This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2014.2339592, IEEE Transactions on Circuits and Systems for Video Technology

Transactions on Circuits and Systems for Video Technology : Spatio-Temporal Flame Modeling and Dynamic Texture Analysis for Automatic Video-Based Fire Detection

Kosmas Dimitropoulos, Panagiotis Barmpoutis and Nikos Grammalidis

***Abstract*—Every year a large number of wildfires all over the world burn forested lands causing adverse ecological, economic and social impacts. Beyond taking precautionary measures, early warning and immediate response are the only ways to avoid great losses. To this end, in this paper we propose a computer vision approach for fire-flame detection to be used by an early-warning fire monitoring system. Initially, candidate fire regions in a frame are defined using background subtraction and color analysis based on a non-parametric model. Subsequently, the fire behavior is modeled by employing various spatio-temporal features such as color probability, flickering, spatial and spatio-temporal energy, while dynamic texture analysis is applied in each candidate region using linear dynamical systems and a bag of systems approach. To increase the robustness of the algorithm, the spatio-temporal consistency energy of each candidate fire region is estimated by exploiting prior knowledge about the possible existence of fire in neighboring blocks from the current and previous video frames. As a last step, a two-class SVM classifier is used to classify the candidate regions. Experimental results have shown that the proposed method outperforms existing state of the art algorithms.**

***Index Terms*—Bag of systems, dynamic textures analysis, fire detection, linear dynamic systems, spatio-temporal modeling**

I. INTRODUCTION

Due to the fact that fire is one of the most harmful natural hazards affecting everyday life around the world, early fire warning systems have attracted particular attention recently. The most advanced approaches in automatic early forest fire detection are based on space borne (satellite), airborne (UAVs - Unmanned Aerial Vehicles) or terrestrial-based systems. Among these, terrestrial systems based on CCD video cameras are considered as the most promising technology for automatic fire detection due to their low cost, high resolution, short time response and easy confirmation of the alarm by a human operator through the surveillance monitor.

For this reason, video-based flame detection techniques have been widely investigated during the last decade. The main challenge in video-based flame detection lies in the modeling of the chaotic and complex nature of the fire phenomenon and the large variations of flame appearance in video. To address this problem many researchers use the motion characteristics of flame as well as the spatial distribution of fire colors in the scene or they try to combine both temporal and spatial characteristics. However, many natural objects have similar behavior with fire, e.g., the sun, various artificial lights or light reflections on various surfaces, dust particles etc, which can often be mistakenly detected as flames. Moreover, scene complexity and low video quality also affect the robustness of vision-based flame detection algorithms, thus, increasing the false alarm rate.

On the other hand, dynamic texture analysis has been successfully applied in the past for the classification of video sequences in multimedia databases. A dynamic texture in video can be simply defined as a texture with motion, i.e., a spatially and time-varying visual pattern that forms an image sequence or part of an image sequence with a certain temporal stationarity [1]. While dynamic texture analysis is also applied to the categorization of sequences containing flame, smoke, steam etc, these general techniques are not used in practical fire detection algorithms due to their high computational cost [2]. Furthermore, most of the existing dynamic texture categorization methods are used to model a complete video sequence or a manually selected image region in a video, hence, they cannot provide to a fire detection system neither any information regarding the exact localization of the fire in the scene nor the time of the fire incident.

Unlike the aforementioned methods, the present work uses both types of flame modeling and makes the following contributions:

- A novel method is proposed for flame detection based on the combination of features extracted from spatio-temporal flame modeling and dynamic texture analysis. In this way, the proposed method not only focuses on the identification of specific fire properties (e.g. color, motion, flickering etc), but exploits also the ability of Linear Dynamical Systems (LDSs) to analyze the temporal evolution of the pixels' intensities in order to

Manuscript submitted 30 October 2013. Revised 10 February, 22 April and 12 June 2014; accepted 24 June.

K. Dimitropoulos, P. Barmpoutis and N. Grammalidis are with Information Technologies Institute, ITI - CERTH, 1st km Thermi-Panorama Rd, Thessaloniki, 57001 GREECE (e-mail: panbar@iti.gr, dimitrop@iti.gr, ngramm@iti.gr).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2014.2339592, IEEE Transactions on Circuits and Systems for Video Technology

increase the robustness of the algorithm.

- The proposed method addresses three significant limitations of LDS-based approaches for dynamic texture categorization. More specifically: i) reduces the computational cost required for the dynamic texture analysis through redundant data reduction, ii) determines the time of the fire incident and iii) localizes the exact position of the fire in the image. To this end, meaningful information is identified by first applying a pre-processing step in order to filter out non-candidate fire regions and then using a sliding time window to exclude candidate blocks located out of the defined subsequence. The exact position of the fire in the current frame can then be defined by combining information from spatio-temporal flame modeling and spatio-temporal consistency energy.
- An efficient flame behavior modeling is introduced in order to identify the color of the fire, the motion characteristics of flame as well as the random variations of its appearance. Specifically, a number of various spatio-temporal features is extracted, such as color probability, flickering, spatial and spatio-temporal energy, for each candidate fire region.
- A novel approach for enhancing the reliability of the algorithm, namely spatio-temporal consistency energy, is introduced by exploiting: i) features extracted from both spatio-temporal flame modeling and dynamic texture analysis and ii) prior knowledge about the possible existence of fire in neighboring blocks from the current and previous video frames. Inspired by MRF approaches, the proposed method estimates an energy cost consisting of two terms: i) a data cost taking into account the features in the current block and ii) a smoothness cost considering the state of neighboring blocks in a 3D image patch.

The remainder of the paper is organized as follows: Related work is presented in detail in Section II. Section III describes the detection of candidate fire regions and presents a general overview of the proposed method. Section IV analyses the spatio-temporal flame modeling, while the dynamic texture analysis is described in Section V. Spatio-temporal consistency energy and classification of the candidate blocks are presented in Section VI and VII respectively. Finally, experimental results are discussed in Section VIII, while conclusions are drawn in Section IX.

II. Related work

To model fire behavior many researchers aim to identify various characteristics of flame. For example, Chen et al. [3] adopted a RGB color model and disorder measurements, while Liu and Ahuja [4] proposed an algorithm that uses spectral, spatial and temporal models of fire regions in visual image sequences. In addition to ordinary motion and color clues, Toreyin et al. [5] detected fire flicker by analyzing the video in the wavelet domain. Quasi-periodic behavior in flame boundaries were detected by performing temporal wavelet transforms. In addition, the color variations in flame regions were detected by computing the spatial wavelet transform of moving fire-colored regions. Furthermore, in [6] Toreyin et al. used a hidden Markov model to mimic the temporal behavior of flame. Specifically, Markov models representing the flame and flame colored ordinary moving objects were used to distinguish flame flicker process from motion of flame colored moving objects, while spatial color variations in flame were also evaluated by the same Markov models. Zhang et al. [7], on the other hand, proposed a contour based forest fire detection method using Fast Fourier Transform and wavelet analysis. The algorithm initially segments fire regions and then uses FFT method to describe the contour, while the calculated Fourier descriptors are analyzed with temporal wavelet.

A different approach was proposed by Celik and Demiral [8], in which a rule-based generic color model for flame pixel classification was introduced. The algorithm uses YCbCr color space to separate luminance from chrominance more effectively than color spaces such as RGB. In addition to translating the rules developed in the RGB and normalized RGB to YCbCr color space, new rules were proposed in YCbCr color space, which further alleviate the harmful effects of changing illumination. Marbach et al. [9] used YUV color model for the representation of video data, where time derivative of luminance component Y was used to declare the candidate fire pixels and the chrominance components U and V were used to classify the candidate pixels to be in the fire sector or not. In addition to luminance and chrominance they have incorporated motion into their work.

In the fire or non-fire classes, Borges and Izquierdo [10] proposed a method that analyzes the frame-to-frame changes of specific low-level features describing potential fire regions such as color, area size, surface coarseness, boundary roughness, and skewness. The behavioral change of each of these features is evaluated, and the results are then combined according to a Bayes classifier for robust fire recognition. Alternatively, Ko et al. [11], proposed hierarchical Bayesian networks for fire-flame detection that contained intermediate nodes. Four probability density functions for evidence at each node were used. These probability density functions were modeled using the skewness of the red color, and three high frequencies obtained from a wavelet transform. Later, Ko et al. also used [12] a fire-flame detection method using fuzzy finite automata (FFA) with probability density functions based on visual features, thereby providing a systemic approach to handle uncertainty in computational systems and the ability to handle continuous spaces by combining the capabilities of automata with fuzzy logic.

