# Video-Based FLame Detection for the protection of Cultural Heritage

K. Dimitropoulos[a], O. Gunay[b], K. Kose[b], F. Erden[b], F. Chaabane[c], F. Tsalakanidou[a], N. Grammalidis[a] and A.E. Cetin[b]

[a]Information Technologies Institute, Centre for Research and Technology Hellas, Greece, (dimitrop,ngramm,filareti)@iti.gr

[b]Department of Electrical and Electronics Engineering, Bilkent University, Turkey,

gunayosman@gmail.com, kkivanc@ee.bilkent.edu.tr, erdenfatih@gmail.com, cetin@bilkent.edu.tr

[c]Ecole Superieure des Communication de Tunis, Sup'Com, Tunisia, ferdaous.chaabene@supcom.rnu.tn

# Video-Based FLame Detection for the protection of Cultural Heritage

K. Dimitropoulos O. Gunay , K. Kose, F. Erden, F. Chaabane, F. Tsalakanidou, N. Grammalidis and A.E. Cetin

**Abstract:**

The majority of cultural heritage and archaeological sites, especially in the Mediterranean region, are covered with vegetation, which increases the risk of fires. These fires may also break out and spread towards nearby forests and other wooded land, or conversely start in nearby forests and spread to archaeological sites. Beyond taking precautionary measures to avoid a forest fire, early warning and immediate response to a fire breakout are the only ways to avoid great losses and environmental and cultural heritage damages. The use of terrestrial systems, typically based on video cameras, is currently the most promising solution for advanced automatic wildfire surveillance and monitoring due to its low cost and short response time. Early and accurate detection and localization of flame is an essential requirement of these systems, however, it remains a challenging issue due to the fact that many natural objects have similar characteristics with fire. This paper presents and compares three video-based flame detection techniques, which were developed within the FIRESENSE EU research project, taking into account the chaotic and complex nature of the fire phenomenon and the large variations of flame appearance in video. Experimental results show that the proposed methods provide high fire detection rates with reasonable false alarm ratios.

**Keywords:** Cultural heritage protection, early warning systems, flame detection

## 1. Introduction

One of the main causes of the destruction of archaeological sites in recent years is wildfires. The increase in seasonal temperatures has caused an explosion in the number of self-ignited wildfires in forested areas. Fanned by the dry winds, and fueled by dry vegetation, some of these fires became disasterous. In addition to possible deliberate actions for harming a particular site, common causes of unintentional fires are human carelessness, exposure to extreme heat and aridity and lightning strikes.

Fire detection systems are the ones that stand to benefit most from technological advances. The most important goals in fire surveillance are quick and reliable detection and localization of fire, since reducing the time between the ignition and the detection of fire is extremely vital for extinguishing it. However, early detection of fire is traditionally based on human surveillance. This can either be done using direct human observation by observers located at monitoring spots (e.g. lookout towers located on highland) [1] or by distant human observation based on video surveillance systems. Relying solely on humans for the detection of forest fires is not the most efficient method. A more advanced approach is automatic surveillance and automatic early forest fire detection using either (i) Space borne (satellite) systems, (ii) Airborne or, (iii) Terrestrial-based systems.

Some advanced forest fire detection systems are based on satellite imagery, e.g. the Advanced Very High Resolution Radiometer [2], launched by the National Oceanic and Atmospheric Administration (NOAA) in 1998 and the Moderate Resolution Imaging Spectroradiometer (MODIS) [3], put in orbit by NASA in 1999, etc. However, there can be a significant amount of delay in communications with satellites, because orbits of satellites are predefined and thus satellite coverage is not continuous. Furthermore, satellite images have relatively low resolution due to the high altitude of satellites, while their geo-referencing is usually problematic due to the high speed of satellites. In addition, the accuracy and reliability of satellite-based systems are largely affected by weather conditions. Clouds and precipitation absorb parts of the frequency spectrum and reduce spectral resolution of satellite images, which consequently degrades the detection accuracy.

Airborne systems refer to systems mounted on helicopters (elevation<1km) or airplanes (up to 2 to 10 km above sea level). They offer great flexibility and short response times and they are able to generate very high-resolution data (typically few cm). Also, geo-referencing is easier and much more accurate compared to satellite based systems. Drawbacks include the increased flight costs, flight limitations by air traffic control or bad weather conditions and limited

coverage. Turbulences, vibrations and possible deviations of the airplane from a pre-planned trajectory due to weather conditions are additional problems. However, recently, a large number of early fire detection projects use Unmanned Aerial Vehicles (UAVs), which can alleviate some of the problems of the airborne systems, e.g. they are cheaper and are allowed to fly in worse weather conditions.

For the above reasons, terrestrial systems based on CCD video are today the most promising solution for realizing automatic surveillance and automatic forest fire detection systems. However, the majority of current wildfire surveillance systems do not realize the full potential offered by current technologies due to the lack of an integrated approach. One of the main objectives of the 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 EU project [4] is to take advantage of multi-sensor surveillance technologies in order to develop an innovative and integrated early warning platform to protect cultural heritage areas from the risk of fire. In this paper, we present and compare three video-based flame detection algorithms using spatiotemporal characteristics of fire, which have been developed and are currently being evaluated within FIRESENSE project for the protection of five cultural heritage test sites: i) Thebes, Greece, ii) Rhodiapolis, Turkey, iii) Dodge Hall, Istanbul, Turkey, iv) Temple of Water, Tunisia, v) Monteferrato-Galceti Park, Prato, Italy.

## 2. Video-Based Fire Detection

### 2.1. Flame Detection

Flame colour is the most identifiable feature used by a video flame detection method. The colour of the flame is not a reflection of the natural light, but it is generated as a result of the burning materials. In some cases, the colour can be white, blue, gold or even green depending on the chemical properties of the burnt material and its burning temperature. However, in the cases of organic materials such as trees and bushes, the fire has a characteristic red-yellow colour. Many natural objects have similar colours as those of the fire (including the sun, various artificial lights or reflections of them on various surfaces) and can often be mistakenly detected as flames, when the decision takes into account only the colour criterion. For this reason, additional criteria have to be used to discriminate between such false alarm situations and real fire.

