

# Smoke detection in video sequences based on dynamic texture using volume local binary patterns

Gaohua Lin<sup>1</sup>, Yongming Zhang<sup>1</sup>, Qixing Zhang<sup>1,\*</sup>, Yang Jia<sup>2</sup>, Gao Xu<sup>1</sup> and Jinjun Wang<sup>1</sup>

<sup>1</sup>State Key Laboratory of Fire Science, University of Science and Technology of China  
Hefei 230026, China

[e-mail: qixing@ustc.edu.cn]

<sup>2</sup>School of Computer Science and Technology, Xi'an University of Posts and Telecommunications  
Xi'an 710121, China

\*Corresponding author: Qixing Zhang

*Received March 4, 2017; revised June 1, 2017; accepted June 27, 2017;  
published November 30, 2017*

---

## Abstract

In this paper, a video based smoke detection method using dynamic texture feature extraction with volume local binary patterns is studied. Block based method was used to distinguish smoke frames in high definition videos obtained by experiments firstly. Then we propose a method that directly extracts dynamic texture features based on irregular motion regions to reduce adverse impacts of block size and motion area ratio threshold. Several general volume local binary patterns were used to extract dynamic texture, including LBPTOP, VLBP, CLBPTOP and CVLBP, to study the effect of the number of sample points, frame interval and modes of the operator on smoke detection. Support vector machine was used as the classifier for dynamic texture features. The results show that dynamic texture is a reliable clue for video based smoke detection. It is generally conducive to reducing the false alarm rate by increasing the dimension of the feature vector. However, it does not always contribute to the improvement of the detection rate. Additionally, it is found that the feature computing time is not directly related to the vector dimension in our experiments, which is important for the realization of real-time detection.

---

**Keywords:** Smoke detection, Video sequences, Volume local binary pattern, Dynamic texture, Support vector machine

---

This work was supported by the Anhui Provincial Key Research and Development Program under Grant No. 1704a0902030, the National Key Research and Development Plan under Grant No. 2016YFC0800100, and the Fundamental Research Funds for the Central Universities under Grant No. WK2320000033 and WK6030000029. The authors gratefully acknowledge all of these supports.

## 1. Introduction

Videos carry very rich and complex information, and are more and more easy to be obtained with the rapid development of devices capable of digital media capture. Computer vision has many applications, e.g., face recognition, human action recognition [1], automatic media classification and annotation [2,3]. Video based fire detection is one of the application areas. Compared to conventional point smoke detector, video based fire detection system shows advantages in being usable in large open spaces, detecting fire immediately, providing more information such as the fire development and location [4]. According to detecting objects, video based fire detection methods can be classified into two categories: flame detection and smoke detection. As we know, smoke often emerges before flames, and the space range of flames is much smaller than smoke. Thus, smoke is a more efficient clue for early fire detection [5].

A considerable amount of work has studied the recognition of smoke in video. Here we overview a few related work but refer the reader to [4,5] for a more complete survey. Generally speaking, video based smoke detection methods distinguish smoke from non-smoke objects based on some distinctive features such as motion, edge, color and texture. Toreyin [6] used spatial wavelet transform to monitor the translucency of smoke. Chen [7] proposed a color decision rule for smoke which usually displays grayish colors. Genovese et al. [8] studied smoke color characteristics in YUV space. Yuan [9] proposed a fast algorithm for smoke detection using motion orientation estimation model. Yu [10] used optical flow computation to calculate the motion features of smoke. Jia [11] proposed a saliency based method for early smoke detection in video sequences.

The local binary pattern (LBP) [12] is one of the most prominent methods in the field of texture analysis with characteristics of gray scale invariance, rotational invariance and low computational complexity. LBP based smoke detection methods have been studied. Yuan [13, 14] proposed a method using the LBP and the variance of the LBP (LBPV) to extract the features of smoke, and used the orientation of gradient over LBP codes to detect smoke. Tian [15] used non-redundant LBP based features to detect smoke.

Previous studies have focused on image based local feature extraction. Zhao [16,17] proposed a LBP based feature descriptor to recognize dynamic texture in video: volume LBP (VLBP). Volume LBP can combine spatial and temporal features of smoke. Chen [18] proposed a video based smoke detection algorithm that used the LBPTOP to extract the smoke feature. Osman Günay et al. [19] proposed a real-time dynamic texture recognition method for fire detection using a randomly sampled subset of pixels in a given spatiotemporal block to reduce the computational cost of VLBP. Furthermore, they used this method to detect smoke in forests [5]. Additionally, high definition (HD) video equipment is becoming increasingly popular in surveillance systems and richer texture information contains in video sequences. Therefore, video based smoke detection based on texture has new expectations.

In this paper, VLBP operators are used to extract the dynamic texture of smoke in videos. Compared with the block processing method, the proposed VLBP dynamic texture extraction method is based on irregular regions. Several kinds of VLBP methods, with different frame intervals and sample points are used to extract the dynamic texture of moving regions and obtain the feature vectors. Then, SVM is used to do classification.

This paper is organized as follows: In Sections 2.1–2.3, texture extraction approaches like LBP, VLBP, LBPTOP and CVLBP are reviewed. In Section 3.1, our experimental video set is introduced. Section 3.2 and 3.3 present the motion region extraction method and the classifier respectively. Section 3.4, 3.5 and 3.6 are the analysis of smoke detection using dynamic texture features based on blocks and irregular regions. Finally, Section 4 is the conclusion of the study.

## 2. Dynamic texture extraction

### 2.1 Local binary pattern

The basic LBP operator proposed by Ojala et al. [12] for texture analysis can be defined as follow:

$$LBP_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c) 2^p, \quad s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}, \quad (1)$$

where  $g_c$  represents the gray value of the center pixel,  $g_p (p = 0, \dots, P-1)$  represents the gray values of the neighboring pixels,  $R$  is the radius of the circle and  $P$  is the total number of sample points in the circular neighborhood. There are four types of patterns: original, uniform, rotation-invariant and rotation-invariant-uniform.