More recently, within the FP7 EU-funded Firesense project [13] various flame detection algorithms were developed. More specifically, Habiboglu et al. [14], proposed a video-based fire detection system, which uses color, spatial and temporal information. The system divides the video into spatio-temporal blocks and uses covariance-based features extracted from these blocks to detect fire. Feature vectors take advantage of both the spatial and the temporal characteristics

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2014.2339592, IEEE Transactions on Circuits and Systems for Video Technology

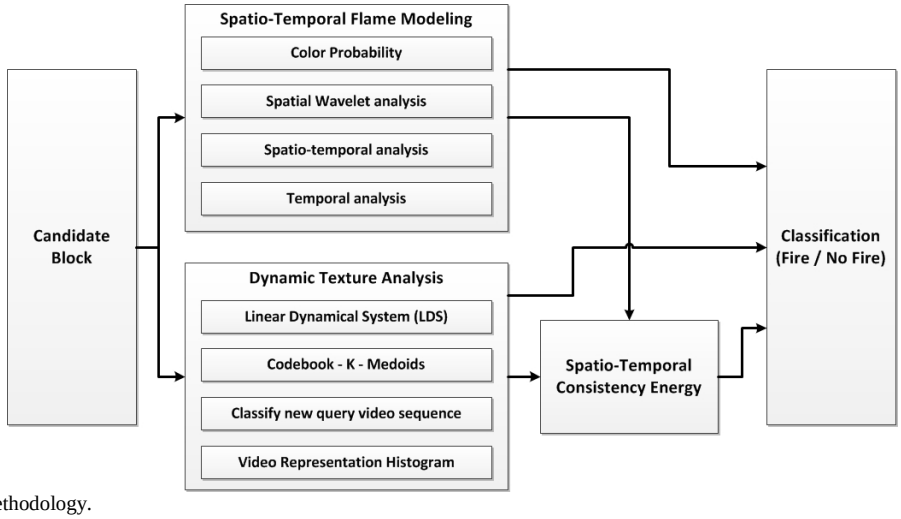


Fig. 1. The proposed methodology.

of flame-colored regions. Furthermore, Dimitropoulos et al. [15] proposed a video flame detection algorithm using a combination of different spatio-temporal features to detect fire. For the discrimination between fire and non-fire regions, two classification methods were investigated: a Support Vector Machine (SVM) classifier and a rule-based approach.

All the aforementioned approaches aim to model the fire behavior based on the identification of high-level cues like flame flickering or spatial color distribution of flame colors and the use of advanced classification techniques to determine whether flames exists in the video images or not. However, problems still exist in many cases, due to the chaotic nature of fire and the large variations in flame appearance in video. Recently three-dimensional dynamic texture analysis techniques have been used for the identification of dynamic phenomena such as sea, smoke, fire, clouds, leaves in the wind, fog or waves especially for the efficient retrieval of video in multimedia databases. Such dynamic textures can be simply defined as spatially and time-varying visual patterns that form an image sequence or part of an image sequence with a certain temporal stationarity [1]. Several such techniques have been proposed for modeling, learning, recognizing and synthesizing dynamic textures [16], [17], [18], [19]. Doreto et al. [1] proposed a method for segmenting a sequence of images of natural scenes into disjoint regions that are characterized by constant spatio-temporal statistics. The spatio-temporal dynamics in each region were modeled using Gauss-Markov models, and their parameters as well as the boundary of the regions were inferred in a variational optimization framework. More recently, Ravichandran et al. [16] proposed a method for the categorization of dynamic textures (e.g., water, fire and smoke). Each video sequence is modeled with a collection of Linear Dynamical Systems (LDSs), each one describing a small spatiotemporal patch extracted from the video. This Bag-of-Systems (BoS) representation is analogous to the Bag-of-Features (BoF) representation for object recognition, except that LDSs are used as feature descriptors.

In this paper, a new flame detection method is proposed, which aims to extend our previous work [15] by modeling the behavior of the fire using various spatio-temporal features and taking advantage of the recent advances in dynamic texture analysis in order to increase the robustness of the algorithm. Initially the algorithm applies background subtraction and color analysis of the moving regions using a non-parametric model to identify the candidate regions (blocks) in the image. Subsequently, the fire behavior is modeled by employing various spatio-temporal features such as color probability, spatial energy, flickering and spatio-temporal energy. Dynamic texture analysis using linear dynamical systems and bag of systems is applied only to the candidate fire regions of each frame in a time-window to reduce the computational cost. In addition, spatiotemporal consistency is estimated for each candidate fire region by taking into account the existence of neighboring: a) candidate fire blocks in the current and previous frames and b) fire blocks in the previous frames. As a last step, a two-class (fire, non-fire) support vector machines (SVM) classifier with a radial basis function (RBF) kernel is used to classify the candidate fire regions.

III. Detection of Candidate Fire Regions

The first step of the proposed method aims to filter out non fire-colored moving regions. For this reason, every frame of the video sequence is divided into NxN blocks (in our experiments $N = 16$). Background subtraction is used as a first step to identify moving objects in the video. Based on the evaluation of thirteen background extraction algorithms in [15], we chose to use the Adaptive Median algorithm, which is fast and very efficient. In the next processing step, color analysis is applied (Fig. 2) and only blocks that contain an adequate percentage of fire-colored moving pixels are selected as candidate "fire" blocks. To filter out non-fire moving pixels we compare their values with a predefined RGB color distribution that is created by non parametric estimation from a number of real fire-samples from a variety of video sequences.

Let $x_1, x_2, \ldots, x_N$ be N fire-colored training RGB samples of the distribution to be approximated. Using these samples, the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2014.2339592, IEEE Transactions on Circuits and Systems for Video Technology

probability density function of a pixel x_t can be non-parametrically estimated using the kernel K_h [20] as:

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

If we select a Gaussian kernel, $K_h = N(0, S)$, where S is a diagonal covariance matrix with different standard deviation σ_j for each color channel j, then the fire color probability can be estimated as:

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

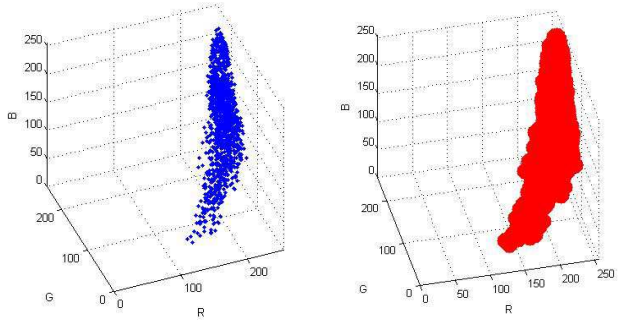


Fig. 2. (a) RGB color distribution of the training samples (b) the final fire-color distribution assuming a global threshold around each sample.

Using this probability estimation, the pixel is considered as a fire-colored pixel if $Pr(x_t) > th$, where the threshold th is a fixed threshold, which can be adjusted to achieve a desired percentage of false positives. Similar to [20] fire color probabilities in equation (2) were calculated very efficiently by using lookup tables based on the intensity value difference and the kernel bandwidths. For this reason this technique was preferred over more advanced iterative techniques using positive and negative samples [21], [22] which would increase the computational cost. Hence, if the pixel has a RGB value, which belongs to the distribution of Fig. 2 (b), then it is considered as a fire-colored pixel. If the percentage of fire-colored pixels within a block is over a specific level (in our experiments equals 12.5%), then, the block is considered as candidate for the next steps. This level is suitable for most cases but if cameras are situated in long distance from fire we can either use smaller blocks or reduce this level to consider it as candidate. For each candidate fire block, a vector of six features is computed in the following steps: a) fire color probability, b) spatial wavelet energy, c) spatio-temporal energy d) flickering energy, e) dynamic texture analysis and f) spatio-temporal consistency energy. The computation of these features is described in detail in Section IV, Section V and Section VI. The final decision (Fire/Non-Fire) is made by an SVM classifier as shown in Fig. 1.

IV. Spatio-temporal Flame Modeling

Since many natural objects have similar colors as those of the fire (e.g. the sun, various artificial lights or light reflections on various surfaces), which can often be mistakenly detected as flames, careful selection of appropriate spatio-temporal features is needed for modeling the behavior of fire. More specifically, in order to accurately model the color space of fire, we use non-parametric color analysis (presented in the previous section), while the spatial energy in each time instant of a candidate region is estimated by applying 2D wavelet analysis only on the red channel of the image. We also estimate the spatiotemporal energy in the candidate block, i.e., the variance of the spatial energy in the region within a temporal window, to identify the irregular changes of the fire's shape. Finally temporal analysis is used for flickering detection.

A. Fire color probability

For the color probability (Fig. 3) of each candidate block, the non-parametrically estimated probabilities of each pixel in the block are used. More specifically, the total color probability of the candidate block is estimated as the average color probability of each pixel (i, j) in the block:

$$P_{block} = \frac{1}{N}\sum_{i,j} P(i,j) \quad (3)$$

where N is the number of pixels in the block and $P(i, j)$ is the fire color probability of each pixel (see equation 2).

