

Flame Detection for Video-Based Early Fire Warning for the Protection of Cultural Heritage

K. Dimitropoulos¹, O. Gunay², K. Kose², F. Erden², F. Chaabene³,
F. Tsalakanidou¹, N. Grammalidis¹, and E. Cetin²

¹ Information Technologies Institute, Centre for Research and Technology Hellas, Greece
{dimitrop, ngramm, filareti}@iti.gr

² Department of Electrical and Electronics Engineering, Bilkent University, Turkey
{gunayosman, erdenfatih}@gmail.com, kkivanc@ee.bilkent.edu.tr,
cetin@bilkent.edu.tr

³ Ecole Supérieure des Communication de Tunis, Sup'Com, Tunisia
ferdaous.chaabene@supcom.rnu.tn

Abstract. Cultural heritage and archaeological sites are exposed to the risk of fire and early warning is the only way to avoid losses and damages. The use of terrestrial systems, typically based on video cameras, is currently the most promising solution for advanced automatic wildfire surveillance and monitoring. Video cameras are sensitive in visible spectra and can be used either for flame or smoke detection. This paper presents and compares three video-based flame detection techniques, which were developed within the FIRESENSE EU research project.

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

1 Introduction

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. 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) [6] 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 [1], launched by the National Oceanic and Atmospheric Administration (NOAA) in 1998 and the Moderate Resolution Imaging Spectroradiometer (MODIS) [10], 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 < 1 km) 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 [7] 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.

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 [8], [12] 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:

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

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 and c) corresponding second order horizontal and vertical derivatives and d) corresponding 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:

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

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.

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. Since we require three neighbouring blocks to reach the highest confidence level, the fire should occupy a region of size 32 by 32 in video.

The proposed method is compared with one of our previous fire detection method [11]. In the decision process, if the confidence level of any block of the frame is greater than or equal than 3, then that frame is marked as a fire containing frame. The method described in [11] has a similar confidence level metric to determine the alarm level. Results are summarized in Table 1 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.

Table 1. Comparison of the proposed method with the previous method proposed in [11] 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			False Alarm Rates		
Video name	Proposed	Old([11])	Video name	Proposed	Old([11])
posVideo1	54.9%	0.0%	negVideo1	3.5%	5.7%
posVideo2	81.0%	0.0%	negVideo2	0.0%	0.0%
posVideo3	81.4%	0.0%	negVideo3	0.0%	0.0%
posVideo4	99.3%	37.9%	negVideo4	7.3%	0.0%
posVideo5	90.5%	73.9%	negVideo5	2.3%	51.9%

Table 1. (continued)

posVideo6	97.7%	0.0%	negVideo6	0.0%	0.0%
posVideo7	98.2%	9.7%	negVideo7	0.0%	0.0%
posVideo8	94.9%	77.0%	negVideo8	0.8%	3.1%

2.1.2 Flame Detection Combining Multiple Features and SVM or Ruled-Based Classification

In this section, another video based flame detection algorithm [14] that was developed within FIRESENSE project is briefly summarized. 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 a simple median average background subtraction algorithm. 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. 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. The probability density function of a moving pixel is non-parametrically estimated, according to the technique proposed in [15].

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.

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. For each blob, spatial wavelet energy is estimated by summing the individual energy of each pixel. 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 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 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” ($0 \rightarrow 1$). 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 $1 \leq M \leq 5$, 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:

$$F(th, f) = \begin{cases} 0 & , f < th \\ 1 & , f > th \end{cases}$$

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) [9] 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. 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 [3]: This model of colour is defined 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 4: $0.05 \leq \frac{B(m, n)}{R(m, n) + 1} \leq 0.45$

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

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

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

Based on the above rules, a binary mask is generated characterizing the flame colour information.

Temporal Intensity Variance [4]: 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 [3]: 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 [3] is calculated and a related threshold is applied to detect fire region candidates.

Shape Variation [3]: 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: $\Delta A_i = |A_i - A_{i-1}| / A_i$, where A_i corresponds to the i^{th} frame object area. A related threshold is applied to discriminate fire objects candidates.

Shape Complexity [13]: In many cases, flame objects correspond to complex shape. This feature can be evaluated by the coefficient $C = L^2 / S$, 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 σ is used for this task: $\sigma(x, y) = xy / (1 - x - y + 2xy)$.

