SMOKE DETECTION USING SPATIO-TEMPORAL ANALYSIS, MOTION MODELING AND DYNAMIC TEXTURE RECOGNITION

Panagiotis Barmpoutis, Kosmas Dimitropoulos and Nikos Grammalidis

Information Technologies Institute, ITI - CERTH
1st km Thermi-Panorama Rd, Thessaloniki, 57001 GREECE
email: panbar@iti.gr, dimitrop@iti.gr, ngramm@iti.gr

ABSTRACT

In this paper, we propose a novel method for video-based smoke detection, which aims to discriminate smoke from smoke-colored moving objects by applying spatio-temporal analysis, smoke motion modeling and dynamic texture recognition. Initially, candidate smoke regions in a frame are identified using background subtraction and color analysis based on the HSV model. Subsequently, spatio-temporal smoke modeling consisting of spatial energy analysis and spatio-temporal energy analysis is applied in the candidate regions. In addition, histograms of oriented gradients and optical flows (HOGHOFs) are computed to take into account both appearance and motion information, while dynamic texture recognition is applied in each candidate region using linear dynamical systems and a bag of systems approach. Dynamic score combination by mean value is finally used to determine whether there is smoke or not in each candidate image region. Experimental results presented in the paper show the great potential of the proposed approach.

Index Terms— Smoke detection, histograms of oriented gradients, histograms of oriented optical flow, dynamic textures analysis, spatio-temporal modeling

1. INTRODUCTION

Video surveillance systems are widely used nowadays in a variety of application fields e.g. security, transportation, military applications etc. for detection, tracking, and event recognition. In recent years, automatic video-based smoke detection is a very promising solution for early warning systems, mainly due to the fact that video cameras have the advantage of small response time in contrast to conventional smoke sensors. Moreover, video-based systems are cost-effective solutions especially for the coverage of large areas, while they are integrated easily into existing closed circuit surveillance systems. Research on video-based smoke detection focuses mostly on the reduction of high false alarm rates produced often by a) natural objects, which have similar characteristics with smoke, b) large variations of smoke appearance in videos and c) environmental changes including clouds, shadows, etc. that complicate smoke detection.

More specifically, Gomez-Rodriguez et al. [1] presented a method that uses wavelets and optical flow for smoke detection and monitoring, while in [2] energy computation from wavelet coefficients is introduced. In [3], features of moving target are extracted and a two-layer back propagation (BP) neural network is introduced for smoke prediction. Furthermore, Yuan [4] proposed an accumulative motion model based on an integral image and fast estimation of the motion orientation of smoke. Later, Calderara et al. [5] proposed a smoke detection method based on the analysis of color and texture features of moving objects, which have been previously identified using background subtraction. The temporal behavior of smoke was modeled by a Mixture of Gaussians (MoG) of the energy variation in the wavelet domain. Furthermore, in [6] a method for video-based smoke detection was presented using multi-scale analysis, local binary patterns (LBPs) and local binary patterns based on variance (LBPVs). On the other hand, Avgerinakis et al. [7] proposed an algorithm in which smoke is detected by using temporal HOGHOF descriptors and energy color statistics. More recently, Kim et al. [8] proposed a smoke detection algorithm using GMM and Adaboost for outdoor videos with different weather conditions.

This paper proposes a novel approach combining a) spatio-temporal smoke analysis, b) smoke motion analysis and c) dynamic texture recognition. For spatio-temporal smoke analysis, we propose the combination of two energies (spatial energy and spatio-temporal energy), while for smoke motion analysis, a HOGHOF descriptor and a bag of visual words is used to model the characteristic motion of smoke. On the other hand, dynamic texture recognition is based on linear dynamical systems and a bag of systems approach aiming to further increase the overall robustness of the algorithm. To address the main limitation of the dynamic texture recognition, i.e. the high computation cost, we propose an approach for redundant data reduction, which considers only meaningful information within a specific subsequence of the video. Finally, results of the aforementioned processing steps are combined using dynamic score combination by mean value [9] for the final classification of the candidate image regions.