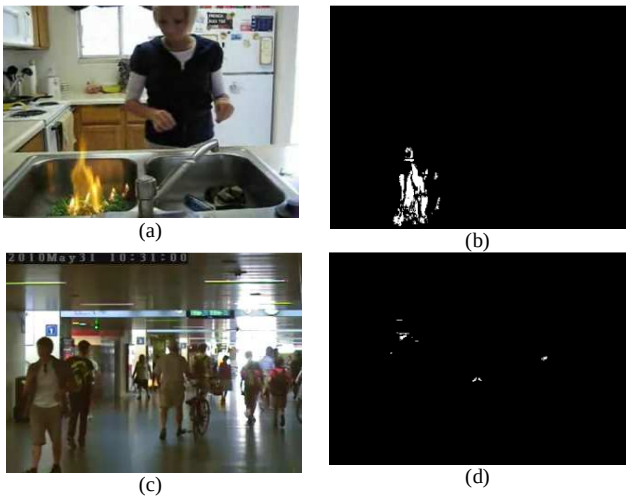


Fig. 3. A fire and non-fire video in (a) and (c) respectively and the corresponding masks in (b) and (d).

B. Spatial wavelet analysis

Image regions containing real fires exhibit a higher spatial variation than those containing fire colored objects. To identify spatial variations in the region various techniques can be adopted such as edge detectors, interest points descriptors, etc. In this paper, wavelet analysis using simple filters (Fig. 4) was used in order to achieve higher computational efficiency, since it can be implemented without any single multiplication, i.e., by simple register shifts. Specifically, a two dimensional wavelet filter is applied on the red channel of each frame and the spatial wavelet energy at each pixel is calculated by adding

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2014.2339592, IEEE Transactions on Circuits and Systems for Video Technology

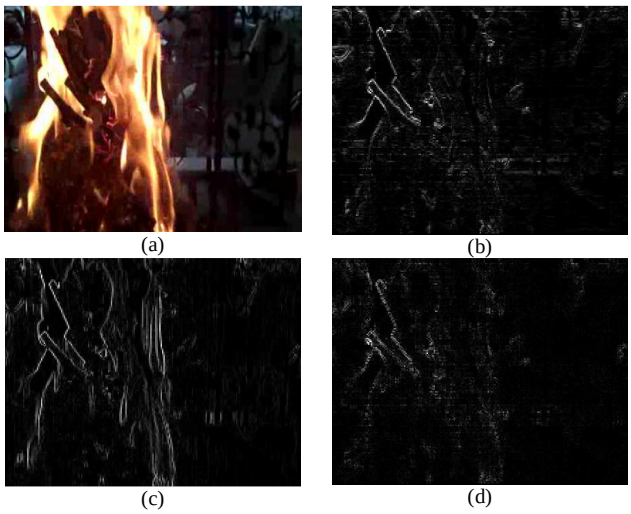
the high-low, low-high and high-high wavelet sub-images according to the following equation:

$$E(i,j) = HL(i,j)^2 + LH(i,j)^2 + HH(i,j)^2 \quad (4)$$

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

$$E_{block} = \frac{1}{N}\sum_{i,j} E(i,j) \quad (5)$$

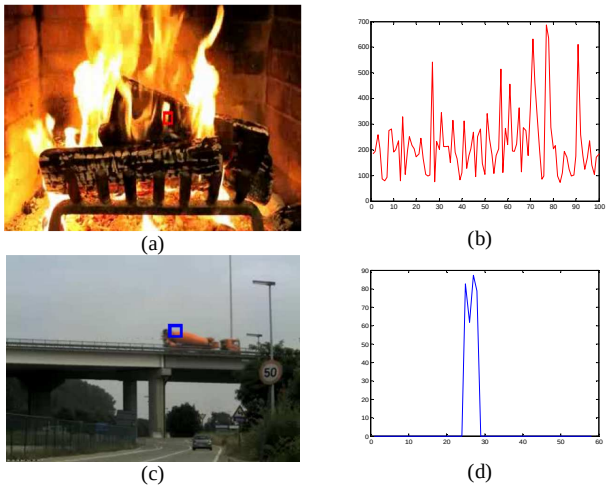
where N, is the number of the pixels in a block.



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

Fig. 4. Two dimensional spatial wavelet analysis for image (a), where (b) HL is the high-low frequencies, (c) LH and (d) HH are the high-frequency subbands of the wavelet decomposition.

For the single stage wavelet transform, the weights of the low pass and high pass filters are [0.25 0.5 0.25] and [-0.25 0.5 -0.25] respectively. Fig. 5 presents the value changes of spatial wavelet energy in a specific block for a) a subsequence of video containing real fire (Fig. 5a-b) and b) a subsequence of video containing sunlight reflections (Fig. 5c-d). As it is clearly shown, the values of spatial wavelet energy for the block containing fire are always higher due to the abnormal shapes formed by the fire.



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

Fig. 5. Changes of spatial wavelet energy: a-b) In the case of video containing actual fire the values of the energy vary between 90-700 c-d) the values of the energy for a video containing a fire colored objects vary between 0-85.

C. Spatio-temporal analysis

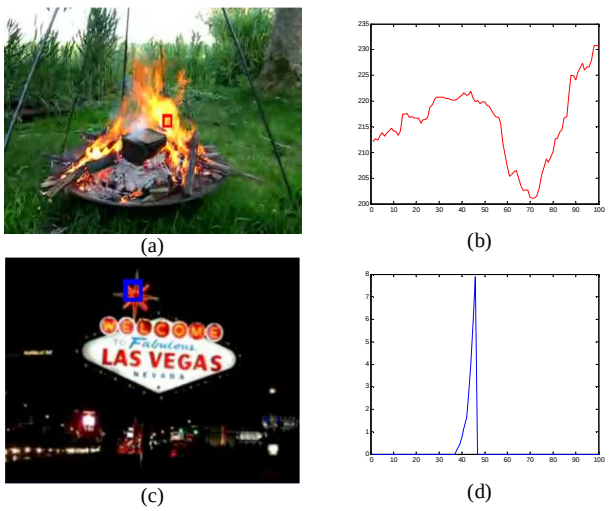
The shape of flame changes irregularly due to the airflow caused by wind or due to the type of burning material. As a result, a real fire causes higher spatial variations within a specific time interval than a fire colored object. On the contrary to the previous feature, which aims to identify high spatial energies in a single frame, this feature aims to indicate the spatio-temporal variations for each block in a sequence of frames. The temporal variance of the spatial energy of pixel (i,j) within a temporal window of T last frames is:

$$V(i,j) = \frac{1}{T}\sum_{t=0}^{T-1}(E_t(i,j) - \bar{E}(i,j))^2 \quad (6)$$

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

$$V_{block} = \frac{1}{N}\sum_{i,j} V(i,j) \quad (7)$$

As shown in the experimental results section, this proposed feature is very important in discriminating between fire and fire colored moving objects. As an example, the values of the spatio-temporal energy for a candidate block in a video containing actual fire (Fig. 6a-b) and a video containing a fire colored object (Fig. 6c-d) are shown in Figure 6.



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

Fig. 6. Changes of spatio-temporal energy: a-b) In the case of a subsequence of a video containing actual fire, the values of the energy vary between 200-235 c-d) the values of the energy for a subsequence containing fire colored objects vary between 0-8.

D. Temporal analysis

Temporal analysis is applied to each candidate block in order to detect flickering effect. Flickering, which is one of the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2014.2339592, IEEE Transactions on Circuits and Systems for Video Technology

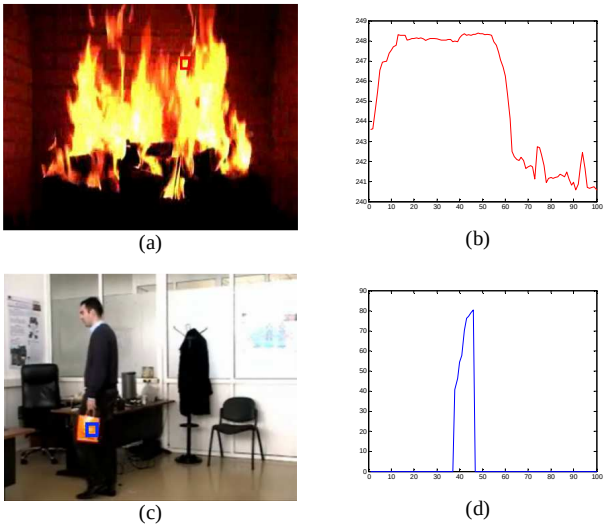
main features of flame, is due to its continuous random motion. To quantify the effect of flickering in a pixel, we estimate the number of transitions $c(i,j)$ from "fire candidate" status, i.e., moving fire colored pixel, to "non-fire candidate" status, i.e. background color or non-moving pixel, and vice-versa within a temporal window of T last frames. Then to estimate "flickering energy" of pixel (i,j) the following formula was used:

$$\mathrm{F(i,j)} = 2^{\mathrm{c(i,j)}} - 1 \tag{8}$$

The flickering feature F_{block} for each block is calculated as the average of the individual flickering contributions of the pixels in the block.

$$\mathrm{F_{block}} = \frac{1}{\mathrm{N}} \sum_{\mathrm{i,j}} \mathrm{F(i,j)} \tag{9}$$

As an example, in Figure 7, the values of F_{block} for a specific candidate block in a video containing fire (Fig. 7a-b) and a video containing a moving fire-colored object are indicatively presented.