The original pattern is defined in Eq. (1) and the other three patterns are deformations of the original pattern.  $LBP_{P,R}^{u2}$  is defined as a pattern that has no more than 2 spatial transitions. The  $U$  value of a LBP pattern is defined as the number of spatial transitions (bitwise 0/1 changes) and can be computed by

$$U(LBP_{P,R}) = |s(g_{p-1} - g_c) - s(g_0 - g_c)| + \sum_{p=1}^{P-1} |s(g_{(p+1)} - g_c) - s(g_p - g_c)|. \quad (2)$$

Patterns with more than 2 spatial transitions ( $U \geq 2$ ) are considered as identical pattern. Rotation invariance is achieved by assigning a unique identifier to each rotation invariant LBP:

$$LBP_{P,R}^{ri} = \min\{ROR(LBP_{P,R}, p) | p = 0, \dots, P-1\}, \quad (3)$$

where  $ROR(x, p)$  performs a circular bitwise right shift on the number  $x$  with  $p$  times.  $LBP_{P,R}^{riu2}$  is a combination of  $LBP_{P,R}^{u2}$  and  $LBP_{P,R}^{ri}$ . It is defined as

$$LBP_{P,R}^{riu2} = \begin{cases} \sum_{p=1}^{P-1} s(g_p - g_c) & \text{if } U(LBP_{P,R}) \leq 2 \\ P+1 & \text{otherwise.} \end{cases} \quad (4)$$

**Fig. 1** shows the basic computation procedure of the LBP with  $P = 8, R = 1$ . For any other value ( $P, R$ ), the gray values of neighbors not located exactly at the center of the pixels are estimated by interpolation.

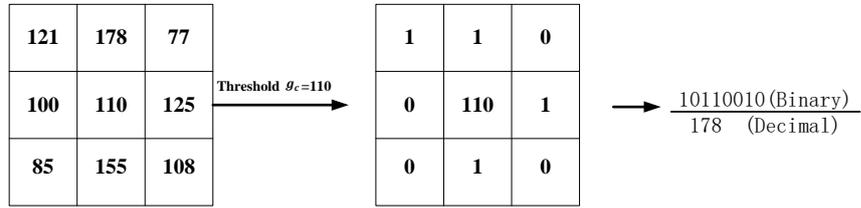


Fig. 1. Basic computation procedure of the LBP with  $P = 8, R = 1$ .

### 2.2 Volume local binary pattern

The VLBP was proposed by Zhao et al. [16,17] for dynamic texture recognition by extending the LBP to the spatiotemporal domain. It is defined as

$$VLBP_{L,P,R} = \sum_{p=0}^{3P+1} s(g_p - g_c) 2^p, \quad (5)$$

where  $g_c$  corresponds to the gray value of the center pixel of the local volume neighborhood,  $g_p$  represents the gray values of  $3P + 2$  pixels in the three frames with interval  $L$  and  $R$  is the radius of the circle in each frame. Fig. 2 shows the entire computing procedure for  $VLBP_{1,4,1}$ .

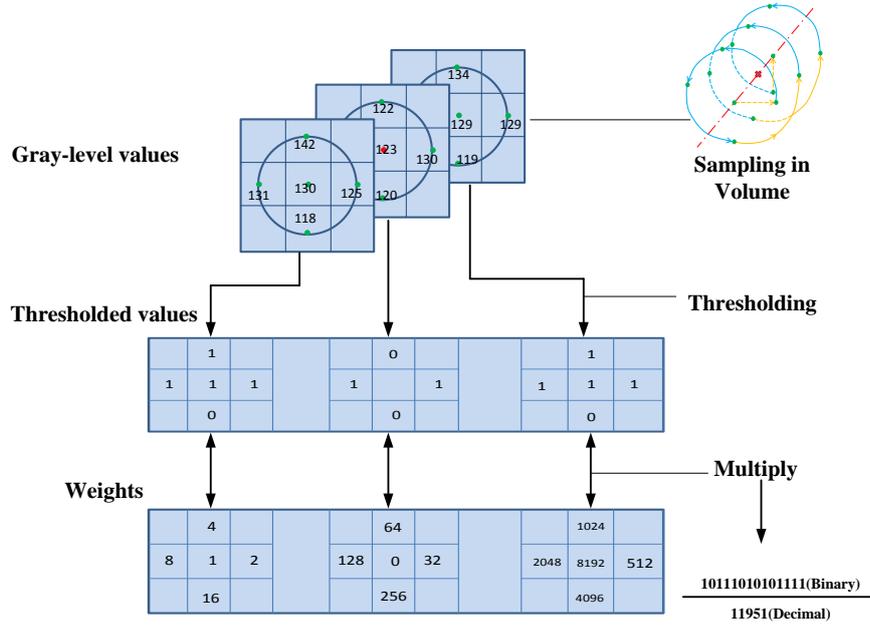


Fig. 2. Computing procedure for  $VLBP_{1,4,1}$ [17].

Similar to the basic LBP, VLBP also has four modes. Note that the rotation invariant VLBP should consider the synchronization of the three frames that rotate simultaneously with the same angles.

The LBPTOP is a simplified VLBP that decreases the dimension of the feature vector by considering three orthogonal planes around the center pixels and calculating the binary number separately for each plane. The final feature vector is obtained by concatenating the

histograms corresponding to each orthogonal plane.