Many researchers use motion characteristics of the flame as well as the special distribution of fire colors in the scene. The use of spatio-temporal criteria in the flame detection algorithm may increase the computational complexity, since multidimensional image processing is

K. Dimitropoulos O. Gunay , K. Kose, F. Erden, F. Chaabane, F. Tsalakanidou, N. Grammalidis and A.E. Cetin

needed. Four dimensions due to position, pixel luminance information and time (x,y,Y,t) exist in the case of grayscale images or 6 dimensions (x,y,r,g,b,t) exist in the case of colour images having red, green and blue components. Therefore, to keep the complexity low, most works in the  literature use either a) purely spatial or b) purely temporal criteria or c) a two-step approach combining results obtained using purely spatial or purely temporal criteria.

More specifically, the methods presented in the literature can be categorized as follows:

- *Change (including Motion) Detection*: In most flame detection algorithms, a pre-processing step exists focusing on regions of interest where there is a temporal change in the scene. This can significantly reduce the computational burden for the subsequent processing by reducing the video processing only in moving regions. Some of the techniques used for this task include a) simple temporal differencing [5], b) background estimation and subtraction [6][7], and c) optical flow based motion detection techniques [8].
- *Colour Detection*: Colour is a very important criterion of the fire which is used in most of the currently available methods. Usually, chromatic analysis of the images to search for regions with fire colors uses one or more decision rules in a colour space. Usually, the RGB colour space is used but other colour spaces as HIS or YUV have also been used in the literature. In some cases look-up tables and/or neural networks are also used for this task. Look-up tables can reduce the complexity but have high memory requirements.
- *Shape/Geometry/Contour Cues*: Specific features of the candidate regions of interest in video such as its shape, geometry and/or contour are examined and an effort is made to identify particular characteristics, patterns or models that are consistent with the presence of flames (e.g. random contour shape etc) [9][10].
- *Temporal Analysis*: Temporal cues due to flame flickering process leading to strong high frequency content in video are identified in image frames forming the video [11] [12][13]. This is a good indication of the presence of a wildfire. The Fast Fourier Transform (FFT), wavelet analysis or simpler mathematical rules can be used for the task of identifying rapidly changing regions in video.
- *Spatial Analysis*: Flame colours may follow certain models or spatial distribution patterns which can be identified, e.g. fractal models, wavelet models in multiple spatial resolutions [14][15][16]. Such models are used to discriminate flames from other natural or man-made moving objects in video.

In the following, new video-based flame detection methods are proposed, either by elegantly combining some of the flame features mentioned above, or by defining new features or variants of them that result in improved performance.

*2.1.1. Flame detection using correlation descriptors:*

The flame detection method presented in this section uses covariance matrix descriptors for feature extraction from video [17], [18] and SVM classification. The video is divided into spatio-temporal blocks before analysis. Each spatio-temporal block is first classified according to its colour content. Blocks that do not contain flame coloured pixels are discarded before further processing. Flame coloured pixels are determined according to two simple rules:

**Condition 1**:  $R \geq G \geq B$

Typically red is the most dominant colour in flames. Therefore any block in which red colour is not dominant is discarded.

**Condition 2**:  $R > R_T$

where $R_T$ is a predefined threshold. The threshold is empirically determined, from a dataset of flame videos.

For each pixel of a video-block containing flame colored pixels, a property vector is defined. The property vector $\varphi_{(i,j,n)}$ of a pixel at location $(i,j)$ in the $n^{th}$ image frame can be defined as:

$$\phi_{(i,j,n)} = [R_{(i,j,n)}, G_{(i,j,n)}, B_{(i,j,n)}, I_{(i,j,n)}, I_{x(i,j,n)},$$
$$I_{y(i,j,n)}, I_{xx(i,j,n)}, I_{yy(i,j,n)}, I_{t(i,j,n)}, I_{tt(i,j,n)}]$$

The individual components that are included in the feature descriptor are as follows: a) colour components (for each channel) and intensity, b) first order horizontal and vertical derivatives of intensity values, c) corresponding second order horizontal and vertical derivatives and d) first and second order temporal derivatives.

The first and the second order derivatives are calculated by convolving the video using the filters [-1,0,1] and [1,-2,1], respectively. After calculating these features, a length-10 descriptor vector for each candidate pixel is defined. The covariance matrix of a spatio-temporal block is estimated as follows:

$$\hat{\Sigma} = \frac{1}{N-1}\sum_i \sum_j \sum_n (\Phi_{i,j,n} - \bar{\Phi})(\Phi_{i,j,n} - \bar{\Phi})^{\text{T}}$$

where $N$ is the number and $\bar{\Phi} = \frac{1}{N}\sum_i \sum_j \sum_n \Phi_{i,j,n}$ is the mean

of the descriptor vectors of the pixels in the block.

K. Dimitropoulos O. Gunay , K. Kose, F. Erden, F. Chaabane, F. Tsalakanidou, N. Grammalidis and A.E. Cetin

In this section details of the covariance matrix computation in video is described. We first divide the video into blocks of size $16 \times 16 \times F_{rate}$ where $F_{rate}$ is the frame rate of the video. Computing the covariance parameters for each block of the video would be computationally inefficient. We use the simple color model to eliminate the blocks that do not contain any fire colored pixels. Therefore, only pixels corresponding to the non-zero values of the following color mask are used in the selection of blocks, using the following function:

$$\Psi(i,j,n) = \begin{cases} 1 & \text{if} & R_{(i,j,n)} > G_{(i,j,n)} \geq B_{(i,j,n)} \text{ and } R_{(i,j,n)} > R_T \\ 0 & \text{otherwise} \end{cases}$$

This is not a tight condition for fire colored pixels. Almost all flame-colored or reddish regions satisfy it. Furthermore, in order to reduce the effect of non-fire colored pixels, only property parameters of chromatically possible fire-pixels are used in the estimation of the covariance matrix, instead of using every pixel of a given block.