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

Fig. 7. Changes of flickering energy: a-b) In the case of a subsequence containing actual fires the values of the energy vary between 240-250 c-d) the values of the energy for a subsequence containing a fire colored moving object vary between 0-80.

V. Dynamic Texture Analysis

Since fire is a spatially and time-varying visual pattern, dynamic texture analysis can be applied in addition to spatiotemporal modeling in order to increase the reliability of the algorithm by modeling the temporal evolution of the pixels' intensities. To this end, we apply linear dynamical systems that initially proposed by Doretto et al. [1] and a bag of dynamical systems approach proposed recently by Ravichandran et al. [16]. However, since these approaches focus on the categorization of video sequences containing natural scenes in multimedia databases, they cannot be applied directly to a fire detection system.

The main limitation that should be addressed is the computational cost needed for the dynamic texture analysis. For the categorization of video sequences, dense sampling is employed, which requires the estimation of LDSs in a large number of sampled image patches. However, as the spatial location and the size of the fire is unknown and varies from a video sequence to a video sequence, many selected image patches may not contain any information related to fire. It is easily conceivable that this number is significantly increased in case of a small fire in the scene, which is mainly the case for an early fire warning system. Since in our case we don't focus on scene categorization, but on the detection of a specific event, such as fire, in the scene, the processing of this redundant information not only increases prohibitively the computational cost, but also may introduce noisy data (i.e. LDSs corresponding to background, non-fire blocks) to the fire detection system, thus affecting the classification process. To reduce the computational burden, we focus only on those image regions for which we have an indication of fire existence. Towards this end, linear dynamical systems are estimated only for the pixels contained in the candidate fire blocks extracted from the first processing step of the proposed algorithm. For each candidate block a 3D image patch of the same size (16x16) and temporal length equal to 16 (i.e., spatiotemporal cubes of size 16x16x16) is formed for the estimation of LDS.

More specifically, given a candidate block of nxn pixels $(n = 16)$ and F frames of the video sequence $(F = 16)$, we can model the pixel intensities of the candidate block $I_{block}(t)$ at each time instant t, where $t = 1 \ldots 16$, assuming that the pixels contained in the 3D image patch can be considered as a linear dynamical system:

$$\mathrm{z(t+1)} = \mathrm{Az(t)} + \mathrm{Bv(t)} \tag{10}$$

$$\mathrm{I_{block}(t)} = \bar{\mathrm{I}}_{\mathrm{block}} + \mathrm{Cz(t)} + \mathrm{w(t)} \tag{11}$$

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

The dynamics of the hidden state is modeled by matrix $A \in \mathrm{R^{nxn}}$, while matrix $C \in \mathrm{R^{pxn}}$ (p is the number of pixels in a candidate block) maps the hidden state to the output of the system. The quantities $w(t)$ and $Bv(t)$ are the measurement and process noise respectively, while $\bar{\mathrm{I}}_{\mathrm{block}}$ is the mean value of the pixels' intensities in a candidate block for the sequence of F frames:

$$\bar{\mathrm{I}}_{\mathrm{block}} = \frac{1}{\mathrm{F}} \sum_{\mathrm{t=1}}^{\mathrm{F}} \mathrm{I_{block}(t)} \tag{12}$$

To identify the system, i.e., to estimate its parameters, a principal component analysis based approach was proposed by Doretto et al. [1]. According to this approach, a singular value decomposition of matrix Y is initially performed:

$$\begin{aligned} \mathrm{Y} &= [\mathrm{I_{block}(1)} - \bar{\mathrm{I}}_{\mathrm{block}}, \ldots, \mathrm{I_{block}(F)} - \bar{\mathrm{I}}_{\mathrm{block}}] \\ &= \mathrm{U\Sigma V^T} \end{aligned} \tag{13}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2014.2339592, IEEE Transactions on Circuits and Systems for Video Technology

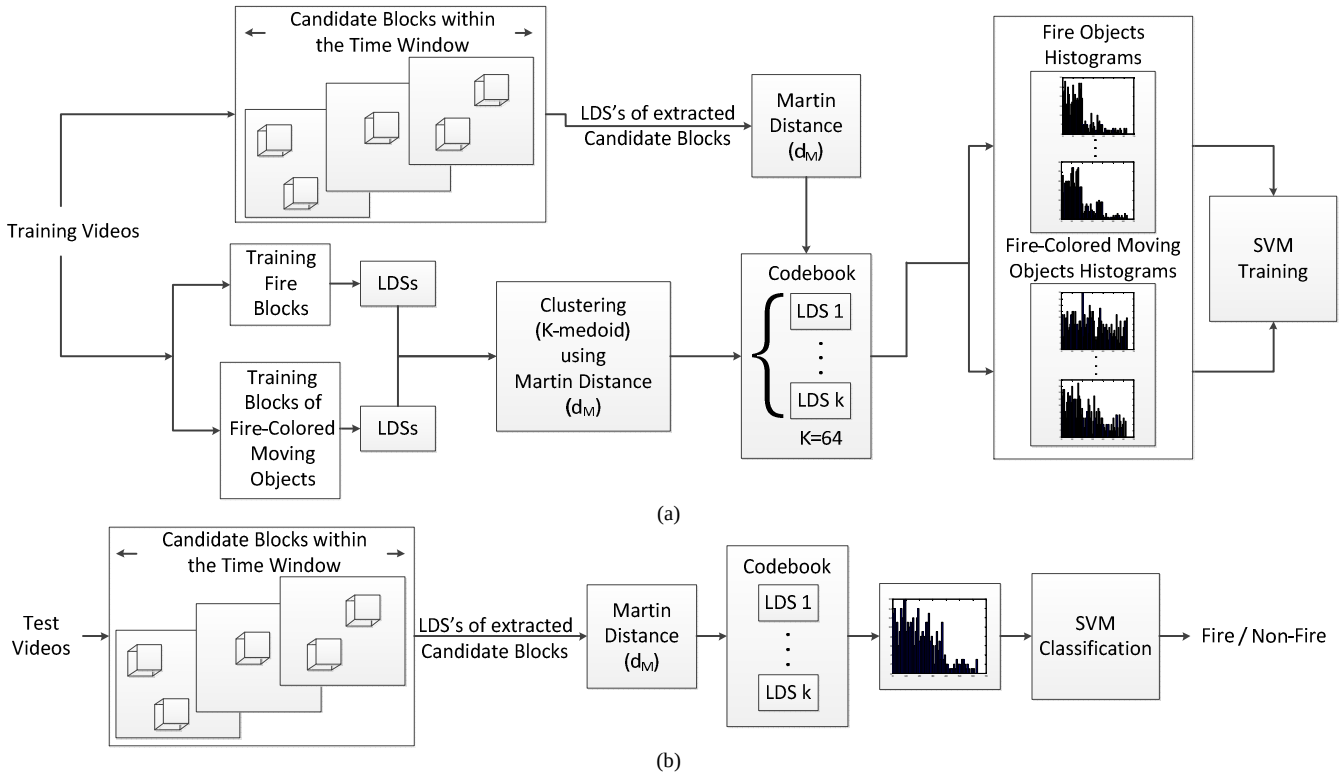


Fig. 8. Methodology for Dynamic Texture model.

Based on the outcome of this factorization process, the temporal evolution of the intensities in a candidate block can be efficiently described using the parameters of the LDS descriptor, i.e., $\mathrm{M_{block}} = (\mathrm{A}, \mathrm{C})$, where $C = U$ and matrix A can be easily computed using least-squares as:

$$\mathrm{A} = [\mathrm{z}(2), \mathrm{z}(3), \ldots, \mathrm{z}(\mathrm{F})][\mathrm{z}(1), \mathrm{z}(2), \ldots, \mathrm{z}(\mathrm{F}-1)]^{+} \quad (14)$$

where $Z = [z(1), z(2), \ldots, z(F)] = \Sigma \mathrm{V}^{\mathrm{T}}$ are the estimated states of the system and Z^{+} represents the pseudoinverse of Z.

Using the extracted LDS descriptors, a codebook can be formed using K-Medoid classification method as proposed in [16]. However, to apply classification we need first to define a similarity metric between LDS descriptors, that is, a clustering approach applicable to the non-Euclidean space of LDSs in order to determine the similarity degree between two descriptors, e.g. $M_1 = (A_1, C_1)$ and $M_2 = (A_2, C_2)$. To overcome the problem, subspace angles between the two LDSs are initially calculated and then a Martin distance between M_1 and M_2 is used as a comparison metric [16], [23].