2. METHODOLOGY

The proposed methodology initially identifies candidate smoke image regions in each frame by applying color analysis and background subtraction and then extracts different features of smoke in each region to distinguish it from smoke-colored moving objects. More specifically, each frame of the video sequence is divided into blocks of $N \times N$ size. In our experiments, N was set equal to 16 as this size has already been used by other researchers in the past [7], [10]. To identify smoke in each candidate region, we apply three processing steps: a) Spatio-temporal smoke analysis, which aims to model both spatial and spatio-temporal energy in each candidate block. This modeling is driven by the fact that smoke regions are usually characterized by low spatial energies. b) Smoke motion analysis, which aims to model the characteristic motion of smoke (usually directed upwards depending on the wind direction). c) Dynamic texture recognition, as smoke textures exhibit certain stationarity properties in time. In the final step, dynamic score combination by mean value is used to determine whether the candidate image region is actually smoke or not.

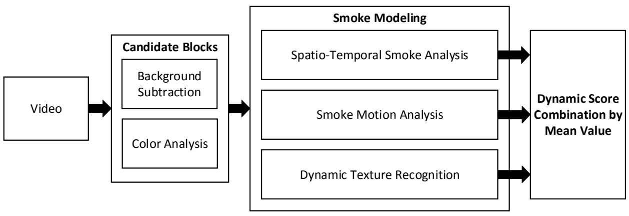


Fig. 1. Methodology of proposed algorithm.

2.1. Identification of Smoke Candidate Blocks

The first step of the proposed approach aims to filter out non fire-colored moving regions. To determine whether individual blocks are part of the background or the foreground we apply an Adaptive Median algorithm [11], which is fast and very efficient. Moreover, to filter out non-smoke colored pixels we use color analysis based on HSV model. The HSV model, defines a color space in terms of three constituent components, namely Hue, Saturation and Value. Given that smoke's color is whitish-blue to white, we can detect smoke colored pixels by thesholding the Saturation and the Value values, which are computed as follows:

$$V = \max(R, G, B)$$

$$S = \begin{cases} \dfrac{V - \min(R, G, B)}{V} & if\ V \neq 0 \\ 0 & otherwise \end{cases} \quad (1)$$

Smoke colored pixels are then detected when:

$$S < Th_1 \text{ and } V > Th_2 \quad (2)$$

where the values of the thresholds Th_1, Th_2 were experimentally determined ($Th_1 = 0.28$, $Th_2 = 108$) using a number of training videos. Candidate blocks are considered only those blocks containing an adequate number of moving and smoke colored pixels. In the proposed method, if at least 10% of the block's pixels are both moving and smoke colored then the block is considered as a candidate smoke region.

2.2. Spatio-Temporal Smoke Analysis

The energy of high spatial frequencies is usually lower in scenes containing smoke than in scenes without smoke. This is due to the fact that smoke introduces a smoothing effect to the scene as gradually coarse image edges become less visible and after some time they may disappear from the scene when smoke becomes thicker. To calculate the energy corresponding to high frequencies for each candidate block, we apply both spatial analysis in the current frame and spatio-temporal analysis in a subsequence of the video to examine the temporal variance of smoke energy. The above features are provided to a SVM classifier to extract the smoke existence probability.

2.2.1 Spatial Analysis

Image regions containing smoke are generally characterized by a smooth appearance, therefore, they exhibit a lower spatial energy than those containing smoke colored objects. To identify spatial energy in the region various techniques can be adopted such as edge detectors, interest points descriptors, etc. In this paper, we used wavelet analysis in order to achieve higher computational efficiency, since it can be implemented without any single multiplication i.e. by simple register shifts. Therefore, a two dimensional wavelet filter is applied and the spatial wavelet energy (corresponding to high frequencies) for each pixel is calculated by following formula:

