Fire Safety Journal 73 (2015) 91–101

Contents lists available at ScienceDirect

Fire Safety Journal

journal homepage: www.elsevier.com/locate/firesaf

Dynamic texture based smoke detection using Surfacelet transform and HMT model

Wei Ye [a], Jianhui Zhao [a,c,*], Song Wang [b], Yong Wang [a], Dengyi Zhang [a,c], Zhiyong Yuan [a]

[a] Computer School, Wuhan University, Wuhan, Hubei, China
[b] Department of Computer Science and Engineering, University of South Carolina, USA
[c] Suzhou Institute of Wuhan University, Suzhou, Jiangsu, China

ARTICLE INFO

Article history:
Received 11 March 2014
Received in revised form
7 October 2014
Accepted 1 March 2015
Available online 12 March 2015

Keywords:
Surfacelet transform
HMT model
Dynamic texture descriptor
Smoke detection

ABSTRACT

To detect smoke regions from video clips, a novel dynamic texture descriptor is proposed with Surfacelet transform and hidden Markov tree (HTM) model. The image sequence is multi-scale decomposed by a pyramid model, and the signals are decomposed to different directions using 3D directional filter banks. Then a 3D HMT model is built for obtained coefficients from Surfacelet transform with both Gaussian mixture model and scale continuity model. Parameters of the HMT model are estimated through expectation maximization algorithm, and the joint probability density is determined as the dynamic texture feature value. Support vector machine (SVM) classifier is trained with samples including smoke and non-smoke videos. For input image sequence, the joint probability density of each divided unit 3D block is taken as the input of SVM to decide whether there is smoke. The new dynamic texture descriptor takes image sequence as a multidimensional volumetric data, i.e., considering both spatial and temporal information of coefficients into one model. In experiments, existing texture descriptors of gray level co-occurrence matrix (GLCM), local binary pattern (LBP) and Wavelet are implemented and used for comparison. Results from many real smoke videos have proved that the new dynamic texture descriptor can obtain higher detection accuracy.

1. Introduction

Since fire can easily bring huge losses to economics and lives [1], vision based automatic fire recognition becomes very important to help alarm and prevent fire disaster. Smoke is inherently related with fire, i.e., smoke usually exists before and during fire. In most cases, smoke spreads faster and will occur much faster in the field of view of the cameras [2]. Therefore, smoke extraction and analysis by computer vision based techniques are very valuable, and prompt identification of smoke plays a vital role in quick fire detection. However, video smoke detection still has great technical challenges [3], mainly due to the following reasons: variability in smoke density, lighting, diverse background, interfering non-rigid objects etc.; none of the primitive image features such as intensity, motion, edge and obscuration characterizes smoke well; visual pattern of smoke is difficult to model.

Until now there are already some methods for smoke detection, during which related features are adopted, such as color, shape, motion, transparency, texture, etc. Yu et al. [4] proposed a smoke detection method using both color and motion features. Background estimation and color based decision are used to determine candidate smoke regions. Lucas Kanade algorithm is used to calculate the optical flow of candidate regions. Then a back propagation neural network is used to classify the smoke features. Therefore, the results depend too much on the selected statistical values for training. Toreyin et al. [5] proposed a contour based method for smoke detection by Wavelets from video captured with stationary camera. In their approach, periodic behavior in smoke boundaries is analyzed using a hidden Markov model (HMM) mimicking the temporal behavior of smoke, boundary of smoke regions is represented in Wavelet domain, and high frequency nature of the boundaries of smoke regions is also used as a clue to model the smoke flicker. Yuan [6] presented an accumulative motion model based on the integral image by fast estimating the motion orientation of smoke. Since smoke is assumed to usually drift upward, the model is insufficient in some specialized cases. Later Yuan [7] proposed one double mapping framework to extract partition based features with AdaBoost. The first mapping is from an original image to block features, represented by concatenating histograms of edge orientation, edge magnitude and

* Corresponding author at: Computer School, Wuhan University, Wuhan, Hubei, China.
E-mail address: jianhuizhao@whu.edu.cn (J. Zhao).

http://dx.doi.org/10.1016/j.firesaf.2015.03.001
0379-7112/

LBP bit, and densities of edge magnitude, LBP bit, color intensity and saturation. The second mapping is from the block features to statistical features. The statistical features, such as mean, variance, skewness, kurtosis, are computed on all partitions to form a feature pool. AdaBoost is then used to select the discriminative shape invariant features of smoke. To detect smoke in a tunnel environment, Han and Lee [8] presented motion-candidate region detection and a false region elimination method for fast and robust smoke detection. Ho [9] provided a machine vision-based method for real-time early flame and smoke detection. It uses the motion history detection algorithm to register the possible flame and smoke position in a video, then analyze the spectral (color histogram model), spatial (relation of perimeter and area) and temporal (flickering area) characteristics of regions in image sequences, and then the continuously adaptive mean shift (CAMSHIFT) tracking algorithm is employed to provide real-time position of flame and smoke. The method needs to be further improved for increasing the reliability in complex environments, and tracking multiple fire regions concurrently. Tung and Kim [10] proposed a four-stage smoke detection algorithm. First, an approximate median method is used to segment moving regions. Second, fuzzy c-means (FCM) method is used to cluster candidate smoke regions from these moving regions. Third, a set of parameters are extracted from both spatial and temporal characteristics of the candidate smoke regions. Fourth, the extracted parameters are used as input feature vectors to train a SVM classifier. Long et al. [11] provided a novel approach to detect smoke using transmission. The preliminary smoke transmission is estimated using dark channel prior and then the result is refined through soft matting algorithm. According to the calculated transmission value, smoke pixels can be detected accurately and the detailed information about thickness distribution of smoke region can also be provided directly. However, the estimation method of smoke transmission is preliminarily based on dark channel prior, so it is limited in detecting gray–white smoke. Tian et al. [12] presented an image separation approach inspired by the transparency property of smoke. A mathematical image formation model that linearly blends an amount of smoke and background image forms the basis of the approach. Using several constraints, viz. sparse representation and local smoothness, the blending parameter is estimated and the smoke component is separated. However, there are still rooms for improvement including: considering constraints based on the nature of smoke to achieve better separation in the presence of high level of noise, understanding how and in what way sparse coefficient vector encodes the characteristics of smoke, investigating more features for smoke to obtain better characterization.

