REAL-TIME WILDFIRE DETECTION USING CORRELATION DESCRIPTORS

Y. Hakan Habiboglu, Osman Gunay, and A. Enis Cetin

Dept. of Electrical and Electronics Engineering,
Bilkent University 06800, Ankara, Turkey
phone: + (90)312-290-15-25, fax:+(90)312-266-4192,
email:{yhakan,osman,cetin}ee.bilkent.edu.tr
web: www.ee.bilkent.edu.tr/~signal

ABSTRACT

A video based wildfire detection system that based on spatio-temporal correlation descriptors is developed. During the initial stages of wildfires smoke plume becomes visible before the flames. The proposed method uses background subtraction and color thresholds to find the smoke colored slow moving regions in video. These regions are divided into spatio-temporal blocks and correlation features are extracted from the blocks. Property sets that represent both the spatial and the temporal characteristics of smoke regions are used to form correlation descriptors. An SVM classifier is trained and tested with descriptors obtained from video data containing smoke and smoke colored objects. Experimental results are presented.

1. INTRODUCTION

Most surveillance systems already have built-in simple detection modules (e.g. motion detection, event analysis). In recent years there has been significant interest in developing real-time algorithms to detect fire and smoke for standard surveillance systems [1]-[7]. Video based smoke detection can be used to replace traditional point sensor type detectors, since a single camera can monitor a large area from a distance and can detect smoke earlier than a traditional point detector if a robust detection algorithm is used. Although video based smoke detection is a promising alternative to traditional smoke detectors, it has some drawbacks that need to be resolved before a perfect system is realized. Smoke is difficult to model due to its dynamic texture and irregular motion characteristics. Unstable cameras, dynamic backgrounds, obstacles in the viewing range of the camera and lighting conditions also pose important problems for smoke detection. Therefore current wildfire detection systems require human assistance and there is always room for improvement.

Smoke plume observed from a long distance and observed from up close have different spatial and temporal characteristics. Therefore, generally different algorithms are designed to detect close range and long range smoke plume.

Jerome and Philippe [1, 2] implemented a real-time automatic smoke detection system for forest surveillance stations. The main assumption for their detection method is that the energy of the velocity distribution of smoke plume is higher than other natural occurrences except for clouds which, on the other hand have lower standart deviation than smoke. In the classification stage they use fractal embedding and linked list chaining to segment smoke regions. This method was used in the forest fire detector “ARTIS FIRE”, commercialized by “T2M Automation”.

Another smoke detection method with an application to wildfire prevention was described in [3]. This method takes the advantages of wavelet decomposition and optical flow algorithm for fire smoke detection and monitoring. The optical flow algorithm is used for motion detection. Wavelet decomposition based method was used to solve the aperture problem in optical flow. After the smoke is detected and segmented, smoke characteristics such as speed, dispersion, apparent volume, maximum height, gray level and inclination angle of the smoke can be extracted using the video frames or image sequences.

Damir et. al. [4] investigated different colour space transformations and feature classifiers that are used in a histogram-based smoke segmentation for a wildfire detection system. They provide evaluations of histograms in YCrCb, CIELab, HSI, and modified HSI colour spaces. They use look up tables and two different naive Bayes classifiers with different density estimation methods to classify the histograms. The best performances are achieved with HSI and RGB colour spaces when using the Bayes classifier. The method described is one of the algorithms used in the Intelligent Forest Fire Monitoring System (iForestFire) that is used to monitor the coastline of the Republic of Croatia.

Qinjuan et. al. [5] proposed a method for long range smoke detection to be used in a wildfire surveillance system. The method uses multi-frame temporal difference and OTSU thresholding to find the moving smoke regions. They also use colour and area growth clues to verify the existence of smoke in the viewing range of the camera.

In [6] a real-time wildfire detection algorithm is developed based on background subtraction and wavelet analysis. In [7] an algorithm for long range smoke detection is developed to be used in a wildfire surveillance system. The algorithm is an online learning method that updates its decision values using the supervision from an oracle (security guard at the watch tower). The main detection algorithm is composed of four sub-algorithms detecting (i) slow moving objects using adaptive background subtraction, (ii) gray regions using YUV colour space, (iii) rising regions using hidden Markov models (HMM), and (iv) shadows using RGB angle between image and the background. Decisions from sub-algorithms are combined using the Least Mean Square (LMS) method in the training stage.