A total of 10 property parameters are used for each pixel satisfying the color condition. This requires $\dfrac{10*11}{2} = 55$ covariance computations. To further reduce the computational cost we compute the covariance values of the pixel property vectors:

$$\Phi_{color}(i,j,n) = \begin{bmatrix} R_{(i,j,n)} & G_{(i,j,n)} & B_{(i,j,n)} \end{bmatrix}^T$$

and

$$\Phi_{ST}(i,j,n) = \begin{bmatrix} I_{(i,j,n)} \\ I_{x(i,j,n)} \\ I_{y(i,j,n)} \\ I_{xx(i,j,n)} \\ I_{yy(i,j,n)} \\ I_{t(i,j,n)} \\ I_{tt(i,j,n)} \end{bmatrix}$$

in a separate manner. Therefore, the property vector $\Phi_{color}(i,j,n)$ produces $\dfrac{3*4}{2} = 6$, while the property vector $\Phi_{ST}(i,j,n)$ produces

$\frac{7*8}{2} = 28$ covariance values. Therefore, instead of 55, only 34 covariance parameters are used in the training and testing of the SVM. The system is tested using both 34 and 55 covariance parameters to see if there is any difference between detection performance and computational cost.

After the spatio-temporal blocks are constructed, features are extracted and used to form a training set. The features are obtained by averaging the correlations between the 34-parameters for every $F_{rate}/2$ frames. A support vector machine (SVM) with a radial basis function (RBF) kernel is trained for classification [2]. The SVM is first trained with seven positive and seven negative video clips given in Figures 1 and 2. The SVM model obtained from this training is tested with the three remaining negative video clips. Sample frames from these clips are given in Figure 3. As it is seen from the figure, these clips contain scenes that usually cause false alarms. So it is beneficial that these non-fire blocks that are classified as fire are added to the negative training set and that the SVM is re-trained with the new dataset.

We have a total of 69516 feature vectors in the training set. Of these 29438 of those are from positive videos while 40078 are from negative videos. In the training, we randomly select half of the elements of the negative set and one-tenth of the elements of the positive set. The SVM model obtained from these vectors is then tested using the whole dataset. We found that using the RBF kernel with γ = 1/N, where N is the number of features, and Cost (c) = 18000 gives the best training results [19].
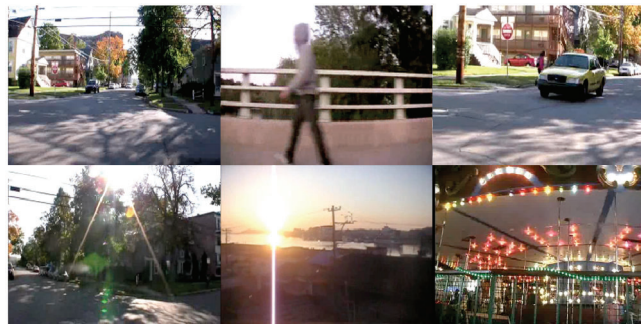


Figure 1: Sample image frames from the positive training video clips.

K. Dimitropoulos O. Gunay , K. Kose, F. Erden, F. Chaabane, F. Tsalakanidou, N. Grammalidis and A.E. Cetin

Figure 2: Sample image frames from the negative training video clips.

Figure 3: Sample image frames from the training video clips that are used to reduce false alarms.

When the system is trained with 34 parameter feature vectors, we get the confusion matrix in Table 1 for the whole set, in which, the number of support vectors is 656. When the system is trained with 55 parameter feature vectors, we get the confusion matrix in Table 2 for the training set, where the number of support vectors is 1309. When we remove the temporal derivatives $I_{t(i,j,n)}$ and $I_{tt(i,j,n)}$ from the property set and obtain 21-parameter feature vectors, we get the confusion matrix in Table 3, with the number of support vectors being 2063. The true detection rates are 96.5%, 95.5% and 91.7% for the 34 parameter, 55 parameter and 21 parameter cases respectively. When 55 parameters are used the detection rate decreases and the number of support vectors increases for our training set. When 21-parameters are used we lose the temporal flickering characteristics of flame-colored blocks and the true detection rate decreases. Therefore, we will use the SVM model obtained using the 34 parameter case in the experiments.

| | | Predicted Labels | |
|---|---|---|---|
| | | Not Fire | Fire |
| Actual | Not Fire | 40069/100.0% | 9/ 0.0% |
| Labels | Fire | 1020/ 3.5% | 28418/96.5% |

Table 1.  Confusion matrix when 34-parameter feature vectors are used

| | | Predicted Labels | |
|---|---|---|---|
| | | Not Fire | Fire |
| Actual | Not Fire | 40023/99.9% | 55/ 0.1% |
| Labels | Fire | 1314/ 4.5% | 28124/95.5% |

Table 2.  Confusion matrix when 55-parameter feature vectors are used

| | | Predicted Labels | |
|---|---|---|---|
| | | Not Fire | Fire |
| Actual | Not Fire | 40007/99.8% | 71/ 0.2% |
| Labels | Fire | 2429/ 8.3% | 27009/91.7% |

Table 3.  Confusion matrix when 21-parameter feature vectors are used

During the implementation, in each spatio-temporal block, the number of chromatic fire-pixels, $\sum_i \sum_j \sum_n \Psi(i,j,n)$, is found. If this number is higher than or equal to two-fifths of the number of elements in the block ($16 \times 16 \times F_{rate}$), then that block is classified as a flame-colored block.

We also assume that the size of the image frames in video is 320x240. If not the video is scaled to 320 by 240 in order to run the fire detection algorithm in real-time.

In the proposed method, $16 \times 16 \times F_{rate}$ blocks are extracted from various video clips. The temporal dimension of the blocks is determined by the frame rate parameter $F_{rate}$ which ranges between 10 and 25 in our training and test videos. These blocks do not overlap in spatial domain but there is fifty percent overlap in time domain. This means that classification is not performed for each frame of the video. A support vector machine (SVM) is used for classification. The resulting system runs in real-time in a PC that has a Core 2 Duo 2.2 GHz processor and the video clips are generally processed at around 20 fps when image frames of size 320 by 240 are used. The detection resolution of the algorithm is determined by the video block size.