For texture based smoke detection, texture descriptors of GLCM, LBP, Wavelet and SVM classifier are employed popularly. Calderara et al. [13] presented a smoke detection system which uses a common CCD camera sensor. A background model is firstly proposed to extract moving regions. Then a Bayesian approach is adopted to detect smoke regions in the scene analyzing image energy by means of Wavelet transform coefficients and color information. Kim et al. [14] detected moving regions through an approximate median background subtraction method using the diffusion behavior of smoke. Then the moving regions are segmented by FCM algorithm based on color features. An object tracking approach is employed to detect candidate smoke objects, and then parameters of image texture are extracted using GLCM. Finally, the neural network is used to discriminate smoke and non-smoke objects. Lee et al. [15] developed a smoke detection approach based on a block processing technique. Motion features are used to extract the candidate regions. Texture and color features of the candidate regions are then analyzed in their spatial, temporal, and spatial-temporal domains before all the features are further combined using a SVM classifier. Further improvements are required to deal with existing critical problems, e.g., light reflections from wet ground, and continuous adjustments of the exposure value by camera. Toreyin et al. [16] proposed a Wavelet based algorithm for real-time smoke detection in video, assuming that the camera of scene monitoring is stationary. It is mainly based on determining the edge regions whose Wavelet subband energies decrease with time. These regions are then analyzed along with their corresponding background regions with respect to their RGB and chrominance values. The flicker of the smoke and convexity of smoke regions are also set as clues for the final decision. Ferrari et al. [17] provided a real-time detection approach of steam in video images. The assumption is that the presence of steam acts as a blurring process, which changes the local texture pattern of an image while reducing the amount of details. A statistical HMT model derived from the coefficients of the dual-tree complex Wavelet transform (DT-CWT) is used to characterize the steam texture pattern. The parameters of HMT model are used as an input feature vector to a specially tailored SVM. To improve the algorithm, further study is needed for changes in the chrominance values of pixels, information about the dynamics of steam, and evaluation of the degree of opacity. Maruta et al. [18] presented a method of smoke detection in open areas. Moving objects are detected from gray-scale image sequences, and the noise is removed with image binarization and morphological operation. Then the smoke feature is extracted with texture analysis, and taken as time series data. To estimate the area information of smoke, a more accurate method is needed to detect the smoke region. Yu et al. [19] used texture analysis for real-time fire smoke detection based on GLCM, and smoke features can be distinguished from other kinds of none fire disturbances. But there does exist some false alarms, because that the fire monitoring scenes have complicate background. Gubbi et al. [20] proposed a method for smoke detection using Wavelets and SVM. Characterization of smoke is carried out by extracting Wavelet features from approximate coefficients and three levels of detailed coefficients. Maruta et al. [21] presented a smoke detection method based on texture analysis and SVM. Firstly, moving objects are extracted as candidate smoke regions. Secondly, texture analysis is used to extract feature vectors of images. Then the moving regions are classified as smoke or non-smoke by SVM with texture features as input. Since extraction of moving objects is sometimes easily affected by environmental conditions, the result of SVM classification is accumulated with time. Gonzalez et al. [22] proposed a method to detect smoke from outdoor forest video sequences. The first step is an image preprocessing block applying a bi-cubic interpolation algorithm. The second step is a smoke detection algorithm performing a stationary Wavelet transform (SWT) to remove high frequencies on horizontal, vertical, and diagonal details. The final step is a smoke verification algorithm determining whether the ROI is increasing its area or not. Yuan [23] presented a smoke detection method using a histogram sequence of pyramids. Firstly, a 3-level image pyramid is constructed by multi-scale analysis. Secondly, LBP are extracted at each level of the image pyramid to generate an LBP pyramid. Thirdly, LBP based on variance (LBPV) with the same patterns are adopted to produce an LBPV pyramid. Fourthly, the histograms of LBP and LBPV pyramids are computed and concatenated into an enhanced feature vector. A neural network classifier is then trained and used for smoke discrimination. Of course, the system performance will drop obviously if a video contains too many objects which are not included in the training set.

Based on the above references, main features adopted in the existing methods are summarized in Table 1. It can be found that texture is the mostly used main feature, and motion is the less used. We can also find that none of the 20 methods [4–23] uses

Table 1
Main features used for vision based smoke detection.

Main features	References	Numbers
Color	[4,9,13–15]	5
Shape	[5, 7, 9]	3
Motion	[4,6,8–10,14,15]	7
Transparency	[11,12]	2
Texture	[13–23]	11

only one feature. In fact, each method tries the combination of different features and different classifiers. As for the popularly utilized features of texture and motion, although texture correlation among consecutive frames can be analyzed from image sequence, it is still not enough to describe the inherent spatial and temporal properties of smoke.

Comparatively, dynamic texture is an extension of ordinary texture to the temporal domain, and is suitable for sequences of images of moving scenes that exhibit certain stationary properties in time [24]. There are a lot of dynamic textures in the real world, such as sea waves, smoke, flame and whirlwind. The existing approaches for dynamic texture mainly include four types [25,26]: methods of optical flow, methods of spatiotemporal geometric properties, methods of local spatiotemporal filtering, methods of model parameters estimation. Dynamic texture is one approach for 3D data description, while Surfacelet transform [27,28] is multi-resolution transform for efficient representation of multi-dimensions, which makes it suitable to deal with dynamic texture as a model based method. Lu and Do [27] proposed a new family of directional filter banks for arbitrary N-dimensional signals (NDFB). NDFB is built upon an efficient tree-structured construction, which leads to a low redundancy ratio and refinable angular resolution. By combining NDFB with multi-scale pyramid, Surfacelet transform is constructed. NDFB and Surfacelet transform have applications in various areas that involve the processing of multidimensional volumetric data, including video processing, medical image analysis, etc. Based on fewer Surfacelets rather than image pixels, Huang et al. [28] presented a strategy of inverse Surfacelet transform to recover images with constraints, which is more suitable for efficient reconstruction of 3D images utilizing prior knowledge. After Surfacelet transform, the numerous coefficients of 3D signals are obtained. Thus they are not suitable for direct analysis, and need to be properly represented for further processing. To model the Wavelet coefficients and the persistence of large/small coefficients across scale, 2D HMT was employed by Choi and Baraniuk [29] for multi-scale image segmentation. The energy compaction property of Wavelet transform implies that the transform of most real-world images consists of a small number of large coefficients and a large number of small coefficients. Thus the population of large coefficients is considered as outcomes of a probability density function (PDF) with a large variance, while the collection of small coefficients is considered as outcomes of a PDF with a small variance. Then the PDF of each Wavelet coefficient is approximated by a two-density Gaussian mixture model (GMM). Helfroush and Taghdir [30] provided methods for image segmentation based on HMT in Wavelet domain. Contourlet coefficients are first computed, and then a HMT model is trained for each texture. For the test phase, decision is made for each block of input image based on likelihood criterion.