### 2.3 Completed Volume local binary pattern

The conventional VLBP operator only takes the sign information of local differences into consideration. The CVLBP adds a global texture feature that uses center pixel information combined with a global mean difference as a threshold to modify the VLBP. The CVLBP framework is illustrated in Fig. 3. The temporal volume is represented as its volume center gray level ( $C$ ) and the local difference of the volume center pixel with circularly symmetric neighborhoods. The local difference is then divided into the sign ( $S$ ) and magnitude ( $M$ ) components. VLBP\_S is the same as the basic VLBP operator given in Eq. (5). VLBP\_M is defined as

$$VLBP_{L,P,R\_M} = \sum_{p=0}^{3P+1} s((g_p - g_c) - th) 2^p, \quad (6)$$

where  $th$  is the mean of the local differences ( $g_p - g_c$ ) over the entire volume.

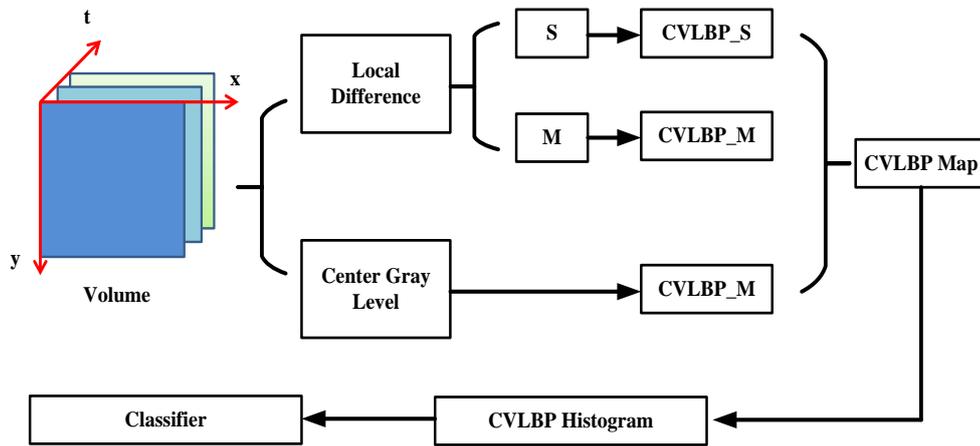


Fig. 3. Framework of the CVLBP.

## 3. Smoke recognition based on dynamic texture

### 3.1 Experimental video

To carry out the study, a fixed video set which can greatly affect the result of video based fire detection is indispensable for the comparison between the algorithms. Existing fire smoke videos for research are essentially low resolution, for example, the videos provided by Bilkent University [20]. HD video network cameras were used in our study to shoot a group of smoke videos and non-smoke videos with a size of  $1920 \times 1080$ , as shown in Fig. 4 and Fig. 5. The videos can be downloaded on our website (<http://smoke.ustc.edu.cn/datasets.htm>).

The videos used for training consisted of 10 smoke videos and 5 non-smoke videos named Video 1–15. Table 1 provides a simple summary of the content of each training video. Additionally, the videos used for testing contained another 10 smoke videos that are similar to the training set and the same 5 non-smoke videos in the training set. The testing videos were named Video A–O. The smoke videos contained black and white smoke with varied density.

The non-smoke videos were mainly indoor interference sources, including walking pedestrians, swinging curtains and flickering light.



Fig. 4. Training videos.



Fig. 5. Testing videos.

### 3.2 Extraction of the motion regions

The background subtraction method was used for moving region extraction, which is shown in Algorithm 1. Then to eliminate interference and reduce computation, dynamic textures only from the moving regions of the video were extracted. The background frame was updated by coefficient  $\alpha$ . After the foreground frame was extracted, morphological operations were applied to reduce noise.

---

#### Algorithm 1 Background subtraction method

---

$n$ : Frame index     $(i, j)$ : Pixel position     $threshold=8$      $\alpha=0.1$

$I_B$ : Background image,     $I_F$ : Foreground image

**if**  $abs(I^n(i, j) - I_B^{n-1}) > threshold$  **then**

$I_F^n=1$ ;

**Else**

$I_F^n=0$ ;

**End**

$I_B^n = I_B^{n-1} \times (1 - \alpha) + I^n \times \alpha$

$erode(I_F^n)$

$dilate(I_F^n)$

---

### 3.3 SVM

Support Vector Machines (SVM) was selected in our analysis as the classifier for dynamic texture features which have been the dominate classifier option for over a decade before the boom of Deep Learning [21]. The SVM introduced by Vapnik [22, 23] shows its advantages in solving the data classification problem in high dimensional space and for the nonlinearity problem in pattern recognition, and also applies to the condition where the feature dimension of the data is higher than the number of samples.

The SVM evolved from the concept of structural risk minimization in statistical learning theory. By simultaneously minimizing the empirical classification error and maximizing the geometric margin, SVM constructs a hyperplane that optimally separates the samples into two categories. SVM was originally used to solve linearly separable problems using an optimal separating hyperplane. However, many problems cannot be expressed by linear functions in the real world. Mapping the original finite dimensional space to a higher dimensional or infinite space with the application of a nonlinear function will make the samples linearly separable.

### 3.4 Block based dynamic texture extraction

We divided the video image into  $100 \times 100$  non-overlapping blocks. A total of 4,805 smoke blocks and 11,842 non-smoke blocks were labeled in video samples 1–15. Fig. 6 is a schematic of some samples. The number of labeled blocks contained in each video is listed in Table 1. Then the  $LBPTOP_{1,8,1}$  operator was used to extract the dynamic texture features of each block. A total of 2,366 smoke blocks and 5,989 non-smoke blocks were randomly selected from the labeled blocks and used as the training data for the SVM.



Fig. 6. Positive samples(left) and negative samples(right) in Videos 1–15.

