2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)
Banff Center, Banff, Canada, October 5-8, 2017

An Adaptive Threshold Deep Learning Method for Fire and Smoke Detection

Xuehui Wu[1, 2], Xiaobo Lu[1, 2,*]
1. School of Automation, Southeast University
2. Key Laboratory of Measurement and Control of CSE, Ministry of Education, Southeast University, Nanjing 210096，China

Henry Leung[3]
3. Department of Electrical and Computer Engineering, University of Calgary, 2500 University Dr N.W., Calgary, AB, T2N 1N4, Canada

Abstract*—This paper proposes a novel method for fire and smoke detection using video images. The ViBe method is used to extract a background from the whole video and to update the exact motion areas using frame-by-frame differences. Dynamic and static features extraction are combined to recognize the fire and smoke areas. For static features, we use deep learning to detect most of fire and smoke areas based on a Caffemodel. Another static feature is the degree of irregularity of fire and smoke. An adaptive weighted direction algorithm is further introduced to this paper. To further reduce the false alarm rate and locate the original fire position, every frame image of video is divided into 16×16 grids and the times of smoke and fire occurrences of each part is recorded. All clues are combined to reach a final detection result. Experimental results show that the proposed method in this paper can efficiently detect fire and smoke and reduce the loss and false detection rates.

Keywords*—smoke and fire detection, motion detection, feature extraction, deep learning, Caffe, weighted value of direction, irregularity degree, original fire position

I. Introduction

Forests are indispensable resources to human survival and social development. They also protect the ecological balance on earth. However, due to some human activities that are unregulated and abnormal natural factors, the forest fires occur frequently which cause huge losses. Therefore, many researchers have recently studied forest fire and smoke detection with computer vision methods by. Yuan et al. [1, 2] analyzed the law of early fire and proposed a fire and smoke detection method based on motion accumulation. Den Breejen et al. [3] developed an autonomous forest fire detection method based on temporal contrast differences with natural background and spatial characteristics of the smoke plume. True et al. [4] explored various ways of combining many feature detection algorithms such as colour, motion, shape, growth and smoke behaviour. Kang [5] used image enhancement, colour image segmentation based on neural networks and GICA fast motion detection methods in detecting early agricultural and forest fires. Chen et al. [6] and Huang et al. [7] distinguished the flame by determining the stability of candidate regions' center of mass. Rong et al. [8] extracted the spatial and time features from the selected flame region, and further developed a pattern analysis method for the statistical landscape feature of the flame region. Zhang et al. [9] presented a new method using FFT (Fast Fourier Transformation) and wavelet transform for the contour analysis of forest fire images in video. Lu et al. [10] proposed an algorithm for testing early fire detection on video clips. Toreyin et al. [11], [12], [13] succeeded in detecting fire in a real-time video using different methods such as hidden Markov models and wavelet transform. In the method presented by Gubbi et al. [14], some statistical features, such as arithmetic mean, geometric mean, standard deviation, skewness, kurtosis, and entropy, are computed on each sub-band of 3-level wavelet transformed images. Then, the SVM (Support Vector Machine) light implementation was used for detection of smoke. Cai et al. [15] used static and dynamic features includeing high frequency energy, compactness and motion direction in fire and smoke direction in video.

Deep learning as a new method using deep convolutional neural networks, presents a modern way to process video and image. There are different tools that make possible working with these network models: Theano [16] framework for Python, Torch [17] built on Lua and Caffe [18] compiled with C++. As Caffe shows a fastest result in deep learning with deep convolutional neural networks, Bc. Tom´aˇs Poledn´ık [19] chose this framework to detect fire in images and videos. He created a sampler that was used to sample the fire recording according to a given approach. The model was created, trained and tested using Caffe [18]. Hohberg [20] trained a C3D convolutional neural network for recognizing wildfire smoke.