Based on Surfacelet transform and 3D HMT model, a novel texture descriptor is proposed for vision based smoke detection. The new descriptor performs better than the existing GLCM, LBP, Wavelet, and can be applied through combination of other features and classifiers. The rest of our paper is organized as follows: Surfacelet transform for image sequence is presented in Section 2, 3D HMT model and parameter estimation are proposed in Section 3, new texture descriptor based smoke detection is provided in Section 4, experimental results are shown and analyzed in Section 5, and then the conclusion is given in Section 6.

2. Surfacelet transform for image sequence

Surfacelet transform [27] is composed of 2 parts, the multi-direction decomposition by 3D directional filter banks (3D-DFB) and the multi-scale decomposition, thus it can capture the singularity changes of signal, such as the varying smoke in video. 3D-DFB decomposes signal into different directional sub-bands, while the multi-scale decomposition refers to a tower structure generating high frequency and low frequency parts from signal. Suppose there is an image sequence of one smoke video clip, having 64 frames with the resolution of 272×480 pixels, as illustrated in Fig. 1. After decomposition of Surfacelet transform, there are 2 layers of high frequency components (e.g., 2 times down-sampling between neighbor layers) with 3 dominant directions in every layer. Having 4 different directions in each dominant direction, there are 12 sub-bands generated for the high frequency signal of one layer. Each sub-band captures the singularity changes of signal in certain scale and direction, which may correspond to texture or boundary of smoke region.

2.1. Multi-scale decomposition

Multi-scale decomposition (illustrated in Fig. 2) of frequency is implemented by one pyramid shaped model with the following steps:

Step 1. Decompose the image sequence through a high pass filter $H()$ and a low-pass filter $L()$ to obtain high frequency and low frequency components of the signal, while high frequency signal may represent smoke texture or boundary in video clip.
Step 2. Determine whether the signal has been decomposed to a user predefined layer (usually 2 or 3 layers are used), if so terminate the decomposition, if not go to Step 3.
Step 3. Perform two times up-sampling $U(2)$ on the low frequency component of signal.
Step 4. Cancel the aliasing caused by the up-sampling operation through an anti-aliasing filter $S()$.
Step 5. Perform three times down-sampling $D(3)$ to obtain the 1.5 times down-sampling (up-sampling by 2 followed by down-sampling by 3).
Step 6. Take the obtained signal from Step 5 as new signal, and then go to Step 1.

2.2. Multi-direction decomposition

For the high frequency component of signal in each layer from multi-scale decomposition, suppose n_1, n_2 and n_3 are 3 principal signal axes orthogonal to each other, multi-direction decomposition (illustrated in Fig. 3) is realized by 3D-DFBs with the following steps:

Step 1. Decompose each 3D high-frequency signal by a 2D-DFB along n_1n_3 plane to obtain the taper sub-bands, as shown in the left of Fig. 3.
Step 2. Decompose the 3D high-frequency signal by another 2D-DFB along n_1n_2 plane to obtain the taper sub-bands, as shown in the middle of Fig. 3.
Step 3. Intersect above taper sub-bands to generate the sub-bands with hourglass shapes along the n_1 axis, i.e., constructing one 3D-DFB with two 2D-DFBs, as shown in the right of Fig. 3.

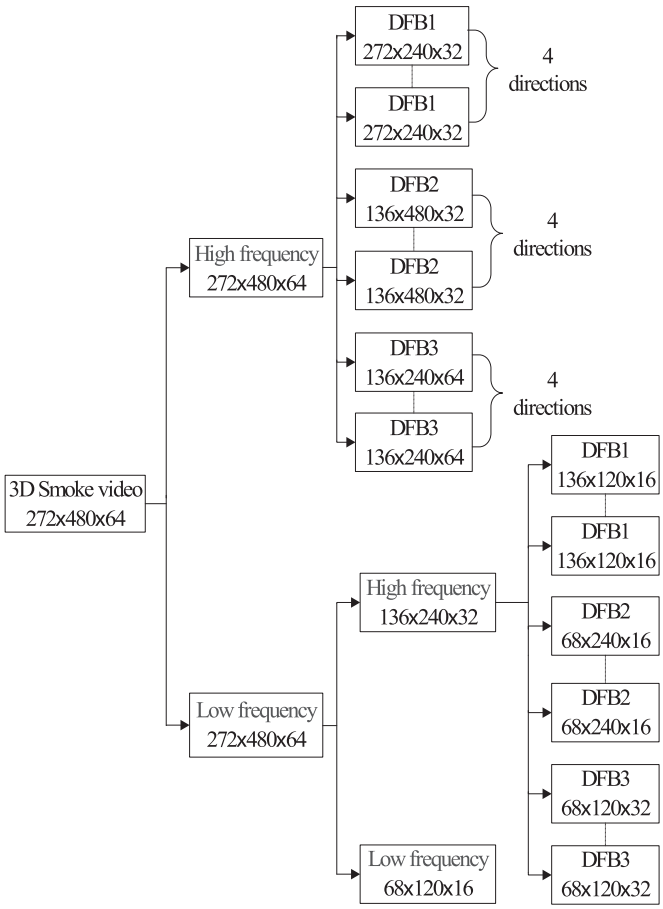

Fig. 1. Surfacelet transform of image sequence.