The proposed method is compared with one of our previous fire detection methods [6]. At the final step of the flame detection method, a confidence value is determined according to the number of positively classified video blocks and their positions. After every block is classified, if there is no neighbor block classified as fire, the confidence level is set to 1. If there is a single neighbor block, which is classified as fire, then the confidence level is set to 2. If there are two or more neighbor blocks classified as fire, then the confidence level of that block is set to 3. In the decision process, if the confidence level of any block of the frame is greater than or equal than 3, then that frame

Table 4. Comparison of the proposed
method with the previous method
proposed in [6] in terms of true
detection rates in video clips that
contain fire and false alarm rates in
video clips that do not contain fire.

| True Detection Rates | | |
|---|---|---|
| Video name | Proposed | Old([6]) |
| posVideo1 | 54.9% | 0.0% |
| posVideo2 | 81.0% | 0.0% |
| posVideo3 | 81.4% | 0.0% |
| posVideo4 | 99.3% | 37.9% |
| posVideo5 | 90.5% | 73.9% |
| posVideo6 | 97.7% | 0.0% |
| posVideo7 | 98.2% | 9.7% |
| posVideo8 | 94.9% | 77.0% |

| False Alarm Rates | | |
|---|---|---|
| Video name | Proposed | Old([6]) |
| negVideo1 | 3.5% | 5.7% |
| negVideo2 | 0.0% | 0.0% |
| negVideo3 | 0.0% | 0.0% |
| negVideo4 | 7.3% | 0.0% |
| negVideo5 | 2.3% | 51.9% |
| negVideo6 | 0.0% | 0.0% |
| negVideo7 | 0.0% | 0.0% |
| negVideo8 | 0.8% | 3.1% |

is marked as a fire containing frame. The method described in [6] has a similar confidence level metric to determine the alarm level. Results are summarized in Table 4 in terms of the true detection and the false alarm ratios, respectively. The true detection rate in a given video clip is defined as the number of correctly classified frames containing fire divided by the total number of frames which contain fire. Similarly, the false alarm rate in a given test video is defined as the number of misclassified frames, which do not contain fire divided by the total number of frames which do not contain fire.

Compared to the previous method the new method has higher true detection rate in all of the videos that contain actual fires. In some of the videos that do not contain fire, the older method has a lower false alarm rate than the new method. Some of the positive videos in the test set are recorded with hand-held moving cameras and since the old method assumes a stationary camera for background subtraction it cannot correctly classify most of the actual fire regions.

*2.1.2. Flame detection combining multiple features and SVM or ruled-based classification:*

In this section, another video based flame detection algorithm [20] that was developed within FIRESENSE project is presented. The proposed algorithm initially applies background subtraction and colour analysis processing to identify candidate flame regions on the image

and subsequently distinguishes between fire and non-fire objects based on a set of extracted features such as colour probability, contour, wavelet energy, spatio-temporal energy and flickering.

More specifically, in the first processing step, moving pixels are detected using Adaptive Median method [21]. The second processing step aims to filter out non fire- coloured moving pixels. Only the remaining pixels are considered for blob analysis, thus reducing the required computational time of the whole processing. To filter out non-fire moving pixels, we compare their values with a predefined RGB colour distribution created by a number of pixel-samples from video sequences containing real fires.

Let $x_1$, $x_2$,...,$x_N$ be $N$ fire-coloured samples of the predefined distribution. Using these samples, the probability density function of a pixel $x_t$ can be non-parametrically estimated using the kernel $K_h$ [21] as:

$$\Pr(x_t) = \frac{1}{N}\sum_{i=1}^{N} K_h(x_t - x_i)$$

If we choose our kernel estimator function, $K_h$, to be a Gaussian kernel, $K_h=N(0,S)$, where $S$ represents the kernel function bandwidth, and we assume a diagonal correlation matrix $S$ with a different kernel bandwidths $\sigma_j$ for the $j^{th}$ colour channel, then the density can be estimated as:

$$\Pr(x_t) = \frac{1}{N}\sum_{i=1}^{N}\prod_{j=1}^{3} \frac{1}{\sqrt{2\pi\sigma_j^2}} e^{-\frac{1}{2}\frac{(x_{t_j}-x_{i_j})^2}{\sigma_j^2}}$$

After the blob analysis step, the colour probability of each candidate blob is estimated by summing the colour probabilities of all pixels in the blob.

Using this probability estimation, the pixel is considered as a fire-coloured pixel if $\Pr(x_t)<th$, where the threshold th is a global threshold for all samples of the predefined distribution and can be adjusted to achieve a desired percentage of false positives. Hence, if the pixel has a RGB value which belongs to the distribution of Figure 4(b), then it is considered as a fire-coloured pixel as shown in Figure 5(a-b). After the blob analysis step, the colour probability of each candidate blob is estimated by summing the colour probability of each pixel in this blob.

K. Dimitropoulos O. Gunay , K. Kose, F. Erden, F. Chaabane, F. Tsalakanidou, N. Grammalidis and A.E. Cetin

Figure. 4: (a) RGB colour distribution
and (b) the colour distribution with a
global threshold around each
sample.
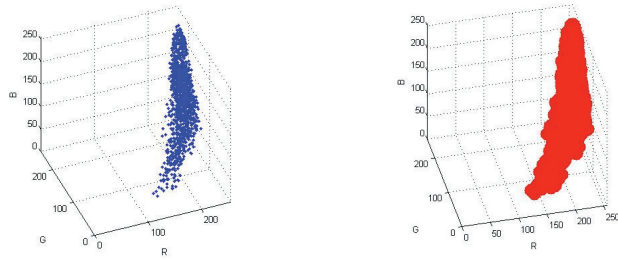
$$\mathrm{Pr}_{blob} = \frac{1}{N}\sum_{i=1}^{N}\mathrm{Pr}(x_i)$$

where N is the number of the pixels in this blob and Pr($x_i$) is the probability estimation of each pixel.

Figure 5: (a) Initial image (b) colour mask.



(a)            (b)

The next processing step concerns the contour of the blob. In general, shapes of flame objects are often irregular, so high irregularity/variability of the blob contour is also considered as a flame indicator. This irregularity is identified by tracing the object contour, starting from any pixel on it.

The third feature concerns the spatial variation in a blob. Usually, there is higher spatial variation in regions containing fire compared to fire-coloured objects. To this end, a two-dimensional wavelet is applied on the red channel of the image, and the final mask is obtained by adding low-high, high-low and high-high wavelet sub-images.

$$E(i,j) = \frac{1}{N}(HL(x,y)^2 + LH(x,y)^2 + HH(x,y)^2)$$

where N is the total number of pixels.

For each blob, the spatial wavelet energy is estimated by summing the individual energy of each pixel belonging to the blob.

$$E_{blob} = \frac{1}{N_b} \sum_{i,j}^{1} E(i,j)$$

where $N_b$ is the number of pixels in a blob.