This operator belongs to fuzzy Context Independent Variable Behaviour (CIVB) classifiers [2]. It has a variable behaviour according to x and y values:

- A conjunctive behaviour (severe), if $\max(x, y) \geq 0.5$ then $\sigma(x, y) < \min(x, y)$, providing a result which confirms the event more than each individual information.
- A disjunctive behaviour (indulgent), if $\min(x, y) \leq 0.5$ then $\sigma(x, y) > \max(x, y)$, resulting in a stronger disconfirmation than each individual information.
- A compromise if $x \leq 0.5 \leq y$ then $x \leq \sigma(x, y) \leq y$; and the reverse in equality holds if $y \leq 0.5 \leq x$; 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 (Fig. 2) are calculated for four fire video sequences.

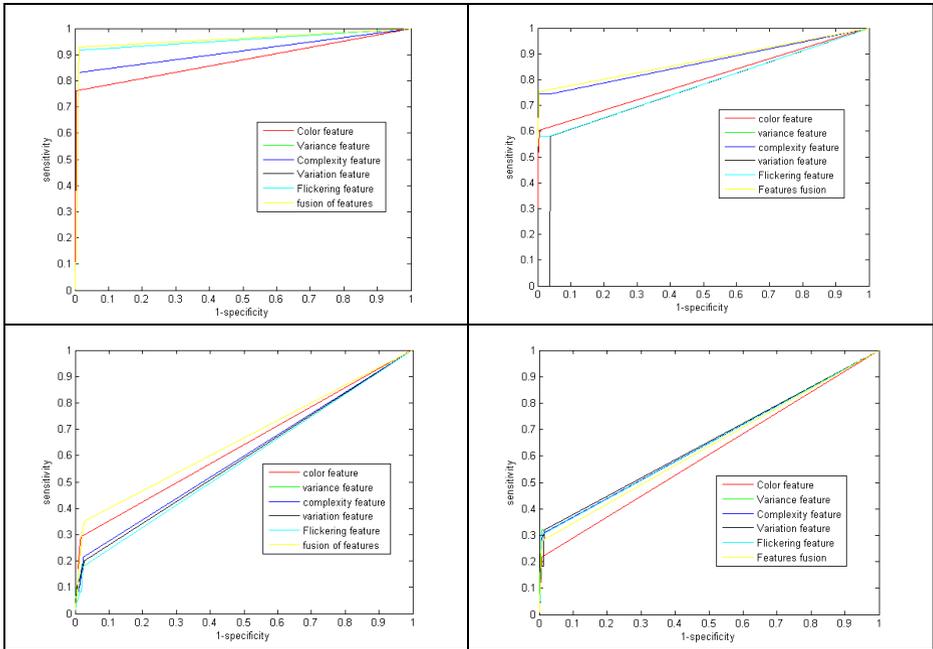


Fig. 1. ROC curves of features and fusion result

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 complementarity of these features in the flame detection procedure. The computational time of the flame detection algorithm is about 60ms for a 320x240 size image, which is considered sufficient for real time detection.

2.1.4 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 [5], i.e. a data set of fire and non-fire videos has been made available to the research community. 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 Fig. 3(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%).

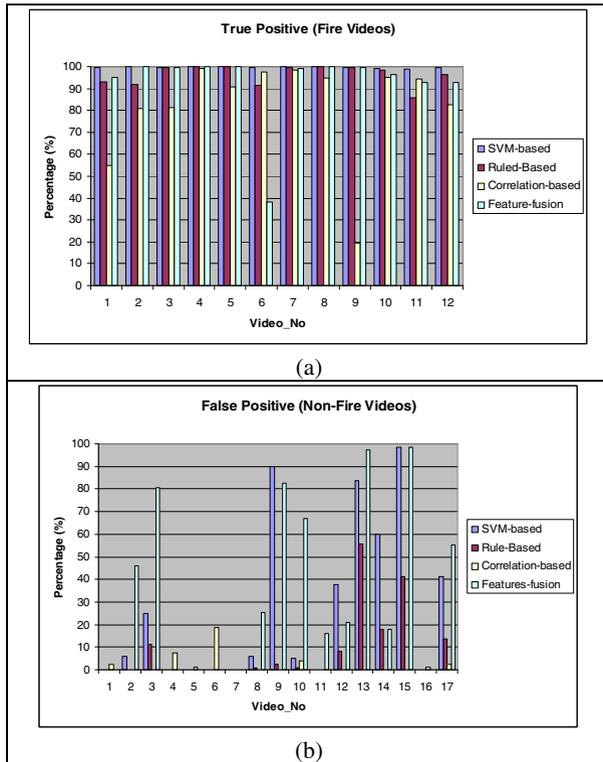


Fig. 2. Evaluation results of flame detection algorithms. (a) True positive rates in videos containing fire and (b) False positive rates in non-fire test videos.