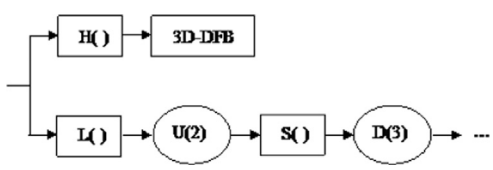

Fig. 2. The multi-scale decomposition.

Step 4. Obtain hourglass shaped sub-bands along the n_2 axis and the n_3 axis in the same way.
Step 5. Take the pixels in decomposed sub-bands of different scales and directions as the coefficients from Surfacelet transformation.

The number of hourglass shaped sub-bands can be controlled by parameter k in a dominant direction, and $k = 2$ for the situation of Fig. 3. Therefore, there are $2^2 = 4$ sub-bands in one of n_1, n_2 and n_3 directions, while there are $3_*2^2 = 12$ sub-bands for all the 3 dominant directions.

3. 3D HMT model and parameter estimation

When GLCM, LBP and Wavelet are used for 2D image or block, features of them are calculated to describe the texture of related pixels. Similarly, for the pixels in decomposed sub-bands of 3D video clip after Surfacelet transformation, feature value has to be computed to represent the dynamic texture. There are huge numbers of coefficients (pixels) after Surfacelet transform, and it is difficult to analyze them directly. The obtained coefficients belong to various scales and directions, while there are hierarchical relations between neighbor scales along the same direction. Fortunately, HMT model has the ability to capture all inter-scale, inter-direction and inter-location dependencies. Therefore, we extend 2D HMT for image [29,30] to 3D HMT for video, and construct it for coefficients from Surfacelet transformation. Then parameters of 3D HMT are estimated and used to obtain the corresponding dynamic texture feature.

3.1. 3D HMT modeling

In HMT, each coefficient is represented with a hidden state of one node. It turns the unknown distribution problem of coefficients into a deterministic problem of hidden states. In other words, it uses the Markov chain to model the hidden states of coefficients, but not model the coefficients directly. Distribution of coefficients is determined after the distribution of their hidden states is determined. Since the coefficients from Surfacelet transformation are 3D data, a 3D HMT is adopted to represent them. As illustrated in Fig.4, neighbor scales of the same direction have hierarchical relation, which is described by an octree. For one sub-band of 272x240 × 32 in direction DBF1 and layer 1 of Fig. 1, it has neighbor sub-bands of 136x120 × 16 in DBF1 and layer 2, i.e., 1 parent node corresponds to 8 child nodes. To build the 3D HMT model, characteristics of coefficients have to be properly described, and they are: energy compact support property in each sub-band, hidden Markov property between adjacent scales in the same direction.

(1) GMM distribution modeling
For each sub-band, energy is concentrated in only limited number of coefficients. Thus all coefficients from one sub-band have the shape of sharp peak with heavy tails in

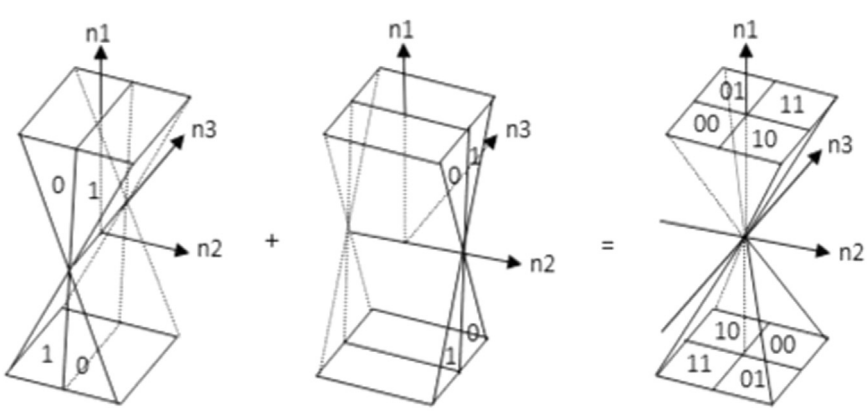

Fig. 3. The multi-direction decomposition.

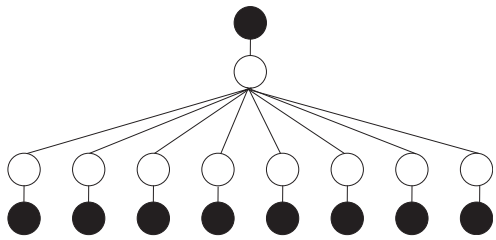

Fig. 4. Neighboring nodes in 3D HMT.

statistical distribution. The distribution is energy compact support, and is suitable to be modeled using Gaussian mixture model (GMM). For our work, there are 2 single Gaussian distribution functions in GMM.

Each coefficient is assigned with a hidden state variable $S_i \in \{0, 1\}$, and the value of S_i indicates which Gaussian function the current coefficient x_i belongs to. State 0 corresponds to the Gauss function with 0 mean and small variance, while state 1 corresponds to the Gauss function with 0 mean and large variance. The single Gauss distribution can be expressed as

$$g(x_i;\, \mu,\, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right) \tag{1}$$

where μ is the mean of coefficients in the sub-band of x_i; σ^2 is the variance of them.

The Gauss probability density functions of state 0 and 1 can be expressed as

$$f(x_i|S_i = 0) = g\left(x_i;\, 0,\, \sigma_0^2\right) \tag{2}$$

$$f(x_i|S_i = 1) = g\left(x_i;\, 0,\, \sigma_1^2\right) \tag{3}$$

Thus the Gauss probability density function for coefficient x_i is

$$f(x_i) = \sum_{m=0,1} P_{S_i}(m) f(x_i|S_i = m) \tag{4}$$

where $P_{S_i}(m)$ is the probability function for state value m, and $P_{S_i}(0) + P_{S_i}(1) = 1$.

(2) Scale continuity modeling

The continuity among coefficients of different scales in the same direction captures the persistence across scales, which is modeled with the following steps:

Step 1. Construct an octree for the coefficients (white circles in Fig. 4) from sub-bands of different scales in one direction.

Step 2. Connect the hidden state of each coefficient with corresponding hidden states (black circles in Fig. 4) of its 8 child nodes, and obtain the hidden Markov probability tree.

