This is a repository copy of *Autonomous Flame Detection in Videos with a Dirichlet Process Gaussian Mixture Color Model*.

White Rose Research Online URL for this paper:
http://eprints.whiterose.ac.uk/123156/

Version: Accepted Version

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

Autonomous Flame Detection in Videos with a Dirichlet Process Gaussian Mixture Color Model

Zhenglin Li, *Student Member, IEEE,* Lyudmila S Mihaylova, *Senior Member, IEEE,* Olga Isupova and Lucile Rossi

Abstract—This paper proposes a flame detection framework based on the color, dynamics and flickering properties of flames. The distribution of flame colors is modelled by a Gaussian Mixture Model whose number of Gaussian component is estimated by a Dirichlet process from training data rather than set empirically. The proposed approach estimates the flame color distribution more accurately as it can determine the number of Gaussian components of the mixture model automatically. Additionally, a probabilistic saliency analysis method and a one-dimensional wavelet transform are used to extract motion saliency and filtered temporal series as features, describing the dynamics and flickering properties of flames. The developed Dirichlet Process Gaussian Mixture Model based approach for autonomous flame detection is tested on various videos and achieves frame-wise accuracy higher than 95%.

Index Terms—Flame detection, Dirichlet Process Gaussian mixture model, saliency analysis.

I. INTRODUCTION

FIRE detection techniques have drawn increasing attention in the last decades due to the great loss caused by fires. To reduce injuries as well as financial loss, fire detection systems are usually required to provide quick and accurate alarms [1].

Early works, which mostly employ smoke or heat sensors for fire detection, have several disadvantages. One of them is that these methods are limited to indoor detection and have a significant drop in their performance if applied to large geographical areas. Additionally, environmental factors have a high impact on the performance of these techniques.

Video based fire detection methods become increasingly popular because of the limitations of traditional fire monitoring techniques [2]. Different from sensor based approaches, the computer vision based ones mostly employ information extracted from optical videos rather than detecting smoke or heat. They are not limited to indoor environments and are suitable for large spaces. Moreover, environmental changes hardly influence their performance. Third, faster and more accurate results are enabled by the way information is captured and processed. Finally, compared with expensive sensors, the video based methods cost much less because they can be combined with existing monitoring systems [3].

Z. Li and L. Mihaylova are with the Department of Automatic Control and Systems Engineering, University of Sheffield, Sheffield, S1 3JD, UK e-mail: (ZLI80@sheffield.ac.uk).

O. Isupova is with University of Liverpool, Liverpool, UK

L. Rossi is with UMR CNRS 6134 SPE, University of Corsica, 202250 Corte, France.

The computer vision based fire detection research includes flame detection and smoke detection (or wildfire detection) [1]. Our work in this paper focuses on the former.

To achieve good detection performance, existing work mostly employs properties of color [4]–[6], texture [7], shape [8] and dynamics [6], [7], [9], [10] as features. Specifically, flames are mostly of particular colors, special textures and irregular shapes, and all the properties above vary significantly and dynamically with time. These features are usually combined to obtain more reliable results [11].

Features of colors work effectively and efficiently in detecting flame pixels. Flame chromatic models aim at detecting as many flame pixels as possible while filtering out non-flame ones at the same time. Candidate pixels are further processed based on other features and the falsely detected ones are discarded. There are two main types of color models employed in the literature, e.g. [4], [9], namely empirical inequality models with experimental thresholds and statistical models trained by real data. Two widely accepted empirical inequality models are proposed by Chen et al. [4] and Celik et al. [5]. They work quite well in detecting real flame pixels, but not in filtering out the disturbance in some areas. Comparatively, statistical models work better if a proper model is selected and trained with enough data, such as the Gaussian Mixture Model (GMM) based flame color model proposed by Torenyin et al. [9], [12].

The GMM is able to approximate any arbitrary distribution theoretically and thus is suitable for flame pixels as well. However, the number of mixture components is not known in advance and it is not reasonable to set it empirically. Additionally, the color spaces in which the models are established are crucial as well. RGB models are susceptible to luminance changes as none of the channels is independent on the light intensity. To overcome these disadvantages, researchers have also established their flame chromatic models in YCbCr [5] and HSV [6] spaces to relieve the influence of luminance [13]. Alternatively, the color features are transformed to a new space with a conversion matrix trained by the particle swarm optimization (PSO) with both flame and non-flame pixels, to enhance the classification performance [14]. However, the colors of non-flame pixels are in a very wide range and not easy to be covered by training data. Comparisons of different color models are made in [15].

Since color features only are not enough for accurate flame detection, dynamics as well as foreground detection are widely employed for further verification. Areas of flames are not static because of air flows. Therefore, many existing methods [9], [16]–[18] employ a motion detection step first to

prune out static regions. It not only reduces the computational complexity but also mitigates the interference due to noise and distractors. Widely used methods include background subtraction [9], [16], the motion history image [17] and the adaptive background estimation based on GMM [18].

After processed by background detectors and/or color models, features of the dynamic property of candidate pixels are extracted for discarding falsely detected ones. The method in [7] divides the video into spatio-temporal blocks and employs covariance matrices as features. It works well in most cases although the covariance matrix-based feature fails to distinguish flame-colored moving objects from flames sometimes. Additionally, two novel optical flow estimation methods are specifically designed for two different kinds of flames in [10]. Besides, some research methods analyze the dynamic features in another domain [9] to better reveal the differences between flames and distractors. Toreyin et al. [9] represent the variations of pixels in the wavelet domain to reflect the flickering property. The method works well for frame-wise detection, but sometimes falsely discards pixels in the central regions of flames.

Apart from color and dynamic properties of flames, textures [11], shapes [19] and other features are also employed for flame detection, but mostly together with other properties.

These features are widely used together with machine learning types of classifiers, for example, the support vector machine (SVM) [20] and neural networks [10] and lead to efficient fire detection. However, false detection rates are not low enough when these classifiers are employed with commonly used features such as the scale-invariant feature transform (SIFT) and histogram of oriented gradients (HOG), due to the variety in the appearance of flames.

Many methods work well in the detection rate of fires currently, e.g. [3], [7]. However, most of them suffer from high false alarm rates, which significantly hinders the applications of these techniques. Therefore, the challenge of flame detection lies in achieving reliable accuracy as well as low false alarm rates.

To solve or relieve the above problems, we propose a hybrid flame detection framework based on Dirichlet Process-Gaussian Mixture Model (DPGMM) [21], [22], saliency analysis and one-dimensional (1-D) wavelet transform in this paper. The proposed color model first assigns each pixel a probability describing how likely it is to be a part of flames according to the color. Subsequently, a saliency map is obtained based on the optical flow magnitude of each pixel with a probabilistic approach. The saliency map is combined with the results of the color model to decide candidate flame pixels with two independent experimental thresholds. Furthermore, the framework prunes out pixels whose intensities are not larger than the mean of the current frame. Candidate pixels obtained by the above three steps are further processed by a 1-D wavelet analysis step based on the flickering characteristic of flames. A frame-wise decision is made according to the number of finally detected flame pixels of each frame.