3 Conclusions

Early detection of fire is crucial for the suppression of wildfires and minimization of human losses and damages. Especially for the protection of archaeological sites, video based systems can be efficiently used, enabling the full coverage of the surrounding area with a small number of PTZ cameras. In this paper, three video-based flame detection techniques, which were developed within the FIRESENSE EU research project, are presented and compared. In the future, these techniques will be further evaluated under real operation condition in five selected cultural heritage test sites.

Acknowledgement. The research leading to these results has received funding from the European Community's FP7 under grant agreement no FP7-ENV-244088 "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".

References

1. Advanced Very High Resolution Radiometer – AVHRR, <http://noaaasis.noaa.gov/NOAASIS/ml/avhrr.html> (accessed May 20, 2010)
2. Bloch, I.: Information Combination Operators for Data Fusion: A Comparative review with Classification. *Systems, Man and Cybernetics* 26(1), 52–67 (1996)
3. Borges, P.V.K., Mayer, J., Izquierdo, E.: Efficient Visual Fire Detection Applied For Video Retrieval. In: 16th European Signal Processing Conference (EUSIPCO 2008), Lausanne, Switzerland, August 25-29 (2008)
4. Chen, J., He, Y., Wang, J.: Multi-Feature Fusion Based Fast Video Flame Detection. *Building and Environment* 45, 1113–1122 (2010)
5. FIRESENSE Database (2011), <http://www.firesense.eu/>
6. Fleming, J., Robertson, R.G.: Fire Management Tech Tips: The Osborne Fire Finder. T. R. 1311-SDTDC, USDA Forest Service (2003)
7. Grammalidis, N., Cetin, E., Dimitropoulos, K., Tsalakanidou, F., Kose, K., Gunay, O., Gouverneur, B., Torri, D., Kuruoglu, E., Tozzi, S., Benazza, A., Chaabane, F., Kosucu, B., Ersoy, C.: A Multi-sensor Network for the Protection of Cultural Heritage. In: 19th European Signal Processing Conference (EUSIPCO 2011), Special Session on Signal Processing for Disaster Management and Prevention, Barcelona, Spain, August 29-September 2 (2011)
8. Habiboglu, H., Gunay, O., Cetin, A.E.: Covariance matrix-based fire and flame detection method in video. *Machine Vision and Applications*, 1–11 (September 2011), doi:10.1007/s00138-011-0369-1
9. Kang, S., Paik, J., Koschan, A., Abidi, B., Abidi, M.A.: Real-Time video tracking using PTZ cameras. In: Proc. of SPIE 6th International Conference on Quality Control by Artificial Vision, vol. 5132, pp. 103–111. Gatlinburg, TN (2003)
10. Modis Web Page, <http://modis.gsfc.nasa.gov>
11. Töreyin, B.U., Dedeoglu, Y., Güdükbay, U., Çetin, A.E.: Computer vision based method for real-time fire and flame detection. *Pattern Recognition Letters* 27(1) (2006)
12. Tuzel, O., Porikli, F., Meer, P.: Region Covariance: A Fast Descriptor for Detection and Classification. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3952, pp. 589–600. Springer, Heidelberg (2006)
13. Zhang, D., Han, S., Zhao, J., Zhang, Z., Qu, C., Ke, Y., Chen, X.: Image Based Forest Fire Detection Using Dynamic Characteristics With Artificial Neural Networks. In: International Joint Conference on Artificial Intelligence, Pasadena, California, USA (2009)
14. Dimitropoulos, K., Tsalakanidou, F., Grammalidis, N.: Flame Detection For Video-Based Early Fire Warning Systems And 3D Visualization of Fire Propagation. In: 13th IASTED International Conference on Computer Graphics and Imaging (CGIM 2012), Crete, Greece (2012)
15. Elgammal, A., Harwood, D., Davis, L.: Non-parametric Model for Background Subtraction. In: Vernon, D. (ed.) ECCV 2000. LNCS, vol. 1843, pp. 751–767. Springer, Heidelberg (2000)