Spatial-Temporal Structural and Dynamics Features for Video Fire Detection

Hongcheng Wang*, Alan Finn, Ozgur Erdinc, Antonio Vincitore
United Technologies Research Center (UTRC), East Hartford, CT, USA

*wangh1@utrc.utc.com

Abstract

We present a new Video Fire Detection (VFD) system for surveillance applications in fire and security industries. The system consists of three modules: pixel-level processing to identify potential fire blobs, blob-based spatial-temporal feature extraction, and a Support Vector Machine (SVM) classifier. The proposed novel spatial-temporal features include a spatial-temporal structural feature and a spatial-temporal contour dynamics feature. The spatial-temporal structural features are extracted from an accumulated motion mask (AMM) and an accumulated intensity template (AIT), capturing the concentric ring structure of fire intensity. The spatial-temporal dynamics features are based on the Fourier descriptor of contours in space and time, capturing the dynamic properties of fire. These global blob-based features are more robust and effective in rejecting false alarms and nuisance sources than pixel-wise features. In addition, extraction of the spatial-temporal features is very efficient, and no tracking of blobs or contours is needed. We also present a new multi-spectrum fire video database for algorithm testing. We evaluate the effectiveness of the proposed features on fire detection on the video database and obtain very promising results.

1. Introduction

When a fire occurs, minimum detection latency is critical to minimize loss of human lives and property. Traditional sensor-based methods usually use heat sensors, optical sensors (ultraviolet, infrared), ion sensors, or thermocouples as fire detectors. These sensors suffer from the transport delay of the fire (heat or smoke particles) to the sensor, and thus increase detection latency. As point detectors, the range of detection is also limited. Other methods such as beam or aspirated smoke detectors try to reduce the detection latency, but do not really solve the problem. In addition, all these methods usually lack effective human-in-the-loop validation capabilities.

Video Fire Detection (VFD) is a relatively new technology that detects fire through intelligent analysis of video using advanced analytical algorithms. There are several advantages of VFD over traditional methods. First, video camera is a volume sensor. It potentially monitors a larger area and has much higher probability of successful early detection of smoke or flame. VFD is suitable for cases where traditional point sensors are difficult to deploy such as large open facilities and outdoor locations, e.g., power plants, hotel lobby, tunnels, warehouse, etc. Second, a video camera has the capability of remote verification. In case of alarms, human operators are kept in the loop for further verification to minimize false alarms. Finally, with the growth of surveillance and security industries, more and more surveillance cameras have been installed for various security and safety applications. VFD can be easily embedded into existing systems to enable more intelligent functionality.

VFD has been received wide attention from industrial and academic researchers. VFD relies on a set of visual features for detection of fire, such as color, motion, flickering, edge blurring, and textures. The large variations of these features pose great challenges of VFD algorithms due to lighting, wind, reflections, and environment changes. The major problem of current VFD systems is robustness to handle false alarm sources, for example, waving red tree leaves in the fall, reflections of sun on water, strobe lighting, moving people with red shirts, etc.

In this paper, we present a new VFD system based on spatial-temporal patterns of fire as seen in conventional surveillance applications. The contributions of this paper are as follows. First, we propose novel spatial-temporal features for fire detection, including spatial-temporal structural features and spatial-temporal contour dynamics features. These features rely on the global spatial-temporal properties of fire patterns, and thus are more robust and effective in false alarm rejection compared with pixel-level (local) processing. Second, the extraction of the spatial-temporal features is very efficient since it is based on an accumulated motion mask and an accumulated intensity template; no tracking of blobs or contours is needed. Finally, we have acquired a multi-spectrum fire video database. Promising results are presented using the database and our proposed features.
This paper is organized as follows. In Section 2, we briefly review relevant literature on video fire detection. In Section 3 we propose our VFD algorithm, including a spatial-temporal structural feature and a contour dynamics feature, along with a Support Vector Machine (SVM) classifier. Section 4 presents the multi-spectrum fire video database and fire detection results using the proposed algorithm. Finally, Section 5 concludes the paper.

