An Effective Method for Forest Fire Smoke Detection

Luxing Qin ¹,², Xuehui Wu ¹,², Yichao Cao¹,² and Xiaobo Lu ¹,²,a

¹School of Automation, Southeast University, Nanjing 210096, China
²Key Laboratory of Measurement and Control of Complex Systems of Engineering, Ministry of Education, Southeast University, Nanjing 210096, China
⁰Corresponding author: xblu2013@126.com

Abstract. This paper focuses on smoke detection in forest environments. In this paper, dark channel prior and OTSU based multi-threshold are used to find the disturbances such as sky and haze. These regions are blocked to reduce false alarm rate. A motion detection method based on frame difference is adapted to find the motion objects. Color moments, HOG and LBP are chosen as features of smoke and SVM is used as the classification. To reduce more false alarms, the motion regions are classified for several consecutive frames. They won’t be regarded as smoke regions unless M frames of them are classified as smoke ones. Experiment results showed that the proposed method can detect smoke in video effectively and work real-timely.

1. Introduction

Forest fire happens every year around the world, which causes serious damage to ecological environment, animals and humans. Traditional smoke detecting systems are made up of sensors. These systems either have high false alarms rate or cost too much as plenty of sensors are required. Besides, sensors have many other limitations, such as communication and control. In order to detect the forest fire as soon as possible, many different forest fire detection algorithms were proposed.

Forest fire smoke detection algorithm can be divided into several categories. Wavelet transform and other signal processing algorithms to a single image is an approach [1]. These methods don’t need training before detection, but perform badly on change of lightness and other disturbances. Methods based on feature extraction are widely adapted. Usually, a background model of the scene should be built to find the motion area and avoid unnecessary calculation. The features are then put into a classifier such as support vector machine (SVM). There are lots of features which have been proved to be useful to describe smoke. Static features such as color feature, LBP, HOG and compactness perform well. Dynamic features include movement directions, area change and so on. Cai et al. [2] combined static features and dynamic features and got better results. In recent years, deep learning makes a great success in computer vision. Convolutional neural network (CNN) gets much higher accuracy rate on images classification and recurrent neural network (RNN) can reinforce the knowledge for a set of sequential frames. Tao et al. [3] showed that CNN works better on smoke detection. A.F. et al. [4] applied CNN and RNN to video sequences. Though deep learning performs better, the speed of training and working is not satisfactory. In practice, it’s difficult to work real-time without a high performance GPU.

The real forest environment can be complicated. The features of smoke are changeable and disturbances such as haze, sky and change of lightness can also be captured by the motion detection algorithm. All of them increase the false positive rate significantly. In this paper, Dark channel prior and multi-threshold segmentation are used to remove the disturbances such as white house, haze and sky before detection. After motion object segmentation, the regions which are moving but not blocked are
activated. Finally, features of motion cells are extracted and support vector machine is used to classify the motion cells.

2. Method
The proposed method consists of three major steps:

(1) Extracting mask regions with dark channel prior and multi-threshold segmentation.
(2) Motion detection.
(3) Extracting features and classifying using SVM.

2.1. Dark channel prior
The dark channel prior theory [5] is a method based on statistics on a huge amount of outdoor pictures. It’s found that in most non-sky areas, there is always at least one of the three color channels with a low value. This channel with the minimum value is called dark channel.

For an input image $J$, the dark channel can be expressed as:

$$J_{\text{dark}}(x) = \min \{ \min(J(x)) \}$$

Where $J$ is a color channel of input image $J$, and $\Omega(x)$ is a local patch centered at $x$. In practical calculation, we usually find the minimum value of the three channels of every pixel and store them in a grayscale image whose size is the same as the original picture. Then a filter is adapted to smooth the grayscale image. The radius is a changeable parameter. Usually the window size is $2 \times \text{Radius} + 1$.

According to the dark channel theory, $J_{\text{dark}}(x)$ should be small for close object but relatively large for sky and haze. As is shown in Figure 1, the dark channel values of the image pixels at different distances differ greatly.

In computer vision, the graph formation is used widely:

$$I(x) = J(x)t(x) + A(1-t(x))$$

Where $I(x)$ is the original image that is captured by the camera, $J(x)$ is the result picture, $A$ is the global atmosphere light component and $t(x)$ is the transmission. The aim is to calculate $J(x)$ with $I(x)$ already known.

In practice, we select the top 0.1% pixels with high brightness in the dark channel which is the most opaque. Among these pixels, the highest intensity pixel mean in the input image is selected as $A$. Assuming that $t(x)$ is a constant value in each window, the estimated value of transmission can be derived:

$$t(x) = 1 - \min \{ \frac{I(x)}{A_{\text{dark}}} \}$$

In actual, the atmosphere is not absolutely free of any particles. For a camera used in forest fire detection, it’s required to be installed on a high place to cover a large area, even several kilometers. The existence of haze is the basis of human perception of the depth of the scene, while it greatly increases the difficulty of smoke detection. It’s hard to distinguish fire smoke and haze at a far distance.