Compared with the approach proposed in [23], the main contribution of this paper lies in the novel DPGMM based flame color model. The model employs a GMM to represent the flame color distribution, with the number of Gaussian components automatically estimated by the Dirichlet Process (DP) from training data [22]. The DPGMM can estimate the distribution of flame colors well since it learns the component number of a GMM from training data rather than setting it empirically as in existing methods [9], [12]. An improperly preset Gaussian component number will lead to poor estimation of other parameters, i.e. means and covariance matrices, which usually results in imprecise estimation of the color distribution of flames. Therefore, the DPGMM significantly enhances the detection performance of the color model, which will contribute to high final detection rates. Additionally, the bright property of flames is employed by discarding pixels whose intensities are smaller than the average value of the processed frame, instead of using a grayscale saliency map in [23]. It describes the flame characteristics better and thus contributes to more accurate detection results.

This paper is organized as follows: the DPGMM flame color model is presented in Section II and we describe the detection approach in Section III. Experiments and discussions of results are provided in Section IV while conclusions are drawn in Section V.

II. DPGMM FLAME COLOR MODEL

The DP estimates the number of GMM components by inferring the posterior of data assignments to clusters, with the assumption that there is an infinite number of latent clusters, but only a finite number of them is used to generate the observed data. The Dirichlet Process is widely used in topic modelling [22], abnormal detection [24] and other areas, but it has not been applied to fire detection yet.

A. Dirichlet Process and Chinese Restaurant Process

1) Dirichlet Process: The Dirichlet Process (DP) works on problems of exchangeable observations [22]. Each observation is denoted as $\mathbf{x}_i$ and is generated from a distribution with the parameter $\boldsymbol{\theta}_i$ ($\mathbf{x}_i$ and $\boldsymbol{\theta}_i$ can be either a scalar or a vector). Different $\boldsymbol{\theta}_i$s are exchangeable and may not be of distinct values. The parameter $\boldsymbol{\theta}_i$ is generated from a prior distribution G. Thus, we have the model as follows:

$$\boldsymbol{\theta}_i | G \sim G \qquad \text{for each } i \qquad (1)$$
$$\mathbf{x}_i | \boldsymbol{\theta}_i \sim F(\boldsymbol{\theta}_i) \qquad \text{for each } i, \qquad (2)$$

where $F(\boldsymbol{\theta}_i)$ is the distribution of $\mathbf{x}_i$ given $\boldsymbol{\theta}_i$. It is assumed that each parameter $\boldsymbol{\theta}_i$ is conditionally independent given the distribution G.

Given a measurable space and a probability measure G_0 on the space [25], a Dirichlet Process is defined as a distribution of a probability measure G over the space. It satisfies the condition that for any finite measurable partition $(A_1, ..., A_r)$ of the space, $(G(A_1), ...G(A_r))$ follows a Dirichlet distribution with parameters of $(\alpha_0 G_0(A_1), ..., \alpha_0 G_0(A_r))$, where α_0 is a positive real parameter, i.e.

$$(G(A_1), ...G(A_r) \sim \text{Dir}(\alpha_0 G_0(A_1), ..., \alpha_0 G_0(A_r)). \quad (3)$$

When G follows a Dirichlet process, we denote it as $G \sim DP(\alpha_0, G_0)$ with the parameter α_0 and a base distribution G_0.

2) Chinese Restaurant Process: The Chinese Restaurant Process (CRP) is a distribution over partition of integers with a parameter α_0 [26]. It is another perspective of the DP. For $G \sim DP(\alpha_0, G_0)$, the CRP focuses on samples from G.

Consider a boundless Chinese restaurant with an infinite number of tables and each table can serve an unlimited number of customers. A sequence of customers $\theta_1, \theta_2, ...$ (a metaphor for a sequence of exchangable random variables drawn from G) comes into the restaurant and chooses tables to sit at. The i-th customer θ_i can either sit at an existing table or choose a new one, following the distribution given below

$$p(c_i|c_1,...,c_{i-1}) = \begin{cases} \dfrac{m_{k,-i}}{i-1+\alpha_0} & \text{at an occupied table } k \\ \dfrac{\alpha_0}{i-1+\alpha_0} & \text{at a new table,} \end{cases} \quad (4)$$

where c_i is an indicator variable specifying on which table customer θ_i sits and $m_{k,-i}$ is the number of customers already at table k (not including θ_i).

After all the customers have taken their seats, a partition plan of those customers (variables $\theta_1, \theta_2, ...$) is obtained. When related with the DP, the customers (random variables) at the same table share a parameter vector drawn from the base distribution G_0. The discrete values of the table related parameters are denoted by $\phi = \{\phi_1, \phi_2, ...\}$. The CRP thus has a naturally clustering property and tables here in the CRP correspond to clusters. The cluster number is influenced by the concentration parameter α_0, as it decides how likely a customer chooses a new table relatively to customers already in the restaurant. Additionally, since $\theta_1, \theta_2, ...$ are exchangeable, each customer can be treated as the last one.

B. DPGMM Based Flame Color Model

We train a GMM to model the flame color distribution in the RGB space. As the training data is extracted from images of various illumination, the DPGMM based flame color model is robust to different lighting conditions. Denote the color vector of a flame pixel i as $\mathbf{x}_i = [R_i, G_i, B_i]^T$. Then we have

$$p(\mathbf{x}_i|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \sum_{k=1}^{K} w_k \mathcal{N}(\mathbf{x}_i|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k), \quad (5)$$

where $\mathcal{N}(\cdot)$ is the Gaussian distribution. Here $\boldsymbol{\mu}_k$, $\boldsymbol{\Sigma}_k$ and w_k denote the mean, covariance and the weight of the k-th Gaussian component respectively. Then the GMM is charactersized with the parameters: $\boldsymbol{\mu} = \{\boldsymbol{\mu}_1, ..., \boldsymbol{\mu}_K\}$ and $\boldsymbol{\Sigma} = \{\boldsymbol{\Sigma}_1, ..., \boldsymbol{\Sigma}_K\}$. The parameters of the k-th component is denoted as $\phi_k \triangleq \{\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\}$, which corresponds to table k in the CRP metaphor. As the component number K is not intuitively known, we employ the DPGMM related approach to find it for more reliable results.