This is a review article describing our ongoing research in FP-7 FIRESENSE project [8]. Most smoke detection systems first find the moving regions using background subtraction. These regions are then analyzed spatially and temporally to detect the characteristics of smoke. In this work, we use a different approach by combining color, spatial and temporal domain information in feature vectors for each spatio-temporal block using region covariance descriptors [9, 10]. The blocks are obtained by dividing the smoke colored regions into 3D regions that overlap in time. Classification of the features is performed only at the temporal boundaries of blocks instead of every frame. This reduces the computational complexity of the method.

In the following sections we describe the building blocks of our algorithm.

2. BUILDING BLOCKS OF WILDFIRE DETECTION ALGORITHM

Watch towers are widely available in forests all around the world to detect wildfires. Surveillance cameras can be placed in these surveillance towers to monitor the surrounding forestal area for possible wildfires. Furthermore, they can be used to monitor the progress of the fire from remote centers.

Cameras, once installed, operate at forest watch towers throughout the fire season for about six months which is mostly dry and sunny in Mediterranean region. It is usually not possible to view flames of a wildfire from a camera mounted on a forest watch tower unless the fire is very near to the tower. However, smoke rising up in the forest due to a fire is usually visible from long distances. A snapshot of a typical wildfire smoke captured by a watch tower camera from a distance of 5 km is shown in Fig. 1.

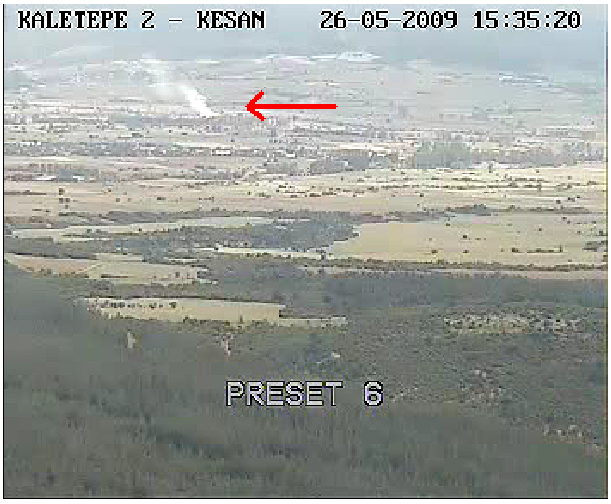

Figure 1: Snapshot of a typical wildfire smoke captured by a forest watch tower which is 5 km away from the fire.

Smoke at far distances exhibits different spatio-temporal characteristics than nearby smoke and fire [11]-[12]. Therefore different methods should be developed for smoke detection at far distances rather than using nearby smoke detection methods described in [13].

The proposed wildfire smoke detection algorithm consists of three main sub-algorithms: (i) slow moving object detection in video, (ii) smoke-colored region detection, (iii) correlation based classification.

2.1 Slow Moving Region Detection

For moving object detection we use a Gaussian mixture model (GMM) based background subtraction method [14]. For a few seconds we update the background very fast and after this learning duration we update the background very slowly so that we can detect small and slow moving objects. We also use a second GMM background model that is optimized to detect fast moving objects and use it to discard ordinary moving objects.

2.2 Smoke Color Model

Smoke colored regions can be identified by setting thresholds in YUV color space [7]. Luminance value of smoke regions should be high for most smoke sources. On the other hand, the chrominance values should be very low in a smoke region.

The conditions in YUV color space are as follows:

Condition 1 $Y > T_Y$

Condition 2 $|U - 128| < T_U$ & $|V - 128| < T_V$

where Y, U and V are the luminance and chrominance values of a pixel. The luminance component Y takes real values in the range $[0,255]$ in an image and the mean values of chrominance channels, U and V are increased to 128 so that they also take values between 0 and 255. The threshold T_Y is an experimentally determined value and taken as 128 on the luminance (Y) component in this work. T_U and T_V are both taken as 10.