Table 1. Content of training videos

Video	1	2	3	4	5	6	7	8	9
Content	White smoke plume					Diffuse white smoke		Thin black smoke	Black smoke
Smoke blocks	124	126	388	461	595	343	650	412	870
Non-smoke blocks	324	402	463	589	579	247	206	414	446
Video	10	11	12	13	14	15	Total	Training	Testing
Content	black smoke	Walking pedestrians			Swing curtains	Lighting changes	/	/	/
Smoke blocks	836	0	0	0	0	0	4805	2366	4085
Non-smoke blocks	445	2177	2206	348	1556	1440	11842	5989	11842

Firstly, a block was considered as the object to conduct the test. All the labeled blocks were used to test the SVM classification. The results are shown in Fig. 7 for the detection rate (DR) and false alarm rate (FAR). It can be observed that the blocks in Videos 1–5 were classified acceptably because they were distinct white smoke plumes. Some smoke blocks in diffuse

white smoke Videos 6 and 7 were identified mistakenly, and some non-smoke blocks in Video 6 that contained extremely thin smoke caused a high FAR. Video 8 contained a thin black smoke plume, and the blocks could not frame out the smoke perfectly, thus the detection performance was the worst in this case.

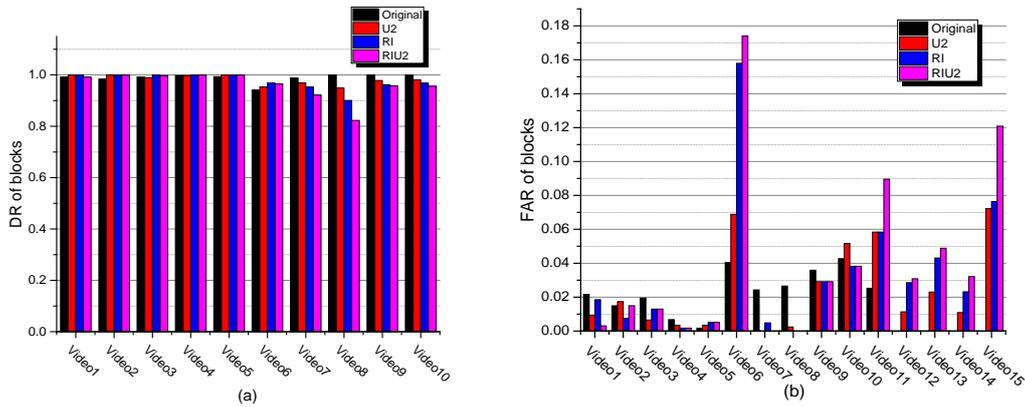


Fig. 7. DR and FAR of smoke block detection.

The black smoke plumes in Videos 9 and 10 have obvious shadows, resulting in that some non-smoke blocks were classified mistakenly. An unstable illumination condition also caused some false alarms. Generally, the volume local binary pattern can be used for smoke dynamic texture detection based on blocks, especially for distinct smoke. However, the robustness is still poor and affected by the shadow, illumination condition, and so on.

After the blocks were classified, the next procedure was to determine the category of each whole frame. For an HD video frame, the number of blocks is large, for example, 190 blocks with size  $100 \times 100$  pixels. It is not reasonable to classify a frame as a smoke frame when just one smoke block appears. We used the neighboring block rule [15] to issue a smoke alarm based on a frame, as shown in Fig. 8. When smoke appears in a frame, the area of smoke typically occupies multiple interconnected blocks. Thus, when the central block in Fig. 8 was detected as a smoke block, the frame was determined to be a smoke frame by verifying that the neighboring blocks of this smoke block had at least one smoke block. Otherwise, the central block was considered as a misclassification and no smoke alarm was issued.

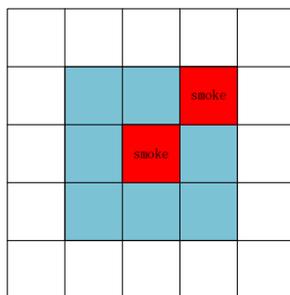


Fig. 8. Neighboring block rule.

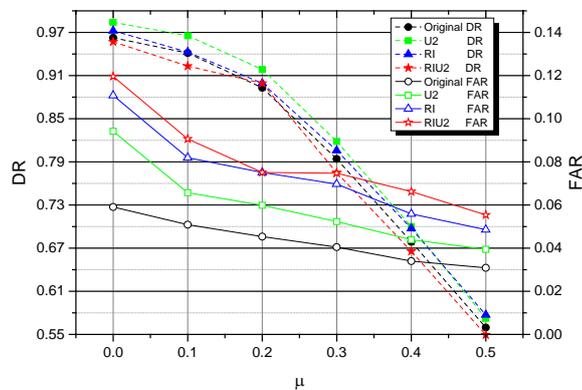


Fig. 9. DR and FAR of frames based on blocks.

For testing video A–O, we first extracted a motion region using Algorithm 1 and selected the candidate blocks from the extracted region subsequently. As the detection result is sensitive to motion area ratio threshold, we set the thresholds as 0, 0.1, 0.2, 0.3, 0.4 and 0.5 to select the candidate block. A motion area ratio of zero means that all the blocks in the frame are treated as the candidate blocks. The total DR and FAR of frames for testing video A–O against different thresholds are shown in Fig. 9. Obviously, a low threshold causes more blocks recognized as candidate smoke blocks which resulted in high DR and FAR of frames. The high FAR will decrease with increase in the motion area ratio accompanied by a lower DR simultaneously.

Compared to the smoke blocks with relatively consistent dynamic texture features, the non-smoke blocks had a variety of feature histograms, with the result that the larger the feature dimension is, the more non-smoke blocks could be excluded. Thus, it can be found that the original mode had the lowest FAR compared with the other three modes in Fig. 9 distinctly.