According to the mixture model theory [26], each $\mathbf{x}_i$ is generated by first choosing a component indexed by c_i which is distributed according to $\mathbf{w} = [w_1, ..., w_K]$. Afterwards, the observation $\mathbf{x}_i$ is generated from the chosen Gaussian component with the parameter $\theta_i = \phi_{c_i} \triangleq \{\boldsymbol{\mu}_{c_i}, \boldsymbol{\Sigma}_{c_i}\}$. However, the distribution weight $\mathbf{w}$ is not available only

with known observations, so we assume that θ_i is distributed according to a DP. Thus, the generative model is

$$G \sim DP(\alpha_0, G_0) \quad (6)$$
$$\theta_i|G \sim G \quad (7)$$
$$\mathbf{x}_i|\theta_i \sim \mathcal{N}(\boldsymbol{\mu}_{c_i}, \boldsymbol{\Sigma}_{c_i}). \quad (8)$$

However, neither the GMM parameters nor the data allocations are known with only training data $\mathbf{X} = \{\mathbf{x}_1, ..., \mathbf{x}_N\}$ available. The collapsed Gibbs sampling [27] is employed here to obtain the assignments of the data $\mathbf{X}$ and other parameters can be estimated based on them. As θ is distributed according to G, the distribution of c_i conditional on $\{c_1, ..., c_{i-1}\}$ is induced by the CRP. Therefore the posterior is as follows [28].

$$\begin{aligned} &p(c_i = k|\mathbf{c}_{-i}, \mathbf{X}, \alpha_0, G_0) \\ &\propto p(c_i = k|\mathbf{c}_{-i}, \alpha_0) \cdot p(\mathbf{x}_i|\mathbf{X}_{-i}, c_i = k, \mathbf{c}_{-i}, G_0), \end{aligned} \quad (9)$$

where $k \in \{1, ..., t, k^*\}$ and t denotes the number of occupied tables while k^* means choosing a new table. Besides, $\mathbf{X}_{-i}$ and $\mathbf{c}_{-i}$ are referred to as the training data except $\mathbf{x}_i$ and their allocations.

To compute the conditional distribution of $c_i = k$ based on assignments of other observations, θ_i is treated as the last customer according to the exchangeability. Thus, the first term of Eq. (9) can be obtained from Eq. (4)

If θ_i sits at an existing table, the second term of Eq. (9) is

$$p(\mathbf{x}_i|\mathbf{X}_{-i}, c_i = k, \mathbf{c}_{-i}, G_0) \quad (10)$$
$$= p(\mathbf{x}_i|\mathbf{X}_{k,-i}, G_0) \quad (11)$$
$$= \frac{p(\mathbf{x}_i, \mathbf{X}_{k,-i}|G_0)}{p(\mathbf{X}_{k,-i}|G_0)} \quad (12)$$
$$= \frac{\int p(\mathbf{x}_i|\boldsymbol{\theta}_k) \left[\prod_{j \neq i, c_j = k} p(\mathbf{x}_j|\boldsymbol{\theta}_k) \right] G_0(\boldsymbol{\theta}_k) \mathrm{d}\boldsymbol{\theta}_k}{\int \left[\prod_{j \neq i, c_j = k} p(\mathbf{x}_j|\boldsymbol{\theta}_k) \right] G_0(\boldsymbol{\theta}_k) \mathrm{d}\boldsymbol{\theta}_k}, \quad (13)$$

where $\mathbf{X}_{k,-i} = \{\mathbf{x}_j : j \neq i, c_j = k\}$ denotes the other customers assigned to table k not including $\mathbf{x}_i$.

Similarly, if the i-th customer chooses a new table, we have

$$p(\mathbf{x}_i|\mathbf{X}_{-i}, c_i = k^*, \mathbf{c}_{-i}, G_0) = p(\mathbf{x}_i|G_0) \quad (14)$$
$$= \int p(\mathbf{x}_i|\boldsymbol{\theta}) G_0(\boldsymbol{\theta}) \mathrm{d}\boldsymbol{\theta}. \quad (15)$$

Based on the collapsed Gibbs sampling for the CRP described from Eq. (9) - Eq. (15), allocation plans are obtained after convergence. Then we can estimate the GMM parameters with the training data and their assignments to clusters.

With the trained color model, each pixel is assigned a probability describing how likely it is part of flames according to its color. Flame pixels will obtain higher probabilities while non-flame regions are likely to have lower ones with an accurate estimation of the flame color distribtution. Given an appropriately chosen threshold, several candidate pixels are obtained for further processing.
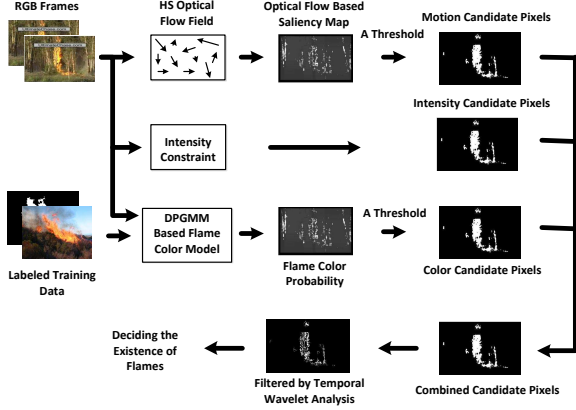
Fig. 1. Flow Chart of the Proposed System.

Fig. 2. Schematic Diagram of Sliding Window.

III. HYBRID FLAME DETECTION SYSTEM BASED ON FUSION OF DIFFERENT FEATURES

A motion saliency map of each frame is obtained with the method in [23], [29], [30]. Candidate pixels with small motions are discarded based on the selected threshold, while others are combined with the results of the color model in Section II. Subsequently, the framework discards pixels whose intensities (I channel values of the HSI color space) are smaller than the mean value of the current frame. Additionally, candidate pixels are further checked by a wavelet transform based analysis step [9]. A frame-wise decision is made according to the final number of candidate flame pixels. The flow chart of the framework is provided in Fig. 1.

A. Probabilistic Saliency Analysis

This step aims at measuring the motion saliency of each pixel. It is based on a probabilistic interpretation of the semi-local feature contrast. A sliding rectangular window W, shown in Fig. 2, is employed in the approach, divided into an inner kernel K and a border area B. The widths and heights of the window W and kernel K are denoted as w_W, h_W, and w_K, h_K, respectively. Let (x, y) be a point inside the window W with $F(x, y)$ (the optical flow magnitude in our framework) as its feature value. The Horn-Schunck method [31] is employed for the estimation of optical flows.

Two hypotheses are proposed as H_0: the point is not salient and H_1: the point is salient. The prior probabilities $p(H_0)$ and $p(H_1)$ satisfy $p(H_0) = 1 - p(H_1)$. It is assumed at the beginning that H_0 is valid for the points in B while H_1 for those in K.

