

Video-based smoke detection by using motion, color, and texture features

Zheng Zhao^a, Guihua Cui^{*a}, Donghui Li^a, Qifan Chen^a

^aCollege of Electrical and Electronic Engineering, Wenzhou University, Wenzhou, China;

* Corresponding author: guihua.cui@foxmail.com

ABSTRACT

Video smoke detection benefits life safety and environment protection, its early warning is of great importance. In response of many disadvantages of traditional smoke detectors, a method of video smoke detection based on dynamic, color and texture features is proposed. Firstly, the motion area is extracted through improved Vibe algorithm. Then, the suspected area corresponding to smoke is identified in CIELAB color space and segmented using a color filtering method. Finally, uniform local binary mode and gray level cooccurrence matrix are extracted from the image within suspected area, and used to form the input vector of machine learning classifier for recognizing smoke. The classifier is tested with 400 images, and the results show that the detection system based on random forest algorithm has better performance, the selected smoke features have high recognition accuracy.

Keywords: smoke detection; multi-features fusion; machine learning;

1. INTRODUCTION

The fire accident usually causes economical and ecological damage as well as endangering people's lives. Automatic visual object detection has been fully used for alarming fire disaster as soon as possible. Generally, most of the objects will generate smoke before it catches fire and this motivates that the smoke detection is employed to provide alarm of fire accident. However, the traditional smoke detection system is usually based on smoke sensors and difficult to play a complete role in outdoor environment, and it can't provide more essential information about the process of fire accident. With the development of image processing technology, smoke detection based on vision technology has attracted researchers' attention.

Several algorithms have been reported in the literature regarding video smoke detection. Emmy Prema et al. [1] converted the RGB color space to YUV space to determine the suspected smoke area. Ding et al. [2] proposed a method according to spectral features used in large and complex scenes. However, the accuracy of smoke detection is not very well for lack of features. Wu et al. [3] developed a new fire smoke model based on Gabor wavelet, and used the smoke change energy model and direction angle distribution model to analyze the dynamic properties of texture changes. Wang et al. [4] proposed a smoke detection method based on the swing and diffusion judgment model by analyzing the relevant characteristics of the gray level co-occurrence matrix (GLCM). Li et al. [5] used the MSER tracking algorithm to determine the smoke candidate regions in the video, and use static and dynamic features to identify smoke. James [6] used the optical flow method to extract smoke candidate regions and neural networks for classification and recognition. Liu et al. [7] divided the smoke into overlapping blocks, extracted the wavelet texture features and HSV color features of each image, obtained the latent semantic analysis. Some classical algorithms have been used for smoke detection, for example, Yuan used local binary pattern [8], Islam et al. used Gaussian Mixture Model [9].

Liu et al. [10] separated suspected smoke area in YUV color space, used the Local Binary Patterns (LBP) and discrete wavelet transform to extract smoke features, the AdaBoost classifier is used for image classification. But when the volume fraction is small, the system cannot quickly identify smoke. Wang [11] proposed an image enhancement method based on fuzzy logic to alleviate the interference of low and uneven lighting in coal mine, and used support vector machine (SVM) to classify smoke by features of perimeter and area ratio, area randomness and drift characteristics. And the convolutional neural network CNN has been widely used in video smoke detection [12-14].

In this paper, multi-feature fusion method is used for smoke detection. Based on the motion area extracted by the improved Vibe algorithm, the suspected smoke area is segmented in the CIELAB color space by color filtering rules, and

then the texture features of smoke are mainly discussed. The uniform local binary pattern codes (ULBP), and GLCM are used. Machine learning classifiers SVM and random forest are used for model training to distinguish smoke and non-smoke. The optimal algorithm is selected from the analysis results, and the test is carried out through the simulated fire smoke monitoring video to verify the accuracy of the detection method.

In the following sections, the method including motion objects detection, color detection and texture detection in detail is described in Section II. Section III gives experimental results while conclusions are drawn in Section IV.