However, the spatial energy within a blob region changes, since the shape of fire changes irregularly due to the airflow caused by wind or the type of burning material. For this reason, another (fourth) feature is extracted considering the spatial variation in a blob within a temporal window of N frames. The variance of a pixel's spatial energy is estimated as follows:

$$V(i,j) = \frac{1}{N} \sum_{t=1}^{N} (E_t(i,j) - \bar{E}(i,j))^2$$

where N is the size of the temporal window, $E_t$ is the spatial energy of the pixel in time instance t and $\bar{E}$ is the average value of its spatial energy. For each blob, the total spatio-temporal energy, $S_{blob}$, is estimated by summing the individual energy of its pixels:

$$S_{blob} = \frac{1}{N_b} \sum_{i,j}^{1} V(i,j)$$

where $N_b$ is the number of pixels in a blob.

The final feature concerns the detection of flickering within a region of a frame. In our approach, we use a temporal window of N frames (N equals 50 in our experiments), yielding an 1-D temporal sequence of N binary values for each pixel position. Each binary value is set to 0 or 1 if



Figure 6: Flame creates high spatio-temporal energy values.

K. Dimitropoulos O. Gunay , K. Kose, F. Erden, F. Chaabane, F. Tsalakanidou, N. Grammalidis and A.E. Cetin

the pixel was labelled as "no flame candidate" or "flame candidate" respectively after the background extraction and colour analysis steps. To quantify the effect of flickering, we traverse this temporal sequence for each "flame candidate" pixel and measure the number of transitions between "flame candidate" and "no flame candidate" (0->1). The number of transitions can directly be used as a flame flickering feature, with flame regions characterized by a sufficiently large value of flame flickering.

For the classification of the 5-dimensional feature vectors, we employed a Support Vector Machines (SVM) classifier with RBF kernels. The training of the SVM classifier was based on approximately 500 feature vectors extracted from 500 frames of fire and non-fire video sequences. In addition to SVM, a second classification approach, which is based on a number of thresholds and rules, was also adopted. More specifically, a threshold $th_i$ is empirically defined for each feature $i$ after a number of experiments (Colour probability: $th_1$ = 0.002, Spatial wavelet energy: $th_2$ =100, Temporal energy: $th_3$ =20, Spatio-temporal variance: $th_4$ =30, Contour: $th_5$ =0.8). Then, the following rule is applied for each feature vector: If C>M with 1£M£5, then the feature vector is classified as a fire, otherwise it is considered as a false alarm i.e. non-fire (in our experiments M=3). The value of metric C for each feature vector $f_i$ is given by the following equation:

$$C = \sum_{i=1}^{5} F(th_i . f_i)$$

where $F$ is a function defined as follows:

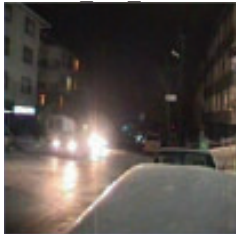Figure 7: Experimental results with videos containing real fires
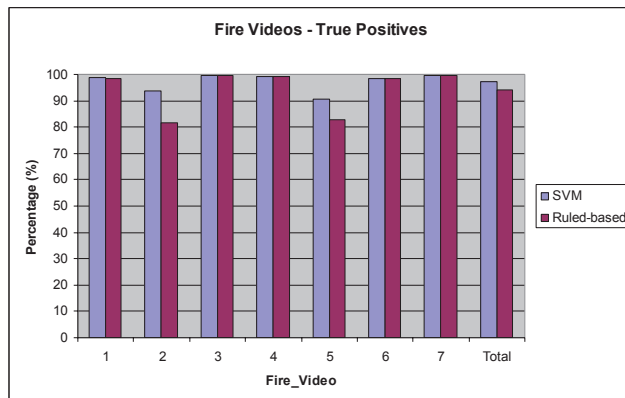
Fire_video1



Non_fire_video1



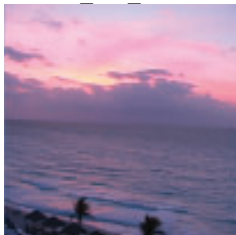Fire_video2



Non_fire_video2



Fire_video3



Non_fire_video3



Fire_video4



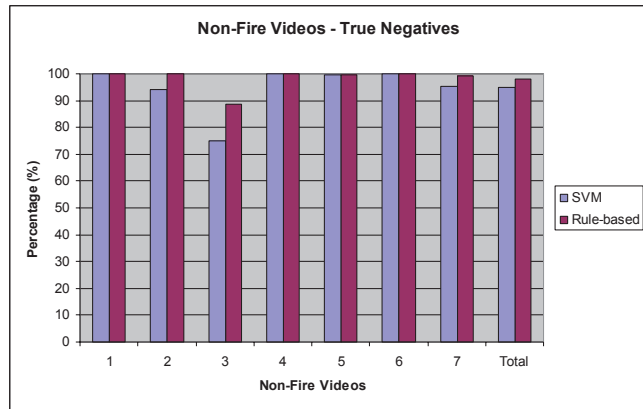Non_fire_video4



Fire_video5



Non_fire_video5



Fire_video6



Non_fire_video6

Figure 8: Experimental results with
videos containing fire coloured
objects.



**Non-Fire Videos - True Negatives**

*2.1.3. Flame detection using features fusion based on a fuzzy classifier:*

Another technique for video-based fire detection in video sequences
that was developed within the FIRESENSE project is based on feature
fusion using a fuzzy classifier.

Initially, to reduce the computational cost, a moving object
detection step is applied to minimize the number of fire objects
candidates. In the literature, many background extraction techniques
exist to estimate and update the background and the foreground
(moving objects) on each frame.  In this case, we adopted the
Adaptive Background with Persistent Pixels (ABPP) [23] method. It is
based on updating background by pixels whose intensity is stable over
*N* consecutive frames. The ABPP method is not the most efficient one
in terms of detection, but rather in terms of computation time, which is
needed for this application. Still, the ABPP detection performance is
very sufficient. After detecting moving objects, a hard study was
performed to define some criteria which better characterize the flame
by extracting the most relevant features.