2. Literature Review

Many video fire detection algorithms provide pixel-wise decisions based on color, motion, flickering, and other similar features. Phillips et al. [10] create a Gaussian-smoothed color histogram to detect fire-colored pixels, and then use a temporal variation of pixels to determine which of these pixels are actually fire pixels. Chen et al. [3] apply simple rules of pixel color and dynamics for flame detection. Fire color pixels are first extracted by the saturation component in HSI color space. A fire alarm will be issued if the number of extracted fire pixels is increasing with time and is greater than some threshold during a time interval. Similarly, Huang et al. [6] choose color and dynamic information for flame detection. The dynamic features are based on the growth of fire region and the invariability of the centroid of the flame. Töreyin et al. [13] propose motion, color, and flicker cues for flame detection. The flicker process is modeled using a hidden Markov model and a spatial wavelet analysis of moving fire regions captures color variations in the pixel values. They extend this work to flame detection in infrared videos [14]. Kolesov et al. [8] propose to combine Optimal Mass Transport (OMT) optical flow and color as features for fire detection. The pixel-wise decision is based on a single-hidden-layer neural network. Ko et al. [7] use a background and color model of flames to detect the candidate fire pixels, and then apply a probabilistic fire model based on the expectation that fire pixel values in consecutive frames change constantly. For surveillance videos, one of the problems of pixel-wise decision is its robustness in false alarm rejection due to low quality of surveillance videos.

Another category of VFD work is region or blob based, i.e. the final decision of fire detection is based on the properties of fire blobs. The work by Xiong et al. [16] studies fire detection based on the turbulent characteristic of fire. It consists of motion segmentation, flickering extraction, and contour classification. The contour classification is based on the findings in flow physics, which state that the shape complexity of turbulent phenomena can be characterized by a dimensionless edge/area or surface/volume measure. Borges and Izquierdo [1] propose fire detection for video retrieval applications. The method analyzes the frame-to-frame changes of features like color, area size, texture, boundary roughness, and skewness of the estimated potential fire regions. The frame-to-frame features are input to a Bayes classifier. Liu and Ahuja [9] use Fourier coefficients to represent the contour of fire regions. The temporal changes in these coefficients are used as the temporal signatures of the fire region. An autoregressive (AR) model of the Fourier coefficient series is determined. The Fourier descriptors and AR model parameters are input to a two-class Support Vector Machine (SVM) classifier for fire region detection. They capture only the contour dynamics features of fires and the AR model fitting is computationally expensive. A recent work [5] is using covariance-based features extracted from spatio-temporal video blocks to detect fire.

There is also some work on infrared video fire detection. Under the “Future Naval Capabilities Program” at ONR, in a dozen papers [11], Rose-Pehrsson et al. evaluate several types of volumetric sensors for fire detection in the visible spectrum, near infrared spectrum (700-1000nm), ultraviolet(UV), and acoustic signatures. Their multi-sensor prototype systems are shown to outperform the commercial systems for flaming and smoldering fires and have a high immunity to nuisance sources. Verstockt et al. [15] use visible and LWIR thermal images for flame detection. The hot objects in thermal images and moving objects in visible images are segmented out. Some ad hoc features are extracted, such as bounding box disorder, principal orientation disorder, and histogram roughness. Flames are detected based on features from both modalities.

3. Video Fire Detection Algorithm

To overcome some of the limitations of existing VFD systems, we propose new detection features and test those features on a new multispectral fire database. A flowchart of our VFD system is shown in Figure 1. It consists of the following steps: (1) Pixel-level processing: This step is to identify potential fire pixels with motion and large intensity. Identified pixels may include many non-fire pixels. The potential fire pixels are grouped into blobs. (2) Blob-level feature extraction: The blobs are accumulated within a time buffer to generate an accumulated motion mask and an accumulated intensity template, from which spatial-temporal features are extracted, including spatial-temporal structural features and spatial-temporal contour dynamics features.
The blob-level features are critical for rejection of false alarm sources. (3) Classifier: The extracted features are input to a classifier. We apply a Support Vector Machine (SVM) classifier here. One of the advantages of SVM is its better generalization performance over other methods, such as traditional Bayesian Networks and Neural Networks. In the following sections, we will describe these steps in detail.