Similarly to the sequence categorization problem, where the training set consists of video sequences from all classes, we used LDSs corresponding to candidate blocks containing both fire and fire-colored moving objects. As shown in Figure 8a, these training LDSs are fed into K-Medoid algorithm for the creation of a codebook consisting of K representative LDSs, i.e., codewords. Various numbers of codewords can be used for this purpose, however, previous studies have shown that using more than 64 clusters does not significantly change the categorization performance [16]. For this reason and in order to keep the computational cost as low as possible, the number of K was set equal to 64 in our experiments. Given a codebook of LDSs, a full video sequence can be represented using this vocabulary. While this is the case in a video categorization problem, in the case of a real fire detection system, this would create two significant problems: i) The system wouldn't be able to determine the time of the fire incident. However, in each time instant there should be a decision whether there is a fire or not in the scene and ii) as time passes the volume of data is continuously increased. To address the problem, we

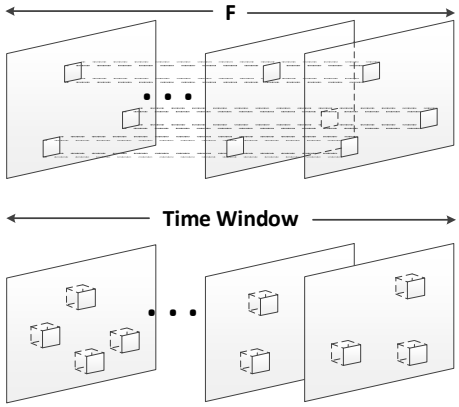


Fig. 9. (a) A 3D image patch of size 16x16xF (where F=16) is formed for each candidate block (three candidate blocks in this figure). (b) LDSs are estimated only for the candidate blocks of a sliding time window of T frames (in our experiments we used a time window of 100 frames). 3D image patches are formed for the candidate blocks of each frame of the time window.

divided the video into equally sized subsequences (see Figure 9) using a sliding time window T (in our experiments $T = 100$) and we then represent each subsequence as a term

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2014.2339592, IEEE Transactions on Circuits and Systems for Video Technology

frequency histogram $h = [h_1, h_2, \dots, h_K]^T$ of the predefined codewords. As shown in Fig. 8a, such subsequence representations are extracted from the training data to create two distinctive classes. The first class contains histograms representing subsequences with real fires, while the second one consists of histograms corresponding to fire colored moving objects. These distributions of codewords are used by the next step for the training of the SVM classifier.

For the classification of a new query, as shown in Figure 8b, the distribution of codewords for a specific subsequence is estimated and the extracted histogram is send to the SVM classifier to infer the class label D. This label is assigned as a feature D_{block} to the candidate blocks of the current frame, i.e., the last frame of the subsequence, indicating the possibility of fire existence in this time instant. However, this processing step cannot provide any information related to the location of the fire in the scene. Towards this end, the information extracted from dynamic texture analysis should be finally fused with that of spatio-temporal flame modeling and spatio-temporal consistency energy in the classification step of the proposed method. This combination of data increases also the reliability of the system, as shown in the experimental results section.

VI. Spatio-Temporal Consistency Energy

In this section we aim to propose a method for enhancing the reliability of the algorithm, by taking into account the features extracted from both spatio-temporal flame modeling and dynamic texture analysis as well as prior knowledge about the possible existence of fire in neighboring blocks in a 3D image patch.

Towards this end, we propose the estimation of a consistency energy inspired by approaches based on Markov Random Fields (MRF). MRF approaches are iterative methods that apply global optimization to solve labeling problems based on a data and a smoothness term [24], [25]. However, such techniques cannot directly be applied to an early fire warning system due to the increased computational cost required for the energy minimization procedure. Hence, in this step of the algorithm we do not attempt to address the problem through an energy minimization approach, but rather estimate a "consistency energy" feature that can be used as an indication of fire in the block taking into account the state of neighboring blocks in a 3-D image patch.

More specifically, we estimate a "consistency energy" cost C_{block} for each candidate block consisting of two terms: i) a data cost taking into account the features in the current block and ii) a smoothness cost considering the state of neighboring blocks:

$$\mathrm{C_{block}} = \mathrm{E_{data}} + \mathrm{E_{smoothness}} \tag{15}$$

The "data cost" is defined as the sum of the previously defined features for the candidate block in the previous sections:

$$\mathrm{E_{data}} = \mathrm{P_{block}} + \mathrm{E_{block}} + \mathrm{V_{block}} + \mathrm{F_{block}} + \mathrm{D_{block}} \tag{16}$$

The smoothness term $E_{smoothness}$ consists of two terms depending on a) the number of "candidate fire blocks" in the current frame and the previous frame and b) the number of "fire-labeled" blocks in the three previous frames, as shown in Fig. 10. For the estimation of the smoothness term, a 3x3 spatial neighborhood around the candidate block is used.

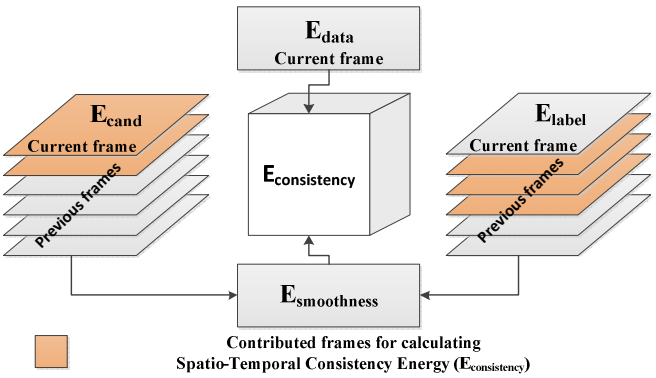


Fig. 10. Calculation of $E_{consistency}$ using neighboring candidate and fire labeled blocks from current and previous frames.

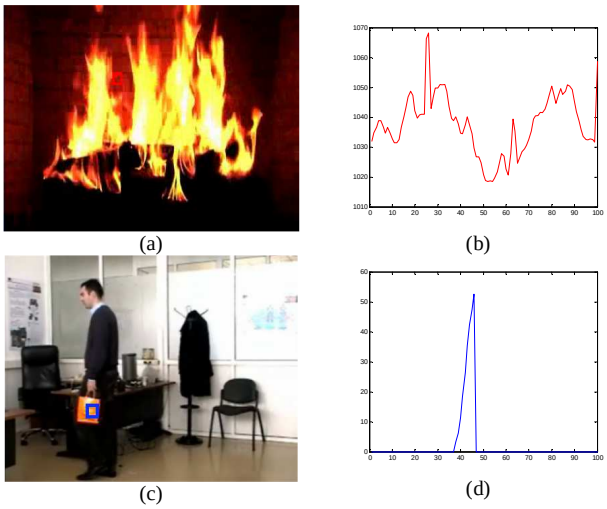


Fig. 11. Changes of spatio-temporal consistency energy: a-b) In the case of a subsequence containing actual fire the values of the energy vary between 1010-1070 c-d) the values of the energy for a subsequence containing a fire colored moving object vary between 0-50

More specifically, the smoothness term can be defined as:

$$\begin{aligned}\mathrm{E_{smoothness}} = \mathrm{E_{cand}} + \mathrm{E_{label}} = \\ = \sum_{\mathrm{i=0}}^{\mathrm{n}} [\mathrm{a_i N_{cand}}(\mathrm{t-i})] \\ + \sum_{\mathrm{i=1}}^{\mathrm{m}} [\mathrm{b_i N_{label}}\,(\mathrm{t-i})]\end{aligned} \tag{17}$$

where $n = 1$ (i.e., only current and previous frame are considered), $m = 3$ (for the three previous frames), $N_{cand}(t)$ is the number of candidate fire-blocks in the neighborhood of the candidate block in frame t and $N_{label}(t)$ is the number of fire-labeled blocks in the neighborhood of the candidate block in frame t. For our experiments, the following weights were

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2014.2339592, IEEE Transactions on Circuits and Systems for Video Technology

used $a_0 = a_1 = 1$, while $b_1 = 2$, $b_2 = 1.5$ and $b_3 = 1$. By allocating a larger weight ($b_i > a_i$) to the blocks that permanently labeled by the algorithm as fire blocks, we improve the efficiency and reduce the number of false alarms.

As is shown in the experimental results section, spatio-temporal consistency energy feature contributes significantly to the robustness of the algorithm, as it combines the prior knowledge about the possible existence of fire in neighboring blocks with the spatio-temporal and dynamic texture analysis features of the candidate block. Fig. 11 presents the value of spatio-temporal consistency energy in a specific block of videos containing actual fire (Fig. 11a-b) and fire colored object (Fig. 11c-d).