2.3 Correlation Method

2.3.1 *Correlation Descriptors for Videos*

Covariance descriptors are proposed by Tuzel, Porikli and Meer to be used in object detection and texture classification problems [9, 10]. We propose temporally extended correlation descriptors to extract features from video sequences.

Covariance descriptors provide very good description of a given image region when the property set of a region in an image can be described by a wide-sense stationary multivariate normal distribution [9]. Wide-sense stationarity is a reasonable assumption for a smoke colored image regions because such regions do not contain strong edges in video. Therefore, covariance descriptors can be used to model spatial characteristics of smoke regions in images. It is experimentally observed that wide-sense stationarity assumption is valid temporally as well. To model the temporal variation in smoke regions we introduce temporally extended and normalized covariance descriptors in this article. To the best of our knowledge spatio-temporal parameters have not been used to construct covariance descriptors by other researchers.

Temporally extended correlation descriptors are designed to describe spatio-temporal video blocks. Let $I(i,j,n)$ be the intensity of $(i,j)^{th}$ pixel of the n^{th} image frame of a spatio-temporal block in video and *Luminance*, *ChrominanceU*, *ChrominanceV* represent the color values of pixels of the block. The property parameters defined in Equation (1) to Equation (8) are used to form a covariance matrix representing spatial information. In addition to spatial parameters we introduce temporal derivatives, *It* and *Itt* which are the first and second derivatives of intensity with respect to time, respectively. By adding these two features to the previous property set, correlation descriptors can be used to define spatio-temporal blocks in video.

$$Y_{i,j,n} = Luminance(i,j,n), \quad (1)$$

$$U_{i,j,n} = ChrominanceU(i,j,n), \quad (2)$$

$$V_{i,j,n} = ChrominanceV(i,j,n), \tag{3}$$

$$I_{i,j,n} = Intensity(i,j,n), \tag{4}$$

$$Ix_{i,j,n} = \left| \frac{\partial Intensity(i,j,n)}{\partial i} \right|, \tag{5}$$

$$Iy_{i,j,n} = \left| \frac{\partial Intensity(i,j,n)}{\partial j} \right|, \tag{6}$$

$$Ixx_{i,j,n} = \left| \frac{\partial^2 Intensity(i,j,n)}{\partial i^2} \right|, \tag{7}$$

$$Iyy_{i,j,n} = \left| \frac{\partial^2 Intensity(i,j,n)}{\partial j^2} \right|, \tag{8}$$

$$It_{i,j,n} = \left| \frac{\partial Intensity(i,j,n)}{\partial n} \right|, \tag{9}$$

$$Itt_{i,j,n} = \left| \frac{\partial^2 Intensity(i,j,n)}{\partial n^2} \right| \tag{10}$$

2.3.2 Computation of Correlation Values in Spatio-temporal Blocks

In this section details of correlation features computation in video is described. We first divide the video into blocks of size $10 \times 10 \times F_{rate}$ where F_{rate} is the frame rate of the video. Computing the correlation parameters for each block of the video would be computationally inefficient. We use the first two sub-algorithms to find the candidate smoke regions. Therefore, only pixels corresponding to the non-zero values of the following mask are used in the selection of blocks. The mask is defined by the following function:

$$\Psi(i,j,n) = \begin{cases} 1 & \text{if } M(i,j,n) = 1 \\ 0 & \text{otherwise} \end{cases} \tag{11}$$

where $M(.,.,n)$ is the binary mask obtained from the first two sub-algorithms. In order to reduce the effect of non-smoke colored pixels, only property parameters of pixels that are obtained from the mask used in the estimation of the correlation based features, instead of using every pixel of a given block.