3.1. Pixel-Level Processing

3.1.1 Motion Extraction

Foreground objects are detected using background subtraction method. We apply one of the state-of-the-art algorithms, adaptive Gaussian Mixture Model (GMM) originally proposed by Stauffer and Grimson [12]. In practice, a pixel process can be approximated by a number of Gaussian models. The probability of observing X at frame t is $P(X_t) = \sum_{i=1}^{K} w_{i,t} * G(X_t, \mu_{i,t}, \sigma_{i,t}^2)$, where K is the number of Gaussians in the mixture, $w_{i,t}$ is the weight for the $i^{th}$ mode, and G is a Gaussian probability density function. In this process, for each new pixel, the parameters $w_{i,t}$, $\mu_t$ and $\sigma^2$ are updated. We then determine which Gaussians from the mixtures represent the background pixels. Pixels that do not match one of the pixel’s background Gaussians are identified as foreground pixels.

3.1.2 Hot Spot Detection

For infrared or gray-level videos, hot spot detection is used to identify pixels with high intensity values. For color videos, the input images are first transformed into HSI color space. The hue channel represents the color type (such as red, green, or yellow), the saturation channel represents the intensity of the color, and the intensity channel represents the brightness. It is convenient to describe flame properties in this color space since flame usually has high intensity, red-yellow color and high color saturation.

The pixels resulting from motion extraction and hot spot detection are grouped into blobs using connected component analysis. This procedure is effective in determining the whole moving object, from which global spatial-temporal features are extracted.

3.2. Blob-Level Feature Extraction

3.2.1 Spatial-Temporal Structural Feature

We accumulate the detected blobs to obtain long term statistics of blob motion and structure of blobs over some period of time, $T$. For this purpose, we propose to generate two patterns: an accumulated motion mask (AMM) and an accumulated intensity template (AIT), as shown in Figure 2.

Figure 2. Spatial-temporal structural feature extraction

The two patterns capture physical properties of flames, i.e., flames usually show ringing profile in space due to temperature variation, and they also exhibit more flicker in the peripheral areas than in the center. The spatial-temporal structural features of fire are then extracted from the two patterns, including the Histogram of Gradient (HoG) of AMM and the spatial concentric ring structure from AIT.

Histogram of Gradient (HOG) of AMM

For some time period $T_t$, the accumulated motion mask is defined as: $AMM : M = \sum_{m}^{T} m_t$, where $m_t$ is the binary motion mask at time $t$ obtained from background subtraction in Section 3.1.1. The accumulated motion mask essentially captures the flickering properties of flames. The gradient of the accumulated motion mask is then computed, $
abla M = \left( \frac{\partial M}{\partial x}, \frac{\partial M}{\partial y} \right)$. For fire blobs, the histogram of the gradient magnitude (HoG), $\left( \frac{\partial M}{\partial x} \right)^2 + \left( \frac{\partial M}{\partial y} \right)^2$, has the following properties as illustrated in Figure 3:

- There are large number of small-gradient pixels. The majority of the small gradient pixels come from the uneven flickering of fire boundaries.
- There are small number of large-gradient pixels and zero-gradient pixels. These pixels mostly come from the core area of fire due to background subtraction and saturation.

Because we believe the HoG for a fire will be heavy tailed, we propose to apply a Generalized Extreme Value Distribution (GEVD) to fit the HoG. The density function of GEVD $f(x | \lambda, \mu, \sigma)$ is:

$$f(x | \lambda, \mu, \sigma) = \frac{1}{\sigma} \left(1 + k \frac{x - \mu}{\sigma} \right)^{-1 - \frac{1}{\lambda}} \exp \left(- \left(1 + k \frac{x - \mu}{\sigma} \right)^{-\frac{1}{\lambda}} \right)$$

where $\mu$ is the location parameter, $\sigma$ the scale parameter and $k$ the shape parameter. The three parameters are features for fire detection. The shape parameter $k$ controls the tail behavior of the distribution. Three subfamilies, Gumbel, Fréchet and Weibull, correspond to $k = 0, k > 0$ and $k < 0$, respectively. A type I (Gumbel) distribution has a tail decreasing exponentially, A type II (Fréchet) distribution has a tail decreasing as a polynomial, and a type III (Weibull) distribution has a finite tail. We find that the
Spatial Concentric Ring Structure of AIT

For some time period $T$, we define the accumulated intensity template as: $AIT: I = \sum_{t=1}^{T} I_t$, where $I_t$ is the intensity region at time $t$ corresponding to its binary motion mask. The accumulated intensity template averages the intensity regions over time $T$, and captures the ring structure of fire with different temperature regions. The ring structure is constructed from the quantization of AIT by rounding $[2 \log(J)]$ to the nearest integer. The quantization is helpful to reduce noise effects and increase feature reliability. For non-fire blobs, the concentric ring structure does not occur or is much less evident. Figure 4 illustrates the spatial structures of a human motion AIT and a fire AIT.