VII. Classification

As a last step, classification is used to obtain the final decision about whether a block is a fire block or not based on the extracted features described in the previous sections. To this end a feature vector consisting of six features $f = [P_{block}, E_{block}, V_{block}, F_{block}, D_{block}, C_{block}]$ is created. This vector is fed as input to a two-class (fire, non-fire) Support Vector Machines (SVM) classifier with a radial basis function (RBF) kernel to classify the candidate fire regions. The training set consisted of 5000 randomly selected candidate blocks from four fire and four non-fire video sequences of the Firesense dataset. Specifically, the training set of candidate blocks was obtained from the positive video sequences posVideo2, posVideo3, posVideo9, posVideo10 (the total number of candidate blocks in these video sequences is 16796, 14669, 63806 and 6125 respectively) and from the negative video sequences negVideo2, negVideo5, negVideo6 and negVideo8 (the total number of candidate blocks in these video sequences is 1150, 742, 122 and 501 respectively) of the Firesense dataset. In total, the candidate blocks used for the training of the SVM algorithm is the 4.81% of the total number of candidate blocks in the eight fire and non-fire video sequences.

VIII. Experimental Results

In this section we present a detailed experimental evaluation of our method using both fire and non-fire videos from two datasets: the Firesense dataset [26] and a set of video sequences used in Ko's experiments [12]. Both datasets have been used in the past for the comparison of flame detection algorithms. Specifically, three state of the art algorithms [13] have been tested with the first dataset, while three other [5][11][12] with the second dataset. To examine the proposed fire-flame detection algorithm, we estimated the number of correctly detected flame frames out of the total number of flame frames (true positive) and the number of non-flame frames that erroneously recognized as flame frames out of the total number of non-flame frames (false positive). For comparison reasons, a frame is labeled as a fire frame if it contains at least one fire block, as in [13]. The average frame rate of the proposed method was 5.2 fps for video sequences with resolution 320x240, which is considered adequate for an early fire warning system. The experiments were performed with a PC that has a Core i5 2.4 GHz processor.

Finally, to further analyze the detection efficiency of the proposed algorithm, we provide a detailed study regarding the contribution of each feature in the classification process. Specifically, we apply the proposed algorithm to both datasets excluding each time one feature from the classification process. In each evaluation test, we follow the same training strategy for the SVM classifier, i.e., the same training candidate blocks are used, however, the excluded feature (or a set of features) is ignored each time from the training process of the SVM.

A. Firesense Dataset

The Firesense dataset consists of eleven fire videos and ten non-fire videos (Fig. 12). The performance of the proposed method was compared with three existing flame detection algorithms that were tested with this dataset in [13]: i) Correlation descriptors [14], ii) Multiple features and rule-based classification [15], and iii) Spatio-temporal consistency energy [27].

(a)

(b)

Fig. 12. Firesense Dataset: Screenshots of video sequences containing a) actual fires b) fire colored moving objects.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2014.2339592, IEEE Transactions on Circuits and Systems for Video Technology

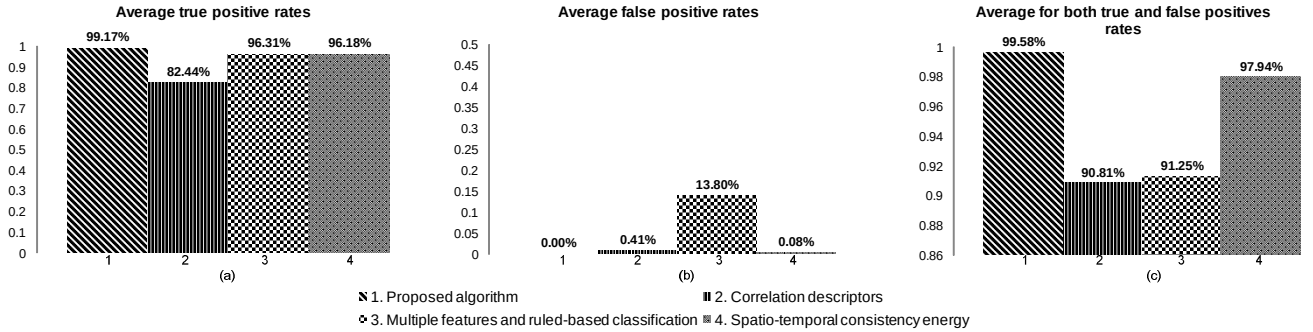


Fig. 13. Comparison of the proposed algorithm with Correlation descriptors [14], Multiple features and rule-based classification [15] and Spatio-temporal consistency energy [27] algorithms using Firesense dataset. (a) Average true positive rates, (b) average false positive rates and (c) average for both true and false positive rates.

As shown in Fig. 13, the proposed method outperforms with an average true positive rate of 99.17%, while it didn't produce any false positive due to the coupling of spatio-temporal modeling and dynamic texture analysis. A small number of missed detections was only observed in the last video of Fig. 12a, where the size of the fire is extremely small in the first frames of the sequence. On the other hand, the rule-based classification [15], which is based only on spatiotemporal features, gives high fire detection rates, however, it is significantly vulnerable to false alarms produced by fire-colored moving objects. The estimation of the consistency energy of the spatio-temporal features, proposed in our previous work [27], and its inclusion in the classification process, shows that improves significantly the robustness of the algorithm, reducing, however, slightly its detection ability. On the contrary, the use of 3D spatio-temporal image patches instead of 2D blocks and the estimation of the covariance matrix descriptors in [14] produce low false alarms rates, affecting, however, the detection rate of the algorithm.

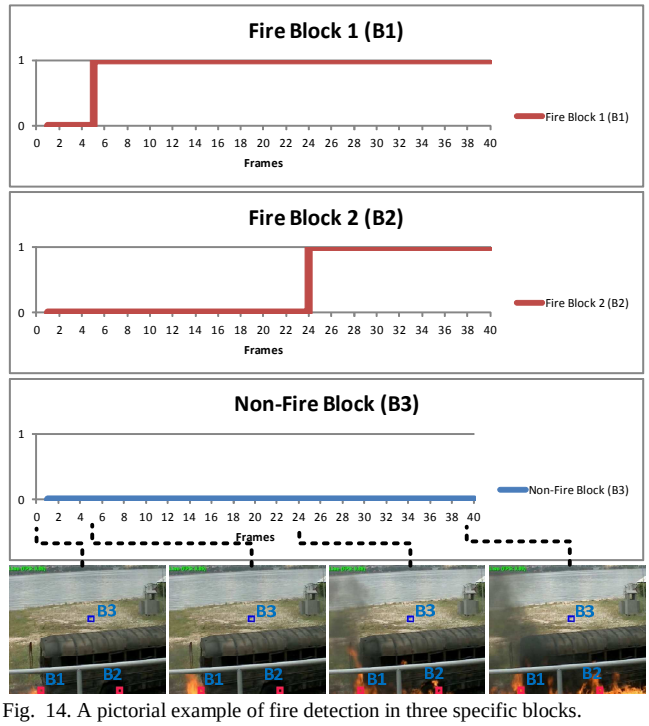


Fig. 14. A pictorial example of fire detection in three specific blocks.

Regarding the computational cost of the dynamic texture analysis, as it was mentioned in Section V, only pixels contained in the candidate fire blocks are considered for the estimation of LDSs. This approach reduces significantly the computational burden in comparison to the dense sampling used for the categorization of video sequences. The computational gain heavily depends on the size of fire in the image. For instance, in the fifth video sequence in Fig. 12a, in which the fire covers a large part of the image, the computational cost is reduced by 72.04%, while in case of small fires (e.g., the sixth video in Fig. 12a) the reduction is even higher, i.e., 98.65%, and can reach to 99.54% in non-fire videos (e.g., sixth video in Fig. 12b). Fig. 14 presents a pictorial example of fire detection (first video in Fig. 12a) in three specific blocks of the image with respect to time.

B. Video sequences used in Ko's experiments

To further evaluate the performance of the proposed method, we tested our algorithm on the video set (Fig. 16) used in Ko's experiments, which consists of eight fire videos and eight non-fire videos. Its performance was compared with three state of the art algorithms: i) Toreyin's algorithm [5], ii) Ko's algorithm based on hierarchical Bayesian Networks [11] and iii) Ko's algorithm based on Fuzzy Finite Automata [12].