2. FEATURES EXTRACTION

2.1 Motion objects detection

Aiming at the relationship between the upward movement characteristics of smoke and fixed surveillance cameras, the moving target detection method under static background is used to extract the moving area. The Vibe algorithm introduces the randomization method into the background model for the first time, which is fast, has a small calculation and a certain robustness to noise. The detection area is more complete, and it can quickly achieve background extraction and moving target detection [15].

However, the original Vibe algorithm has the following problems when extracting the motion area: the algorithm is updated too fast, and sometimes the initial smoke movement is slow, which will absorb the smoke area as the background, resulting in no smoke detected; if the algorithm is updated too slowly, it will leave shadow; misclassification of pixels may be caused due to parameter setting issues. Aiming at the problems existing in the original Vibe algorithm, the improved algorithm is proposed.

Firstly, the sampling range is extended from 8 neighborhoods to 12 neighborhoods. Enlarging the sampling range can avoid a pixel from being sampled multiple times and improve the quality of the initial model.

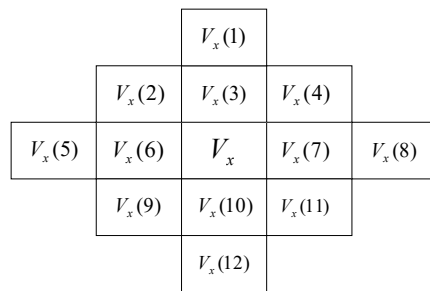


Figure 1. 12 neighborhoods model

The original Vibe algorithm is performed by calculating the current pixel and the sample set the Euclidean distance D between each pixel for foreground and background segmentation. As is shown in Figure2, set $v(x)$ (current point) as the center, R as radius, $v_1, v_2, v_3, \dots, v_n$ is grayscale values in sample set. Count each distance of sample set with $v(x)$, if the count of distance which is less than threshold T , then take $v(x)$ as foreground, otherwise, take it as background. In original Vibe algorithm, $R = 20, T = 2, n = 20$.

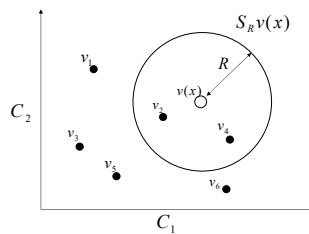


Figure 2. Pixel classification in Euclidean distance

On the basis of the original Vibe algorithm, the OTSU is introduced to calculate the optimal segmentation threshold of the current frame, and the foreground points are double-discriminated [16]. OTSU is defined as:

$$\sigma^2 = \omega_0(\mu_0 - \mu)^2 + \omega_1(\mu_1 - \mu)^2 = \omega_0\omega_1(\mu_0 - \mu_1)^2 \quad (1)$$

Where the subscript 0, 1 and without subscript indicate foreground, background and entire image, respectively. ω indicate the proportion of the number of pixels, μ are the average grayscale value. Set the segmentation threshold be t , and obtained when the σ^2 reaches the maximum.

In the detection results of the traditional vibe algorithm, N background pixels are randomly selected, the average gray value V_N of these background points is calculated, and the secondary discrimination is performed:

$$f(x) = \begin{cases} \text{foreground} & v(x) > t \\ \text{background} & v(x) \leq t \end{cases} \quad V_N \leq t \quad (2)$$

$$f(x) = \begin{cases} \text{foreground} & v(x) < t \\ \text{background} & v(x) \geq t \end{cases} \quad V_N > t$$

Secondly, a dynamic threshold is proposed: taking the average distance $meanD(x, y)$ from each pixel to the background sample set first. The dynamic threshold is calculated as follows:

$$R = \begin{cases} r - \alpha_1 meanD(x, y), & \beta meanD(x, y) < r \\ r + \alpha_2 meanD(x, y), & \beta meanD(x, y) \geq r \end{cases} \quad (3)$$

In this paper, $\alpha_1=0.5$, $\alpha_2=0.25$, $\beta=3$, r is original radius parameter 20. The larger the $meanD(x, y)$ is, the more complex the background is, so a larger R is needed, otherwise a smaller R is required. The parameter β is adjusted to an appropriate value according to the background change, and take $\beta meanD(x, y)$ as the threshold judgment standard.