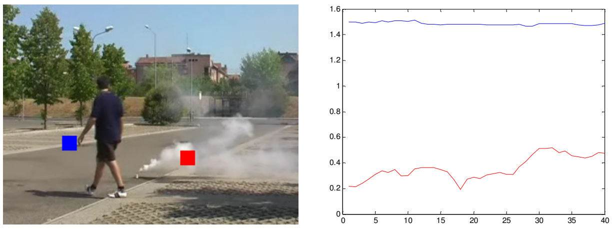
$$\mathrm{E(i,j)} = \mathrm{HL(i,j)}^2 + \mathrm{LH(i,j)}^2 + \mathrm{HH(i,j)}^2 \quad (3)$$

where HL, LH and HH are the high-frequency sub-bands of the wavelet decomposition. For each block, the spatial wavelet energy is estimated as the average of the energy of the pixels in the block.

$$\mathrm{E_{block}} = \frac{1}{\mathrm{N}^2} \sum_{\mathrm{i,j}} \mathrm{E(i,j)} \quad (4)$$

where *NxN* is the size of block.

As an example, the values of the spatial energy, for a candidate block containing smoke and another containing smoke colored object are shown in Fig.2.



(a) (b)

Fig. 2. Variation of spatial energy: i) a block with smoke (red block-line) ii) a block with a smoke colored moving object (blue block-line).

2.2.2 Spatio-Temporal Smoke Analysis

The shape and thickness (i.e. transparency) of smoke change irregularly due to the airflow caused by wind and as a result, smoke causes higher spatial variations within a specific time interval than a smoke colored object. These variations are particularly high at the edges of smoke or when smoke is at an early stage and become lower at the center of smoke (however, they still remain higher than those caused by smoke colored objects, see Fig.3). In contrast to the previous feature, which aims to identify high spatial energies in a single frame, this feature aims to measure the spatio-temporal variations for each block in a sequence of frames.

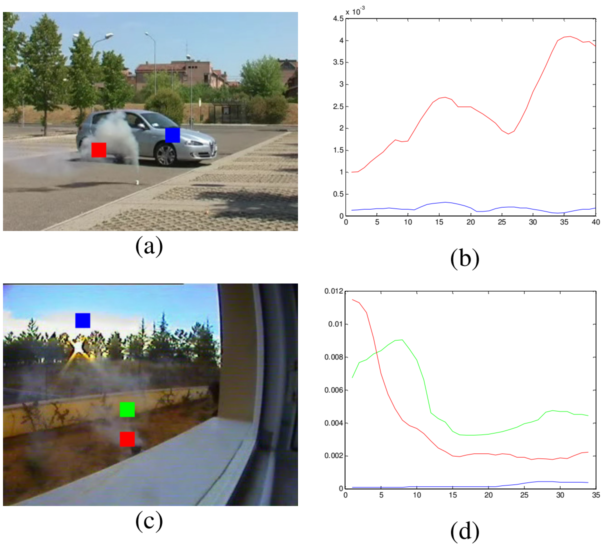


Fig. 3. (a)-(b) Variation of spatio-temporal energy for a block containing i) smoke (red block and line) and ii) a moving smoke colored object i.e. car (blue block and line), (c)-(d) variation of spatio-temporal energy for a block i) at the centre of smoke (red block and line), ii) at the edge of smoke (green block and line) iii) without smoke but with clouds (blue block and line).

The temporal variance of the spatial energy of pixel (i, j) within a temporal window of T last frames is:

$$\mathrm{V}(\mathrm{i},\mathrm{j}) = \frac{1}{\mathrm{T}}\sum_{\mathrm{t}=0}^{\mathrm{T}-1}(\mathrm{E_t}(\mathrm{i},\mathrm{j}) - \overline{\mathrm{E}}(\mathrm{i},\mathrm{j}))^2 \quad (5)$$

where E_t is the spatial energy of the pixel in time instance t and $\bar{E}$ is the average value of this energy. For each block, the total spatio-temporal energy V_{block}, is estimated as the average energy of all pixels in the block:

$$\mathrm{V_{block}} = \frac{1}{\mathrm{N}^2}\sum_{\mathrm{i},\mathrm{j}}\mathrm{V}(\mathrm{i},\mathrm{j}) \quad (6)$$

As an example, the values of the spatio-temporal energy, for candidate blocks containing both smoke and moving smoke colored object are shown in Fig.3.