The posterior probability $p(H_1|F(x, y))$ reflects the saliency $S(x, y)$ of the point (x, y) according to its feature $F(x, y)$. That is

$$S(x, y) = p(H_1|F(x, y)). \qquad (16)$$

Using Bayes' theorem

$$p(H_1|F(x, y)) = \frac{p(H_1)p(F(x, y)|H_1)}{p(H_1)p(F(x, y)|H_1) + p(H_0)p(F(x, y)|H_0)}. \qquad (17)$$
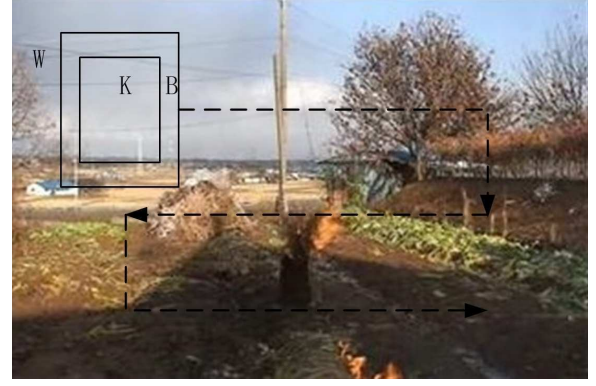
It is reasonable to estimate the priors according to the area ratios of K and B. The likelihoods $p(F(x, y)|H_1)$ and $p(F(x, y)|H_0)$ are estimated using histograms of $F(x, y)$ computed in K and B. To enhance the robustness, the obtained histograms are smoothed with a Gaussian blur function before normalization. Mathematically,

$$\hat{p}(F(x, y)|H_1) = Norm(g(F) * hist_K(F))$$
$$\hat{p}(F(x, y)|H_0) = Norm(g(F) * hist_B(F)), \qquad (18)$$

where $\hat{p}(F(x, y)|H_1)$ and $\hat{p}(F(x, y)|H_0)$ denote the estimated likelihoods respectively, $g(F)$ stands for the Gaussian blur function and $Norm$ denotes the normalization operation.

The sliding window W centering at point j is denoted by $W(j)$. When K is sliding with a step s_W, windows at different positions may overlap. If it happens, $S(x, y)$ is calculated as follows

$$S(x, y) = \max_j \{S_j(x, y)|(x, y) \in W(j)\}. \qquad (19)$$

Different step and window scales are employed to reduce the influence of step and window sizes. Using the saliency estimation approach mentioned above, a motion saliency map of each frame is obtained by setting optical flow magnitudes as features.

B. 1-D Wavelet Transform Based Analysis

As flames flicker with frequencies around 10Hz [9], different from most distractors, the proposed framework applies the 1-D wavelet transform to analyze this property of flame pixels, especially those on the boundaries [9]. Denote $r_k(x, y)$ as the R channel value of a pixel located at (x, y) in the k-th frame. Then we set $R_k(x, y) = [r_k(x, y), r_{k+1}(x, y), \cdots, r_{k+N-1}(x, y)]$ as a temporal series of R values for N frames. To reveal the temporal characteristics, a 1-D discrete wavelet transform is performed on $R_k(x, y)$ as

$$T_k(x, y) = DWT(R_k(x, y)), \qquad (20)$$

where $DWT(\cdot)$ represents the 1-D discrete wavelet transform with a high-pass and low-pass filter of [-0.25 0.5 -0.25] and [0.25 0.5 0.25] in our experiments. High frequency wavelet series T_k of pixels in flame regions and those in moving object areas differ significantly, as shown in Fig. 3 and Fig. 4.
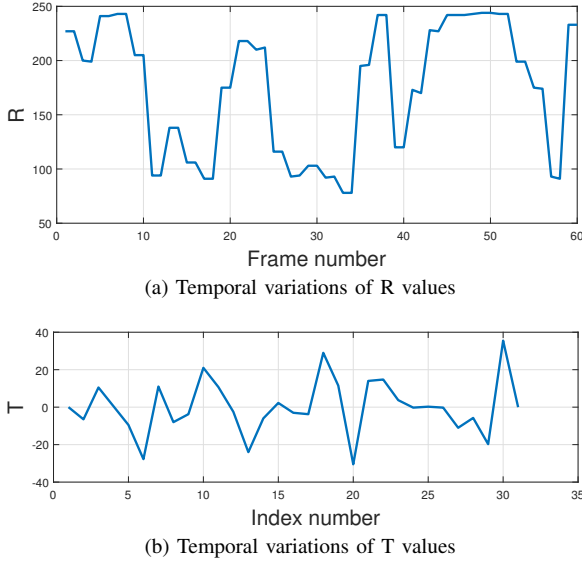
(a) Temporal variations of R values

(b) Temporal variations of T values

Fig. 3. Temporal variations of a flame pixel.

(a) Temporal variations of R values

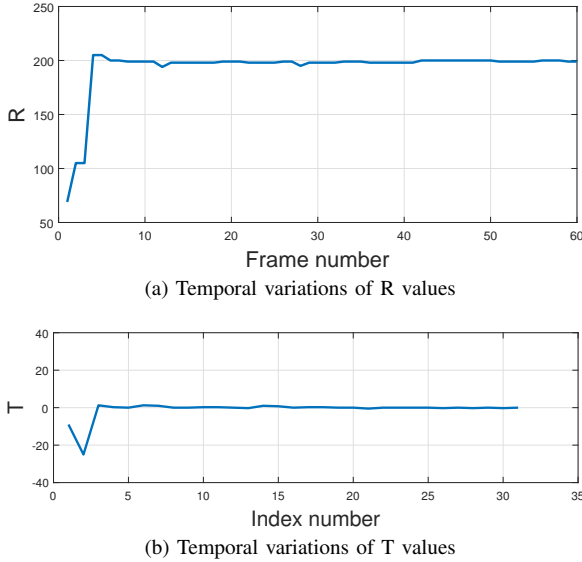(b) Temporal variations of T values

Fig. 4. Temporal variations of a nonfire pixel of a moving object.

It is obvious that T_k series of a flame pixel fluctuates with several zero-crossings and values are relatively large. Instead, the values of T_k series in moving object regions are around zero with only one spike. Therefore, the number of spikes or zero crossings can be employed to distinguish flame pixels with non-flame ones. In our experiment, candidate pixels with less than five zero-crossings are discarded as non-flame ones.

IV. EXPERIMENTS AND DISCUSSION

A. Benchmarking Database and Performance Evaluation Methods

We validate the performance of both the introduced DPGMM color model and the entire detection framework. The proposed color model is tested on 50 images from the database of [2]. The frame-wise performance of the framework is tested with videos of various scenes (downloaded from [3], [32]).

There are 4468 frames altogether and are different from the training ones. Table I describes the testing videos briefly.

The performance of the DPGMM model is illustrated pixel-wise with the Receiver Operating Characteristic (ROC) curve [33]. Comparisons with other models are also shown in the ROC curve. Different from the color model, the framework is evaluated frame-wisely as the detection of fire existence is more important than marking flame regions. A widely accepted way is with true positive rate (TPR) and true negative rate (TNR) [33], which reflect the sensitivity and specificity of models, respectively. However, the TPR and TNR are usually competing. Therefore, flame detection methods need to balance the TPR and TNR to achieve satisfactory performance.

A natural logarithmic threshold of -14.4441 is selected for the color probabilities obtained from the DPGMM color model in the experiment. Moreover, the motion saliency threshold is set as 0.21. Besides, a frame is considered as a flame one when the detected flame pixels in it are more than 25.