This step is very important since it is directly related to fire
identification step. Five different features have been defined and
chosen to identify flame regions:

**Colour [24]**: Flame color is very remarkable and specific but there
are similar flame objects (sun and lights) that can share the same color
model and cause false alarms. Thus the colour descriptor is very
constructive but not sufficient. The selected colour model is introduced
to overcome lighting change and low quality recording conditions
problems. Let *R, G* and *B* be the red, green and blue channels of
pixel (m,n):

**Rule 1**: : $R(m,n) > R_T$

**Rule 2**: : $R(m,n) > G(m,n) > B(m,n)$

Rule 3: : $0.25 \leq \dfrac{G(m,n)}{R(m,n)+1} \leq 0.65$

Rule 4: : $0.20 \leq \dfrac{B(m,n)}{G(m,n)+1} \leq 0.60$

Rule 5: : $0.20 \leq \dfrac{B(m,n)}{G(m,n)+1} \leq 0.60$

The normalization ensures the contribution of this model by overcoming the lighting change. Based on the above rules, a binary mask is generated characterizing the flame colour information.

**Temporal intensity variance [25]:** An important flame characteristic is that inside the object, the intensity changes randomly and quickly. This variation can be calculated by a temporal variance: if the tested object presents a high temporal variance value, it will be considered as a flame candidate and surpass this test. Let $I(x, y, t)$ be the pixel $(x, y)$ intensity (gray scale or mean of three channels), if the brightness value changes remarkably between two frames $\Delta I = |I(x, y, t) - I(x, y, t - 1)| > T_1$, a counter called SUM is incremented. A pixel is then regarded as part of the frame flickering flame if its oscillation registration counter SUM exceeds a threshold.

**Spatial intensity variance [24]: Fire** regions are characterized by a significant amount of texture because of their random nature. This characteristic can discriminate flame from fire coloured objects e.g. a car lights. For each moving object, the spatial intensity variance feature described in [24] is calculated and a related threshold is applied to detect fire region candidates.

**Shape variation [24]:** Fire objects are characterized by a significant change of their areas between two consecutive frames because of their random nature. Non-fire areas have a less random change in the area size. This feature is quantized by the term of the $i^{th}$ frame:

$$\Delta A = \frac{\left| A_i - A_{i-1} \right|}{A_i}$$

Where $A_i$ corresponds to the $i^{th}$ frame object area. A related threshold is applied to discriminate fire objects candidates.

**Shape complexity [26]:** In many cases, flame objects correspond to complex shape. This feature can be evaluated by the coefficient C:

$$C = \frac{L^2}{S}$$

Where L corresponds to the shape perimeter and S to the shape surface.

**Features fusion:** After extracting the aforementioned fire characteristics, feature fusion is performed to extract flame objects. A symmetric and associative fusion operator $\sigma$, defined from [0,1]x[0,1] to [0,1] is used for this task:

$$\sigma(x, y) = \frac{xy}{1 - x - y + 2xy}$$

This operator belongs to fuzzy Context Independent Variable Behaviour (CIVB) classifiers [27]. It has a variable behaviour according to *x* and *y* values:

- A conjunctive behaviour (severe), if *max(x, y) ≥ 0.5* then *σ (x,y) < min(x, y)*, providing a result which confirms the event more than each individual information.
- A disjunctive behaviour (indulgent), if *min(x,y) ≤ 0.5* then *σ (x, y) > max(x, y)*, resulting in a stronger disconfirmation than each individual information.
- A compromise if *x ≤ 0.5 ≤ y* then *x ≤ σ (x, y) ≤ y;* and the reverse in equality holds if *y ≤ 0.5 ≤ x*; the sign of the result depends on the strength of disconfirmation (respectively, confirmation) of each individual information.

To study fusion detection efficiency according to the contribution of each feature in discriminating fire objects, ROC curves (Figure 9) are calculated for four fire video sequences.

We can notice that the curve related to the features fusion (yellow curve) is almost all of the time over the others curves, which means that it takes contributions from all features and presents the best detection. This also confirms the complementarities of these features in the flame detection procedure. Table 6 shows that all the features have potential contributions, in dependence to the sequence scene.

The computational time of the flame detection algorithm is about 60ms for a 320x240 size image, which is considered sufficient for real time detection.
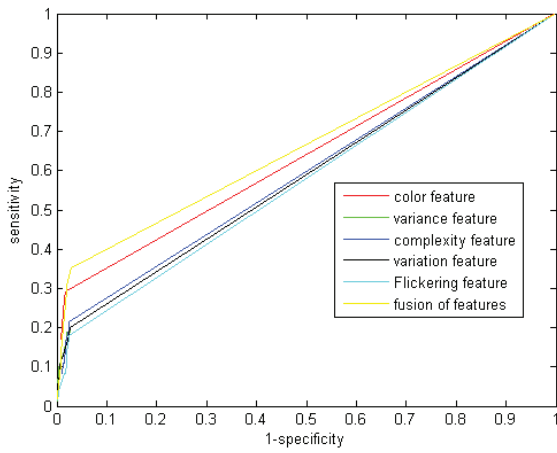
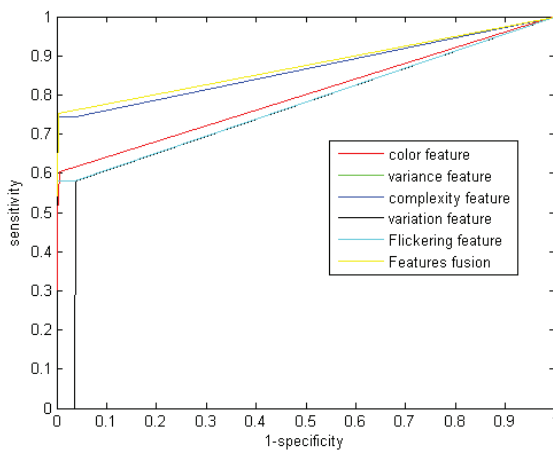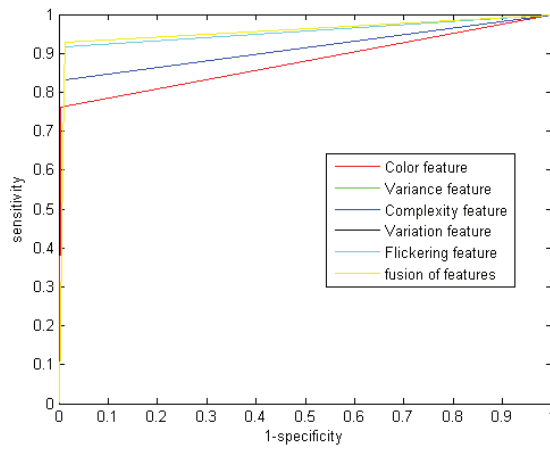Figure 9: ROC curves associated to fire individual features and fusion results.