After morphological processing operations such as dilation, erosion, opening operation and closing operation, the detection result of smoke video motion area is shown in the figure3.

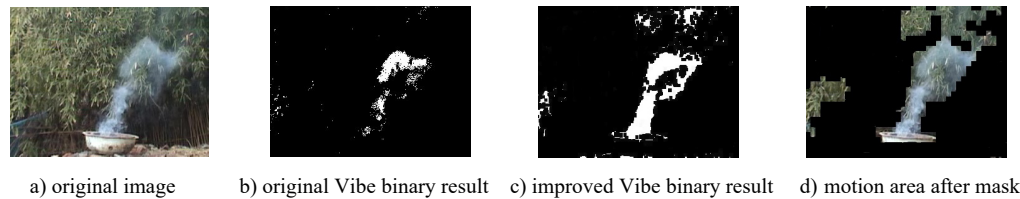


Figure3. The result of motion detection

The motion area extracted by the motion detection method includes not only the smoke diffusion area, but also may moving vehicles, walking person, etc. Therefore, further segment is needed to obtain a more accurate smoke motion area.

2.2 Color Extraction

The color attributes have an important role in smoke detection. Color is a powerful descriptive factor in human vision, it carries a large amount of visual information. Although the smoke area has been determined by the motion extraction, there have non-smoke motion areas in these areas that are different in color with smoke. It would cost much time if feature extraction is performed directly on this result. Therefore, color segmentation is performed on the motion region. CIELAB [17] is a color system of CIE that is used to confirm the numerical information of color. It can express colors based on human vision system. The CIELAB values cannot be obtained directly from RGB, but can be obtained roughly from sRGB [18]:

$$R = \text{gamma}\left(\frac{sR}{255}\right), G = \text{gamma}\left(\frac{sG}{255}\right), B = \text{gamma}\left(\frac{sB}{255}\right)$$

$$\text{gamma}(x) = \begin{cases} \left(\frac{x + 0.055}{1.055}\right)^{2.4}, & x \geq 0.04045 \\ x / 12.92, & x < 0.04045 \end{cases} \quad (4)$$

$$[X \ Y \ Z]^T = \begin{bmatrix} 0.412453 & 0.357580 & 0.180423 \\ 0.202671 & 0.715160 & 0.072169 \\ 0.019334 & 0.119193 & 0.950227 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

$$X = \frac{X}{X_n}, Y = \frac{Y}{Y_n}, Z = \frac{Z}{Z_n}$$

Among this, $X_n = 0.950456$, $Y_n = 1$, $Z_n = 1.088754$.

$$L = 116f(Y) - 16$$

$$a = 500 * (f(X) - f(Y))$$

$$b = 200 * (f(Y) - f(Z))$$

$f(X), f(Y), f(Z)$ can be expressed as the filter:

$$f(m) = \begin{cases} m^{\frac{1}{3}}, & m > (24/116)^3 \\ \frac{841}{108}m + \frac{16}{116}, & m \leq (24/116)^3 \end{cases} \quad (7)$$

By extracting and analyzing the color components of the smoke image in CIELAB color spaces, a color statistical model based on the Lab components is proposed.

$$40 < L < 80$$

$$C = \sqrt{a^2 + b^2} < 15 \quad (8)$$

C stands for chroma which indicates the vividness of the color.

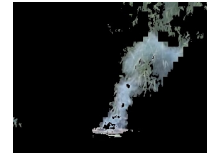


Figure4. Image before and after color filtering

Figure5. Suspected smoke area before and after filtering

After using the CIELAB color filtering to segment the suspected smoke area, although most of the interference between the color and the smoke moving area can be eliminated (see Fig 4 and 5), there may still be very few moving areas similar to the smoke color. Therefore, in-depth analysis based on more essential characteristics of the smoke image should be carried out.

2.3 Texture detection

Texture feature is a global feature, which reflects the organization of slow transformation and periodic change on the surface of the object. Better results can be achieved by texture feature in recognizing smoke image. LBP and GLCM are the most widely used method in texture feature extraction.