Environment of forest is complicated, fire and smoke features are changeable, and other moving objects can also be disturbance for detection. Taking advantages of these above methods into consideration, we combine deep learning and traditional detection method, dynamic and static features are used to recognize the exact fire and smoke areas. Adaptive method of locating original fire position improves the efficiency and accuracy of fire and smoke detection and reduces the loss and false detection.

II. Motion Detection

Before identifying the fire and smoke in video collected by a camera, it is necessary to find out motion areas on all the moving objects in video and extract the dynamic area from the background. We use ViBe (O. Barnich et al. [21]) method to extract the background of the whole video. Each background point has 1/ φ probability to update the values of its model samples and the model samples of its neighbour points. We randomly select a sample value to update, so that we can guarantee the smoothness of sample values, probability of a sample not being updated at time t is the $(N - 1)/N$ (N is the sum of frames). Assume that time is continuous, the probability of sample value being remained after dt time is

$$P(t, t+dt) = e^{-\ln\left(\frac{N}{N-1}\right)dt} \quad (1)$$

Different from the traditional ViBe algorithm, because fire and smoke are both continuous processes from their generation to disappearance, we need not only use the background for static feature extraction and classification, their dynamic characteristics of its diffusion should also be made full use of. Comparing every two adjacent frame images, we use the original background model got from ViBe and frame differences to update the background frame by frame. Every dynamic region of each frame will be used as a moving object to extract its static features and be classified. Dynamic feature extraction and recognition will be carried out by using the correlation of similar dynamic regions of adjacent frames.

Figure 1 shows the motion detection of a frame image. The left one is the binary image. Green rectangles of right image enclose all the motion areas while the red rectangles show the smoke area which is detected by our proposed method.

Fig. 1. Motion detection of smoke area

III. FEATURES EXTRACTION AND RECOGNITION

A. Static Features Extraction and Recognition

Smoke and fire in video are different from rigid objects with specific contours and features. An effective feature extraction method combined with deep learning and traditional feature extraction is used in this paper. For deep learning, many type of fire and smoke images have been chosen as our training data. However, other effective and special features of fire and smoke should also be extracted separately besides the trained deep learning model. Unlike cars that have a fixed shape, the shape of fire or smoke is very irregular, we need to calculate a quantitative data of irregularity degree, so that we can set a threshold to distinguish smoke and fire with high irregularity degree from other rigid objects with low irregularity degree such as cars. The Caffemodel was trained out from using massive amount of data. The dynamic feature of smoke and fire in video should also be extracted independently. We assign different direction of smoke with different weights and define an area as a smoke area when pre-selected values of direction reaches a threshold. To further distinguish fire and smoke from other moving objects, it is used to locate the original fire position. Figure 2 shows the architecture of fire and smoke detection in this paper. A motion region will be recognized as fire and smoke when its features satisfy all the conditions.

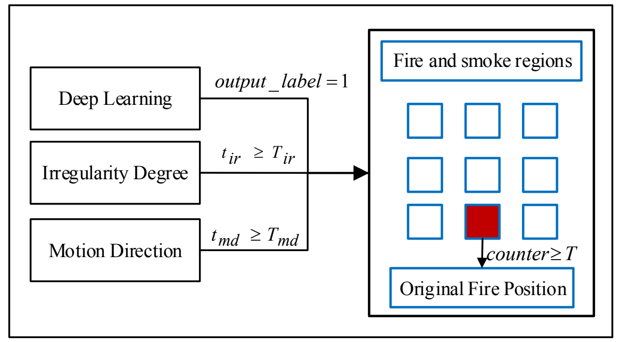


Fig. 2. Detection architecture

1) Fire and smoke detection with Caffe

The framework used to train model is the same structure of the deep convolutional neural network (CNN) that was proposed in the work of Krizhevsky et al. [22] for Imagenet classification. In the Caffe environment, this network structure is referred to as Alexnet.