K. Dimitropoulos O. Gunay , K. Kose, F. Erden, F. Chaabane, F. Tsalakanidou, N. Grammalidis and A.E. Cetin

Table 6. Features contributions in fusion process.

| Features/Sequences | Seq N°1 | Seq N°2 | Seq N°3 | Seq N°4 | Seq N°5 |
|---|---|---|---|---|---|
| Color | | X | | X | |
| Temporal intensity variance | X | X | | | |
| Spatial intensity variance | | | X | | X |
| Shape variation | | | X | | X |
| Shape complexity | X | X | | X | |

## 3. Experimental evaluation of flame detection algorithms:

The last two algorithms are based on background subtraction for flame detection and for this reason they are not applicable to moving camera scenes. On the other hand, the technique based on correlation descriptors does not employ background subtraction; therefore, camera movement does not cause any problem for this method. Another minor problem is the shake of cameras due to the effect of the wind. In this case, image registration techniques can be effectively applied to address the problem. To evaluate the performance of flame detection algorithms, we used a set of video sequences from the FIRESENSE database [28], i.e. a data set of fire and non-fire videos has been made available to the research community (Figure 10). The true positive rate is the number of frames in which fire is correctly detected out of the total number of frames in a fire test video, while the false positive rate is defined as the number of frames in which fire was erroneously detected out of the total number of frames in a non-fire test video. Some non-fire video sequences (1, 4, 5, 6, 7 and 16) contain moving camera scenes, and therefore algorithms based on background subtraction are not applicable. As seen in Figure 11(a-b), the average true positive rates for the proposed algorithms in video sequences containing fire are: correlation descriptors (82,43%), SVM-based (99.7%), ruled-based (96.31%) and features-fusion (fuzzy-based) (92.77%). Similarly, the false positive rates in non-fire video sequences are: correlation descriptors (2.17%), SVM-based (41.13%), ruled-based (13.8%) and feature fusion (fuzzy-based) (55.17%).
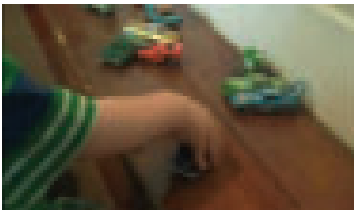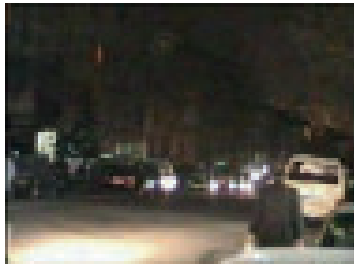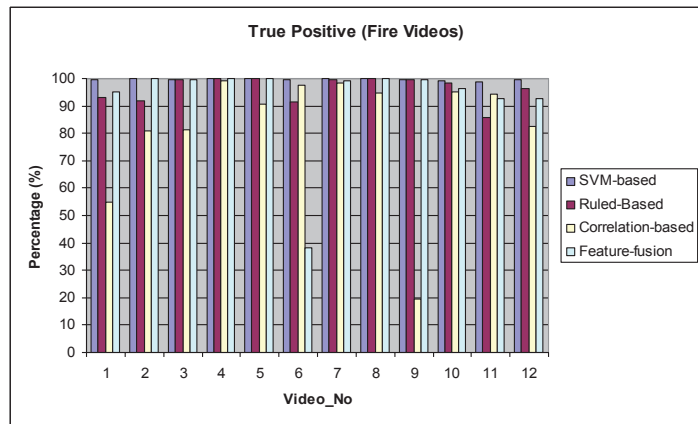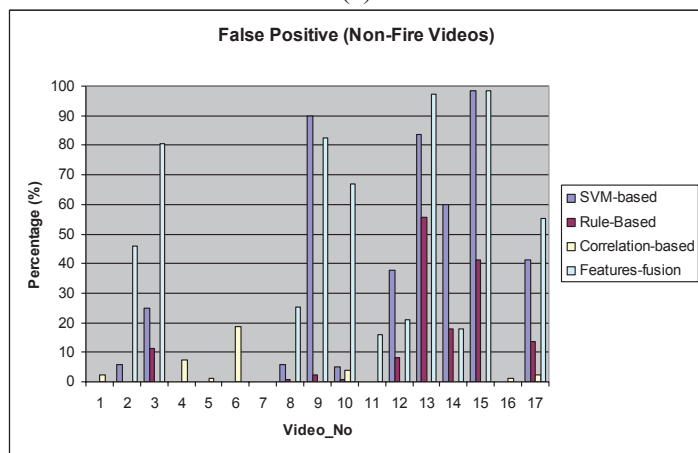
Figure 10: Indicative fire and non-fire videos from the FIRESENSE database.

Figure 11: Evaluation results of flame detection algorithms. (a) True positive rates in videos containing fire and (b) False positive rates in non-fire test videos.

**True Positive (Fire Videos)**



(a)

**False Positive (Non-Fire Videos)**



(b)

## 4. Conclusions

Early detection of fire is crucial for the suppression of wildfires and minimization of human losses and damages to cultural heritage and archaeological sites. In this paper, three video-based flame detection techniques, which were developed within the FIRESENSE EU research project, are presented and compared. Currently, within FIRESENSE project, these techniques are further evaluated so as to be used for the protection of five FIRESENSE test sites in Greece, Turkey, Italy and Tunisia.

## 5. Acknowledgement

# References

[1] Fleming, J. , Robertson, R. G., 2003. Fire Management Tech Tips: The Osborne Fire Finder. T. R. 1311-SDTDC, USDA Forest Service, 2003.

[2] Advanced Very High Resolution Radiometer – AVHRR, http://noaasis.noaa.gov/NOAASIS/ml/avhrr.html (accessed 5 Feb 2013).

[3] Modis Web Page. http://modis.gsfc.nasa.gov, (accessed 5 Feb 2013).