2.3. Smoke Motion Modeling

Smoke is typically directed upwards (upwards-right or upwards-left depending on the wind direction), while other object motions can be towards any direction. To model the smoke motion, we use HOGHOF descriptors and a visual vocabulary, which is built by applying hierarchical k-means clustering on these descriptors. HOGs and HOFs are used since they take into account both appearance and motion information. Values of HOGs represent the presence of edges and corners and HOFs values represent the motion orientation. HOGs and HOFs features are normalized and concatenated into one. Using k-means classification, a codebook of 16 codewords is formed from the extracted HOGHOFs descriptors. Term Frequency is used to represent each video sequence using the generated vocabulary for T previous frames. An SVM classifier is used to predict whether candidate blocks contain smoke or not. The distribution of codewords is estimated for each new video sequence and is provided to the SVM classifier. The training and evaluation procedure of HOGHOF descriptors is illustrated in Fig. 4.

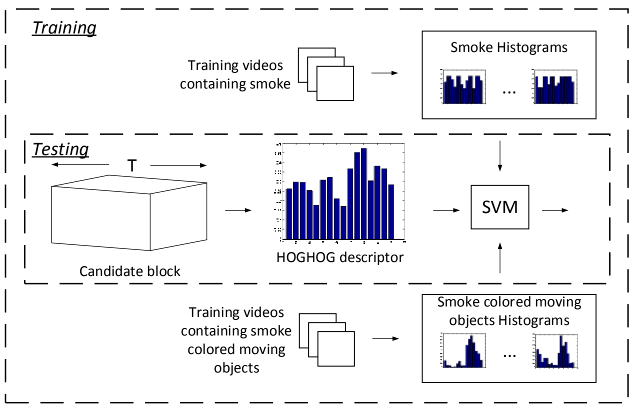


Fig. 4. Training and evaluation procedure of HOGHOF descriptors.

2.4. Dynamic Texture Analysis

Dynamic textures are sequences of images of moving scenes that exhibit certain stationarity properties in time [12]. Most of the existing dynamic texture categorization methods are applied to complete video sequences, therefore, they do not provide any information about the exact localization of the smoke in the image and the time of the incident. In the proposed method, we apply linear dynamical systems that initially proposed by Doretto et al. [12] and a bag of systems approach proposed recently by Ravichandran et al. [13]. However, these methods require the estimation of LDSs (Linear Dynamics Systems) in a large number of sample image patches. To reduce the computational burden, we focus only on those image regions for which we have an indication of smoke existence. Towards this end, LDSs are estimated only for the pixels contained in the candidate smoke blocks extracted from the first processing step of the proposed algorithm.

For each candidate block a 3D image patch is formed for the estimation of LDS. More specifically, given a candidate block of $\mathrm{N} \times \mathrm{N}$ pixels and F frames of the video sequence ($\mathrm{F} = 16$), we can model the pixel intensities of the candidate block $\mathrm{I_{block}(t)}$ at each time instant t, where $\mathrm{t} = 0 \dots 15$, assuming that the pixels contained in the 3D image patch can be considered as a linear dynamical system:

$$\mathrm{z(t+1) = Az(t) + Bv(t)} \tag{7}$$

$$\mathrm{I_{block}(t) = \bar{I}_{block} + Cz(t) + w(t)} \tag{8}$$

where $z(t) \in R^n$ is the hidden state at time t.

The dynamics of the hidden state are modeled by matrix $\mathrm{A} \in \mathrm{R^{nxn}}$, while matrix $\mathrm{C} \in \mathrm{R^{pxn}}$ (p is the number of pixels in a candidate block) maps the hidden state to the output of the system. The quantities $\mathrm{w(t)}$ and $\mathrm{Bv(t)}$ are the measurement and process noise respectively, while $\mathrm{\bar{I}_{block}}$ is the mean value of the pixels' intensities in a candidate block for the sequence of F frames. The LDS descriptors, i.e. M=(A,C), are estimated following a principal component analysis as proposed in [12]. Subsequently, similarly to [13], a codebook of 64 codewords is formed from the extracted LDS descriptors using a K-Medoid classification method.