Motion regions will be obtained by motion detection, the size of which is not big or even small because they are just some parts of an image. These small parts will be classified as test data with Caffemodel, so the training data size should also not be large. We use images of size 64×64×3 as training data. Figure 3 is the training process of fire and smoke Caffemodel.

The size of input data is 56×56×3 cropped from the original images (Every input image will be cropped to 5 images from its top left, top right, bottom left, top right and centre, and then rot them with 180°). At the first convolution layer, the input images are convolved with 96 kernels of size 4×4×3 combined with bias to form the output. And then non-saturating nonlinearity function- Rectified Linear Units (ReLUs: $f(x) = \max(0, x)$) [23] is put through as activation function. In general, we have that

$$x_j^{\ell} = f\left(\sum\nolimits_{i \in M_j} x_j^{\ell} \times k_{ij}^{\ell} + b_j^{\ell}\right) \quad (2)$$

where M_j represents a selection of input maps, ℓ is the number of layers and k_{ij}^{ℓ} is value of convolution kernels. Each output map is given an additive bias b_j^{ℓ}. The weight initialization of kernels is given by

$$w = \sqrt{\frac{6}{I_k + I_{k+1}}} \quad (3)$$

where I_k and I_{k+1} are the sizes of the layer before and

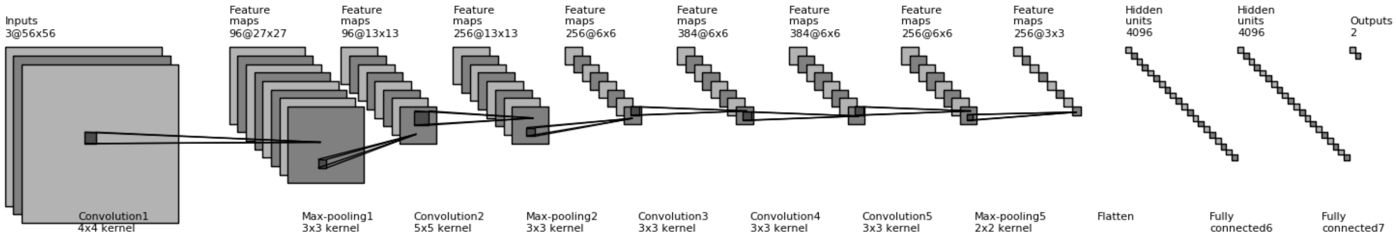


Fig. 3. CNN structure proposed in the work of Krizhevsky et al. [22] for Imagenet classification in this paper

after the weight matrix. After that, the size of first convolution layer output is 27×27×96 ((56-4)/2+1=27).

Local response normalization (LRN) is applied after applying ReLU nonlinearity.

$$x_j = x_i \Big/ \left(1+(\alpha/n)\sum_i x_i^2\right)^{\beta} \tag{4}$$

Where n is the local size of adjacent kernel maps and N is the total number of kernels in this layer.

Overlapping max pooling method is used after LRN to progressively reduce the spatial size of the representation to reduce the amount of parameters and computation in the network, and hence to also control overfitting.

The second convolutional layer takes the local response normalized and pooled output of first layer as input and filters it with 256 kernels of size 5 ×5 ×96 with 2 pads and stride of 1 pixels. The third, fourth, and fifth convolutional layers are connected to one another without any intervening pooling or normalization layers. The fully-connected layers have 4096 neurons each, and set the output of each hidden neuron to zero with probability of 0.5, which is called "dropout" [24].