B. Performance Comparison of Flame Color Models and Analysis

The DPGMM flame color model is trained on 293756 flame pixels with concentration parameter α_0 set as a fixed value of 1. The base distribution G_0 is set as the conjugate prior Gaussian-Wishart distribution for computational convenience. With training data of $\mathbf{X} = \{\mathbf{x}_1, ..., \mathbf{x}_N\}$, the hyperparameters of the Gaussian-Wishart distribution are set as

$$G_0 \sim \mathcal{N}(\boldsymbol{\mu}'|\boldsymbol{\mu}_0, (\beta\boldsymbol{\Lambda})^{-1})\mathcal{W}(\boldsymbol{\Lambda}|\mathbf{W}, v) \tag{21}$$

$$\boldsymbol{\mu_0} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i \tag{22}$$

$$\mathbf{W} = s * \mathbf{I}_d \tag{23}$$

$$s = \frac{1}{N * d}\sum_{i=1}^{N}\|\mathbf{x}_i - \boldsymbol{\mu_0}\|_2^2 \tag{24}$$

where d is the dimension of $\mathbf{x}_i$ and $\mathbf{I}_d$ is a $d \times d$ identity matrix. The number of degrees of freedom v is set equally to the data dimension. The scale parameter β is set as 1 in our experiment.

The trained GMM has 22 mixture components after discarding the ones with too small weights (less than 0.001). It is quite different from the predetermined component number 10 of the model proposed in [9], [12]. Their method also assumes that R, G and B are independent and each channel has the same variance for computational convenience. However, our estimated covariances show that the assumption is not so reasonable.

A threshold is needed in the proposed model to turn the obtained probability of each pixel into a binary detection result. The ROC curve of the DPGMM based model is shown in Fig. 5 together with those of some state of the art models introduced by Chen et al. [4], Celik et al. [5] and Toreyin et al. [9], [12]. From it we can see that the DPGMM based model achieves a higher TPR than others with a FPR smaller than 0.05. Specifically, our model outperforms the fixed component number GMM based color model proposed

TABLE I
TESTING VIDEOS

| Video	Burning Objects	Distractors	Positive Frames	Negative Frames	Lighting Condition	Smoke Condition	Location
V1	Trees	None	230	0	Bright	Thin	Outdoor
V2	Trees	None	192	0	Dark	Thin	Outdoor
V3	Branches	A walking man	692	0	Bright	Medium	Outdoor
V4	Grass	None	386	0	Bright	Medium	Outdoor
V5	Papers	A moving light	395	0	Bright	Thin	Indoor
V6	Trees	None	202	0	Bright	Thick	Outdoor
V7	Assemble line	None	571	70	Bright	Thin	Indoor
V8	None	A walking person in red clothes	0	155	Bright	None	Indoor
V9	None	Flashing carlights	0	378	Bright	None	Indoor
V10	None	Moving cars and people	0	943	Bright	None	Outdoor
V11	None	Moving people	0	254	Bright	None	Indoor
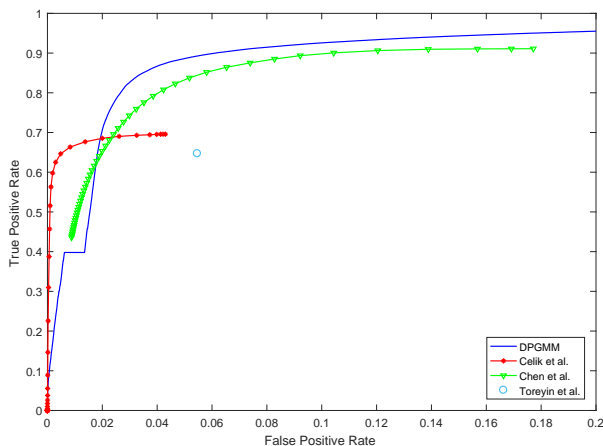
Fig. 5. ROC curves of color models.

by Toreyin et al. [9], [12]. It is because the DPGMM estimates the parameters more accurately and approaches a better estimation of the flame color distribution than a GMM of fixed component number. It proves experimentally the advantages of the DPGMM which learns the component number from training data instead of setting it empirically. The models of [4] and [5] both contain a group of rules, so the false positive rate (FPR) cannot reach 1 no matter how the thresholds are changed. Besides, the model in [9], [12] has no threshold since it conducts a hard classification (pixels within two standard deviations from one of the means are classified as flame ones).

Furthermore, Fig. 6 shows the detected flame color pixels by those methods. From the results, we can see that the proposed approach succeeds in detecting most flame pixels including the ones behind thick smoke. At the same time, it prunes out most pixels of a flame-colored trunk, which works better than the model by Chen et al. [4]. The DPGMM based model works well in discarding artificial red colors with higher saturation values. Thus, it can reduce false alarms caused by distractors like red vehicles or clothes. Although the proposed model misses some pixels of inner parts of flames, it will not influence the final detection results. This can be explained with the fact that contour pixels of flames rather than inner ones reflect the dynamic property better.

C. Detection Performance Evaluation and Discussion

In Fig. 7 and Fig. 8, some examples are given, including both successful detections of flames and excluding the disturbance of non-fire objects. Fig. 7 shows not only the final detected pixels but also intermediate results of each phase of the detection framework. It can be seen that the DPGMM based flame color model detects most flame regions for further processing, which helps enhance the TPR of the final detection results. Though parts of the grounds are detected as candidate pixels based on colors due to the reflection of lights emitted by flames, they are then pruned out by the saliency map and temporal wavelet transform based analysis.

Fig. 9 and Fig. 10 illustrate the frame-wise detecting results of the proposed framework compared with Method 1 in [3] and Method 2 in [23]. We can see that the introduced approach achieves good performance on most of the experiment videos. It overwhelms Method 1 and achieves a significant enhancement compared with Method 2 in TPR, especially in V3 shown in Fig. 7. The challenge lies in the transparent color of weak flames at the beginning and end of the video. The proposed approach works better than benchmark methods because the DPGMM based color model achieves a higher TPR, which means it detects more flame pixels for further processing. Instead, the color models in the compared methods (Chen et al. in Method 1 [3] and Celik et al. in Method 2 [23]) fail to detect pixels of weak flames, resulting in misdetection of fires of semitransparent colors like in V3. As we know, flames are mostly weak at the beginning of fires and thus not easy to be detected. Better performance on these situations means earlier detection of fires which can reduce injuries and financial loss. At the same time, both Method 2 and the proposed system work well in reducing false alarm rates. Although the TNRs of Method 2 are slightly better than the proposed approach in a few videos (both are higher than 99%), our framework works much better in TPR. Generally, the introduced framework achieves better results on all the experiment videos than comparing approaches. The average TPR and TNR are 97.08% and 99.50% respectively. The performance of the proposed framework is shown in Table II.