A total of 10 property parameters are used for each pixel satisfying the color condition. To further reduce the computational cost we compute the correlation values of the pixel property vectors

$$\Phi_{color}(i,j,n) = [\; Y(i,j,n) \quad U(i,j,n) \quad V(i,j,n) \;]^T \tag{12}$$

and

$$\Phi_{ST}(i,j,n) = \begin{bmatrix} I(i,j,n) \\ Ix(i,j,n) \\ Iy(i,j,n) \\ Ixx(i,j,n) \\ Iyy(i,j,n) \\ It(i,j,n) \\ Itt(i,j,n) \end{bmatrix} \tag{13}$$

separately. Therefore, the property vector $\Phi_{color}(i,j,n)$ produces $\frac{3*4}{2} = 6$ and the property vector $\Phi_{ST}(i,j,n)$ produces $\frac{7*8}{2} = 28$ correlation values, respectively and 34 correlation parameters are used in training and testing of the SVM instead of 55 parameters.

During the implementation of the correlation method, the first derivative of the image is computed by filtering the image with [-1 0 1] and second derivative is found by filtering the image with [1 -2 1] filters, respectively. The lower or upper triangular parts of the correlation matrix, $\widehat{C}(a,b)$, that is obtained by normalizing the covariance matrix, $\widehat{\Sigma}(a,b)$, form the feature vector of a given image region. We use the correlation matrix estimation formula given in Equation (15), that can be started to calculate without waiting for the entire data. The feature vectors are processed by a support vector machine (SVM).

$$\widehat{\Sigma}(a,b) = \frac{1}{N-1}\left(\sum_i \sum_j \Phi_{i,j}(a)\Phi_{i,j}(b) - C_N\right)$$
$$\text{where} \tag{14}$$
$$C_N = \frac{1}{N}\left(\sum_i \sum_j \Phi_{i,j}(a)\right)\left(\sum_i \sum_j \Phi_{i,j}(b)\right)$$

$$\widehat{C}(a,b) = \begin{cases} \sqrt{\widehat{\Sigma}(a,b)} & \text{if } a = b \\ \frac{\widehat{\Sigma}(a,b)}{\sqrt{\widehat{\Sigma}(a,a)}\sqrt{\widehat{\Sigma}(b,b)}} & \text{otherwise} \end{cases} \tag{15}$$

We also assume that the size of the image frames in video is 320 by 240. If not the video is scaled to 320 by 240 in order to run the smoke detection algorithm in real-time.

3. TRAINING AND TESTING

For training and testing, $10 \times 10 \times F_{rate}$ blocks are extracted from various video clips. The temporal dimension of the blocks are determined by the frame rate parameter F_{rate} which is between 10 and 25 in our train and test videos. These blocks do not overlap in spatial domain but there is 50% overlap in time domain. This means that classification is not repeated after every frame of the video. After the blocks are constructed, features are extracted and used to form a training set. A support vector machine (SVM) [15] is trained for classification.

The classification is done periodically with the period $F_{rate}/2$. This decreases the cost of classification.

During the implementation, in each spatio-temporal block, the number of smoke colored slow moving pixels, $\sum_i \sum_j \sum_n \Psi(i,j,n)$, is found. If this number is higher than or equal to $\frac{2}{5}$ of the number of the elements of block ($10 \times 10 \times F_{rate}$) then that block is classified as a smoke block. This thresholding is done because only smoke colored pixels according to the YUV color model described in [7] is used in correlation analysis. If the number of possible smoke-pixels is enough, then classification is done by the SVM classifier using the augmented feature vector described in Section 2.3.2.

In this article, 13 positive and 12 negative video clips are used for training. Negative video clips contain smoke colored moving regions. For positive videos (video clips containing smoke) only parts of the video clips that contain smoke are used.

At the final step of our smoke detection method a confidence value is determined according to the number of positively classified video blocks and their positions. After every block is classified spatial neighborhoods of the block are used to decide the confidence level of the alarm. If there is no neighbor block classified as smoke, the confidence level is set to 1. If there is a single neighbor block, which is classified as smoke, then the confidence level is set to 2. If there are more than 2 neighbor blocks classified as smoke then the confidence level of that block is set to 3 which is the highest level of confidence that the algorithm provides.