The classifier used in this network is Softmax. Its input is the output of last fully-connect layer with a size of 4096×1 feature maps. Our goal is to find out fire and smoke in an image, and there are only two classes in output layer—fire and smoke ('1') and non-fire and smoke ('0'), so we set weight matrix size to be 2×4096. Figure 4 shows the details of Softmax classifier.

$$f_{y_i}(x_i;W,b_i) = Wx_i + b_i \tag{5}$$

$$F_{y_i}(z_{y_i}) = \frac{e^{z_{y_i}}}{\sum_j e^{f_j}} \tag{6}$$

where x_i, W, b_i are inputs, weight matrix and bias. f_{y_i} represents the function mapping F_{y_i} is the score as the unnormalized log probability. Softmax classifier encourages the normalized log probability of the correct class to be high while the loss to be low. The loss function is called cross-entropy loss.

$$L_i = -\log\left(\frac{e^{f_{y_i}}}{\sum_j e^{f_j}}\right) \tag{7}$$

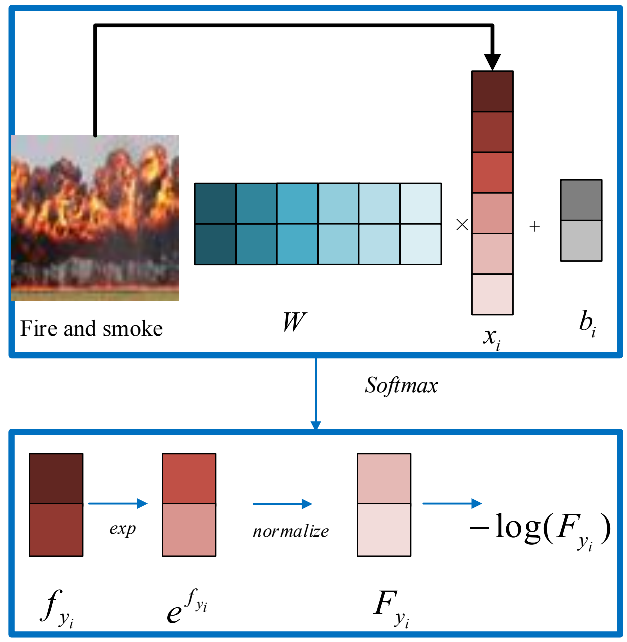


Fig. 4. Details of Softmax layer

We download enough independent non-fire and smoke images and smoke and fire images from the internet and use them as our training data. As shown in Figure 5，the value of loss function become lower and lower with oscillation while the accuracy increases with iterations.

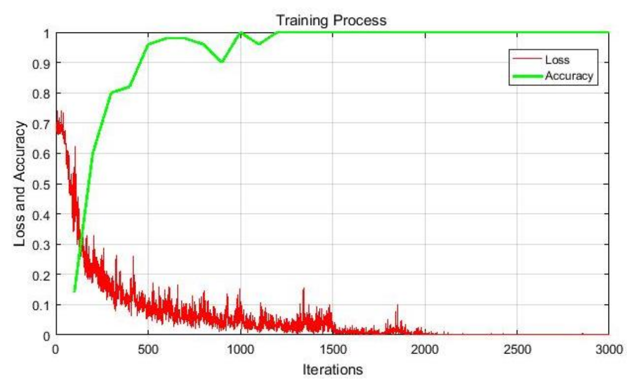


Fig. 5. Value of loss function and accuracy in training Caffemodel

Figure 6 shows the output feature map of first convolution layer, the general characteristics of fire and smoke can be seen clearly and the contour can still be identified as smoke and fire. The main contours will be more and more indistinguishable while more and more details will be learned with the increasing number of layers.