The quantized AITs are normalized to a specified size, e.g., $[20, 20]$, and will serve as features for fire detection. To reduce the dimensionality, Principal Component Analysis (PCA) is applied to reduce the dimension from 400 to 24, which retains 95% of the total energy.

Spatial Temporal Contour Dynamics Feature

Spatial-temporal contour dynamics is another discriminating feature of fire. From the extracted intensity fire regions, $I_t$, corresponding to their binary motion masks, a contour is formed as the boundary of those pixels. The spatial-temporal contour features are then computed through FFTs of spatial contour points and Eigen-projections of the FFT coefficients across time.

A fire contour usually shows large variations in space and time, while other hot work sequences show less contour variations, such as heating (microwaving, cooking), grinding, welding, cutting, etc. Figure 5 illustrates the contours from fire and microwaving sequences.

We apply a Fourier descriptor to represent fire contours. The Fourier descriptor is defined as the Fourier Transform coefficients of the contour shape, and it was first used by Liu and Ahuja [5] for fire contour description. In their work, an autoregressive (AR) model was learned from the Fourier descriptors of contours to capture fire dynamics, based on the assumption of a linear dynamical system. Here we use eigen-projection to reduce the computational complexity of AR model learning.

Assume that we have $N$ points, $(x_k, y_k), k = 0, ..., N-1$, extracted clockwise from the contour boundary from an arbitrary starting point. The Fourier descriptor is defined as the magnitude of the FFT coefficients of these boundary points: $F_k = \sum_{k=0}^{N-1} P_k \exp(-\frac{2\pi i k}{N} k\mu)$, where contour point $P_k = x_k + iy_k$ is expressed in a complex form. With normalized contour size and centered contour location, the Fourier descriptor is scale and translation invariant. Since only the magnitude of the Fourier coefficients are used, the Fourier descriptor is rotation and starting-point invariant.

For the contours within a time period, we project the Fourier descriptors for all contours to the largest eigen-mode. The features, such as variation and flickering frequency, are derived from the eigen-projections. These features are discriminating for fire detection. The flickering frequency can be calculated using FFT, Wavelet Transform, or Mean Crossing Rate (MCR). In our system, we estimate the MCR over the eigen-projections [16].

![Figure 3](image1.png)  (a) Fire sequence; (b) GEVD fitting curve (in red) of the HoG of AMM; (c) Waving red tree leaves; (d) GEVD fitting curve (in red) of the HoG of AMM.

![Figure 4](image2.png)  (a) Human motion AMM; (b) The non-ring structure of (a); (c) Fire AMM; (d) The spatial ring structure of (c).
3.3. Support Vector Machine

A Support Vector Machine (SVM) is one of the supervised machine learning methods for two-class classification problems. We apply one of the widely deployed SVM library by Chang and Lin [2]. Given \( l \) training samples of \( n \)-dimension, \( x_i \in \mathbb{R}^n \), in two classes and class vector \( y_i \in \{1, -1\} \), the goal is to classify the two classes in the feature space by finding the weights \( \alpha_i \) of the dual expression of the separating hyperplane’s vector \( w = \sum_{1 \leq i \leq l} \alpha_i y_i \phi(x_i) \), where \( \phi(x_i) \) maps \( x_i \) into a higher dimensional space. It can be shown [2] that finding the maximum margin separating hyperplane is equivalent to solving the following optimization problem:

\[
\min_{\alpha} \sum_{1 \leq i,j \leq l} \frac{1}{2} \alpha_i \alpha_j y_i y_j K(x_i, x_j) - \sum_{1 \leq i \leq l} \alpha_i \\
\text{s.t. } \sum_{1 \leq i \leq l} \alpha_i y_i = 0, 0 \leq \alpha_i \leq C, i = 1, ..., l
\]