Step 3. Suppose the state of coefficient x_i is only determined by the state of its parent node p_i, the state transfer matrix is defined by

$$\begin{bmatrix} \varepsilon_{x_i,0}^{p_i,0} & \varepsilon_{x_i,1}^{p_i,0} \\ \varepsilon_{x_i,0}^{p_i,1} & \varepsilon_{x_i,1}^{p_i,1} \end{bmatrix} \text{or} \begin{bmatrix} \varepsilon_{x_i,0}^{p_i,0} & 1-\varepsilon_{x_i,0}^{p_i,0} \\ 1-\varepsilon_{x_i,1}^{p_i,1} & \varepsilon_{x_i,1}^{p_i,1} \end{bmatrix} \tag{5}$$

where $\varepsilon_{x_i,m_x}^{p_i,m_p}$ is the state transition probability, i.e., when the hidden state of p_i is m_p, the hidden state of x_i is m_x, while m_p and m_x have the state value of 0 or 1.

Step 4. Obtain the parameter vector from the octree for coefficients of different scales in one direction by

$$\Theta = \left\{P_{S_0}(m),\, \varepsilon_{x_i,m_x}^{p_i,m_p},\, \sigma_{j,m}^2\right\} \tag{6}$$

where $P_{S_0}(m)$ is the state probability of the root node x_0; $\sigma_{j,m}^2$ is the variance of state m in the sub-band of the jth scale.

If the decomposed sub-bands have N directions, the 3D HMT model for all coefficients from Surfacelet transformation is expressed as

$$M = \{\Theta_1,\, \Theta_2,\, \cdots \Theta_N\} \tag{7}$$

3.2. Estimation of HMT parameters

After construction of the 3D HMT model, all parameters are initialized: the mean of each sub-band is 0, the variance is the mean square of coefficients in sub-band, while the state transfer matrix is uniform distribution and each element is set as 0.5. Then the parameters of HMT need to be trained to match a set of training data. Once trained, HMT model can provide a close approximation to the coefficients of sample data from Surfacelet transformation. For sample based learning, the iterative expectation maximization (EM) algorithm is adopted to find the optimal set of model parameters for given set of training data. The 3D HMT parameters are optimized with EM algorithm by the following steps:

Step 1. Initialize the parameter vector Θ^l for the coefficients in one direction, i.e., set the iteration number as $l = 0$.

Step 2. Calculate the joint posterior probability distribution $P(S_i|x_i,\, \Theta^l)$ based on coefficient x_i and the current parameter vector Θ^l, then parameter vector Θ^{l+1} is expressed through the expectation of logarithmic likelihood function

$$Q(\Theta^{l+1}|\Theta^l) = E_s[\,\ln f(x_i,\, S_i|\Theta^{l+1})|x_i,\, \Theta^l] \tag{8}$$

where f is the probability density function, E_s is the expectation function, and Q is the expectation of parameter vector Θ^{l+1}.

Step 3. Compute new parameter vector Θ^{l+1} by the expectation maximization

$$\Theta^{l+1} = \arg\max\{Q(\Theta^{l+1}|\Theta^l)\} \tag{9}$$

Step 4. Stop if the predefined convergence condition is satisfied, otherwise set $l = l + 1$ and go to Step 2 for iterative execution.

In EM algorithm, the maximum number of iterations is set to 300 and the threshold value for difference between two iterations is set to 1.0E−5, i.e., EM optimization stops when any one of the convergence conditions is satisfied. After training, 3D HMT parameters are obtained for the provided sample data, then the parameters can be used to describe the dynamic texture of related data.

4. Dynamic texture descriptor based smoke detection

Since there are N directions of decomposed sub-bands, all the coefficients from Surfacelet transformation can be represented as $x = \{x_1,\, x_2,\, \cdots x_N\}$. Suppose the sub-bands of different directions are independent with each other, the joint probability density for coefficients is determined by

$$f(x|M) = f(x_1|\Theta_1)f(x_2|\Theta_2)\cdots f(x_N|\Theta_N) \tag{10}$$

With the help of estimated parameters M for 3D HMT from training data, the joint probability density is calculated as the

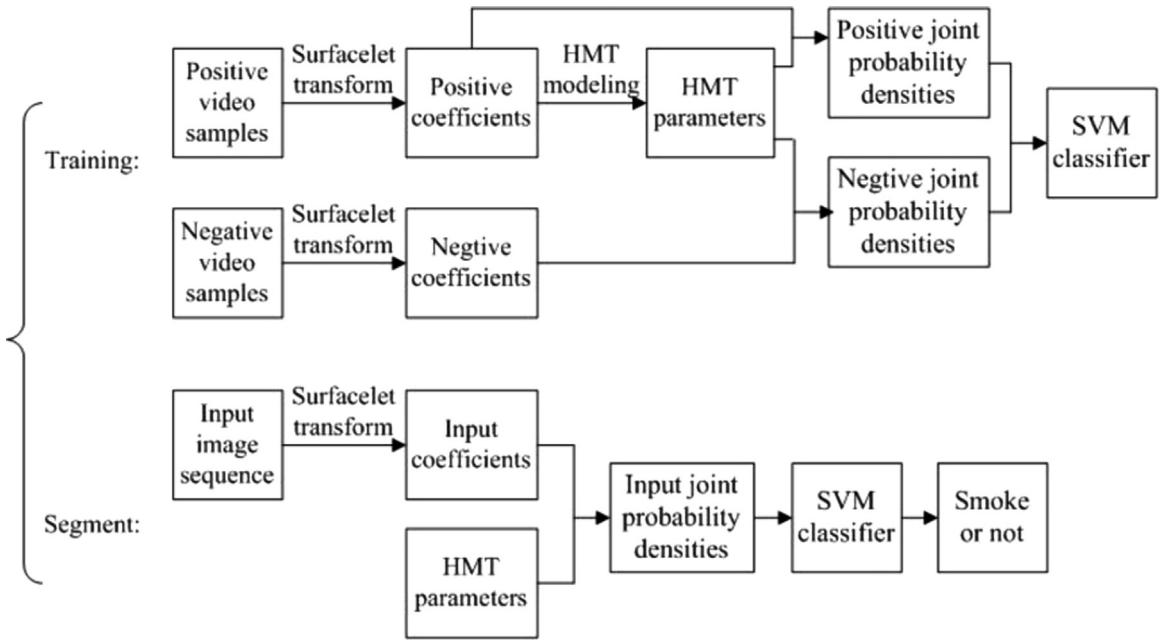

Fig. 5. Application of the dynamic texture descriptor.