V. CONCLUSIONS

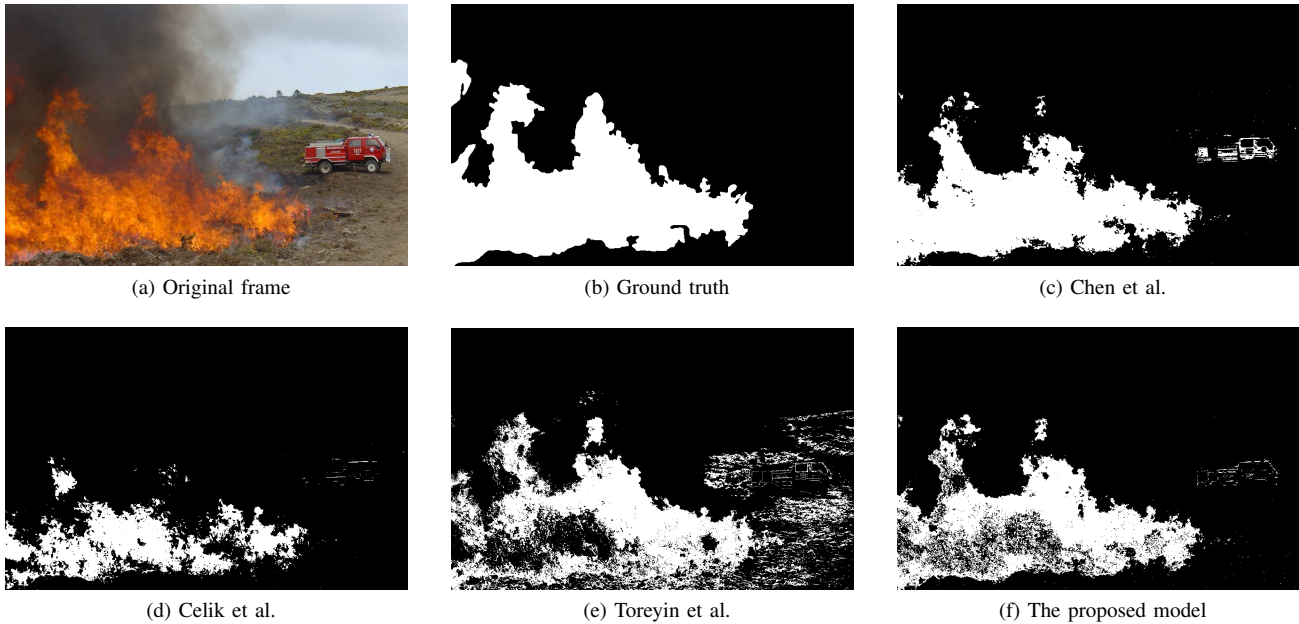This paper proposes a novel flame color model based on the DPGMM, employed together with saliency analysis and tem-

(a) Original frame   (b) Ground truth   (c) Chen et al.

(d) Celik et al.   (e) Toreyin et al.   (f) The proposed model

Fig. 6. Comparison of detection results by different flame color models.

TABLE II
DETECTION PERFORMANCE

Videos	tp	fn	tn	fp	Total positive frames	Total negative frames	TPR	TNR
V1	230	0	0	0	230	0	1	-
V2	192	0	0	0	192	0	1	-
V3	630	62	0	0	692	0	0.9104	-
V4	384	2	0	0	386	0	0.9948	-
V5	392	3	0	0	395	0	0.9924	-
V6	202	0	0	0	202	0	1	-
V7	560	11	70	0	571	70	0.9807	1
V8	0	0	155	0	0	155	-	1
V9	0	0	378	0	0	378	-	1
V10	0	0	937	6	0	943	-	0.9936
V11	0	0	251	3	0	254	-	0.9882
Overall	2590	78	1791	9	2668	1800	0.9708	0.9950

poral wavelet transform for flame detection. The color model approaches the flame color distribution with a GMM whose component number is learned from training data by a Dirichlet process. It avoids the deviations caused by improper number of Gaussians set empirically and thus achieves more accurate estimation of other parameters of the GMM. Experiments show that the proposed approach outperforms existing color models. Together with the saliency analysis and the wavelet transform based temporal feature, the developed color model contributes to final detection results of TPR and TNR higher than 95%, which are better than state of the art approaches. For future work, we aim to solve the multiscale problem of flame detection with super-resolution algorithms.

