IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received October 21, 2019, accepted December 9, 2019, date of publication December 16, 2019,
date of current version December 26, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2960209

Unsupervised Segmentation of Fire and Smoke From Infra-Red Videos

MEENU AJITH AND MANEL MARTÍNEZ-RAMÓN, (Senior Member, IEEE)
Department of Electrical and Computer Engineering, The University of New Mexico, Albuquerque, NM 87106, USA
Corresponding author: Meenu Ajith (majith@unm.edu)

This work has been supported by the NSF S&CC EAGER under Grant 1637092.

ABSTRACT This paper proposes a vision-based fire and smoke segmentation system which uses spatial, temporal and motion information to extract the desired regions from the video frames. The fusion of information is done using multiple features such as optical flow, divergence and intensity values. These features extracted from the images are used to segment the pixels into different classes in an unsupervised way. A comparative analysis is done by using multiple clustering algorithms for segmentation. Here the Markov Random Field performs more accurately than other segmentation algorithms since it characterizes the spatial interactions of pixels using a finite number of parameters. It builds a probabilistic image model that selects the most likely labeling using the maximum a posteriori (MAP) estimation. This unsupervised approach is tested on various images and achieves a frame-wise fire detection rate of 95.39%. Hence this method can be used for early detection of fire in real-time and it can be incorporated into an indoor or outdoor surveillance system.

INDEX TERMS Fire detection, gaussian mixture models, iterated conditional modes, k-means clustering, markov random fields, optical flow.

I. INTRODUCTION

The fire and smoke detectors are an important part of fire-fighting systems and are also widely used in monitoring indoor buildings and outside environments. The conventional detection systems use inbuilt sensors that do not issue the alarm unless the particles reach the sensors to activate them. To obtain high precision, the sensors must be distributed densely nearby. Hence in real-world situations, these are highly inefficient, and the delayed response may cost the life of firefighters and other human lives. As an appropriate alternative to conventional methods, vision-based fire and smoke detection systems were introduced in the past few years.

The vision-based systems either utilize the color information of fire and smoke or it uses the dynamic motion features [1], [2]. The classical approaches operated on RGB [3], YCbCr [4], CIE L*a*b [5] or HSI [6] color spaces and the image pixels were classified according to the appearance model of the fire. But the use of color information gives high false alarm rates due to similar colors present in the surrounding environments. Another approach involving disorder analysis and growth rate was used to minimize these false alarms [7], [8]. Further, one of the most popular methods was wavelet analysis, which used various extracted features for the detection process. The wavelet domain energy analysis was done to analyze the time-variant behavior of the smoke [9]. To improve the performance, an Expectation-Maximization based GMM model was obtained by training the pixels of previously occurred events [10]. On the other hand, [11] used a change detection algorithm by extracting foreground pixels. Nevertheless, the system suffers heavily from a change in illumination and hence requires fine-tuning of the algorithm parameters. In [12], the optical flow vectors were calculated based on the turbulence characteristics of the smoke and were used to eliminate the background. Other methods calculated the motion direction of smoke by assuming grayscale invariance in the optical flow algorithm [13]. Several papers also modeled a Markovian process by considering the motion of the fire [1]. Later, Hidden Markov models [14] were used to distinguish between the flame and flame-colored objects. These models evaluated the spatial color variations in flame to reach a final decision.

The state-of-the-art architecture of a fire/smoke detector can be summarized in three steps such as pixel-wise classification of fire/smoke, region-based segmentation and the analysis of these regions. In this paper, a comparative analysis of different segmentation algorithms is done to find the appropriate one for fire and smoke detection. The experiments are conducted on Infrared (IR) datasets available online.

The associate editor coordinating the review of this manuscript and approving it for publication was Eduardo Rosa-Molinar.

Moreover, different feature extraction methods such as optical flow, SIFT flow and divergence are also evaluated. The feature vector is computed for each of the IR videos and it is further passed on to various segmentation algorithms. In particular, the main methods used in the segmentation of fire and smoke are K-Means Clustering [15], Gaussian Mixture Models (GMM) [16], Markov Random Fields (MRF) [17] and Gaussian Markov Random Fields (GMRF) [18]. Finally, the confusion matrix and accuracy are computed to analyze the efficiency of the system. This work is intended to discriminate fire (meaning the fire flames present in the scene), smoke and background in thermal imaging sequences.