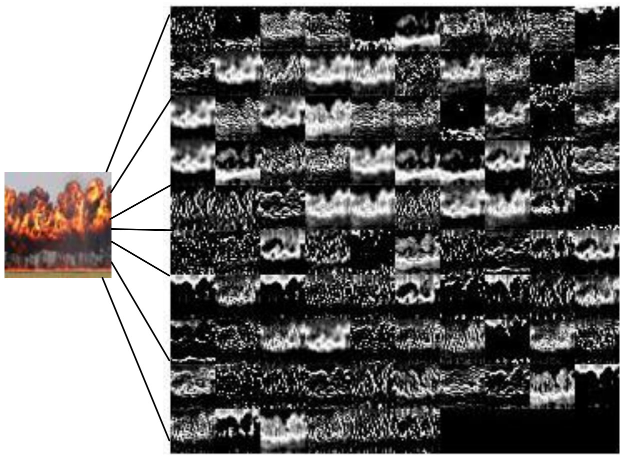


Fig. 6. Original image and feature map of first convolution layer

2) Fire and smoke detection with irregularity degree

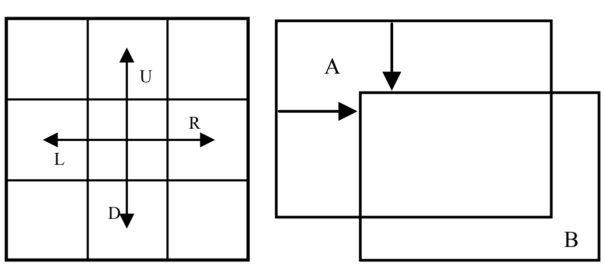
As deep learning is used to extract features, classify and recognize objects, their details will be learned constantly deepening on layers, however, the most obvious feature of fire and smoke is that they do not have fixed feature especially for their shape. Fire or smoke has no definite shape in the process of produce and diffusion, and its shape is irregular, we define this as the irregularity degree. We use length of side and area of irregular region as the quantitative expression of irregularity degree,

$$ID = \frac{c^2}{4\pi s} \tag{8}$$

where *ID* is the irregularity degree, *c* is the length of side and *s* is the area. In a binary image, the detected dynamic region is corroded to form a complete connected region, its every pixel value is 1, and we define the length of side as the sum of edge pixel values, the area as the sum of all pixel values of the whole region.

B. Dynamic Features Extraction and Recognition

Dynamic regions of fire and smoke in similar positions of adjacent frames have similar shape and their motion direction can be detected. Figure 7 (a) shows four motion directions. In Figure 7 (b), region B represents the motion region of current frame and region A the last frame. It shows that the smoke moves downward and rightward.



(a) (b)

Fig. 7. Motion direction

Four values of motion direction are calculated with

$$\begin{aligned} U = {} & I_{k,i}\left(y_1^{(k,i)} : y_2^{(k,i)}, x_1^{(k,i)} : x_2^{(k,i)}\right) \\ & - I_{k-1,i}\left(\left(y_1^{(k,i)} + p\right) : \left(y_2^{(k,i)} + p\right), x_1^{(k,i)} : x_2^{(k,i)}\right) \end{aligned} \tag{9}$$

$$\begin{aligned} D = {} & I_{k,i}\left(y_1^{(k,i)} : y_2^{(k,i)}, x_1^{(k,i)} : x_2^{(k,i)}\right) \\ & - I_{k-1,i}\left(\left(y_1^{(k,i)} - p\right) : \left(y_2^{(k,i)} - p\right), x_1^{(k,i)} : x_2^{(k,i)}\right) \end{aligned} \tag{10}$$

$$\begin{aligned} L = {} & I_{k,i}\left(y_1^{(k,i)} : y_2^{(k,i)}, x_1^{(k,i)} : x_2^{(k,i)}\right) \\ & - I_{k-1,i}\left(y_1^{(k,i)} : y_2^{(k,i)}, \left(x_1^{(k,i)} + p\right) : \left(x_2^{(k,i)} + p\right)\right) \end{aligned} \tag{11}$$

$$\begin{aligned} R = {} & I_{k,i}\left(y_1^{(k,i)} : y_2^{(k,i)}, x_1^{(k,i)} : x_2^{(k,i)}\right) \\ & - I_{k-1,i}\left(y_1^{(k,i)} : y_2^{(k,i)}, \left(x_1^{(k,i)} - p\right) : \left(x_2^{(k,i)} - p\right)\right) \end{aligned} \tag{12}$$

where *k* is the current frame, *i* is the number of connected region, *p* is a constant pixels, and *U,D,L,R* respectively represent the values of upward, downward, rightward and leftward. The main direction of fire and smoke is upward, therefore, we set the weighted value of upward the largest.