[4] Grammalidis, N., Cetin, A.E., Dimitropoulos, K., Tsalakanidou, F., Kose, K., Gunay, O., Gouverneur, B., Torri, D., Kuruoglu, E., Tozzi, S., Benazza, A., Chaabane, F.,  Kosucu, B.  and Ersoy, C.,  2011. A Multi-sensor Network for the Protection of Cultural Heritage, *19th European Signal Processing Conference (EUSIPCO2011)*, Special Session on Signal processing for disaster management and prevention, Barcelona, Spain

[5] Horng W-B, Peng, J-W, Chen, C-Y, 2005. A new image-based real-time flame detection method using color analysis, *Networking, Sensing and Control,  Proceedings 2005 IEEE*, vol., no., pp. 100-105, 19-22 March 2005.

[6] Töreyin, B. U., Dedeoglu, Y., Güdükbay, U. and Çetin A.E., 2006. "Computer vision based method for real-time fire and flame detection", *Pattern Recognition Letters*, 27(1), pp. 49-58.

[7] Collins, R.T., Lipton, A.J., Kanade, T., 1999. A system for video surveillance and monitoring. In: *Proc. American Nuclear Society (ANS) Eighth International Topical Meeting on Robotics and Remote Systems*, Pittsburgh, PA.

[8] Changwoo Ha; Ung Hwang, Gwanggil Jeon, Joongwhee Cho; Jechang Jeong, 2012. Vision-Based Fire Detection Algorithm Using Optical Flow, *Complex, Intelligent and Software Intensive Systems (CISIS), Sixth International Conference on* , pp. 526-530, 4-6 July 2012

[9] Jinhua Zhang, Jian Zhuang, Haifeng Du, 2006. A New Flame Detection Method Using Probability Model, *Computational Intelligence and Security, 2006 International Conference on* , vol.2, no., pp.1614-1617, 3-6 Nov. 2006.

[10] Yamagishi, H., Yamaguchi, J., 2000. A contour fluctuation data processing method for fire flame detection using a color camera, *Industrial Electronics Society, 2000. IECON 2000. 26th Annual Conference of the IEEE* , vol.2, no., pp.824-829 vol.2.

[11] Phillips, W., III, Shah, M., Da Vitoria Lobo, N.,2000. Flame recognition in video, *Applications of Computer Vision, 2000, Fifth IEEE Workshop on.* , vol., no., pp.224-229.

[12] Ko, B.,  Cheong, K., Nam, J., 2009. Fire detection based on vision sensor and support vector machines, *Fire Safety Journal*, Volume 44, Issue 3, pp. 322-329

[13] Günay, O., Taşdemir, K., Töreyin, B. U, Çetin, A. E., 2009. Video based wildfire detection at night, *Fire Safety Journal*, Volume 44, Issue 6, pp. 860-868

[14] Dedeoglu, Y., Toreyin, B.U., Gudukbay, U., Cetin, A.E., 2005. Real-Time Fire and Flame Detection in Video, *Acoustics, Speech, and Signal Processing, 2005. Proceedings. (ICASSP '05). IEEE International Conference on* , vol.2, no., pp. 669- 672.

[15] Che-Bin Liu, Ahuja, N., 2004. Vision based fire detection, *Pattern Recognition, ICPR 2004. Proceedings of the 17th International Conference on* , vol.4, no., pp. 134- 137 Vol.4

[16] Chen Jun, Du Yang, Wang Dong, 2009. An Early Fire Image Detection and Identification Algorithm Based on DFBIR Model, *Computer Science and Information Engineering, 2009 WRI World Congress on*, pp. 229-232.

[17] Habiboglu, H., Gunay, O., Cetin, A. E., 2011. Covariance matrix-based fire and flame detection method in video, *Machine Vision and Applications*, DOI: 10.1007/s00138-011-0369-1, pp.1-11.

[18] Tuzel, O., Porikli, F. and P. Meer, 2006. Region covariance: A fast descriptor for detection and classification, *9th European Conference on Computer Vision, (ECCV 2006)*, pp. 589-600, Graz, Austria.

[19] Chang, C.-C., Lin, C.-J, 2001. LIBSVM: a library for support vector machines. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm (accessed 5 Feb 2013).

[20] Dimitropoulos, K., Tsalakanidou, F. and Grammalidis, N., 2012. Flame Detection For Video-Based Early Fire Warning Systems and 3D Visualization of Fire Propagation, *13th IASTED International Conference on Computer Graphics and Imaging (CGIM 2012)*, Crete, Greece

[21] McFarlane, N. and Schofield, C., 1995. Segmentation and tracking of piglets in images, *British Machine Vision and Applications*, Vol 8, pp. 187-193.

[22] Elgammal, A. Harwood, D., Davis L.,2000. Non-parametric model for background subtraction, Proc. *6th European Conference on Computer Vision*, Vol. 1843, pp. 751-767, Dublin, Ireland.

[23] Kang, S. , Paik, J. , Koschan, A., Abidi, B. and Abidi, M. A., 2003. "Real-Time video tracking using PTZ cameras", *Proc. of SPIE 6th International Conference on Quality Control by Artificial Vision*, Vol. 5132, pp. 103-111, Gatlinburg, TN.

[24] Borges, P. V. K., Mayer, J. and Izquierdo, E., 2008. "Efficient Visual Fire Detection Applied For Video Retrieval", *16th European Signal Processing Conference (EUSIPCO 2008)*, Lausanne, Switzerland.

[25] Chen, J., He, Y., Wang, J., 2010. Multi-Feature Fusion Based Fast Video Flame Detection, *Building and Environment*, Vol. 45, pp.1113-1122.

[26] Zhang, D., Han, S., Zhao, J., Zhang, Z., Qu, C., Ke, Y., Chen, X., 2009. Image Based Forest Fire Detection Using Dynamic Characteristics With Artificial Neural Networks, *International Joint Conference on Artificial Intelligence*, Pasadena, California, USA.

[27] Bloch, I., 1996. Information Combination Operators for Data Fusion: A Comparative review with Classification, *Systems, Man and Cybernetics*, Vol. 26, n° 1, pp. 52-67.

[28] FIRESENSE Database, http://www.firesense.eu/, (accessed 5 Feb 2013).