### 3.5 Dynamic texture extraction based on irregular regions

Both the motion area ratio threshold and frame alarm rule have a great influence on the frame DR and FAR. What even worse is that as HD cameras have wider view coverage and greater monitoring distance, the smoke region in the HD video frame have a large variable range. It is very difficult to determine a reasonable block size. We proposed a dynamic texture feature extraction method based on irregular region to avoid these problems. The moving regions with irregular shape in a frame are obtained using the same motion region extraction method as before. Then, the irregular motion regions in a frame are treated as an entire target, and dynamic texture was only obtained from these regions. The procedure is shown in Fig. 10.

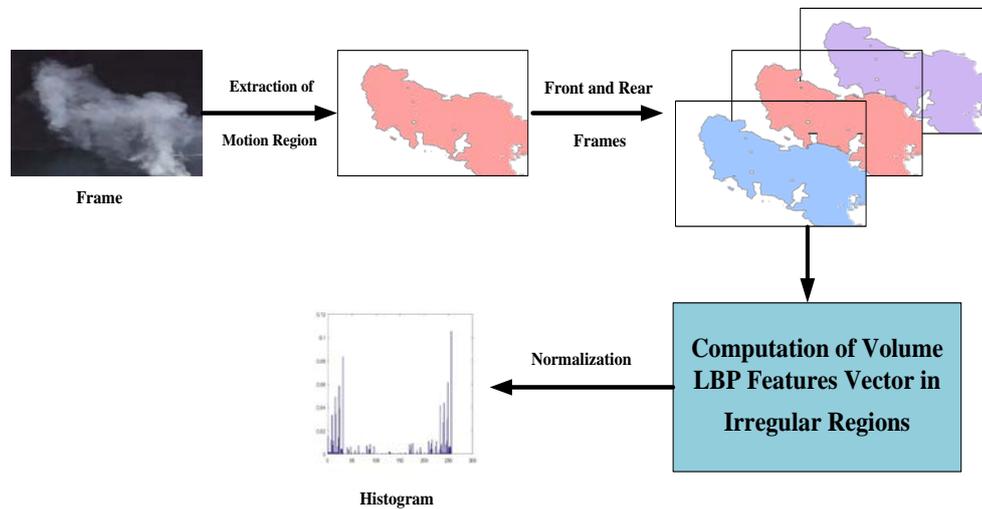


Fig. 10. Extraction of the VLBP texture features from irregular regions.

We also extracted the  $LBPTOP_{1,8,1}$  dynamic texture from irregular regions of the training frames in Video 1–15 and testing frames in Video A–O to conduct a comparison with the blocks based method. SVM was used to train and test the feature vectors of those frames and the result is shown in Fig. 11.

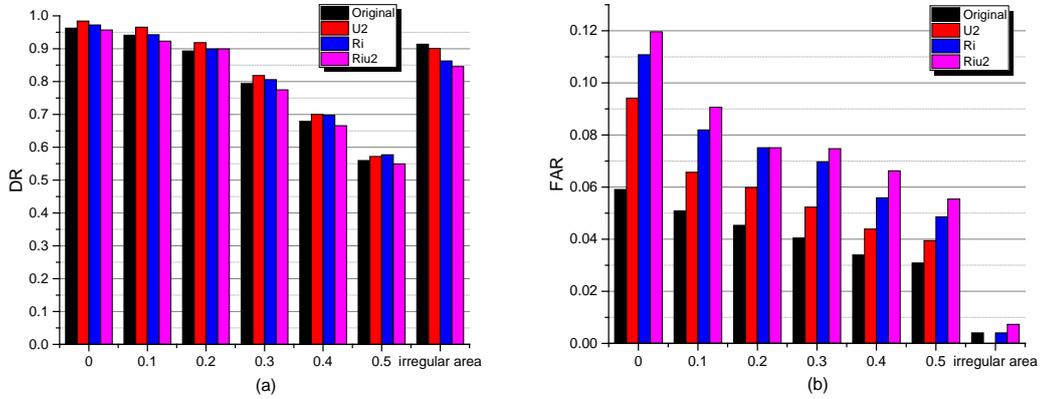


Fig. 11. DR and FAR of the irregular regions based method compared with the blocks based method.

It can be observed that the regions based method could greatly reduce the FAR while keeping the DR at a high level at the same time. Additionally, it is conducive to reduce calculation as the regions based method only extracts texture features from motion regions strictly. So for smoke recognition in HD videos, it is better to extract dynamic texture from irregular regions compared with the dynamic texture extraction based on blocks. The next experiments will be conducted based on irregular regions.

### 3.6 Evaluations of volume local binary patterns

#### 3.6.1 Temporal Information

To evaluate the performance of volume local binary patterns for smoke recognition, we implemented several methods for comparisons. Experimental results are shown in Fig. 12.  $LBP_{1,8,1}$  is the basic LBP operator with  $P = 8$  and  $R = 1$ .  $PLBP_{1,8,1}$  and  $PLBPPEOH_{1,8,1}$  are described in [24]. They are texture extraction methods based on one frame without temporal information. It is obvious that the volume local binary patterns ( $LBPTOP_{1,8,1}$ ,  $VLBP_{1,2,1}$ ,  $CVLBP_{1,2,1}$ ) have better results.