The novelty of this approach is the following:

- It involves the fusion of intensity, divergence, and optical flow-based features to obtain the most discriminative features for segmentation. The divergence and optical flow features are chosen since they calculate the flow at a given point and the displacement of pixels from one frame to another. Hence these motion features are combined with the intensity values representing the variations in temperature to form the most significant feature vector for segmentation. Flames, smoke and background have different dynamic and thermal features, so they form three different clusters. Therefore, the easiest way to detect flames is to discriminate them from smoke and the background.
- The cascading of the feature extraction with an MRF framework for the segmentation of fire and smoke. The main advantage of the proposed system is that it does unsupervised learning by modeling the likelihood of the data. The latent variable is already known and for this experiment, it can have 3 values corresponding to fire, smoke and background. Since these unsupervised algorithms support pre-training, it can be used in real-time for testing. Though various deep learning strategies are currently used to obtain high performance; they are supervised and hence require a large amount of labeled data for training [19], [20]. In the case of real-time firefighting scenarios, the availability of labeled data of both fire and smoke is limited.
- The proposed system has low computational complexity and less trainable parameters. It also requires lesser training data to obtain high accuracy in segmentation. Multi-spectral systems can be developed using these algorithms by mixing IR and UV [21] data. Further, the feature extraction also allows fusing information from different sensors and hence by using this information the fire, smoke, and background can be accurately classified.

II. RELATED WORKS

Several state-of-the-art segmentation methods have been proposed based on the characteristics of fire in different color spaces. One of the earlier methods exploited the YCbCr color space [4] whereas another method combined the saturation channel of HSV with the RGB color space [2]. These techniques were based on decision rules, which were further used to define thresholds for segmentation. Similarly, in [22] a reference color model was introduced to represent smoke in the RGB space. Another feature extraction method that was used for texture analysis was Local Binary Patterns (LBP) [23]. It was used to identify the smoke regions by extracting the blurred edges. Some works also utilized the fractal property [24] to create the feature vector for smoke detection. Another model utilized the probability-based thresholding in the RGB color space [25]. The YUV and RGB color spaces were also used together to develop a 3-D Gaussian model [26].

Further, there are several methods of fire detection based on motion features. Traditionally, background subtraction [27], [28] is done on consecutive frames to identify the motion features whereas, in [29], [30], adaptive GMM was also used for modeling the background. Later, the optical flow was introduced to calculate the motion of smoke [31]. The main disadvantage of these approaches was high computational cost and noise sensitivity. Apart from using the motion and color features fire detection was also done by applying wavelet transform [1]. In [14] both Wavelet transforms, and Hidden Markov Model was combined to identify the flame-colored pixels. But these algorithms perform well only if the fire is in a close range with clear visibility. Hence the detection of fire and smoke is still challenging due to varying lighting and diverse background. The general-purpose cameras which capture RGB images fail to produce relevant information during nighttime. Infrared (IR) images are highly reliable for detecting fire during situations with inadequate lighting because of the transparent nature of smoke. Hence in [32], infrared (IR) images were utilized to distinguish smoke from other smoke like objects. The IR images were used to construct a mask for smoke and the efficiency of smoke detection was enhanced by fusing thermal and visual information. In [33], [34] a thresholding based method is implemented to obtain binary images. Since the pixels with high intensity corresponds to fire in IR images, this helps to minimize the false alarm rates. Decision fusion [35] was another method that utilized IR images for forest fire detection. In [36], fire detection is done in IR images by utilizing the brightness and motion features. Additionally, a histogram-based segmentation is done to extract the target regions. A comparative analysis of different algorithms which uses video-based smoke detection for visible and infrared images are presented in [37]. In the case of algorithms that use infra-red images, the highest detection rate achieved was 91% [38]. Here the different methods also use a combination of features to decrease the false alarm rate. Finally, the comparisons concluded that the IR images are more reliable in smoke detection than the images in the visible range [28].

