objects.

**Conclusion**

The traditional methods based on static features cannot well distinguish smoke from many other similar objects. The traditional methods based on dynamic features could not extract enough temporal features from videos. The features of these methods are usually hand crafted, which are not enough to capture the smoke characteristics. So, multiple model features, including static features such as chrominance, texture, shape, and dynamic features such as rising up, flickering, have to be integrated to improve accuracy. On the one hand, some hypotheses, such as color, rising vertically to sky, and so on, underlie these methods,

| Data sets | Sensitivity |
|-----------|-------------|
| Viso16    | 150³⁰      | 25*         |
| Fesb17    | /           | 16*         |
| Yfn18     | 88³⁹        | 16*         |
| VisiFire19 | 99³⁹, 91³⁹² | 32³²        |

*: no data are provided for these indexes on this data set.
*The sensitiveness of our method based on the extern data sets, and the mean sensitiveness of the proposed method is 23.

![Figure 8](image_url). The contrasting sensitiveness between the proposed method and the others.

![Table 4](image_url). The contrasting sensitiveness between the proposed method and the others.

| Data sets | Sensitiveness |
|-----------|---------------|
| Viso16    | 150³⁰         | 25*          |
| Fesb17    | /             | 16*          |
| Yfn18     | 88³⁹         | 16*          |
| VisiFire19 | 99³⁹, 91³⁹² | 32³²         |

*The sensitiveness of our method based on the extern data sets, and the mean sensitiveness of the proposed method is 23.

![Figure 9](image_url). The testing samples. The title of each subplot is the predicted label with the corresponding probability. For fire, which usually coexists with smoke, two labels and corresponding two probabilities are listed. For these case, the maximum one is usually corresponding to fire and the second one is corresponding to smoke.
but restrict their application fields. On the other hand, such methods are not sensitive to thin smoke.

Consequently, the proposed method has been validated that dynamic characteristics is discriminating enough to classify smoke, and the discriminating capability is even better than the one of the compound methods which integrate static and dynamic features. This also proves that neural network is outstanding to extract compound features in contrast to the hand crafted methods.

Because the fire videos which happened from the embryonic stages to the violent stages are rare, all samples on hand are collected into this set. If exact samples are provided, the network would be more useful.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Natural Science Foundation of Zhejiang Province under Grant LQ19F020005, the Project of Science and Technology Plans of Wenzhou City under Grant 2018ZG021, the Natural Science Foundation of China under Grants 617622442, the Joint Funds of 5th Round of Health and Education Research Program of Fujian Province (No. 2019-WJ-41), the Joint Funds of Scientific and Technological Innovation Program of Fujian Province (No. 2017Y9059) and Natural Science Foundation of Fujian Province under Grants 2018J01574.

ORCID iDs

Dahan Wang https://orcid.org/0000-0002-5901-0778
Li Zhao https://orcid.org/0000-0001-5787-2705