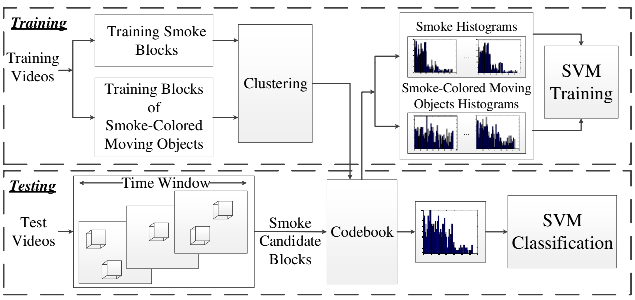


Fig. 5. Methodology for dynamic texture analysis.

A sliding time window T (in our experiments $T = 100$) is used to divide the video into equally sized subsequences. Each subsequence is then represented as a term frequency histogram of the predefined codeword of LDSs. Two distinctive classes are produced with histograms that represent subsequences of smoke and smoke colored moving objects. In the final step, a SVM classifier is trained with the above distributions of codewords. For the classification of a new sequence, the distribution of words is estimated and the extracted histogram is provided to the SVM classifier. The training and evaluation procedures are illustrated in Fig. 5.

2.5. Classification

As a last step, dynamic score combination by mean value is used to obtain the final decision about whether a block is smoke block or not. To this end a feature vector is created consisting of the three smoke probabilities $p_i, i = 1,2,3$ computed from the three previous stages (spatio-temporal smoke analysis, smoke motion analysis and dynamic texture recognition). This vector is fed as input to the dynamic score combination by mean value:

$$a = \frac{1}{3}(\sum_{i=1}^{3} p_i) \tag{9}$$

$$p = (1-a)\min\left(\sum_{i=1}^{3} p_i\right) + a\,max\left(\sum_{i=1}^{3} p_i\right) \tag{10}$$

A candidate block is classified as "smoke" if $p \geq 0.5$ and as "non smoke" otherwise.

3. EXPERIMENTAL RESULTS

In this section we present a detailed experimental evaluation using videos with or without smoke. To evaluate the proposed smoke detection method, experimental tests were performed on twenty video sequences (indoor and outdoor scenarios), which either containing smoke or not. Some of the sequences also contain smoke events and smoke colored moving objects. The video resolution is 240×320 and the average frame rate achieved by the proposed method was 5.7 fps, which is considered adequate for an early smoke warning system. The experiments were performed with a PC that has a Core i5 2.4 GHz processor.

To evaluate the performance of the proposed method, experimental results were obtained and compared with those obtained by the smoke detection method [7]. Towards to this end, videos from the video datasets of Bilkent University [14], [2] and VISOR smoke dataset [10] were used. The proposed algorithm has an average detection rate of 93.37%, while the corresponding value for the algorithm in [7] is 84.08%. However, in video sequences Bilkent/sEmptyR1 and Bilkent/sParkingLot smoke detection rates are quite lower than the average detection rate. In sEmptyR1 video sequence, the smoke is in front of the sun and for this reason the recognition is not correct in the first frames. On the other hand in sParkingLot video sequence the smoke is extremely thin and transparent, something that creates problems to the recognition. Results for the different video sequences are presented in Table 1 and screenshots showing the performance of the proposed algorithm are shown in Fig. 6.