Toreyin's algorithm, which is based purely on spatio-temporal features, produces lower detection rate than the other three algorithms. In addition, Ko's method [11] produces better classification results using hierarchical Bayesian Networks, however, it also seems to miss a significant number of real fire flames. As shown in Fig. 15, both algorithms produce a number of false alarms, which are mainly due to continuous light changes in some videos, especially those containing flickering car lights. On the other hand, the method based on Fuzzy Finite Automata [12] produces high detection rates due to the use of feature probability models for detecting flame flickering over time. The missed detections are owned to the abrupt movements of flame due to the wind, as the method is based on the assumption, which is not always true, that fire has only upward motion. Nevertheless, the method seems to be really robust to false alarms. Finally, the proposed algorithm provides extremely high detection rates, 99.65%, while it is capable to discriminate fire-colored moving objects from real fires without producing any false alarm.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2014.2339592, IEEE Transactions on Circuits and Systems for Video Technology

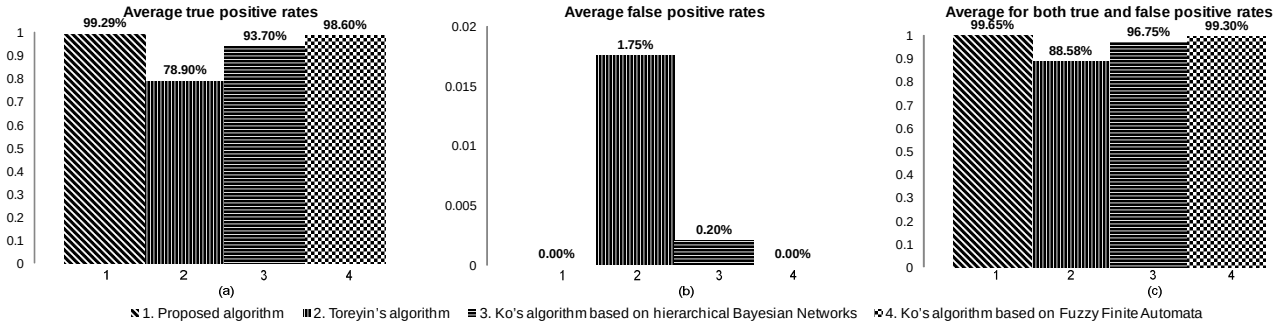


Fig. 15. Comparison of the proposed algorithm with Toreyin's algorithm [5], Ko's algorithm based on hierarchical Bayesian Networks [11] and Ko's algorithm based on Fuzzy Finite Automata [12] using the video set used in Ko's experiments. (a) Average true positive rates, (b) average false positive rates and (c) average for both true and false positive rates for all algorithms.

Fig. 16. Dataset used in Ko's experiments: Screenshots of video sequences containing a) actual fires b) fire colored moving objects.

C. Analysis of features' contribution

In this section, we elaborate a more detailed analysis on the experimental results in terms of the contribution of each feature in the classification process. To have a more clear view, we apply a two-step validation procedure using both Firesense and Ko's datasets: We first analyze the contribution of each one of the three main elements of the proposed method, i.e., spatiotemporal modeling, dynamic texture analysis and spatiotemporal consistency energy, and then we study the role of each individual spatiotemporal feature in the classification approach. As it was also mentioned, in each evaluation test we exclude one feature or a set of features from the classification process. To have comparable results, we also train each time the SVM classifier using the same training data (i.e., the 5000 candidate blocks), but ignoring the excluded feature or features from the training process.

Fig. 17 shows the average detection rate of the algorithm by using each time one of the three main parts of the proposed method. As it is clearly shown, dynamic texture analysis produces lower detection rate, 84.25%, than spatio-temporal modeling, 91.12%, while the detection rate is significantly increased when we couple spatio-temporal modeling with spatio-temporal consistency energy, 97.66%, (in this case only spatio-temporal features are considered in the estimation of the energy). It is worth mentioning that when we exclude spatio-temporal consistency energy (i.e., using only dynamic texture analysis and spatio-temporal flame modeling) the detection rate remains low, i.e., 91.34%. This fact really indicates the significance of the proposed spatio-temporal consistency energy in the classification process. On the other hand, as it is clearly shown in Fig. 17b, the dynamic texture analysis is extremely robust to false alarms, as no false positive is produced. Similarly, spatio-temporal consistency energy contributes significantly to the reduction of the false positive rate of spatio-temporal flame modeling from 4.93% to 0.25%.

The above analysis makes evident that each of the three main elements of the proposed algorithm plays a crucial role in the classification process. More specifically, spatio-temporal features contribute mainly to the increase of the detection rate of the algorithm, while dynamic texture analysis enhances the robustness of the algorithm to false alarms. Finally, the consistency energy seems to improve both true and false positive rates. The exclusion of any of the above elements from the classification process deteriorates significantly the performance of the algorithm, as shown in Fig. 17.

The next validation tests concern the contribution of spatio-temporal flame modeling features to the classification process. Similarly, the SVM classifier is trained with the same training set of 5000 candidate blocks and then the proposed algorithm is evaluated using both FIRESENSE and Ko's video datasets excluding each time a particular feature from both the training and classification process. Experimental results show that spatio-temporal analysis feature proposed in this paper for the identification of the random variations of flame appearance plays a significant role in the classification process.

On the other hand, the detection of fire's random motion through temporal modeling improves the robustness of the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2014.2339592, IEEE Transactions on Circuits and Systems for Video Technology

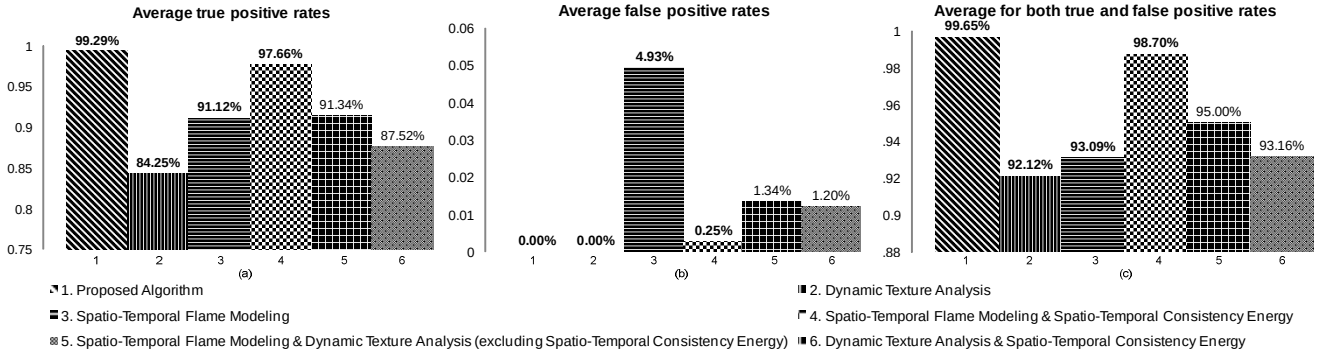


Fig. 17. Contribution of the main three elements of the proposed algorithm to the classification process. (a) Average true positive rates, (b) average false positive rates and (c) average for both true and false positive rates.

algorithm, since many fire-colored moving objects do not cause any flickering effect (however this is not true in the case of car lights, reflections etc., where temporal modeling has less contribution to the classification process). Finally, while color is a significant factor for filtering out regions containing moving objects, it contributes less than the other features to the classification process as all candidate blocks (either containing fire or non-fire objects) contain pixels with color similar to that of fire.

D. Parameters' Analysis

In all experimental results presented in this section (Fig. 18) we used a block size equal to 16x16, which was used in the past by other researchers [14]. By reducing the size of the block, the detection rate can be further increased, since the algorithm is able to detect even smaller fires, however, this also increases the false alarm rate. Similarly, by increasing the size of the block, we can enhance the robustness of the algorithm, reducing, however, its sensitivity. For instance in the last fire video of Firesense dataset, where the size of the fire is small in the first frames of the video, the detection rate is reduced to 64.8% from 93.787% for a block size of 32x32, while it is increased to 95.27% for a block size of 8x8. However, using 8x8 blocks increases the false alarm rate by 2.16% for videos containing car lights. Furthermore, the use of a time window of 100 frames seems to be suitable, since larger values increase the computational cost, while smaller values reduce the detection rate of the algorithm (e.g., values below 50 produce a reduction of 9% to the detection rate). Finally, the size of the codebook was set to 64, since as it was discussed in section V, a larger codebook size does not significantly improve the categorization performance.

IX. Conclusions

In this paper, we proposed an algorithm for real time video-based flame detection. By modeling both the behavior of the fire using various spatio-temporal features and the temporal evolution of the pixels' intensities in a candidate image block through dynamic texture analysis, we showed that we can have high detection rates, while reducing the false alarms caused by fire-colored moving objects. The use of spatio-temporal consistency energy increases the robustness of the algorithm by exploiting prior knowledge about the possible existence of fire in neighboring blocks from the current and previous video frames. Experimental results with thirty seven videos containing actual fire and moving fire colored objects showed that the proposed algorithm outperforms existing flame detection algorithms. Due to the dynamic texture analysis, the computational requirements of the proposed method are higher compared to these of other algorithms based solely on spatio-temporal modeling, however, the average frame rate of the algorithm is still considered adequate for an early fire warning system. Future implementations in FPGAs are expected to increase more the average frame rate of the algorithm.

