Video fire recognition based on multi-channel convolutional neural network

Chen Zhong\textsuperscript{1*}, Yu Shao\textsuperscript{1}, Hongjun Ding\textsuperscript{1} and Ke Wang\textsuperscript{2}

\textsuperscript{1} Shenyang Fire Research Institute of MEM, Shenyang, Liaoning, 110034, China
\textsuperscript{2} State Key Laboratory of Robotics and Systems, Harbin Institute of Technology, Harbin, Heilongjiang, 150080, China
\textsuperscript{*}Corresponding author’s e-mail: zhongchen@efire.cn

Abstract. A video fire detection method based on multi-channel convolutional neural network (CNN) is presented in this paper. With improved OTSU and connectivity analysis, an adaptive threshold segmentation for flame image is implemented. According to the flame combustion characteristics, the feature of flame area is extracted using parameters such as colour, roundness and area change. Using TensorFlow platform, the CNN and the video database is formed. In order to improve the ability of dynamic fire detection, the method of combining random gradient descent and momentum correction is used to train CNN. Experimental results show that the accuracy of CNN fire detection method is 86\%, compared with 68.2\% of the traditional k-nearest algorithm (KNN).

1. Introduction
In recent years, with the wide application of video data acquisition and the development of video image fire detection technology, the research on video fire recognition and its anti-interference performance has been continuously deepened[1-2].

The independent segmentation of flame images and the anti-interference of light sources are the research focuses in this field[3]. Most video fire detection uses the conventional recognition methods in image processing[4]. The key of this kind of method lies in the segmentation of the suspected flame area and the extraction of dynamic and static features of the flame[5]. These segmentation methods mostly work well in simple backgrounds, and it is difficult to achieve independent and accurate segmentation of flame targets in complex backgrounds[6]. In addition, commonly used image features (image moments, textures, etc.) are difficult to fully reflect the combustion characteristics of flame, resulting in insufficient anti-interference performance and high false recognition rate[7].

Using improved OTSU for adaptive flame segmentation, and extracting multi-features such as colour, shape and area change, then a multi-channel CNN based on TensorFlow is built. Compared with the traditional KNN algorithm, the accuracy of the fire detection based on multi-channel CNN adopted in this paper has been significantly improved.

2. Flame segmentation of video images

2.1. Static segmentation based on improved OTSU
According to the OTSU principle[8], it is assumed that the difference between background and foreground is the largest at the optimal threshold, and the extremum of variance between classes is
used as the measurement standard. Cause the gray-histogram of some fire images is flat (that means the foreground and background boundary are fuzzy), so a large threshold uncertainty may lead a detail loss of the binary image, such as flame angle. Therefore, it’s difficult to get a true reflection of the suspected flame contour. For threshold selection, the improvements are as follows:

Assuming the size of the image I is X×Y, the number of pixels whose gray level is less than or greater than the threshold T is denoted as N_s and N_g, respectively. Then the foreground and background pixel ratios \( \omega_0 \) and \( \omega_1 \) are:

\[
\omega_0 = \frac{N_s}{X \times Y} \tag{1}
\]

\[
\omega_1 = \frac{N_g}{X \times Y} \tag{2}
\]

If the average gray level of the foreground and background is \( \mu_0 \) and \( \mu_1 \), the stretching coefficients are \( k_0 \) and \( k_1 \), then the total stretched gray level of the image is:

\[
\mu_l = \omega_0 k_0 \times \mu_0 + \omega_1 k_1 \times \mu_1 \tag{3}
\]

\[
g_l = \omega_0 \left( k_0 \times \mu_0 - \mu_l \right)^2 + \omega_1 \left( k_1 \times \mu_1 - \mu_l \right)^2 \tag{4}
\]

2.2. Dynamic segmentation based on connectivity analysis

The video segmentation of fire scenes often uses the inter-frame difference method [9]. Due to the weak dynamic characteristics for the middle part of the flame, it tends to produce a hollow in the flame area during segmentation (see figure 2(b)).

Therefore, the inter-frame difference method that introduces connectivity analysis is used to achieve dynamic segmentation of the suspected flame area. The algorithm flow is as follows:

- Take two consecutive images in the video image sequence, the previous image notes \( p_{k-1}(x, y) \), and the current image notes \( p_k(x, y) \);
- Calculate the difference \( F_D(x, y) \) between the current frame and the background frame, and the difference \( F_G(x, y) \) between the current frame and the previous frame, respectively;
- Calculate the intersection of the difference image \( F_D(x, y) \) and \( F_G(x, y) \) to get a rough segmented image of the moving target;
- Using mathematical morphology method[10] to seal, continuous, complete and reduce noise in the divided area.