texture feature for any image sequence to be recognized. Using aforementioned techniques, the dynamic texture descriptor based smoke detection is proposed with the following steps:

Step 1. Select sample data including sets of smoke (positive) videos and non-smoke (negative) videos, then estimate 3D HMT parameters for positive samples based on Surfacelet transform and HMT modeling.
Step 2. Calculate the positive/negative joint probability densities from positive/negative coefficients and estimated HMT parameters with Eq. (10).
Step 3. Train a SVM classifier with positive and negative joint probability densities of samples.
Step 4. For any input image sequence, decompose the video clip with Surfacelet transform.
Step 5. Compute the joint probability density for the input image sequence with its decomposed coefficients based on Eq. (10).
Step 6. Take the joint probability density of the image sequence as the input of SVM classifier, and decide whether the image sequence is smoke or not.

Application of the dynamic texture descriptor includes two separated parts: training and segmenting, which is illustrated in Fig. 5. The novelties of our algorithm are: (1) the image sequence is taken as a multidimensional volumetric data, not only a collection of image frames; (2) high frequency components of signal are decomposed in different scales and directions from Surfacelet transformation; and (3) obtained coefficients are properly described by 3D HMT including energy compact support property and hidden Markov property.

5. Experimental results and analysis

A novel dynamic texture descriptor is proposed based on Surfacelet transform and HMT model, and it can work similarly as the existing texture descriptors, e.g., GLCM, LBP and Wavelet. Due to the fact that all published smoke recognition approaches combined different features (e.g., color, shape, motion, transparency, and texture) and different classifiers (e.g., SVM, neural network, AdaBoost, and Bayesian), it is unfair to compare the dynamic texture descriptor with these complete smoke detection methods. Therefore, the popularly used texture descriptors of GLCM, LBP and Wavelet are implemented together with SVM (the same classifier as our descriptor), and then compared with the presented new dynamic texture descriptor in smoke detection, which is more reasonable for performance evaluation. Implementations of the related texture descriptors are described as below.

(1) Texture descriptor of GLCM
Based on the existing GLCM methods [14,19], each image frame is divided into $n \times n$ blocks and GLCM is constructed for each block. Then 5 features of GLCM, including energy, contrast, correlation, entropy and autocorrelation, are calculated to characterize the block texture. A feature vector composed by the 5 features is taken as the input of trained SVM classifier, and the block is determined as smoke region or not.
(2) Texture descriptor of LBP
Based on the existing LBP methods [7,23], each image frame is divided into $n \times n$ blocks. Every pixel of one block is compared with its 8 neighbor pixels in the 3×3 local window to obtain a 8 bit binary number, i.e., its LBP. Then a histogram is calculated and normalized to represent the occurrence probability of each LBP value. The histograms of each block are combined into a feature vector, which is taken as the input of trained SVM classifier. Through the smoke decision of each block, smoke areas are segmented from the image.
(3) Texture descriptor of Wavelet
Based on the existing Wavelet methods [13,16,17,20,22], each image frame is divided into $n \times n$ blocks. Then 2D discrete Wavelet transform is performed on each block, and 10 sub-bands are obtained for 3 directions and 3 layers. For coefficients of every sub-band, 6 features are computed including arithmetic mean, geometric mean, standard deviation, skewness, kurtosis, entropy. Thus a feature vector constructed by the total of 60 features is obtained for each block. Taking the feature vector as input of trained SVM classifier, whether there is smoke in related block can be decided.
(4) Our dynamic texture descriptor.

Similarly, the image frame is divided into $n \times n$ blocks. Considering the image sequence along time axis, there are $n \times n$ 3D blocks. Each 3D block is decomposed by pyramid model with 2 layers, and decomposed by 3D-DFBs along 3 axes, then 24 subbands are obtained with different scales and directions. 3D HMT model is constructed for coefficients of Surfacelet transform, then HMT parameters are estimated with EM algorithm, and then positive/negative joint probability densities are used to train the SVM classifier. For one 3D block of video clip to be processed, joint probability density is computed as the dynamic texture feature for all coefficients in various scales and directions after Surfacelet transform. Taking the feature as input of trained SVM classifier, we can determine whether there is smoke in the 3D block.

To test the texture descriptors effectively, some smoke video clips captured in natural scenes with cluttered backgrounds are selected for experiments. There are 80 video clips, among them 40 are used for training while the other half are used for testing. Each video clip has about 200 frames, and the image frame has resolution of 480×272 pixels. To compare the texture descriptors for smoke detection thoroughly, the divided blocks with different

Fig. 6. Texture based smoke detections on the 3 video clips with 32×32 block size, from 1st to 4th row of each video clip: GLCM, LBP, Wavelet, and our dynamic texture descriptor.

Fig. 7. Texture based smoke detections on the 3 video clips with 24×24 block size, from 1st to 4th row of each video clip: GLCM, LBP, Wavelet, our dynamic texture descriptor.

resolutions, 32×32, 24×24 and 16×16, are tried respectively.

As shown in Fig. 6, GLCM, LBP, Wavelet and our dynamic texture descriptor are tested on three video clips with 32×32 block size. From the experiments it can be found that: GLCM method has obvious error detections, LBP method has obvious missing detections, Wavelet method has a few error and missing detections, while our descriptor has the highest detection accuracy.

As shown in Fig. 7, GLCM, LBP, Wavelet and our dynamic texture descriptor are tested on three video clips with 24×24 block size. From the experiments it can be found that four descriptors have the similar performances as 32×32 block size, but they can obtain more precise smoke regions and edges due to the smaller size of block.