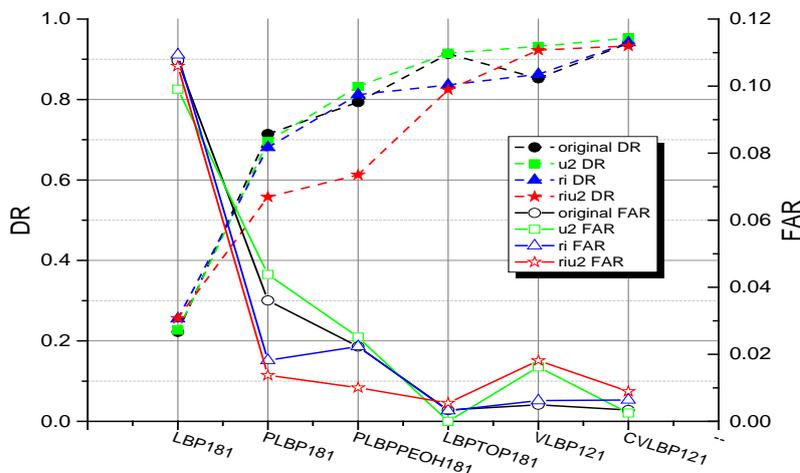


Fig. 12. Performances of different texture extraction methods.

### 3.6.2 Characteristics of volume local binary patterns operators

To study the difference between each dynamic texture extraction method, we used the LBPTOP, VLBP, and CVLBP operators to extract dynamic texture for smoke detection respectively. Fig. 13 shows the results. Generally for the FAR, as the non-smoke samples had various texture distributions, an operator with high feature dimension (more sample points and more complexity) was conducive to excluding non-smoke samples, thereby reducing the FAR.

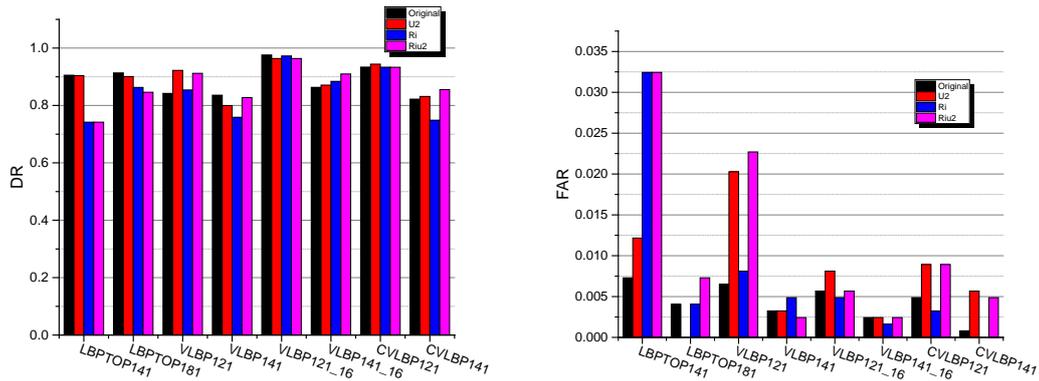


Fig. 13. DR and FAR of different dynamic texture extraction methods.

The DR for the LBPTOP with  $P = 8$  was higher than that with  $P = 4$ . However, increasing the number of sample points  $P = 2$  to  $P = 4$  did not improve the DR for the VLBP operators. It is generally believed that the LBPTOP is a simplification of the VLBP with the purpose of reducing the feature dimension and the computational complexity. We found that there may be some misunderstanding. Table 2 lists some parameters of the operators, where *pixels* represents the number of pixels actually used by the operator, excluding the central pixel, *dim* represents the dimension of the feature vector for each mode, and *t* is the time for calculating the feature vector for an identical video.

Table 2. Feature dimension and computation time.

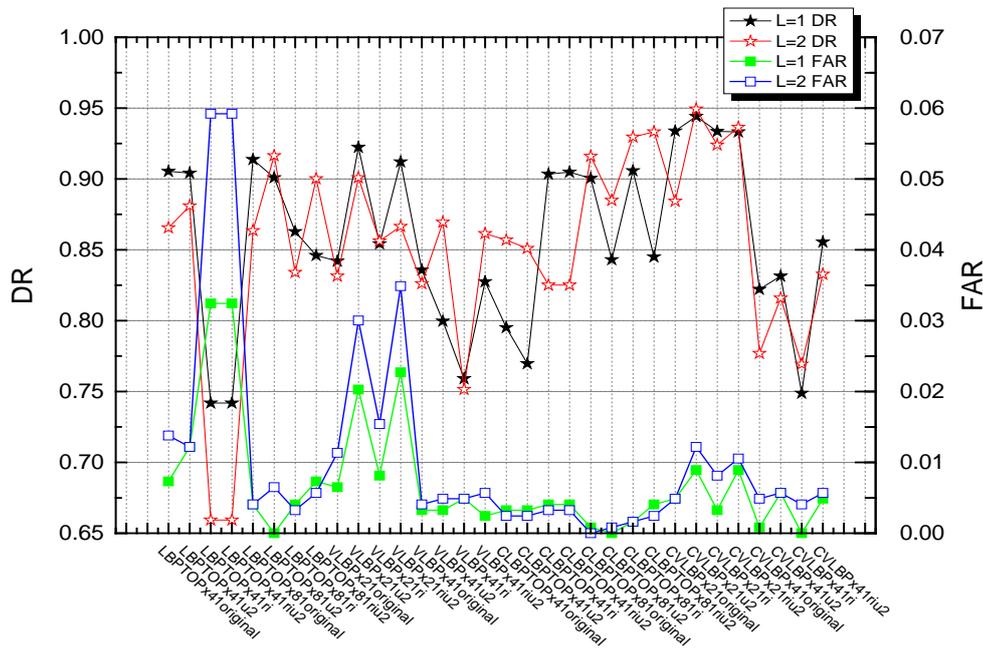
method	$L$	$P$	$R$	<i>pixels</i>	Original		U2		Ri		Riu2	
					dim	t/s	dim	t/s	dim	t/s	dim	t/s
$LBPTOP_{1,4,1}$	1	4	1	6	48	2.94	45	17.57	18	17.69	18	17.52
$LBPTOP_{1,8,1}$	1	8	1	18	768	6.06	177	20.52	108	20.77	30	20.29
$VLBP_{1,2,1}$	1	2	1	6	256	1.56	59	6.301	144	6.50	52	6.33
$VLBP_{1,4,1}$	1	4	1	14	16384	2.88	185	6.82	4176	7.39	148	6.80

It can be found that the computation time is not directly related to the feature dimension, for example, the original dimension of  $LBPTOP_{1,8,1}$  is 768 and the calculation time is 6.06

seconds; however, when the original dimension of  $VLBP_{1,4,1}$  reaches up to 16384, the calculation is only 2.88 seconds. The U2, Ri and Riu2 modes greatly reduce the feature dimension compared with the original mode. However, both of them are calculated based on the original one. Thus, the calculation time is obviously longer. In fact, the feature dimension of the operator only directly affects the training time of the SVM, which is greatly shorter than the time required for feature vector computation.