C. Location of original fire position

A flying bird, however, may also have the same motion features as smoke such as motion direction and its shape can also be changeable and irregular because of its flapping wings, then another obvious feature should be found to distinguish smoke and fire from these objects. Although smoke can appear and disappear in their diffusion, the original fire position will not move and can produce smoke around it constantly, while the flying birds or driving cars will move from one place to another, so that they cannot be detected as fire or smoke for several times at one position. We use this feature to record the number of times the area is determined as fire or smoke area. When the probability of occurrences of fire or smoke reaches our set threshold, we determine that the region is the original fire position, which can further shorten the rescue time, narrow the scope of search and rescue and reduce false alarm rate.

IV. EXPERIMENTS AND ANALYSIS

This experiment was carried out on a PC with an Intel Core i3, 2.40 GHz processor using method proposed in this paper.

In order to find out the optimal thresholds of two traditional feature extraction values, we experiment on two different videos-smoke video and none-smoke video. Figure 8 and Figure 9 respectively show the accuracy changes with threshold to update of irregularity degree and motion direction value. To guarantee the accuracies of both smoke and none smoke detections, threshold at cross point will be chosen.

Four videos with different environments have been tested in this experiment with the trained Caffemodel. Tabel I shows the experimental results using Method 1: traditional feature extraction method proposed by Cai et al. [15], Method 2: detection with only our trained Caffemodel, Method 3: detection with only irregularity degree and motion direction, and Method 4: the proposed method in this paper.

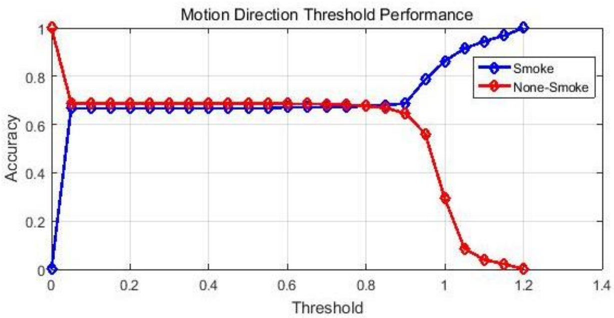


Fig. 8. Motion direction threshold performance in smoke and none-smoke videos detection, t=0.8 at cross point is chosen.

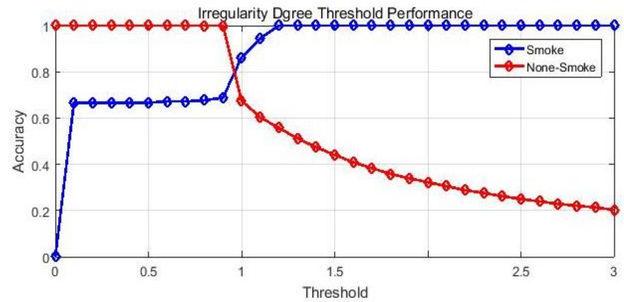
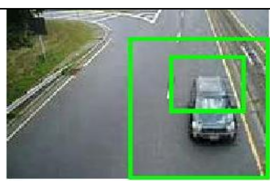
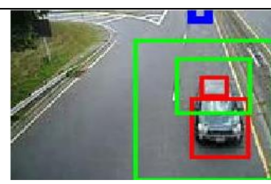
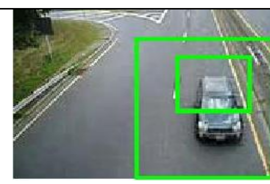
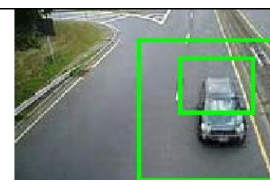
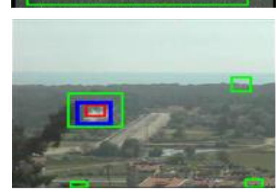


Fig. 9. Irregularity degree direction threshold performance in smoke and none-smoke videos dtection, t=0.95 at cross point is chosen.