As shown in Fig. 8, GLCM, LBP, Wavelet and our dynamic texture descriptor are tested on three video clips with 16×16 block size. From the experiments it can be found that all descriptors have error detections, since the blocks with too small size cannot reflect the texture properties properly. However, our descriptor

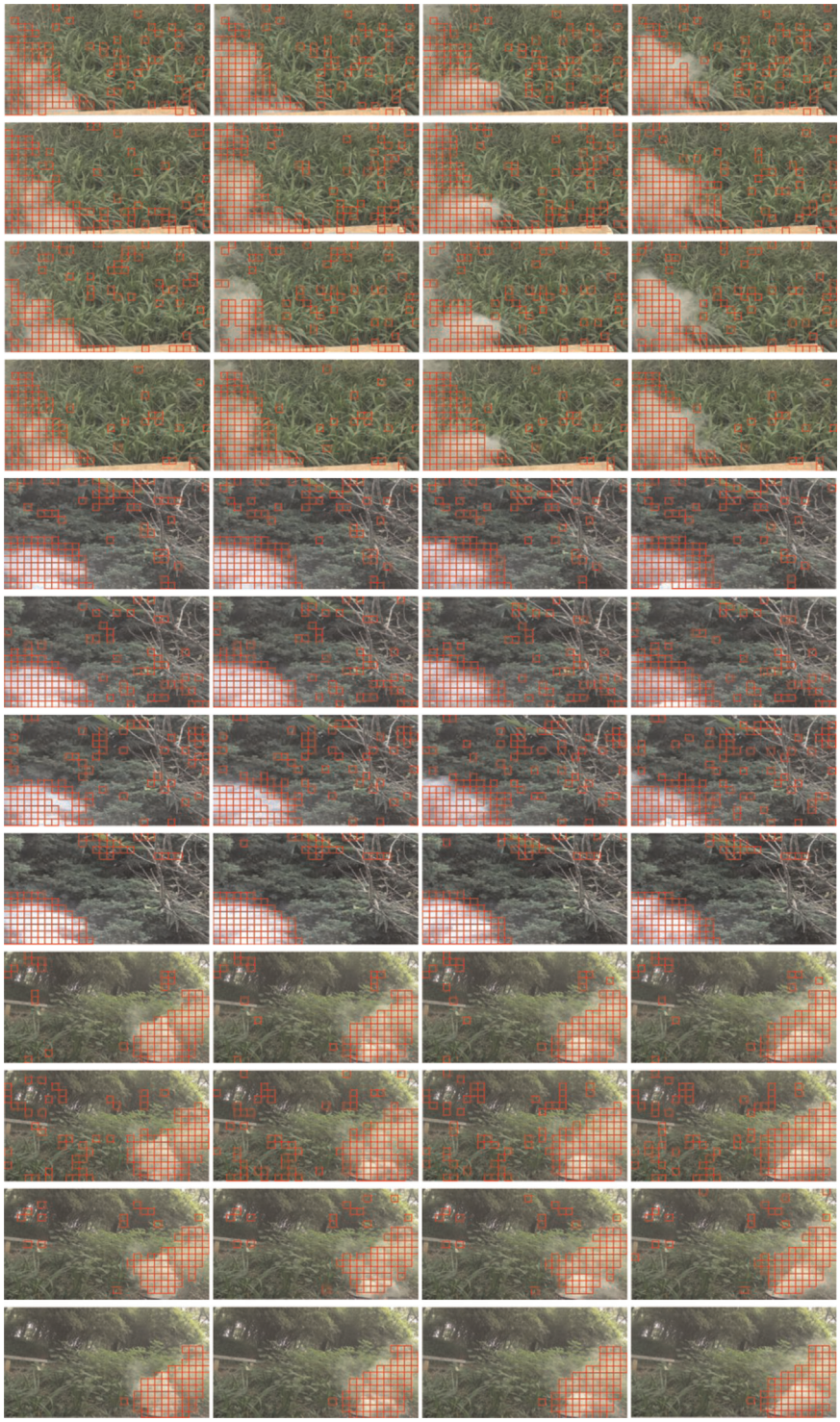

Fig. 8. Texture based smoke detections on the 3 video clips with 16×16 block size, from 1st to 4th row of each video clip: GLCM, LBP, Wavelet, and our dynamic texture descriptor.

still has the best performance even in this situation.

Of course, there is less number of blocks if the size of each block is large, and consequently there is less computing expense. Considering both detecting precision and computational efficiency, the block size of 24×24 is adopted. Then 9 other video clips with more image frames are used to evaluate our dynamic texture descriptor. As shown in Fig. 9, smoke regions are detected effectively from different backgrounds.

6. Conclusion

In this paper, a novel dynamic texture descriptor for smoke detection is proposed with Surfacelet transform and 3D HMT model. Different with the existing texture descriptors, we consider a sequence of image frames as a multidimensional volumetric data. Taking both spatial and temporal information of coefficients into one model, the new texture descriptor is closer to the essential characteristics of dynamic texture. After EM based

Fig. 9. Smoke detections from long video clips with our dynamic texture descriptor.

estimation, HMT model parameters are used to describe the relations among coefficients from Surfacelet transform. Then the joint probability density computed from HMT parameters and coefficients is taken as input texture feature value of trained SVM classifier to judge each divided block.

Experiments are carried out to evaluate the performance of proposed dynamic texture descriptor. Compared with GLCM, LBP and Wavelet methods, our descriptor can obtain smoke detections from different backgrounds with the highest precision, i.e., with the least error detections and missing detections. In real applications, our texture descriptor can be used together with other kinds of features such as color, shape, motion, etc. and other types of classifiers such as neural network, AdaBoost, Bayesian, etc., which may help recognize the smoke areas more accurately from video clips.

The presented dynamic texture descriptor needs computing operations of Surfacelet transform and 3D HMT modeling, thus the lower computational efficiency is its disadvantage. Therefore, parallel computing techniques will be used in the future, such as simultaneously processing the divided blocks of video clip, and HMT modeling the decomposed sub-bands from Surfacelet transform at the same time.

Acknowledgments

This work was supported by Science and Technology Bureau of Suzhou of China (No. SH201115), Natural Science Fund of Hubei province of China, National Basic Research Program of China (973 Program, No. 2011CB707904), Science and Technology Bureau of Wuhan of China (No. 201150124001), Hubei Support Plan for Science & Technology (No. 2014BAA149).

Appendix A. Supplementary materials

Supplementary data associated with this article can be found online at: http://dx.doi.org/10.1016/j.firesaf.2015.03.001.