REFERENCES

[1] A. E. Çetin, K. Dimitropoulos, B. Gouverneur, N. Grammalidis, O. Günay, Y. H. Habibolu, B. U. Töreyin, and S. Verstockt, "Video fire detection–Review," *Digital Signal Process.*, vol. 23, no. 6, pp. 1827–1843, 2013.
[2] T. Toulouse, L. Rossi, M. Akhloufi, T. Celik, and X. Maldague, "Benchmarking of wildland fire colour segmentation algorithms," *Image Process.*, vol. 9, no. 12, pp. 1064–1072, 2015.
[3] A. E. Cetin, "Computer vision based fire detection software [Online]." Available: http://signal.ee.bilkent.edu.trWAR/VisiFire/.
[4] T. Chen, P. Wu, and Y. Chiou, "An early fire-detection method based on image processing," in *Proc. 2004 Int. Conf. Image Process.*, vol. 3, pp. 1707–1710, IEEE, 2004.
[5] T. Celik and H. Demirel, "Fire detection in video sequences using a generic color model," *Fire Safety J.*, vol. 44, no. 2, pp. 147–158, 2009.
[6] H. Yamagishi and J. Yamaguchi, "A contour fluctuation data processing method for fire flame detection using a color camera," in *Proc. 26th Annu. Conf. Ind. Electron. Soc.*, vol. 2, pp. 824–829, IEEE, 2000.
[7] Y. H. Habiboğlu, O. Günay, and A. E. Çetin, "Covariance matrix-based fire and flame detection method in video," *Mach. Vision and Appl.*, vol. 23, no. 6, pp. 1103–1113, 2012.
[8] P. Foggia, A. Saggese, and M. Vento, "Real-time fire detection for video-surveillance applications using a combination of experts based on color, shape, and motion," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 9, pp. 1545–1556, 2015.
[9] B. U. Töreyin, Y. Dedeoğlu, U. Güdükbay, and A. E. Cetin, "Computer vision based method for real-time fire and flame detection," *Pattern Recognit. Lett.*, vol. 27, no. 1, pp. 49–58, 2006.
[10] M. Mueller, P. Karasev, I. Kolesov, and A. Tannenbaum, "Optical flow estimation for flame detection in videos," *IEEE Trans. Image Process.*, vol. 22, no. 7, pp. 2786–2797, 2013.
[11] R. Chi, Z.-M. Lu, and Q.-G. Ji, "Real-time multi-feature based fire flame detection in video," *IET Image Process.*, vol. 11, no. 1, pp. 31–37, 2016.
[12] B. U. Töreyin, *Fire detection algorithms using multimodal signal and image analysis*. PhD thesis, Bilkent Univ., Ankara, Turkey, 2009.
[13] V. B. Celen and M. F. Demirci, "Fire detection in different color models," in *Proc. Int. Conf. Image Process., Comput. Vision, and Pattern Recognit.*, pp. 1–7, 2012.
[14] A. Khatami, S. Mirghasemi, A. Khosravi, C. P. Lim, and S. Nahavandi, "A new PSO-based approach to fire flame detection using K-medoids clustering," *Expert Syst. With Appl.*, vol. 68, pp. 69–80, 2017.
[15] T. Toulouse, L. Rossi, T. Celik, and M. Akhloufi, "Automatic fire pixel detection using image processing: A comparative analysis of rule-based and machine learning-based methods," *Signal, Image and Video Process.*, vol. 10, no. 4, pp. 647–654, 2016.
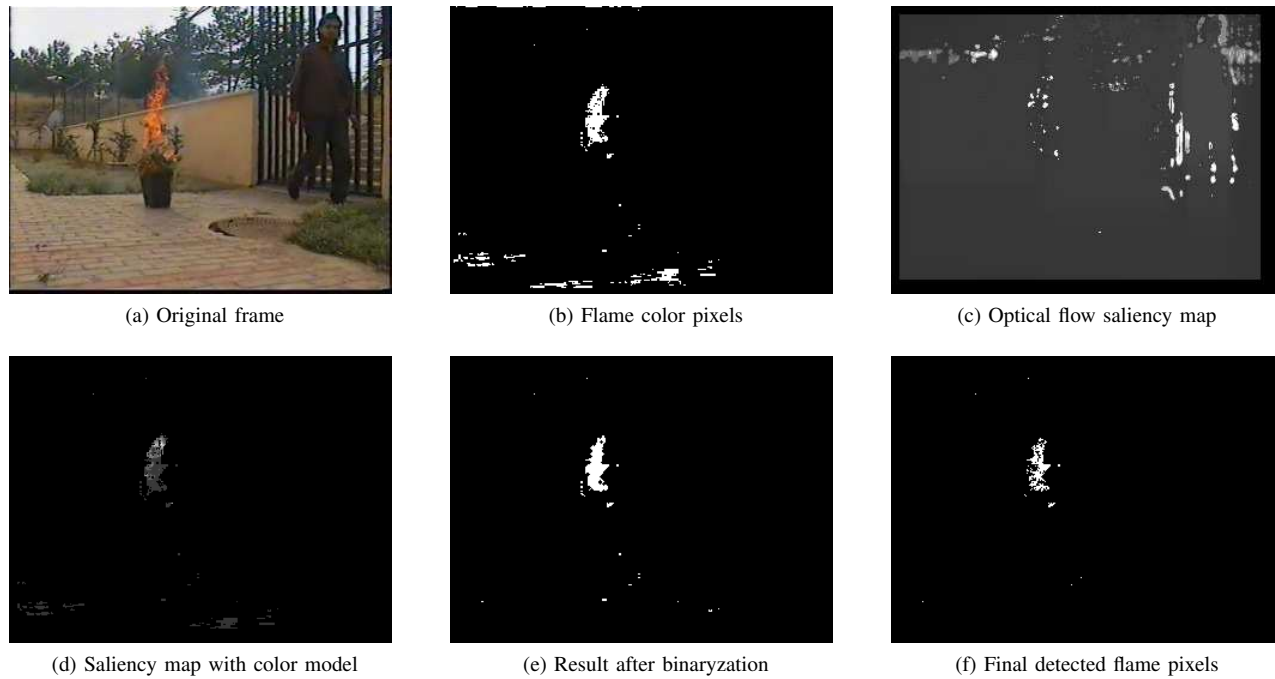
(a) Original frame

(b) Flame color pixels

(c) Optical flow saliency map

(d) Saliency map with color model

(e) Result after binaryzation

(f) Final detected flame pixels

Fig. 7. Flame detection results.

(a) Original frame

(b) Original frame

(c) Original frame

(d) Final detected flame pixels

(e) Final detected flame pixels
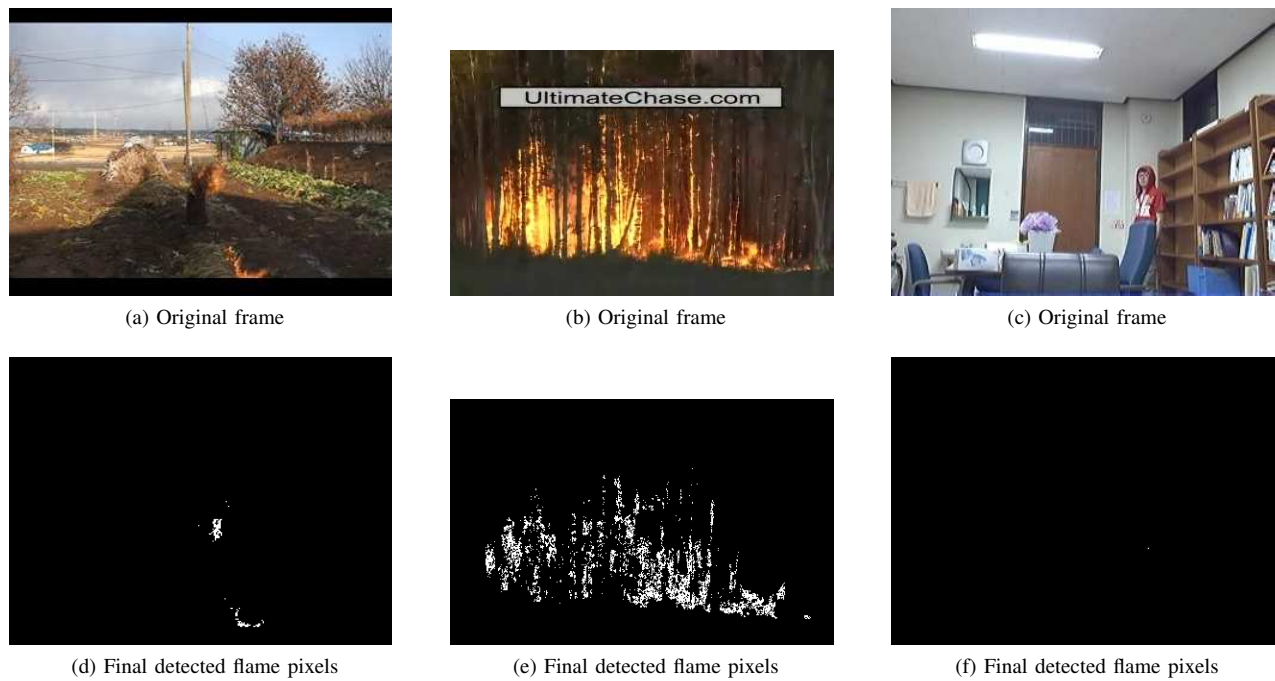

(f) Final detected flame pixels

Fig. 8. Flame detection examples.

[16] T. Celik, H. Demirel, and H. Ozkaramanli, "Automatic fire detection in video sequences," in *Proc. 14th European Signal Process. Conf.*, pp. 1–5, IEEE, 2006.