Smoke Video	**Proposed**	**Smoke Detection Algorithm**[7]
	False Positives	
Bilkent/CarLights1	0	-
Bilkent/CarLights2	0	-
	True Positives	
Bikent/sBehindtheFence	94.44	96.15
Bilkent/sBtFence2	98.71	96.55
Bilkent/sEmptyR1	73.08	80
Bilkent/sEmptyR2	88.60	81.25
Bilkent/smoky	99.78	96.67
Bilkent/sParkingLot	81.56	100
Bilkent/sWasteBasket	99.29	94.23
Bilkent/sWindow	88.52	100
VISOR/movie08	96.65	74.86
VISOR/movie09	98.45	92.51
VISOR/movie10	94.81	58.51
VISOR/movie11	96.52	79.02
VISOR/movie12	93.2	88.52
VISOR/movie13	89.16	79.36
VISOR/movie14	96.14	89.48
VISOR/movie15	87.69	51.28
VISOR/movie16	97.69	100
VISOR/movie17	93.17	58.04
Total Average	93.37	84.08

Table 1. Smoke Detection in Bilkent [14], [2] and VISOR datasets [10].

(a) (b) (c) (d)

Fig. 6. Experimental results of the proposed algorithm: (a-c) true positive detection and (d) true negative detection.

4. CONCLUSIONS

In this paper, we proposed a novel algorithm for real time video-based smoke detection. The framework consists of several processing steps involving background subtraction, color analysis, spatial energy analysis, spatio-temporal analysis, HOGHOF analysis and dynamic texture analysis. To discriminate between smoke and smoke colored moving objects we used dynamic score combination by mean value of spatio-temporal analysis, motion modeling and dynamic texture recognition. Experimental results using twenty videos showed that the algorithm can achieve high detection rates, while increasing the robustness of the algorithm.

5. REFERENCES

[1] F. Gomez-Rodriguez, B. C. Arrue, and A. Ollero, "Smoke Monitoring and Measurement Using Image Processing. Application to Forest Fires," in *SPIE*, Sevilla, 2003, pp. 404-409.

[2] U. B. Toreyin, Y. Dedeoglu, and E. A. Cetin, "Contour based smoke detection in video using wavelets," in *EUSIPCO*, Florence, 2006.

[3] Z. Xu and J. Xu, "Automatic fire smoke detection based on image visual features," *Int. Conf. on Computational Intelligence and Security*, 2007.

[4] F. Yuan, "A fast accumulative motion orientation model based on integral image for video smoke detection," *Pattern Recognition Letters*, vol. 29, pp. 925-932, 2008.

[5] S. Calderara, P. Piccinini, and R. Cucchiara, "Smoke Detection in Video Surveillance: A MoG Model in the Wavelet Domain," *Systems Lecture Notes in Computer Science in Computer Vision*, vol. 5008, 119-128, 2008.

[6] F. Yuan, "Video-based smoke detection with histogram sequence of LBP and LBPV pyramids," *Fire Safety Journal*, vol. 46, pp. 132-139, 2011.

[7] K. Avgerinakis, A. Briassouli, and I. Kompatsiaris, "Smoke Detection Using Temporal HOGHOF Descriptors and Energy Colour Statistics from Video" *International Workshop on Multi-Sensor Systems and Networks for Fire Detection and Management*, 2012.

[8] H. Kim, D. Ryu, and J. Park, "Smoke Detection Using GMM and Adaboost," *International Journal of Computer and Communication Engineering*, vol. 3, no. 2, Mar. 2014.

[9] L. Piras, R. Tronci, and G. Giacinto, "Diversity in Ensembles of Codebooks for Visual Concept Detection," *ICIAP*, 2013, pp. 399-408.

[10] P. Paccinini, S. Calderara, and R. Cucchiara, "Releable smoke detection system in the domains of image energy and color," in *IEEE International Conference on Image Processing, ICIP*, 2008.

[11] C. Stuaffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," in *IEEE CVPR*, 1999, pp. 246-252.

[12] G. Doretto, A. Chiuso, Y. N. Wu, and S. Soatto, "Dynamic Textures," *International Journal of Computer Vision*, vol. 51, no. 2, pp. 91-109, 2003.

[13] A. Ravichandran, R. Chaudhry, and R. Vidal, "Categorizing Dynamic Textures Using a Bag of Dynamical Systems," *IEEE Transactions on PAMI*, vol. 35, no. 2, Feb. 2013.

[14] U. B. Toreyin, Y. Dedeoglu, and E. A. Cetin, "Wavelet based real-time smoke detection in video," in *EUSIPCO*, 2005.