References

[1] B.C. Ko, J.O. Park, J.Y. Nam, Spatiotemporal bag-of-features for early wildfire smoke detection, Image Vis. Comput. 31 (10) (2013) 786–795.
[2] A.E. Çetin, K. Dimitropoulos, B. Gouverneur, N. Grammalidis, O. Günay, Y. H. Habiboglu, B.U. Toreyin, S. Verstockt, Video fire detection – review, Digit. Signal Process. 23 (6) (2013) 1827–1843.
[3] Z. Xiong, R. Caballero, H. Wang, A.M. Finn, M.A. Lelic, P.Y. Peng., IFPA, Fire Suppression and Detection Research and Applications—A Technical Working Conference (SUPDET) 2007, pp. 1–7.
[4] C. Yu, J. Fang, J. Wang, Y. Zhang, Video fire smoke detection using motion and color features, Fire Technol. 46 (3) (2012) 651–663.
[5] B.U. Toreyin, D. Yigithan, A.E. Cetin, Contour based smoke detection in video using wavelets, in: Proceedings of the European Signal Processing Conference, Ankara, Turkey, 2006, pp. 1–5.
[6] F. Yuan, A fast accumulative motion orientation model based on integral image for video smoke detection, Pattern Recognit. Lett. 29 (7) (2008) 925–932.
[7] F. Yuan, A double mapping framework for extraction of shape-invariant features based on multi-scale partitions with AdaBoost for video smoke detection, Pattern Recognit. 45 (12) (2012) 4326–4336.
[8] D. Han, B. Lee, Flame and smoke detection method for early real-time detection of a tunnel fire, Fire Saf. J. 44 (7) (2009) 951–961.
[9] C.C. Ho, Machine vision-based real-time early flame and smoke detection, Meas. Sci. Technol. 20 (2009) 1–13.
[10] T.X. Tung, J.M. Kim, An effective four stage smoke detection algorithm using video images for early fire-alarm systems, Fire Saf. J. 46 (5) (2011) 276–282.
[11] C. Long, J. Zhao, S. Han, L. Xiong, Z. Yuan, J. Huang, W. Gao, Transmission: A New Feature for Computer Vision Based Smoke Detection, AICI, Berlin, Germany (2010) 389–396.
[12] H. Tian, W. Li, L. Wang, Smoke detection in video: an image separation approach, Int. J. Comput. Vis. 106 (2014) 192–209.
[13] S. Calderara, P. Piccinini, R. Cucchiara, Vision based smoke detection system using image energy and color information, Mach. Vis. Appl. 22 (2011) 705–719.
[14] T. Kim, T. Nguyen, J.M. Kim, Multistage optical smoke detection approach for smoke alarm systems, Opt. Eng. 52 (5) (2013) 1–12.
[15] C.Y. Lee, C.T. Lin, C.T. Hong, M.T. Su, Smoke detection using spatial and temporal analyses, Int. J. Innov. Comput. Inform. Control 8 (7) (2012) 4749–4770.
[16] B.U. Toreyin, D. Yigithan, A.E. Cetin, Wavelet based real-time smoke detection in video, in: Proceedings of the European Signal Processing Conference, Antalya, Turkey, 2005, pp. 1–4.
[17] R.J. Ferrari, H. Zhang, C.R. Kube, Real-time detection of steam in video images, Pattern Recognit. 40 (2007) 1148–1159.
[18] H. Maruta, Y. Kato, A. Nakamura, F. Kurokawa, Smoke detection in open areas using its texture features and time series properties, in: Proceedings of the IEEE International Symposium on Industrial Electronics, Seoul, Korea, 5–8 July 2009, pp. 1904–1908.
[19] C. Yu, Y. Zhang, J. Fang, J. Wang, Texture analysis of smoke for real-time fire detection, in: Proceedings of the 2009s International Workshop on Computer Science and Engineering, vol. 2, Qingdao, China, 28–30 October 2009, pp. 511–515.

[20] J. Gubbi, S. Marusic, M. Palaniswami, Smoke detection in video using wavelets and support vector machines, Fire Saf. J. 44 (8) (2009) 1110–1115.
[21] H. Maruta, A. Nakamura, F. Kurokawa, A new approach for smoke detection with texture analysis and support vector machine, in: Proceedings of the IEEE International Symposium Industrial Electronics, Bari, Italy, 4–7 July 2010, pp. 1550–1555.
[22] R.G. Gonzalez, V.A. Aquino, R.R. Romero, O. Starostenko, J.R. Asomoza, J.M.R. Cortes, Wavelet-based smoke detection in outdoor video sequences, in: Proceedings of the 53rd IEEE International Midwest Symposium on Circuits and Systems, Seattle, WA, USA, 1–4 August 2010, pp. 383–387.
[23] F. Yuan, Video-based smoke detection with histogram sequence of LBP and LBPV pyramids, Fire Saf. J. 46 (3) (2011) 132–139.
[24] S. Soatto, G. Doretto, Y.N. Wu., Dynamic textures, in: Proceedings of the Eighth International Conference on Computer Vision, Vancouver, British Columbia, Canada, 7–14 July 2001, pp. 439–446.
[25] D. Chetverikov, R. Peteri, A brief survey of dynamic texture description and recognition, Comput. Recognit. Syst. Adv. Soft Comput. 30 (2005) 17–26.
[26] G. Zhao, M. Pietikainen, Dynamic texture recognition using local binary patterns with an application to facial expressions, IEEE Trans. Pattern Anal. Mach. Intell. 29 (6) (2007) 915–928.
[27] Y. Lu, M.N. Do, Multidimensional directional filter banks and surfacelets, IEEE Trans. Image Process. 16 (4) (2007) 918–931.
[28] W. Huang, Y. Wang, D.W. Rosen, Inverse surfacelet transform for image reconstruction with prior knowledge, in: Proceedings of the the ASME 2013 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference, Portland, Oregon, USA, 4–7 August, 2013, pp. 1–11.
[29] H. Choi, R.G. Baraniuk, Multiscale image segmentation using wavelet-domain hidden markov models, IEEE Trans. Image Process. 10 (9) (2001) 1309–1321.
[30] M.S. Helfroush, N. Taghdir, HMT contourlet image segmentation based on majority vote, Aust. J. Basic Appl. Sci. 5 (10) (2011) 385–397.