where \( K(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle \) is the kernel function, and the regularization parameter \( C > 0 \) showing the trade-off between margin maximization and training error minimization. There are a couple of commonly used kernels, including a linear kernel \( K(x, y) = x \cdot y \), a polynomial kernel \( K(x, y) = (x \cdot y + 1)^d \) of degree \( d \), a Radial Basis Function (RBF) kernel \( K(x, y) = e^{-\gamma \|x-y\|^2} \), and a sigmoid kernel \( \tanh(\gamma x \cdot y + c) \), \( \gamma > 0 \). We tested each of these and results are presented in Section 4.

The final decision function is defined as: \( \text{sgn}(f(x)) = \text{sgn} \left( \sum_{1=1}^{l} y_i \alpha_i K(x_i, x) + b \right) \), where the offset \( b \) is chosen so that the margins between the hyperplane and the two classes of sample images are equal.

4. Results

We captured our multi-spectrum database in the fire test facility at the Fire Science Academy, University of Nevada at Reno. The facility simulates shipboard scenarios. The facility is composed of two rooms: a Class A fire room and a Class B fire room. The Class A fire room is for testing fire from ordinary combustibles, such as paper, wood, hay, cardboard boxes, etc. The Class B fire room is for testing flammable liquids or gases, such as gasoline and methane fuel. We also captured some hidden fires occluded by walls or other objects in the scene. In addition, we captured challenging false alarm and nuisance sources, such as people in motion, hot work (grinding, welding, cutting torch), heating (toaster, flare, microwaving), overheated electrical wiring, in varying environmental conditions such as indoors, outdoor shade and sunlight, and so on.

Four cameras were used to captured the multi-spectrum videos, including visible-spectrum videos by Panasonic Camcorder DVC30, Near Infrared (NIR) videos by Pixellink PL-B741EU with 780nm IR pass filter, Broadband IR videos (3 – 14μm) by Electrophysic PV320, and MWIR videos (3 – 5μm) by FLIR GasFinder 320. Sample videos are shown in Figure 7. We extracted 133 video clips in total.

We generated 685 positive and negative AIT samples
based on pixel-level processing described in Section 3.1. Some samples are shown in Figure 8. We extracted features from the HoG of AMMs, the ring structure of AITs, and the spatial-temporal contour dynamics. The features are normalized to $[0, 1]$ before inputting to the SVM classifiers. The normalization is to avoid attributes in greater numeric ranges dominating those in smaller numeric ranges, and to avoid numerical difficulties during the calculation.

We randomly selected 300 positive and negative samples for training. For most existing fire detection algorithms, it is a challenging task to discriminate the false alarm sources from real fire. We obtained very promising results using our proposed VFD system.

An ROC (Receiver Operating Characteristic) curve is used to evaluate the performance of our approach. With the model trained from the 300 samples, the ROC curve is generated for the un-trained 385 samples. We tested different SVM kernels, including linear, second-order polynomial, Radial Basis Function (RBF), and sigmoid. We find that the RBF kernel generates slightly larger AUC (Area Under Curve) than other kernels. Therefore we chose the RBF kernel in the following results.

We also investigated the contributions of different features for fire detection. We first test the significance of the spatial concentric ring structure features. Figure 9 shows the ROC curves of using different features: combined features, HoG of AMM, concentric ring structure, and contour dynamics features. From the figure we can see that these features are discriminating in fire detection, and the contour dynamics features are more discriminating than the concentric ring structure and HoG of AMM features. The combined features give the best performance ($AUC = 0.951$). The promising results demonstrate the effectiveness of the proposed features on fire detection on the multi-spectrum fire video database.

5. Conclusion

We exploited new spatial-temporal features for video based fire detection. The features capture the physical properties of flames, i.e., the structural pattern of temperature profiles, the flickering pattern, and the temporal dynamics of flames. We showed the efficacy of the extracted features on a multi-spectrum video database. We are working on multi-spectrum feature fusion for further improvement of the fire detection performance.

This work was conducted under ONR Contract No. N00014-10-C-0204. The authors would like to acknowledge the guidance and support of the contract manager Dr. Luise Couchman.

References