4. EXPERIMENTAL RESULTS

The proposed system is compared with the wildfire detection method in [6]. In the decision process, if the confidence level of any block of the frame is greater than or equal to 2 then that frame is marked as a smoke containing frame. Results are summarized in Table 1 and Table 2 in terms of the true detection and the false alarm ratios, respectively. In Tables 1 and 2 the true detection rate in a given video clip is defined as the number of correctly classified frames containing smoke divided by the total number of frames which contain smoke. Similarly, the false alarm rate in a given test video is defined as the number of misclassified frames, which do not contain smoke divided by the total number of frames which do not contain smoke.

Table 1: Correlation based method is compared with the method proposed in [6] in terms of true detection rates in video clips that contain smoke.

	True Detection Rates	
Video name	New Method	Old Method
posVideo1	$\frac{726}{768} = 94.53\%$	$\frac{584}{768} = 76.04\%$
posVideo2	$\frac{215}{260} = 82.69\%$	$\frac{84}{260} = 32.30\%$
posVideo3	$\frac{307}{419} = 73.26\%$	$\frac{64}{419} = 15.27\%$
posVideo4	$\frac{292}{430} = 67.90\%$	$\frac{246}{430} = 57.20\%$
posVideo5	$\frac{774}{1350} = 57.33\%$	$\frac{780}{1350} = 57.77\%$
posVideo6	$\frac{324}{360} = 90.00\%$	$\frac{163}{360} = 45.27\%$
posVideo7	$\frac{124}{210} = 59.04\%$	$\frac{0}{210} = 0.00\%$
posVideo8	$\frac{268}{545} = 49.17\%$	$\frac{5}{545} = 0.91\%$
Average	71.74%	35.59%

15 video clips are used to test the proposed system. First 8 videos contain actual wildfire smoke or artificial test fires that we recorded and the remaining 7 videos do not contain smoke but contain smoke colored moving objects like clouds and shadows. In Table 1 the true detection rates of the two algorithms are presented for the 8 videos containing smoke. In Table 2 the false alarm rates of the two algorithms are presented for the 7 videos that do not contain smoke.

Compared to the previous method the new method has higher true detection rate in all video clips that contain actual smoke plumes. “posVideo7” and “posVideo8” are actual forest fire videos recorded with cameras that are mounted on high poles which shake in the wind when they are zoomed

(a) posVideo1 - new method (b) posVideo1 - old method

(c) posVideo5 - new method (d) posVideo5 - old method

(e) posVideo7 - new method (f) posVideo7 - old method

(g) posVideo8 - new method (h) posVideo8 - old method

(i) negVideo7 - new method (j) negVideo7 - old method

Figure 2: Detection results from test videos.

in. Since the old method [6] assumes a stationary camera for background subtraction it cannot correctly classify most of the actual smoke regions in these videos. Although the true detection rate is low in some videos, we do not need to detect all smoke frames correctly to issue an alarm. It is enough to detect smoke in a short time without too many false alarms. The first detection time is less than 10 seconds in all the test video clips. In most of the videos that do not contain smoke the new method has a lower false alarm rate than the old method.

In Figure 2 the detection results of the new method and

Table 2: Correlation based method is compared with the method proposed in [6] in terms of false alarm rates in video clips that do not contain smoke.

	False Alarm Rates	
Video name	New Method	Old Method
negVideo1	$\frac{100}{6300} = 1.59\%$	$\frac{623}{6300} = 9.88\%$
negVideo2	$\frac{0}{3500} = 0.00\%$	$\frac{81}{3500} = 2.31\%$
negVideo3	$\frac{0}{4000} = 0.00\%$	$\frac{419}{4000} = 10.47\%$
negVideo4	$\frac{30}{1500} = 2.00\%$	$\frac{52}{1500} = 3.46\%$
negVideo5	$\frac{30}{1000} = 3.00\%$	$\frac{10}{1000} = 1.00\%$
negVideo6	$\frac{0}{360} = 0.00\%$	$\frac{0}{360} = 0.00\%$
negVideo7	$\frac{82}{2900} = 2.83\%$	$\frac{92}{2900} = 3.17\%$
Average	1.34%	4.32%

old method are shown on some of the test videos. The new method significantly improved detection results compared to the old method.