Using conventional or connectivity analysis introduced inter-frame difference method, the dynamically segmentation comparison is shown in figure 2.
3. Multi-feature extraction of flame
Flame characteristics mainly include static (colour, shape, etc.) and dynamic (area change, flicker, jitter, etc.) characteristics. Considering calculation complexity, colour, shape and area change are selected.

3.1. Colour feature
RGB (red, green, blue) combined with S (saturation) information of the image is used. Since red and green components of the flame are larger than blue, with a correlation of saturation, the following constraints are established:

\[
\begin{align*}
R & \geq R_T \\
R & \geq G > B \\
S & \geq (255 - R) \times S_T / R_T
\end{align*}
\]

Where \( R_T \) is the red component threshold, \( S_T \) is the saturation threshold, \( R_T \) and \( S_T \) can be set by experience. It is impossible to accurately distinguish between flame and light interference only by colour features (see figure 3), so it is necessary to adopt a multi-feature fusion method.

3.2. Roundness feature
The shape of flame is usually irregular, which different from the interference. Roundness is defined as the degree that the edge of an object is similar to a circle, which can be used to indicate the complexity of the object edge. For calculating the roundness of the i-th frame image \( C_i \):

\[
C_i = \frac{4\pi \times S_i}{L_i^2}
\]

Where \( L_i \) and \( S_i \) are the contour perimeter and coverage area of the suspected flame in the i-th frame, respectively. For a circular object, \( C_i \) takes the maximum value of 1, and the more complex the shape, the smaller the value. A test of roundness feature with flame and interferences is shown in table 1.
### Table 1. Roundness test for flame and interference.

| Image group | 1   | 2   | 3   | 4   | 5   | Mean |
|-------------|-----|-----|-----|-----|-----|------|
| Flame       | 0.23| 0.23| 0.32| 0.32| 0.49| 0.32 |
| Lamp        | 0.82| 0.63| 0.86| 0.64| 0.73| 0.74 |
| Candle      | 0.54| 0.67| 0.63| 0.63| 0.76| 0.65 |
| Red cloth   | 0.50| 0.87| 0.90| 0.73| 0.39| 0.68 |

#### 3.3. Feature of area change

The target area obtained by frame segmentation can be marked as $\Omega_1,...,\Omega_n$, and the corresponding area is $S_1,...,S_n$. As the combustion spreads, the flame area will show a tendency to increase. Take the average flame area of consecutive $N$ frames for comparison (considering the requirement of recognition speed, the number of frames should not be excessive, take 50 frames as an example), $S_{in}$ and $R_{in}$ represent the area and change trend of $\Omega_i$ within the $n$-th frame. When $R_{in}>1$, it means that the current target area has an increasing trend. The calculation for $R_{in}$ is shown as follows:

$$R_{in} = \frac{\sum_{i=1}^{N} S_{in}/N}{\sum_{i=1}^{N} S_{(i-1)}/N}$$

#### 4. Multi-channel CNN model

A three-channel neural network is used to identify the features, as shown in figure 4. The colour and area change features using 3 layers’ neural network, and the roundness feature is recognized by CNN. This CNN consists of 5 convolutional layers, 3 pooling layers and 4 fully connected layers, of which 2 pooling layers are included between the first 3 convolutional layers, then 2 continuous convolutional layers, and then a pooling layer followed by 4 fully connected layers.

![Figure 4. Schematic diagram of CNN model structure.](image)

#### 4.1. Convolutional layer
- The first layer has 96 convolution kernels with size $11\times11$. The convolution step size is 4. The image size for input layer is $227\times227\times3$, and the size after the first convolution is $55\times55\times96$. 

• The second layer has 256 convolution kernels with size 5×5. The convolution step size is 1. The pooled 27×27×96 features becomes 27×27×256 features.
• The 3rd&4th layers have 384 convolution kernels with size 3×3. The convolution step size is 1.
• The 5th layer has 256 convolution kernels with size 3×3. The convolution step size is 1.

4.2. Pooling layer
The pooling layer integrates the output of neighbour neurons in the same feature map. Its input is the feature map calculated by the convolution layer. Each pixel in adjacent area of the flame has a large similarity. The feature dimension is greatly reduced after pooling as well as key information such as shape and corner of the flame is retained. Take the maximum value of this area:

$$P_{i,j} = \max_{(k,l) \in R_{i,j}} M_{k,l}$$

Where $P_{i,j}$ represents the maximum pooling value at position $(i, j)$, and $M$ represents the feature map. $R_{i,j}$ represents the pooling area at $(i, j)$. The first pooling layer sizes 3×3 with step size 1, the second pooling layer sizes 3×3 with step size 2, and the last pooling layer sizes 3×3 with step size 2.