References

1. World Fire Statistics. http://www.ctif.org/ctif/world-fire-statistics.
2. Wang D, Chen T, Yin Y, et al. Smoke detection for early fire-alarming system based on video processing. J Digit Inf Manag 2008; 6(2): 196–200.
3. Chen T, Yin Y, Huang S, et al. The smoke detection for early fire-alarming system based on video processing. In: International conference on intelligent information hiding and multimedia signal processing, Pasadena, California, USA, 18–20 December 2006, pp. 427–430.
4. Ye W, Zhao J, Wang S, et al. Dynamic texture based smoke detection using surfacelet transform and HMT model. Fire Safety J 2015; 73: 91–101.
5. Yu C, Zhang Y, Fang J, et al. Texture analysis of smoke for real-time fire detection. In: 2nd international workshop on computer science and engineering, Qingdao, China, 28–30 October 2009, vol. 2, pp. 511–515.
6. Tian H, Li W, Ogunbona P, et al. Smoke detection in videos using non-redundant local binary pattern-based features. In: 3rd IEEE international workshop on multimedia signal processing, Hangzhou, China, 17–19 October 2011.
7. Kim D. Smoke detection using boundary growing and moments. In: International conference on convergence and hybrid information technology, Daejeon, Korea, 27–29 August 2009, vol. 321, pp. 430–433.
8. Yuan F, Fang Z, Wu S, et al. Real-time image smoke detection using staircase-searching-based dual threshold AdaBoost and dynamic analysis. Inst Eng Technol 2015; 9(10): 849–856.
9. Yuan F. Video-based smoke detection with histogram sequence of LBP and LBPV pyramids. Fire Safety 2011; 46(3): 132–139.
10. Anton M and Olga K. Real-time smoke detection in video sequences: combined approach. In: 9th International conference on pattern recognition and machine intelligence, Kolkata, India, 10–14 December 2013, vol. 8251, pp. 445–450.
11. Yu C, Jun F, Wang J, et al. Video fire smoke detection using motion and color features. Fire Technol 2010; 46(3): 651–663.
12. Rinsurongkawong S, Ekpanyapong M, and Dailey MN. Fire detection for early fire alarm based on optical flow video processing. In: 9th International conference on electrical engineering/electronics, computer, telecommunications and information technology, Phetchaburi, Thailand, 16–18 May 2012.
13. Lu Z, Xie W, Pei J, et al. Dynamic texture recognition by spatio-temporal multiresolution histograms. In: IEEE workshop on motion and video computing, Breckenridge, CO, USA, 5–7 January 2005, vol. 1, pp. 241–246.
14. Calderara S, Piccinini P, and Cucchiara R. Vision based smoke detection system using image energy and color information. Mach Vision Appl 2011; 22(4): 705–719.
15. Luo S, Yan C, Wu K, et al. Smoke detection based on condensed image. Fire Safety J 2015; 75: 23–35.
16. visor. http://imagelab.unimore.it/visor/video_categories.asp.
17. fesb [EB/OL]. http://wildfire.fesb.hr/index.php?option=com_content&view=article&id=65%3Adaptive-estimation-of-detection-parameters-video-database&Itemid=72.
18. yfn. http://staff.ustc.edu.cn/~yfn/vsd.html.
19. VisiFire. http://signal.ee.bilkent.edu.tr/VisiFire/Demo.
20. Lee C, Lin C, and Hong C. Spatio-temporal analysis in smoke detection. In: IEEE international conference on signal and image processing applications (ICSIPA), Malaysia, 18–19 November 2009, pp. 80–83.
21. Sheng L, Jing-Hua X, and Shu-You Z. Decompose image into meaningful regions based on contour detector and watershed algorithm. J Intell Fuzzy Syst 2017; (32): 4259–4271.
22. Xiaqin Z, Di W, Zhengyuan Z, et al. Robust low-rank tensor recovery with rectification and alignment. IEEE T Pattern Anal 2020. DOI: 10.1109/TPAMI.2019.2929043.
23. Xiaqin Z, Qianqian L, Di W, et al. Self-taught semisupervised dictionary learning with nonnegative constraint. IEEE T Ind Inform 2020; 16(1): 532–543.
24. Xiaoqin Z, Jingjing Z, Di W, et al. Exemplar-based denoising: a unified low-rank recovery framework. *IEEE T Circ Syst Vid* 2020. DOI: 10.1109/TCSVT.2019.2927603.

25. Li X, Song W, Lian L, et al. Forest fire smoke detection using back-propagation neural network based on MODIS data. *Remote Sens* 2015; 7(4): 4473–4498.

26. Yuan F. A fast accumulative motion orientation model based on integral image for video smoke detection. *Pattern Recognit Lett* 2008; 29(7): 925–932.

27. Xiaoqin Z, Weiming H, Nianhua X, et al. A robust tracking system for low frame rate video. *Int J Comput Vis* 2015; 115(3): 279–304.

28. Xiaoqin Z, Xingchu S, and Weiming H. Visual tracking via dynamic tensor analysis with mean update. *Neurocomputing* 2011; 74(17): 3277–3285.

29. Xiaoqin Z, Wei L, Xiuzi Y, et al. Robust hand tracking via novel multi-cue integration. *Neurocomputing* 2015; 157: 296–305.

30. Calderara S, Piccinini P, and Cucchiara R. Smoke detection in video surveillance: a MoG model in the wavelet domain. Computer vision systems. In: *6th International conference on computer vision systems*, Santorini, Greece, 12–15 May 2008, pp. 119–128.

31. Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation. In: *27th IEEE conference on computer vision and pattern recognition*, Portland, Oregon, USA, 23–28 June 2013, pp. 580–587.

32. Zhang Q, Xu J, and Xu L. Deep convolutional neural networks for forest fire detection. In: *International forum on management, education and information technology application*, Guangzhou, China, 30–31 January 2016.

33. Sebastien F, Rabeb K, Moez B, et al. Convolutional neural network for video fire and smoke detection. In: *42nd Annual conference of the IEEE industrial electronics society*, Florence, Italy, 23–26 October 2016, pp. 877–882.

34. Sheng L, Xiaoqin Z, Muchou W, et al. A slight smoke perceptual network. *IEEE Access* 2019; 7(4): 42889–42896.

35. Tran D, Bourdev L, Fergus R, et al. Learning spatiotemporal features with 3D convolutional networks. In: *15th IEEE international conference on computer vision*, Santiago, Chile, 13–16 December 2015, pp. 4489–4497.

36. Li S, Wang B, Dong R, et al. A novel smoke detection algorithm based on fast self-tuning background subtraction. In: *Chinese control and decision conference*, Yinchuan, China, 28–30 May 2016, pp. 3539–3543.