The proposed method is computationally efficient. The experiments are performed with a PC that has a Core 2 Duo 2.66 GHz processor and the video clips are generally processed around 15-20 fps when image frames of size 320 by 240 are used. The processing speed might decrease when there are too many smoke colored moving regions since this increases the number of blocks that are classified by the SVM.

The detection resolution of the algorithm is determined by the video block size. Since we require two neighboring blocks to reach the highest confidence level the smoke should occupy a region of size 10 by 20 in video.

5. CONCLUSIONS

A real-time video smoke detection system is proposed that uses correlation descriptors with an SVM classifer. An important contribution of this article is the use of temporal correlation information in the decision process. Most smoke detection methods use color, spatial and temporal information separately, but in this work we use temporally extended correlation matrices to use all the information together. The proposed method is computationally efficient and it can process 320 by 240 frames at 15-20 fps in a standard PC.

Acknowledgement

This work was supported in part by FIRESENSE (Fire Detection and Management through a Multi-Sensor Network for the Protection of Cultural Heritage Areas from the Risk of Fire and Extreme Weather Conditions, FP7-ENV-2009-1244088-FIRESENSE).

REFERENCES

[1] Philippe Guillemant and Jerome Vicente, "Real-time identification of smoke images by clustering motions on a fractal curve with a temporal embedding method," *Optical Engineering*, vol. 40, no. 4, pp. 554–563, 2001.

[2] Jerome Vicente and Philippe Guillemant, "An image processing technique for automatically detecting forest fire," *International Journal of Thermal Sciences*, vol. 41, no. 12, pp. 1113 – 1120, 2002.

[3] F. Gomez-Rodriguez, B. C. Arrue, and A. Ollero, "Smoke monitoring and measurement using image processing: application to forest fires," in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, F. A. Sadjadi, Ed., Sept. 2003, vol. 5094, pp. 404–411.

[4] Damir K., Darko S., and Toni J, "Histogram-based smoke segmentation in forest fire detection system," *INFORMATION TECHNOLOGY AND CONTROL*, 2009.

[5] Qinjuan L. and Ning H., "Effective dynamic object detecting for video-based forest fire smog recognition," in *2nd International Congress on IImage and Signal Processing CISP*, 2009, pp. 1–5.

[6] B. U. Töreyin and A. E. Çetin, "Computer vision based forest fire detection," in *IEEE 16. Sinyal Isleme ve Iletisim Uygulamalari Kurultayi, SIU-2008*, 2008.

[7] B. U. Töreyin and A. E. Çetin, "Wildfire detection using lms based active learning," in *in Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP*, 2009.

[8] FIRESENSE, "Fire detection and management through a multi-sensor network for the protection of cultural heritage areas from the risk of fire and extreme weather conditions, fp7-env-2009-1244088-firesense," 2009, `http://www.firesense.eu`.

[9] O. Tuzel, F. Porikli, and P. Meer, "Region covariance: A fast descriptor for detection and classification," *Computer Vision–ECCV 2006*, pp. 589–600, 2006.

[10] H. Tuna, I. Onaran, and A. Enis Çetin, "Image description using a multiplier-less operator," *Signal Processing Letters, IEEE*, vol. 16, no. 9, pp. 751 –753, sept. 2009.

[11] B. U. Töreyin, Y. Dedeoğlu, and A. E. Çetin, "Flame detection in video using hidden markov models," in *International conference on Image Processing (ICIP)*, 2005, pp. 1230–1233.

[12] B. U. Töreyin, Y. Dedeoğlu, U. Güdükbay, and A. E. Çetin, "Computer vision based system for real-time fire and flame detection," *Pattern Recognition Letters*, vol. 27, pp. 49–58, 2006.

[13] B. U. Töreyin, Y. Dedeoğlu, and A. E. Çetin, "Wavelet based real-time smoke detection in video," in *European Signal Processing Conference (EUSIPCO)*, 2005.

[14] Z. Zivkovic, "Improved adaptive gaussian mixture model for background subtraction," in *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, aug. 2004, vol. 2, pp. 28 – 31 Vol.2.

[15] Chih-Chung Chang and Chih-Jen Lin, *LIBSVM: a library for support vector machines*, 2001, Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`.