A few other works also fuse visual and IR data to improve the robustness of the detection method [39]. Even though these fusion approaches have been implemented, another challenging problem is the optimization of the number

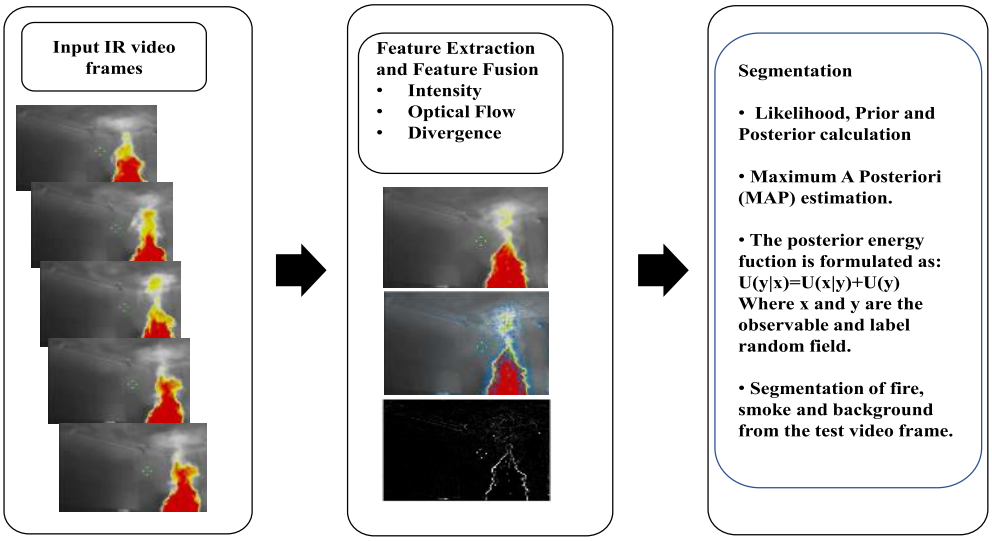

FIGURE 1. Structure of the Proposed Methodology. The frames are first pre-processed to extract the features (intensity, optical flow and divergence in the final algorithm). After this, a segmentation that uses a Bayesian modeling is applied to the features of the pixels. K-Means, GMM, GMRF and MRF are the compared models. The experiments show the superior performance of the MRF segmentation method.

of features. Our work implements a cascaded approach consisting of multiple feature extraction and segmentation for detecting fire and smoke in a computationally inexpensive manner.

III. PROPOSED ALGORITHM

The proposed algorithm has two main stages, which are feature extraction and segmentation. The algorithm is depicted in Fig. 1. The algorithm can be summarized as follows. The first stage (labeled as *feature extraction* in the figure), consists of obtaining different features of every pixel of the sequence of P images. The first set of features is just the intensity of the pixel, which represents the temperature. From the sequence of P images, we extract the velocity field and assign a velocity to each pixel in the sequence. This is an estimation of the speed of the particles moving in the sequence of images. Finally, from the velocity field, we compute its divergence field. We use the magnitude of the velocity and divergence as features. We also compute the SIFT flow features of the images as additional features. The details of the feature extraction methods are included in the following subsections. The second stage of the algorithm (*segmentation*) is an unsupervised segmentation of the pixels as a function of their features. We compare a variety of methods that cluster the data into different classes. Clustering methods can be seen as the construction of a set of class conditional likelihood functions for each of the possible classes. Using these likelihood functions and assuming a parametric prior for the classes, parametric posterior probabilities given the observed pixels can be proposed and then the parameters of both prior and posterior are estimated. After this, the segmentation is completed by deciding on the class of the pixel based on the maximum value for those posteriors (maximum a posteriori criterion). The clustering methods compared in this paper are K-means, Gaussian Mixture Models (GMM) trained with the Expectation- Maximization Algorithm (EM), Markov Random Fields (MRF) and Gaussian Markov Random Field (GMRF).