The earliest LBP descriptor was proposed by Ojala et al. [19] for the texture image classification. Taking the gray value of the center pixel as a threshold for binary quantization, then the neighborhood gray value is encoded as 1 if it is greater than or equal to the gray value of the center point, otherwise, it will be encoded as 0. Then, the LBP value of center pixel is obtained by transforming the 8-bit binary number into decimal number. In the meanwhile, there will be $2^8=256$ patterns.

The uniform LBP [20] was proposed to achieve effective dimensionality reduction, which is defined as when the cyclic binary number corresponding to a certain LBP has at most two jumps from 0 to 1 or from 1 to 0. The binary corresponding to that LBP is called the uniform class, which is proven to be a powerful texture. The rest of the patterns are grouped into another class, called the mixed pattern class. Therefore, the generated binary codes are classified into 58 uniform patterns and one mixed pattern, for reducing computational complexity.

GLCM [21] can reflect the comprehensive information of image gray scale, such as direction, adjacent interval and amplitude of change, but it cannot directly provide the features of different textures. Therefore, statistical parameters that can quantitatively describe texture features are extracted by counting the values of normalized GLCM.

$$\text{energy} = \sum_{i=1}^g \sum_{j=1}^g P^2(i, j, d, \theta) \quad (9)$$

$$\text{entropy} = - \sum_{i=1}^g \sum_{j=1}^g P(i, j, d, \theta) \times \lg P(i, j, d, \theta) \quad (10)$$

$$\text{correlation} = \frac{\sum_{i=1}^g \sum_{j=1}^g ijP(i, j, d, \theta) - \mu_1\mu_2}{\sigma_x\sigma_y} \quad (11)$$

$$\text{contrast} = \sum_{i=1}^g \sum_{j=1}^g [(i-j)^2 \times P^2(i, j, d, \theta)] \quad (12)$$

Where (i, j) is the grayscale value of any point (x, y) in image and another point $(x + \Delta x, y + \Delta y)$, their distance is d and the angle is θ , let the level of gray value be g . For the whole image, count the number of occurrences of every (i, j) and arrange them into a square matrix. Normalize the occurrences to get the probability, and $P(i, j, d, \theta)$ is obtained. Among

$$\text{this, } \mu_1 = \sum_{i=1}^g i \sum_{j=1}^g P(i, j, d, \theta), \mu_2 = \sum_{i=1}^g j \sum_{j=1}^g P(i, j, d, \theta), \sigma_x = \sqrt{\sum_{i=1}^g (i - \mu_1)^2 \sum_{j=1}^g P(i, j, d, \theta)}, \sigma_y = \sqrt{\sum_{j=1}^g (j - \mu_1)^2 \sum_{j=1}^g P(i, j, d, \theta)}.$$

In this paper, energy characteristics, entropy, correlation and contrast, 4 directions and 16 features vectors are selected to further identify smoke.

3. EXPERIMENTS

3.1 Experiment configuration and evaluation index

The first experiment for comparing between different texture features and different classifiers was first conducted based on Windows 11 system with Intel Core i5 CPU; the software is Matlab2018. In the experiment, the smoke dataset from the National Laboratory of Fire Science, University of Science and Technology of China is used [22]. It includes 800 still images of 100×100 pixels, dividing into two groups, 400 images containing smoke and 400 images not containing smoke. 59 uniform LBP patterns and 16 GLCM parameters, totally 75 feature vectors were extracted and feed into a classifier for training to obtain a classification model. SVM and random forest (RF) were compared.

In order to measure the performance of the algorithm quantitatively, the confusion matrix is chosen as the index to evaluate the retrieval performance of the algorithm, TN denote the number of samples which is predicted to be negative from smokeless, FP denote the number of samples which is predicted to be positive from smokeless, FN denote the number of samples which is predicted to be negative from smoke and TP denote the number of samples which is predicted to be positive from smoke. Accuracy (Acc), precision (Prec), recall rate (Rec) and the value of λ in (13)-(16) is used as measures for testing algorithms.