TABLE I. COMPARISON OF TRADITIONAL FEATURE EXTRACTION METHOD PROPOSED BY CAI ET AL. [15] (METHOD 1), DETECTION WITH ONLY CAFFEMODEL (METHOD2), DETECTION WITH ONLY IRREGULARITY DEGREE AND MOTION DIRECTION (METHOD 3)AND PROPOSED METHOD IN THIS PAPER (METHOD 4)

Video	Total motion region	Detection rate				False positive rate				Description
		Method1	*Method2*	*Method 3*	*Method4*	*Method1*	*Method2*	*Method 3*	*Method4*	
Movie1	378	0.1296	0.5212	0.1296	0.0688	0.01296	0.5212	0.1296	0.0688	Cars on highway without smoke and fire
Movie2	9783	0.0154	0.3299	0.3957	0.1441	0	0.0081	0.0127	0	Fire and smoke in a garden
Movie3	4215	0.0230	0.4031	0.4698	0.1661	0	0.0064	0.0071	0	Smoke beside fire-coloured bin
Movie4	13131	0.0177	0.7601	0.1164	0.0646	0	0.0246	0.0037	0	Smoke in forest with some driving cars

TABLE II. FINAL DETECTION RESULTS OF METHOD 1, 2 ,3 AND 4 FOR THE MOVIES IN TABEL1

	Method1	*Method2*	*Method 3*	*Method3*
Movie1				
Movie2				
Movie3				
Movie4				

From the comparison, the performance of fire and smoke recognition with Method 1 has lowest detection rate even though its false positive rate is lower too. Method 2 and 3 can recognize more fire and smoke regions while the false recognition is higher, because they both just use only one algorithm (deep learning or traditional feature extraction), and our Caffemodel should be further adjusted. Therefore, we cannot detect fire and smoke only based on traditional feature extraction or deep learning method. The adaptive thresholds combined with deep learning approves to be more appropriate. Detection results of the experiment is shown in Table Π, our proposed method has a better performance and find the exact original fire position.

V. Conclusion

An adaptive threshold of feature and deep learning method for fire and smoke detection is proposed in this paper. We combine traditional feature extraction algorithm include irregularity degree and sum of weighted values of direction, with the deep learning method based on Caffe framework. Experiments shows that fire and smoke regions are effectively recognized and false alarm rate can be reduced. In the future, we will consider about how to recognize these feature-changeable objects like fire and smoke with deep learning framework, and try to insert static and dynamic features into deep convolutional neural network.

Acknowledgment

This work was supported by the National Key Science & Technology Pillar Program of China (No.2014BAG01B03), the National Natural Science Foundation of China (No.61374194), Key Research and Development Program of Jiangsu Province (No.BE2016739), and a Project Funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions.

References

[1] Yuan F N, Zhang Y M, Liu S X. Video smoke detection based on accumulation and main motion orientation. Journal of Image and Graphics, vol. 13, no. 4, pp. 808-813, 2008.

[2] Yuan F. Motion accumulation and translucence based model for video smoke detection. Journal of Data Acquisition and Processing,vol. 22, no. 4, pp. 396-400, 2007.

[3] E.Den Breejen et al.. Autonomous Forest Fire Detection. Proc. Third Int'l Conf. Forest Fire Research, pp. 2003-2012,1998.

[4] Nicholas True et al.. Computer Vision Based Fire Detection. San Diego: University of California, 2009.

[5] Kang F.The Preliminary Study of Early Agricultural and Forestry Fire Detection Method Based on Visual Features. PhD thesis, Zhejiang University, 2010.

[6] Chen, T.H., Wu, P.H. and Chiou, Y.C. An early fire-detection method based on image processing. Proceedings of IEEE International Conference on Image Processing, Singapore City, pp. 1707-1710, 2004.