2.2. Multi-threshold segmentation
After the calculation of the air transmission of each pixel of the image, we get a grayscale image. In order to find the area that we want to mask, a threshold segmentation method should be used.

The OTSU algorithm [6] is an adaptive method to get the proper threshold and segment the image. Suppose the grayscale space of an image is represented by $L$. For each pixel, $i$ is the grayscale value in the grayscale space. The number of each grayscale value in the grayscale histogram of the image is represented by $n_i$, and the total number of pixel of the image is denoted by $N$, then $N = n_1 + n_2 + \ldots + n_{L-1}$.

If the pixels of the image are divided into 2 classes, the class probability $\theta_{\text{b1}}(t)$ is computed from the histogram:
\[ \omega_0(t) = \sum_{i=0}^{L-1} p(i) \]  
\[ \omega_1(t) = \sum_{i=L}^{L-1} p(i) \]  

Where the probability of each grayscale value is represented by \( p(i) \) and \( t \) is the single threshold of the segmentation.

The total mean value of the image is \( \mu_T \), and the mean values of each of the 2 classes in the image are \( \mu_0 \) and \( \mu_1 \) respectively:

\[ \mu_0(t) = \frac{\sum_{i=0}^{L-1} i p(i)}{\omega_0(t)} \]  
\[ \mu_1(t) = \frac{\sum_{i=L}^{L-1} i p(i)}{\omega_1(t)} \]  
\[ \mu_T(t) = \frac{\sum_{i=0}^{L-1} i p(i)}{\omega_T(t)} \]  

The inter-class variance \( \sigma_b^2 \) and the intra-class variance \( \sigma_w^2 \) of the two classes can be described:

\[ \sigma_b^2(t) = \sigma^2 - \sigma_w^2(t) \]  
\[ \sigma_w^2 = \omega_0(t)\sigma_b^2(t) + \omega_1(t)\sigma_b^2(t) \]  

When using the OTSU method to find the threshold, the optimal threshold can be calculated by maximizing the inter-class variance or minimizing the intra-class variance iteratively:

\[ t_{opt} = \arg \max_{\sigma_b^2(t)} \{ \sigma_b^2(t) \} = \arg \max_{\sigma_w^2(t)} \{ \sigma_w^2(t) \} \]  

As it’s shown in Figure 2, the forest scene usually can be made up of three or more regions, so it’s necessary to extend to multi-threshold method [7]. The grayscale space is still \([0, L-1]\) with \( N = n_1 + n_2 + \cdots + n_{L-1} \). For multi-threshold segmentation, the image should be segmented into \( n \) classes, so there are \( n-1 \) thresholds \( t_0, t_1, \ldots, t_{n-1} \). The total mean value of the image is still \( \mu_T \), and the mean values of the regions corresponding to \( n \) classes are \( u_0, u_1, \ldots, u_{n-1} \). Similar to the single threshold method, the optimal threshold for multi-threshold algorithm should follow the principle of maximizing inter-class variance or minimizing intra-class variance.

Figure 1. (a) Source image (b) Dark channel image (c) Multi-threshold segmentation results (d) Mask regions of the images

### 2.3. Motion detection
Motion detection is an important part of video smoke detection. A good motion detection algorithm can locate the moving objects accurately and avoid unnecessary calculation. Usually, a background model should be built and updated every frame. Then the foreground can be separated from the difference between current frame and the background model.

Gaussian mixture model (GMM) [8] and ViBe [9] have been widely used as motion detection methods before smoke detection. GMM needs several frames to build the background model and requests various calculations. Compared to GMM, ViBe needs less modeling process and costs less time. Recently, RPCA has been used in background modeling and performs better [10]. However, this approach requests a large amount of memory space and complicated calculations, which is hard to apply to 720P and 1080P videos.

In forest environments, the field of a camera vision is wide. Therefore, the motion detection method should be highly sensitive to catch every possible motion. Besides, it should be fast enough to satisfy the need for real time.

In this paper, we use an improved method based on frame difference, which is showed in Figure 2:

![Proposed background modeling flow chart](image)

**Figure 2. Proposed background modeling flow chart**

Experiments were carried out to test the efficiency of the proposed method, GMM and ViBe. Figure 3 shows that our method locates the motion regions effectively. Table 1 shows that our method is faster than both of the others.
2.4. Feature extraction and smoke detection

As a widely used feature, color feature doesn’t need much calculations. For smoke detection, color is a very useful feature. Color moments can measure color distribution of an image. It is usually that the first three color moments are used as features. Color histogram counts the color information of the entire image and quantifies the proportion of different colors in the entire image. In this paper, color moments in HSV color space is chosen to describe the color characteristic of the smoke.

Local binary patterns (LBP) is a type of visual descriptor used for classification in computer vision. There are many improvements of original LBP methods. In this paper, original circle LBP is adapted as a characteristic of the smoke.

The histogram of oriented gradients (HOG) is also a feature descriptor used in computer vision and image processing for the purpose of object detection. As a statistical feature, it counts occurrences of gradient orientation in localized portions of an image.