$$\text{Acc} = \frac{TP + TN}{TP + FN + FP + TN} \times 100\% \quad (13)$$

$$\text{Prec} = \frac{TP}{TP + FP} \times 100\% \quad (14)$$

$$\text{Rec} = \frac{TP}{TP + FN} \times 100\% \quad (15)$$

$$F_1 = 2 \times \frac{\text{Rec} \times \text{Prec}}{\text{Rec} + \text{Prec}} \times 100\% \quad (16)$$

3.2 Experiment results

Three different features, LBP, GLCM and the combination of LBP and GLCM, and two classifiers, SVM and random forest (RF), were compared in the experiment. 400 random images, half with smoke and half without smoke, were selected as the training images and test images respectively. The comparison result is presented in Table I and Figure 6 for 3 texture features and two classifiers. In Table I and Figure 6, LG denotes the combination of LBP and GLCM features, RF denotes random forest.

Table 1. Evaluation of different method

Index	LBP		GLCM		LG	
	SVM	RF	SVM	RF	SVM	RF
Acc	83.0	90.5	91.0	87.0	81.0	96.0
Prec	81.7	86.5	79.8	77.5	79.2	95.1
Rec	85.0	96.0	91.0	93.0	84.0	97.0
F_1	83.3	91.0	85.0	84.5	81.5	96.0

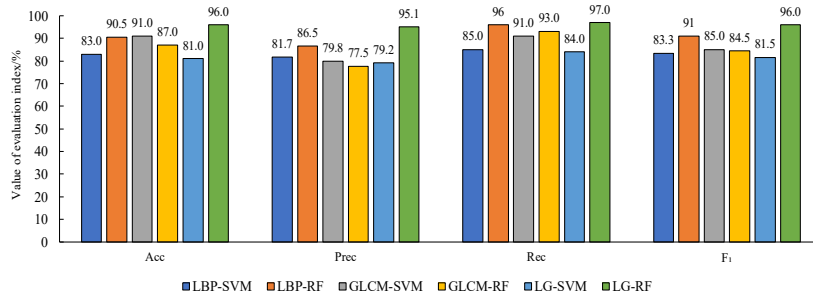


Figure6. Comparison of evaluation index of classification

It can be seen from Table I that the Acc, Prec, Rec, and value of SVM is better than those of RF when only classifying GLCM features. However, when it comes to LBP and the combination of LBP and GLCM, the result of RF is better than SVM. By analyzing the characteristics of the SVM, it can be seen that when there are too many features, it is difficult to find an optimal hyperplane to divide two types of samples, the reliability of segmentation is not enough, and the generalization ability is insufficient. Therefore, SVM has better effect on small sample training, but if the number of image features is large, the classification effect is poor. By contrast, RF has a high degree of self-adaptation and good generalization ability.

In the second experiment, the RF classifier that has been trained is used to test smoke video. In the experiment, the proposed Vibe algorithm and color filter were implemented to segment suspected smoke area, and then combination of LBP and GLCM features were extracted and feed into trained RF classifier. The results were compared that from with Liu's method [23] as is shown in the Table II. It used background modeling and color information to mix the static and dynamic features, then SVM was used to discriminate texture features. On the whole, the performances of the proposed method outperform Liu's method. Liu's algorithm is easy to be influenced by surroundings, and some object with subtle texture, blurred color and irregular movement are likely to be detected as smoke.

Table 2. Smoke video detection results

Smoke videos	Number of frames	Proposed method	Liu's method
Smoke With gray walls	517	97.8	100
Smoke With walking people	2886	98.4	75.4
Smoke With green leaves	1200	99.6	93

4. CONCLUSION

This paper proposes an effective smoke detection method based on surveillance video images. Combined with the proposed motion feature and color filtering, the complete smoke area is segmented first to improve detection. The eigenvectors extracted by the uniform LBP and the GLCM can effectively express the essential characteristics of smoke; the random forest classifier is selected for distinguishing smoke and non-smoke. The test results show that the method has higher detection efficiency, compared with traditional smoke detectors.

ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China (No.61775170).