[7] Huang, P.H., Su, J.Y. and Lu, Z.M. A fire-alarming method based on video processing. Proceeding of 2006 International Conference on Intelligent Information Hiding and Multimedia Signal Processing, Pasadena, pp. 359-364, 2006.

[8] Jianzhong R, Wei Y, Wei G, et al.. Fire Video Detection Method Based on Multi-Features Fusion. Journal of Combustion Science and Technology, 3: 007. 2013.

[9] Z. Zhang, J. Zhao, D. Zhang, C. Qu, Y. Ke, and B. Cai, Contour based forest fire detection using FFT and wavelet. In Proc. Int. Conf. CSSE, Wuhan, China, pp. 760–763, 2008, Dec.

[10] X. Zhou, F. Yu, Y. Wen, Z. Lu, and G. Song, Early fire detection based on flame contours in video. Inf. Technol. J., vol. 9, no. 5, pp. 899–908, 2010.

[11] B. U. Toreyin, Y. Dedeoglu, and A. E. Cetin, Flame detection in video using hidden Markov models, in Proc. IEEE ICIP, pp. II-1230-1233, 2005, Sep.

[12] B. U. Toreyin, Y. Dedeoglu, U. Gudukbay, and A. E. Cetin, Computer vision based method for real-time fire and flame detection, Pattern Recognition. Lett., vol. 27, no. 1, pp. 49-58, 2006, Jan.

[13] B. U. Toreyin and A. E. Cetin, Online detection of fire in video. In Proc. IEEE Conf. CVPR, pp. 1-5, 2007.

[14] Jayavardhana Gubbi, Slaven Marusic, and Marimuthu Palaniswami. Smoke detection in video using wavelets and support vector machines. Fire Safety Journal, vol. 44, pp. 1110 -1115, 2009, Feb.

[15] Min Cai,Xiaobo Lu, Xuehui Wu,Yifei Feng. Intelligent Video Analysis-based Forest Fires Smoke Detection Algorithms. Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), 2016 12th International Conference on,pp. 1504-1508, 2016, Aug.

[16] James Bergstra, Olivier Breuleux, Fr´ed´eric Bastien, Pascal Lamblin, Razvan Pascanu, Guillaume Desjardins, Joseph Turian, David Warde-Farley, and Yoshua Bengio. Theano: a CPU and GPU math expression compiler. In Proceedings of the Python for Scientific Computing Conference (SciPy), June 2010. Oral Presentation.

[17] R. Collobert, K. Kavukcuoglu, and C. Farabet. Torch7: A MATLAB-like environment for machine learning. In BigLearn, NIPS Workshop, 2011.

[18] Jia Y, Shelhamer E, Donahue J, et al. Caffe: Convolutional architecture for fast feature embedding. In Proceedings of the 22nd ACM international conference on Multimedia, pp. 675-678, 2014.

[19] Bc. Tom´aˇs Poledn´ık, Detection of Fire in Images and Video.Excel @FIT 2015.

[20] Simon Philipp Hohberg. Wildfire smoke detection using convolutional neural networks. 21 Technical report, Freie Universitt Berlin, Berlin, Germany, 2015, Sep.

[21] O. Barnich and M. Van Droogenbroeck. ViBe: A universal background subtraction algorithm for video sequences. IEEE Trans. Image Process. vol. 20,no. 6, pp. 1709-1724, 2011.

[22] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C.J.C. Burges, L. Bottou, and K.Q. Weinberger, editors, Advances in Neural Information Processing Systems, pp. 1097–1105, 2012.

[23] V. Nair and G. E. Hinton. Rectified linear units improve restricted boltzmann machines. In Proc. 27th International Conference on Machine Learning, 2010.

[24] G.E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R.R. Salakhutdinov. Improving neural networks by preventing co-adaptation of feature detectors. arXiv preprint arXiv:1207.0580, 2012.