In [19], the volume block for VLBP dynamic texture extraction contained 64 frames. Given that the time and storage space required for the calculation increase with the number of frames, it is difficult to achieve a real-time detection. In this paper, we used only 3 frames to extract dynamic texture with irregular regions. A comparison between the detection results based on 3 frames and 16 frames is shown in Fig. 13. It is obvious that more frames result in a better detection capacity with higher DR and lower FAR. Meanwhile, more frames means more calculation. Thus, the detection requirements and hardware performance should be considered synthetically to determine the optimal number of frames for dynamic texture extraction.

Compared with the VLBP, the CVLBP contains not only the sign information of local difference, but also the magnitude features of local difference. As the magnitude (M) and center average gray (C) do not directly use the volume center pixel for thresholding, they are more robust to noise. As shown in Fig. 13,  $CVLBP_{1,2,1}$  has achieved the highest DR in the operators based on 3 frames, and is more computationally efficient than  $VLBP_{1,2,1-16}$ .



influences dynamic texture extraction. We used  $L = 2$  for all operators mentioned above to conduct the training and testing. The DR and FAR results are shown in [Fig. 14](#). We can observe that the general trends of FAR for the two conditions are similar, Whereas the amplitudes for  $L = 1$  are smaller in most cases. For the DR of the smoke frames, neither  $L = 1$  nor  $L = 2$  has obvious advantages over the other. The optimal frame interval value is depends on the operator.

## 4. Conclusion

In this paper, a video based smoke detection method extracting dynamic texture feature with volume local binary pattern was studied. Using dynamic texture features to detect smoke was an effective method, especially for clear white smoke plume with an extremely high DR. Diffusion and thinning of smoke reduced the texture discrimination of smoke.

HD cameras have wider view coverage and greater monitoring distance. HD videos provide more details for smoke detection, but the size of smoke in HD videos also become more changeable. Thus the detection method based on blocks was limited by the size of the block, the frame alarm rule, and so on. The threshold of the candidate smoke block greatly affected the smoke DR and FAR. A method based on irregular regions to extract the dynamic texture feature from motion areas directly was used, which ensured a relatively high DR and greatly reduced the FAR. This method also reduced the feature extraction time and improved the robustness of the detection algorithm.

In dynamic texture feature extraction, the frame interval  $L$ , number of sample points  $P$ , radius  $R$  and selected operator had effects on detection performance. Generally, because the non-smoke samples had various texture distributions, an operator with high feature dimension (achieved by increasing the number of sample points or increasing the complexity of the method) was more conducive to excluding non-smoke samples, thereby reducing the FAR. However, in this experiment, it is found that the same operations were not necessarily conducive to the improvement of DR. Additionally, the feature extraction computing time of an operator was not directly related to the size of the vector dimension. In fact, the Original and U2 modes were more suitable for real-time smoke detection in most cases.

## References

- [1] Wang, Yang, and G. Mori. "Human Action Recognition by Semilattent Topic Models," *IEEE Transactions on Pattern Analysis & Machine Intelligence* 31.10, 1762-74, 2009. [Article \(CrossRef Link\)](#)
- [2] Fu, Y., Hospedales, T., Xiang, T., & Gong, S. "Attribute learning for understanding unstructured social activity," *European Conference on Computer Vision* Springer-Verlag, 530-543, 2012. [Article \(CrossRef Link\)](#)
- [3] Fu, Y., Hospedales, T. M., Xiang, T., & Gong, S. "Learning multimodal latent attributes," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 36(2), 303-316, 2014. [Article \(CrossRef Link\)](#)
- [4] Cetin, Enis A, Günay, et al. "Video fire detection – Review," *Digital Signal Processing*, 23(6):1827-1843, 2013. [Article \(CrossRef Link\)](#)
- [5] A. Enis Cetin, Bart Merci, Osman Günay, Behçet Uğur Töreyn and Steven Verstockt, *Methods and Techniques for Fire Detection*, 39-42, 2016.
- [6] Töreyn, B. Ugur, Yigithan Dedeoglu, and A. Enis Cetin. "Contour based smoke detection in video using wavelets," in *Proc. of Signal Processing Conference, 2006 14th European*. IEEE, 2006. [Article \(CrossRef Link\)](#)