REFERENCES

- [1] C. Emmy P and S. S. Vinsley and S. Suresh. "Multi Feature Analysis of Smoke in YUV Color Space for Early Forest Fire Detection," *Fire Technology*, vol. 52, pp.1319-1342, 2016.
- [2] DING Xiong, LU Yan. "Early smoke detection of forest fires based on SVM image segmentation. *Journal of Forest Science*," vol. 65, pp.150-159, 2019.
- [3] WU Zhangxian, YANG Guotian, LIU Xiangjie, "A new method for fire smoke identification based on Gabor wavelet," *Journal of Instrumentation*, vol.31, pp.1-7, 2010.
- [4] Shidong W, "Early smoke detection in video using swaying and diffusion feature," *Journal of Intelligent & Fuzzy Systems*, vol.26, pp.267-275, 2014.
- [5] S. Li, B. Wang, L. Gong, "A novel smoke detection algorithm based on MSER tracking," in *The 27th Chinese Control and Decision Conference (2015 CCDC)*, 2015, pp. 5676-5681
- [6] James M. "Flame and smoke estimation for fire detection in video based on optical flow and neural networks," *International Journal of Research in Engineering and Technology*, vol.03, pp.324-328, 2014.
- [7] LIU Ying, GU Xiaodong, LI Daxiang, "Fire and Smoke Detection Algorithm Based on LSA and SVM," *Journal of Xi'an University of Posts and Telecommunications*, vol.19, pp. 6-10, 2014.
- [8] Yuan, F., Shi, J., Xia, X., Yang, Y., Fang, Y., Wang, R., "Sub oriented histograms of local binary patterns for smoke detection and texture classification," *KSII Trans. Internet Informat. Syst. (TIIS)*, vol.10 (4), pp.1807-1823, 2016.
- [9] Islam, M.R., Amiruzzaman, M., Nasim, S., Shin, J, "Smoke object segmentation and the dynamic growth feature model for video-based smoke detection systems," *Symmetry*, vol.12, pp.1075-1093,2020.
- [10] LIU Kai, LIU Xiang, CHANG Liping, "Video smoke detection based on YUV color space and multiple feature fusion," *Chinese Journal of Sensors and Actuators*, vol.32, pp. 237-244, 2019.
- [11] WANG Yuanbin, "Research on fire detection and recognition in coal mine based on image features," *Xi'an University of Science and Technology*, 2013.
- [12] Frizzi S, Kaabi R, Bouchouicha M, "Convolutional neural network for video fire and smoke detection," in *Conference of the IEEE Industrial Electronics Society*, 2016, pp.877-882.
- [13] Gagliardi A, Gioia F D, Saponara S, "A real-time video smoke detection algorithm based on Kalman filter and CNN," in *Journal of Real-Time Image Processing*, 2021, pp.2085-2095.

- [14] Hohberg S P, "Wildfire Smoke Detection using Convolutional Neural Networks," Freie Universität Berlin, 2015.
- [15] Barnich O, Droogenbroeck M V. "ViBe: A Universal Background Subtraction Algorithm for Video Sequences," in Image Processing of the IEEE Transactions, vol.20, pp.1709-1724, 2011.
- [16] Otsu N. "A threshold selection method from Gray-Level Histograms ", in Transactions on System Man and Cybernetic of the IEEE, vol.9, pp62-66, 1979.
- [17] CIE. Colorimetry. CIE 15:2004, CIE Central Bureau, Vienna, 2004.
- [18] Stokes M, Anderson M, Chandrasekar S, Motta R. "A Standard Default Color Space for the Internet — sRGB", 1996.
- [19] T. Ojala, M. Pietikainen and D. Harwood, "A Comparative Study of Texture Measures with Classification Based on Feature Distributions," Pattern Recognition, vol.29, pp.51-59, 1996.
- [20] T. Ojala, M. Pietikainen and T. Maenpaa, " Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," IEEE Trans. Pattern Anal. Mach. Intell. vol.24, pp. 971-987, July 2002.
- [21] HARALICK R M, "Statistical and structural approaches to texture, " Proceedings of the IEEE, vol.67, pp.786, 2005.
- [22] <http://staff.ustc.edu.cn/~yfn/vsd.html>
- [23] LIU Minjie, "Design and implementation of video-based fire smoke detection," Nanjing University of Posts and Telecommunications, 2019.