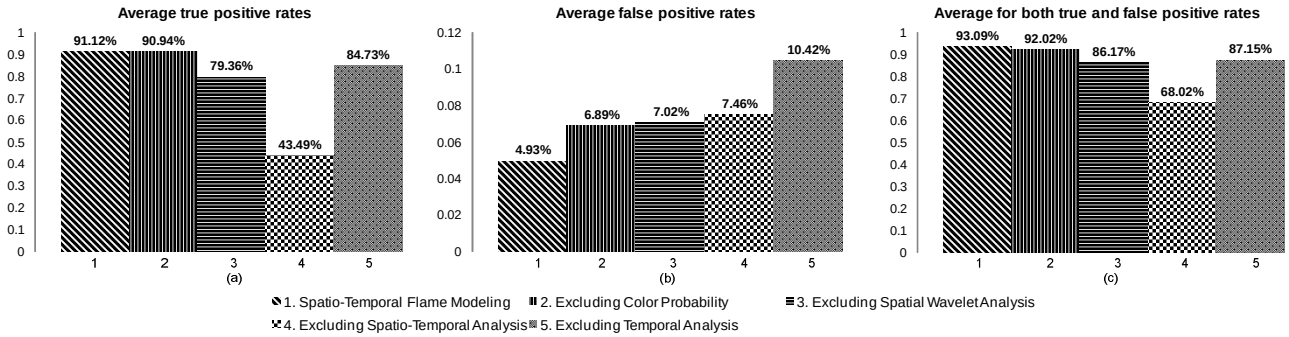


Fig. 18. Contribution of features participating in the Spatio-Temporal Flame Modeling to the classification process. (a) Average true positive rates, (b) average false positive rates and (c) average for both true and false positive rates.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2014.2339592, IEEE Transactions on Circuits and Systems for Video Technology

REFERENCES

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

[2] E. A. Cetin, K. Dimitropoulos, B. Gouverneur, N. Grammalidis, O. Gunay, H. Y. Habiboglu, U. B. Toreyin and S. Verstockt, "Video fire detection – Review," Digital Signal Processing, vol. 23, no. 6, pp. 1827-2843, December 2013.

[3] T.-H. Chen, P.-H. Wu and Y.-C. Chiou, "An early fire-detection method based on image processing," in ICIP, 2004.

[4] C.-B. Liu and N. Ahuja, "Vision based fire detection," in ICPR, 2004.

[5] B. Toreyin, Y. Dedeoglu, U. Gugukbay and A. Cetin, "Computer vision based method for real-time fire and flame detection," Pattern Recognition Letter, vol. 27, no. 1, pp. 49-58, 2006.

[6] B. Toreyin, Y. Dedeoglu, U. Gudukbay and A. E. Cetin, "Flame detection in video using hidden markov models," in IEEE International Conference on Image Processing, 2005.

[7] Z. Zhang, J. Zhao, D. Zhang, C. Qu, Y. Ke and B. Cai, "Contour based forest fire detection using FFT and wavelet," in CSSE, 2008.

[8] T. Celik and H. Demirel, "Fire detection in video sequences using a generic colour model," Fire Safety Journal, vol. 44, no. 2, pp. 147-158, February 2009.

[9] G. Marbach, M. Loepfe and T. Brupbacher, "An image processing technique for fire detection in video images," Fire Safety Journal, vol. 41, no. 4, pp. 285-289, June 2006.

[10] P. Borges and E. Izquierdo, "A probabilistic approach for vision-based fire detection in videos," IEEE Transactions on Circuits and Systems for Video Technology, vol. 20, no. 5, pp. 721-731, May 2010.

[11] B. Ko, K. Cheong and J. Nam, "Early fire detection algorithm based on irregular patterns of flames and hierarchical bayesian networks," Fire Safety Journal, vol. 45, no. 4, pp. 262-270, 2010.

[12] B. Ko, S. Ham and J. Nam, "Modeling and formalization of fuzzy finite automata for detection of irregular fire flames," IEEE Transactions on Circuits and Systems for Video Technology, vol. 21, no. 12, pp. 1903-1912, December 2011.

[13] K. Dimitropoulos, O. Gunmay, K. Kose, F. Erden, F. Chaabene, F. Tsalakanidou, N. Grammalidis and E. Cetin, "Flame detection for video-based early fire warning for the protection of cultural heritage," in 4th International Euro-Mediterranean Conference on Cultural Heritage (EuroMed 2012), Lemesos, 2012.

[14] H. Habiboglu, O. Gunay and E. A. Cetin, "Covariance matrix-based fire and flame detection method in video," Machine Vision and Applications, vol. 23, no. 6, pp. 1103-1113, November 2012.

[15] K. Dimitropoulos, F. Tsalakanidou and N. Grammalidis, "Flame detection for video-based early warning systems and 3D visualization of fire propagation," in IASTED International Conference on Computer Graphics and Imaging, 2012.

[16] A. Ravichandran, R. Chaudhry and R. Vidal, "Categotizing dynamic textures using a bag of dynamical systems," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 35, no. 2, pp. 342-353, February 2013.

[17] Z. Bar-Joseph, R. El-Yaniv, D. Lischinski and M. Werman, "Texture mixing and texture movie synthesis using statistical learning," IEEE Transactions on Visualization and Computer Graphics, vol. 7, no. 2, pp. 120-135, April 2001.

[18] V. Kwatra, A. Schodl, I. Essa, G. Turk and A. Bobick, "Graphcut textures: image and video synthesis using graph cuts," ACM Trans. Graphics, vol. 22, no. 3, pp. 277-286, July 2003.

[19] M. Szummer and R. W. Picard, "Temporal texture modeling," in IEEE Int'l Conf. Image Processing, 1996.

[20] A. Elgammal, D. Harwood and L. Davis, "Non-parametric model for background subtraction," in 6th European Conference on Computer Vision, Dublin, Ireland, 2000.

[21] N. Oudjane and C. Mousso, "L2-Density estimation with negative kernels," in Image and Signal Processing and Analysis, 2005.

[22] S. Mika, G. Raetsch, J. Weston, B. Schoelkopf and K. R. Mueller, "Fisher discriminant analysis with," Neural Networks for Signal Processing, vol. 9, pp. 41-48, 1999.

[23] K. D. Cock and B. D. Moor, "Subspace angles and distances between ARMA models," System and Control Letters, vol. 4, pp. 265-270, 2002.

[24] Y. Boykov, O. Veksler and R. Zabih, "Fast Approximate Energy Minimization via Graph Cuts," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, no. 11, pp. 1222-1239, 2001.

[25] K. Dimitropoulos, T. Semertzidis and N. Grammalidis, "3D content generation for autostereoscopic displays," in 3DTV CON, Berlin, 2009.

[26] FIRESENSE, "FIRESENSE Project Protection of Cultural Heritage," [Online]. Available: http://www.firesense.eu/.

[27] P. Barmpoutis, K. Dimitropoulos and N. Grammalidis, "Real time video fire detection using spatio-temporal consistency energy," in 10th IEEE International Conference on Advanced Video and Signal-Based Surveillance, Krakow, Poland, 2013.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2014.2339592, IEEE Transactions on Circuits and Systems for Video Technology

Dr. Kosmas Dimitropoulos received his B.Sc degree in Electrical and Computer Engineering from Democritus University and his Ph.D. degree in Applied Informatics from Macedonia University of Thessaloniki in 2001 and 2007 respectively.

He is currently a post-doctoral research fellow at the Information Technologies Institute of the Centre for Research and Technology Hellas (ITI-CERTH) and a visiting lecturer at the University of Macedonia. His main research interests include computer vision, pattern recognition, 3D motion analysis from multiple depth cameras, 3D graphics and visualization.

He has participated in several European and national research projects and he has served as a regular reviewer for a number of international journals and conferences.

Panagiotis Barmpoutis received his B.Eng. & M.Eng. in Electrical and Computer Engineering from the Aristotle University of Thessaloniki in 2009. He also received his MSc in Forestry Informatics and his MSc in Medical Informatics from the Aristotle University of Thessaloniki in 2012 and 2013 respectively.

From 2012, he is a Research Assistant in the Information Technologies Institute at Centre for Research and Technology Hellas (CERTH). His current research interests lie in the areas of computer vision and applications, real-time image processing, analysis and visualization, pattern recognition and machine learning.

Nikos Grammalidis is a Senior Researcher (Researcher Grade B) at the Information Technologies Institute - Centre of Research and Technology Hellas. He received the B.S. and Ph.D. degrees in Electrical and Computer Engineering from the Aristotle University of Thessaloniki, in 1992 and 2000, respectively.

Prior to his current position, he was a researcher in 3D Imaging Laboratory at the Aristotle University of Thessaloniki. His main research interests include computer vision, signal, image and video processing, stereoscopy and multiview image sequence analysis and coding.

His involvement with those research areas has led to the co-authoring of more than 25 articles in refereed journals and more than 75 papers in international conferences. Since 1992, he has been actively involved in more than 25 EC and National projects. He has served as a regular reviewer for a number of international journals and conferences.