- [7] Chen, Thou-Ho, Ping-Hsueh Wu, and Yung-Chuen Chiou. "An early fire-detection method based on image processing," in *Proc. of Image Processing, 2004. ICIP'04. 2004 International Conference on*. Vol. 3. IEEE, 2004. [Article \(CrossRef Link\)](#)
- [8] Genovese, A., Labati, R. D., Piuri, V., & Scotti, F. "Wildfire smoke detection using computational intelligence techniques," in *Proc. of Computational Intelligence for Measurement Systems and Applications (CIMSA), 2011 IEEE International Conference on*. IEEE, 2011. [Article \(CrossRef Link\)](#)
- [9] Yuan, Feiniu. "A fast accumulative motion orientation model based on integral image for video smoke detection," *Pattern Recognition Letters* 29.7, 925-932, 2008. [Article \(CrossRef Link\)](#)
- [10] Chunyu, Y., Jun, F., Jinjun, W., & Yongming, Z. "Video fire smoke detection using motion and color features," *Fire technology* 46.3, 651-663, 2010. [Article \(CrossRef Link\)](#)
- [11] Jia, Y., Yuan, J., Wang, J., Fang, J., Zhang, Q., & Zhang, Y. "A saliency-based method for early smoke detection in video sequences," *Fire Technology* 5.52, 1271-1292, 2016. [Article \(CrossRef Link\)](#)
- [12] Ojala, Timo, Matti Pietikainen, and Topi Maenpaa. "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on pattern analysis and machine intelligence* 24.7, 971-987, 2002. [Article \(CrossRef Link\)](#)
- [13] Yuan, Feiniu. "Video-based smoke detection with histogram sequence of LBP and LBPV pyramids," *Fire safety journal* 46.3, 132-139, 2011. [Article \(CrossRef Link\)](#)
- [14] 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 Transactions on Internet and Information Systems (TIIS)* 10.4, 1807-1823, 2016. [Article \(CrossRef Link\)](#)
- [15] Tian, H., Li, W., Ogunbona, P., Nguyen, D. T., & Zhan, C. "Smoke detection in videos using non-redundant local binary pattern-based features," in *Proc. of Multimedia Signal Processing (MMSP), 2011 IEEE 13th International Workshop on*. IEEE, 2011. [Article \(CrossRef Link\)](#)
- [16] Zhao, Guoying, and Matti Pietikainen. "Dynamic texture recognition using local binary patterns with an application to facial expressions," *IEEE transactions on pattern analysis and machine intelligence* 29.6, 915-928, 2007. [Article \(CrossRef Link\)](#)
- [17] Zhao, Guoying, and Matti Pietikainen. "Local binary pattern descriptors for dynamic texture recognition," in *Proc. of Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*. Vol. 2. IEEE, 2006. [Article \(CrossRef Link\)](#)
- [18] Chen, Junzhou, Yong You, and Qiang Peng. "Dynamic analysis for video based smoke detection," *International Journal of Computer Science Issues* 10.2, 298-304, 2013.
- [19] Günay, Osman, and A. Enis Çetin. "Real-time dynamic texture recognition using random sampling and dimension reduction," in *Proc. of Image Processing (ICIP), 2015 IEEE International Conference on*. IEEE, 2015. [Article \(CrossRef Link\)](#)
- [20] <http://signal.ee.bilkent.edu.tr/VisiFire/Demo/SampleClips.html>
- [21] Jiang, Y. G., Wu, Z., Wang, J., Xue, X., & Chang, S. F. "Exploiting feature and class relationships in video categorization with regularized deep neural networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017. [Article \(CrossRef Link\)](#)
- [22] Boser, Bernhard E., Isabelle M. Guyon, and Vladimir N. Vapnik. "A training algorithm for optimal margin classifiers," *Proceedings of the fifth annual workshop on Computational learning theory*, ACM, 1992. [Article \(CrossRef Link\)](#)
- [23] Cortes, Corinna, and Vladimir Vapnik. "Support-vector networks," *Machine learning* 20.3, 273-297, 1995. [Article \(CrossRef Link\)](#)
- [24] Li H. D., Yuan F. N. "Image based smoke detection using pyramid texture and edge features," *Journal of Image and Graphics*, 20(6):0772-0780, 2015. [Article \(CrossRef Link\)](#)



**Gaohua Lin** received his B.Eng. degree in safety engineering from University of Science and Technology of China, Hefei, in 2009. He is currently a Ph.D. candidate in the State Key Laboratory of Fire Science, University of Science and Technology of China. His research interests include fire detection principles and techniques, image processing and pattern recognition.



**Yongming Zhang** received his B.S., M.S. and Ph.D. degrees in the Department of Modern Physics Nuclear Electronics & Information Processing at the University of Science and Technology of China, where he served as a teacher from 1983 through 2000 and a professor from 2001 through 2004. He was a visiting scholar at the Italian National Institute of Nuclear Physics from 1993 through 1995. He is currently a professor in the State Key Laboratory of Fire Science, University of Science and Technology of China. His research interests include fire detection principles and techniques, special environment fire detection, physical information detection and processing technology, nuclear detection.



**Qixing Zhang** received his B.Eng. and Ph.D. degree in safety engineering from University of Science and Technology of China, Hefei, in 2006 and 2011 respectively. He joined the 38th Research Institute of China Electronic Technology Group Corp. after graduation. Currently, he is an associate professor in State Key Laboratory of Fire Science, University of Science and Technology of China. His research interests include video smoke detection, signal processing, fire detection in harsh environments, and optical properties of smoke.



**Yang Jia** received the B.Eng. and M.E. degrees in Signal and Information Processing from the Xi'an University of Architecture and Technology, Xi'an, China, in 2010 and 2013, respectively, and the Ph.D. degree in safety engineering from the University of Science and Technology of China, Hefei, in 2016. She is currently a lecturer in Xi'an University of Posts and Telecommunications. Her research interests include image processing and pattern recognition.



**Gao Xu** received his B.Eng. degree in safety engineering from China University of Geosciences, Wuhan, in 2015. He is currently a Ph.D. candidate in the State Key Laboratory of Fire Science, University of Science and Technology of China. His research interests include intelligent surveillance systems, computer vision and deep learning.



**Jinjun Wang** received his B.Eng. and M.S. degree in safety engineering from University of Science and Technology of China, where he served as a teaching assistant from 1999 through 2005. He is currently an engineer in State Key Laboratory of Fire Science, University of Science and Technology of China. His research interests include fire detection, improvements of optical smoke detector.