[17] D. Han and B. Lee, "Flame and smoke detection method for early real-time detection of a tunnel fire," *Fire Safety J.*, vol. 44, no. 7, pp. 951–961, 2009.

[18] T. X. Truong and J. M. Kim, "Fire flame detection in video sequences using multi-stage pattern recognition techniques," *Eng. Appl. Artificial Intell.*, vol. 25, no. 7, pp. 1365–1372, 2012.

[19] D.-c. Wang, X. Cui, E. Park, C. Jin, and H. Kim, "Adaptive flame detection using randomness testing and robust features," *Fire Safety J.*,

vol. 55, pp. 116–125, 2013.

[20] B. C. Ko, K.-H. Cheong, and J.-Y. Nam, "Fire detection based on vision sensor and support vector machines," *Fire Safety J.*, vol. 44, no. 3, pp. 322–329, 2009.

[21] D. Görür and C. E. Rasmussen, "Dirichlet process Gaussian mixture models: Choice of the base distribution," *J. Comput. Sci. and Technol.*, vol. 25, no. 4, pp. 653–664, 2010.

[22] Y. W. Teh, M. I. Jordan, M. J. Beal, and D. M. Blei, "Hierarchical Dirichlet processes," *J. Amer. Statistical Assoc.*, vol. 101, no. 476, pp. 1566–1581, 2006.

[23] Z. Li, O. Isupova, L. Mihaylova, and L. Rossi, "Autonomous flame
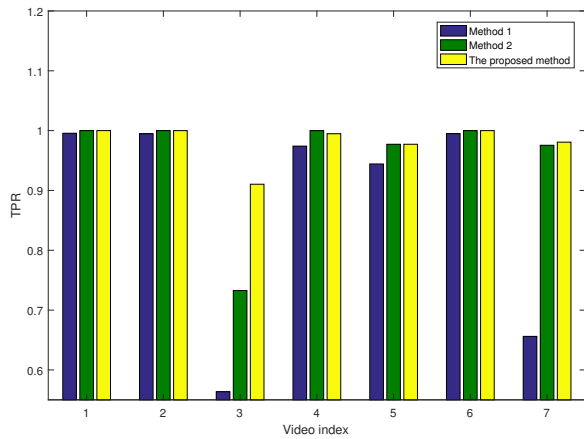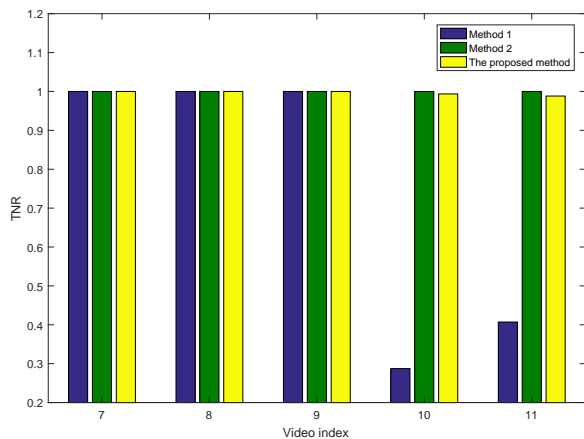
Fig. 9. Comparison of TPR with other methods.

Fig. 10. Comparison of TNR with other methods.

detection in video based on saliency analysis and optical flow," in *Proc. Int. Conf. Multisensor Fusion and Integration*, pp. 218 – 223, IEEE, 2016.

[24] O. Isupova, D. Kuzin, and L. Mihaylova, "Dynamic hierarchical dirichlet process for abnormal behaviour detection in video," in *19th Int. Conf. Inform. Fusion*, pp. 750–757, July 2016.

[25] V. I. Bogachev, *Measure Theory*, vol. 1. Springer Science & Business Media, 2007.

[26] S. J. Gershman and D. M. Blei, "A tutorial on Bayesian nonparametric models," *J. Math. Psychology*, vol. 56, no. 1, pp. 1–12, 2012.

[27] R. M. Neal, "Markov chain sampling methods for Dirichlet process mixture models," *J. Computational and Graphical Statist.*, vol. 9, no. 2, pp. 249–265, 2000.

[28] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*. MIT press, 2012.

[29] Y. Jia, J. Yuan, J. Wang, J. Fang, Q. Zhang, and Y. Zhang, "A saliency-based method for early smoke detection in video sequences," *Fire Technol.*, pp. 1–22, 2015.

[30] E. Rahtu and J. Heikkilä, "A simple and efficient saliency detector for background subtraction," in *Proc. 12th Int. Conf. Comput. Vision Workshops*, pp. 1137–1144, IEEE, 2009.

[31] B. K. Horn and B. G. Schunck, "Determining optical flow," *Artificial Intell.*, vol. 17, no. 1-3, pp. 185–203, 1981.

[32] Computer Vision and Pattern Recognition Laboratory, "KMU Fire and Smoke Database." http://cvpr.kmu.ac.kr/.

[33] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognition Lett.*, vol. 27, no. 8, pp. 861–874, 2006.

Zhenglin Li is a PhD candidate in the department of Automatic Control and Systems Engineering, University of Sheffield, United Kingdom. She got her master degree on Signal and Information Processing in 2015, from Dalian University of Technology, China. Her research is on computer vision, Bayesian nonparametrics and fire detection based on videos.

Lyudmila Mihaylova (M'98, SM'2008) is Professor of Signal Processing and Control at the Department of Automatic Control and Systems Engineering at the University of Sheffield, United Kingdom. Her research is in the areas of machine learning and autonomous systems with various applications such as navigation, surveillance and sensor network systems. She has given a number of talks and tutorials, including the plenary talk for the IEEE Sensor Data Fusion 2015 (Germany), invited talks University of California, Los Angeles, IPAMI Traffic Workshop 2016 (USA), IET ICWMMN 2013 in Beijing, China. Dr. Mihaylova is an Associate Editor of the IEEE Transactions on Aerospace and Electronic Systems and of the Elsevier Signal Processing Journal. She was elected in March 2016 as a president of the International Society of Information Fusion (ISIF). She is on the board of Directors of ISIF and a Senior IEEE member.

Olga Isupova is a postdoctoral researcher at the Department of Computer Science at the University of Liverpool. She got her PhD at the Department of Automatic Control and Systems Engineering at the University of Sheffield, 2017. She received the Specialist (eq. to M.Sc.) degree in Applied Mathematics and Computer Science, 2012, from Lomonosov Moscow State University, Moscow, Russia. Her research is on machine learning, Bayesian nonparametrics, anomaly detection.

Lucile Rossi is an assistant professor at the University of Corsica since 1998 and HDR since 2011. Her research is dedicated to the image and signal processing for the study and the modeling of environmental physical phenomena. From 2001 to 2004, she worked on the development of an acoustic device to deter dolphins from fishing nets. In 2007, she created a thematic entitled Image processing and vision for the study and modeling of forest fires. Her current research activities concern image processing and vision for the development of metrology instruments based on 3D vision dedicated to forest fires.