As smoke moves slowly and won’t disappear in a short time, the moving regions won’t be regarded as a potential smoke region if it can’t be detected by the motion detection in consecutive N frames. The mask regions calculated from dark channel image and multi-threshold segmentation will not be detected, for it’s useless to detect these regions. Color moments, LBP and HOG are adapted as features for image classification. In practice, we classify the motion cells for consecutive 50 frames. If M of them are regarded as smoke, this region will be classified as smoke regions.

3. Experiment results
Figure 4. The result images of different videos

The judgment of the performance of video smoke detecting algorithm is not easy to define. Smoke fire detection is not the same as car or face. Actually, a cloud of smoke can be divided into several cells, and a correct detection of one of the cells is enough. From this point, we can judge the performance of the algorithm by whether detecting smoke from the video. However, it will be more reliable if improving the recognition rate of every cell.

The experiment was carried out on a PC with an Intel Core i3, 3.40GHz processor using the method proposed in this paper. Figure 3 shows images of smoke detection. Red box regions are smoke cells and regions circled by red contours are mask regions.

Twenty 720P videos which are captured by a practical camera used for forest fire detection have been tested in this experiment with a trained SVM classifier. A half of them are with smoke while others are not but can easily cause error detection. The classifier is trained with about 5000 picture samples in total. Color moments, LBP and HOG are adapted as the features to train the classifier and detect smoke. To prove the efficiency of the proposed method, Cai’s method [2] is tested on the same videos, too. The experiment results of non-smoky videos and smoky videos with mask or not and Cai’s method are showed in Table 2 and Table 3 respectively.

In forest smoke detection, a single false alarm region will cause a false alarm video. As is showed in the table, the false positive rate of regions and videos decreases significantly, while the true positive rate only decreases a little. Actually, all of the smoke videos do detect smoke. Besides, the speed of the method can reach about 18fps. As for Cai’s method, some of the smoke regions are detected but the false alarm rate is higher. Besides, for some videos whose smoke regions are small and far away from the camera, smoke is not detected. As is showed in Table 4, in order to use dynamic features, multi-object tracking costs too much calculation, which makes it much slower than static ones.

| Method | Motion regions | Detected regions | False alarm rate of regions | Smoky videos |
|--------|----------------|------------------|-----------------------------|--------------|

Table 2. Comparison of non-smoky videos
Table 3. Comparison of smoky videos

| Method                  | Motion regions | Actual regions | Detected regions | Detection rate | Smoky videos |
|-------------------------|----------------|----------------|------------------|----------------|--------------|
| Classification only     | 1736           | 88             | 70               | 79.5%          | 10           |
| With mask regions       | 1711           | 88             | 61               | 69.3%          | 10           |
| Cai’s method            | 2118           | 121            | 84               | 69.4%          | 6            |

Table 4. Comparison of speed

| Method                  | Size of video | Time per frame |
|-------------------------|---------------|----------------|
| Classification only     | 720P          | 57.49ms        |
| With mask regions       | 720P          | 57.50ms        |
| Cai’s method            | 720P          | 257.69ms       |

4. Conclusion
In this paper, a method for forest smoke detection is proposed. In terms of the specific forest scene, Dark channel prior and multi-threshold segmentation are used to mask the region that may easily cause error detection. The results show that the false positive rate decreases significantly while the true positive rate almost keep the same. The method can be real-time on a PC with no GPU.

Acknowledgment
This work was supported by the National Natural Science Foundation of China (No.61871123), Key Research and Development Program in Jiangsu Province (No.BE2016739) and a Project Funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions.

References
[1] G.R. et al.. Wavelet-Based Smoke Detection in Outdoor Video Sequences [C]. MWSCAS, pp.383-387(2010)
[2] M. Cai, et al.. Intelligent Video Analysis-based Forest Fires Smoke Detection Algorithms [C]. ICNC-FSKD. pp.1504-1508(2016)
[3] C. Tao, et al.. Smoke detection based on deep convolutional neural networks [C]. ICIICII, pp.150–15(2016)
[4] A.F., L.K., K. Jo. Smoke Detection on Video Sequences Using Convolutional and Recurrent Neural Networks. University of Ulsan, Ulsan, Republic of Korea(2017)
[5] He K, Sun J, Tang X. Single Image Haze Removal Using Dark Channel Prior [J]. IEEE Trans Pattern Anal Mach Intell, 33(12):2341-2353(2016)
[6] Otsu N. A Threshold Selection Method from Gray-Level Histograms [J]. IEEE Trans.syst.man.& Cybern, 9(1):62-66(2007)
[7] Y. Cao et al.. Single Image Depth Level Estimation Using Dark Channel Prior [C]. (to be published)
[8] T. B. et al.. Background Modeling using Mixture of Gaussians for Foreground Detection – A Survey [J]. Recent Patents on Computer Science, pp.219-237(2008)
[9] Olivier Barnich, Marc Van Droogenbroeck. ViBe-a universal background subtraction algorithm for video sequences [J]. IEEE Trans Image Processing, 20(6):1709-1724(2011)
[10] M. Wu, Y. Sun et al.. Multi-component group sparse RPCA model for motion object detection under complex dynamic background [J]. Neurocomputing, pp.120-131(2018)