4.3. Fully connected layer
A 4-layer fully connected layer with lengths of 9216, 1024, 256 and 2 is built. Multi-channels features are merged by the SoftMax function in the output layer.

5. Experimental results

5.1. Database
The experimental samples were taken from internet, laboratory, etc. The details are:
• 124 indoor fire videos, a single duration is from 8 to 150 seconds, total about 80 minutes, 30 frames per second;
• 20 indoor non-fire videos, a single duration is from 2 minutes to 10 minutes, total about 100 minutes, 25 frames or 30 frames per second.

After intercepting images form the videos, the images are then mirrored, rotated and cropped. The final sample size is: 250,000 training sets and 100,000 testing sets.

5.2. Fire recognition based on KNN
In KNN-based flame feature extraction and video classification (with or without fire):
• The training set is: 15 videos, ranging from 10 to 500 seconds, total about 1000 seconds, 30 frames per second;
• The testing set is: 8 videos, ranging from 10 to 200 seconds, total about 500 seconds, 25 frames or 30 frames per second.

The recognition accuracy shifts with the values of k. When k=5, it meet the peak accuracy 68.2%. Experimental results are shown in Figure 5.

![Figure 5. The testing result with KNN.](image-url)
5.3. Fire recognition based on CNN

Based on TensorFlow framework, training and classification with CNN is tested. 350,000 image samples are trained in a training time of 42 minutes.

The experimental results (see figure 6) show that the accuracy of the training set keeps a rising trend, while the accuracy of the testing set decreases after reaching 8 epoch, which due to overfitting. Therefore, the recognition accuracy of CNN should be the extreme value of 86% at 8 epoch, which is much higher than 68.2% of the KNN-based method.

![Graph of accuracy vs epoch for training and testing sets](image)

(a) Accuracy of training set

(b) Accuracy of testing set

Figure 6. Training and testing results with CNN.

6. Conclusion

This paper proposes a multi-channel CNN based video fire detection method. A combination of static segmentation based on improved OTSU and dynamic segmentation based on connectivity analysis is adopted. With features of colour, shape, and area change, a multi-channel CNN is built. Collecting videos and images with fire and non-fire scenes, a training set of 250,000 pictures and a testing set of 100,000 pictures are chosen for test. Stochastic gradient descent combined with momentum correction is adopted for training, and Dropout method is used to prevent overfitting. The final recognition accuracy rate is 86%, which is much higher than 68.2% of the comparison group based on KNN.

Acknowledgments

In this paper, the research was sponsored by National Key R&D Plan Project(Project No. 2017YFF0207004), and Science and Technology Program of Liaoning Province(2019-ZD-0865).

References

[1] Hongliang, L., Qing, L., Sunan, W. (2012) A novel fire recognition algorithm based on flame's Multi-features fusion. In: International Conference on Computer Communication and Informatics. Coimbatore. pp. 1-6.

[2] Celik, T., Demirel, H., Ozkaramanli, H., & Uyguroglu, M. (2007) Fire detection using statistical color model in video sequences. Journal of Visual Communication and Image Representation, 18(2), 176-185.

[3] Na, G., DONGAi-hua. (2008) Study on the method of flame image segmentation in fire detection. Journal of Henan Polytechnic University (Natural ence).
[4] Wang, Y., Ren, J. (2019). Application of KNN Algorithm Based on Particle Swarm Optimization in Fire Image Segmentation. Journal of Electrical Engineering & Technology, 14(4), 1707-1715.

[5] Bohush, R., Brouka, N. (2013) Smoke and flame detection in video sequences based on static and dynamic features. signal processing algorithms architectures arrangements and applications. 20-25.

[6] Liu, C., & Ahuja, N. (2004) Vision based fire detection. In: international conference on pattern recognition. pp. 134-137.

[7] Li, X., Hua, Y., Xia, N. (2013) Fire Detecting Technology based on Dynamic Textures. Procedia Engineering, 186-195.

[8] Xu, X., Xu, S., Jin, L., Song, E. (2011) Characteristic analysis of Otsu threshold and its applications. Pattern Recognition Letters, 32(7), 956-961.

[9] Cheng, Y. H., Wang, J. (2014) A Motion Image Detection Method Based on the Inter-Frame Difference Method. Applied Mechanics and Materials, 1283-1286.

[10] Haralick, R. M., Sternberg, S. R., Zhuang, X. (1987) Image Analysis Using Mathematical Morphology. IEEE Transactions on Pattern Analysis and Machine Intelligence, 9(4), 532-550.