Apart from the probabilistic interpretation of these methods, a reason to choose Bayesian methods for clustering is that they have no free parameters to adjust. The price to pay for this advantage is that an assumption on the class of the likelihood functions and the prior probabilities have to be assumed. In three of the methods, we choose a Gaussian assumption. K-means further assumes that the covariance is proportional to an identity (isotropic). The MRF assumes that the likelihood is a general Gibbs (exponential) form. Other methods not exploiting the Markovian do not need a probabilistic model, as the classical self-organizing map [40] or they may present lower computational burden, as in the case of Gradient descent K-means [41], but in exchange, they have learning parameters and structure dimensions that have to be cross-validated and this is why they haven't been considered in this paper.

A. FEATURE EXTRACTION

The first feature is the pixel intensity itself. The subsequent features (velocity, divergence, and SIFT features) are computed using sequences of two consecutive images. This feature extraction sequence is summarized below with illustrative examples of the result of each one of the extracted features.

1) OPTICAL FLOW

The first stage of feature extraction in the proposed algorithm is optical flow computation. Our proposed fire detection algorithm takes advantage of one of the visually detectable characteristics of fire, i.e. motion. Here the motion estimation

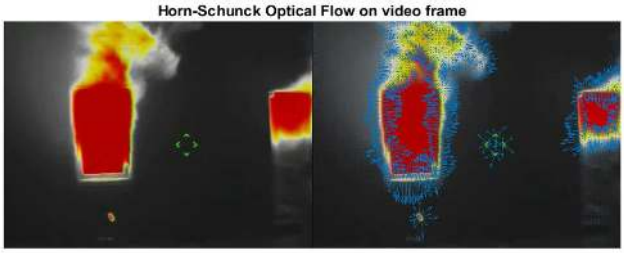

FIGURE 2. Left: A raw image frame with fire and smoke. Right: The quiver plot of the optical flow vectors on that frame, represented as blue arrows.

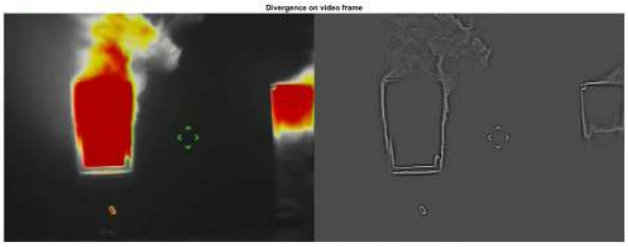

FIGURE 3. Left: Raw image. Right: Divergence of the velocity vector field.

of fire is done using the Horn-Schunck Optical Flow method [42]. This algorithm makes use of the flow vectors of moving objects over time to detect moving regions in an image. It computes a 2-dimensional vector known as the motion vector which indicates the velocities as well as the directions of each pixel of two consecutive frames in a time sequence. The algorithm is based on two approximations: The intensity of a moving object is preserved between two frames, and all objects in a frame will be present in the next frame.

If they are applied to the intensity $I(\mathrm{x}, \mathrm{y}, \mathrm{t})$ of an image, then one can state

$$I(\mathrm{x}, \mathrm{y}, \mathrm{t}) = I(\mathrm{x} + \Delta\mathrm{x}, \mathrm{y} + \Delta\mathrm{y}, \mathrm{t} + \Delta\mathrm{t}) \tag{1}$$

Using the Taylor series approximation on the above equation and by assuming the movement to be small, we obtain the following first order approximation:

$$I(\mathrm{x} + \Delta\mathrm{x}, \mathrm{y} + \Delta\mathrm{y}, \mathrm{t} + \Delta\mathrm{t}) = I(\mathrm{x}, \mathrm{y}, \mathrm{t}) + \frac{\partial I}{\partial \mathrm{x}}\Delta\mathrm{x} + \frac{\partial I}{\partial \mathrm{y}}\Delta\mathrm{y} + \frac{\partial I}{\partial \mathrm{t}}\Delta\mathrm{t} + \mathcal{O}(t) \tag{2}$$

where $\mathcal{O}$ is the bgf O Landau function. Combining equations (1) and (2) gives expression

$$\frac{\partial I}{\partial \mathrm{x}}V_\mathrm{x} + \frac{\partial I}{\partial \mathrm{y}}V_\mathrm{y} + \frac{\partial I}{\partial \mathrm{t}} = 0 \tag{3}$$

where $V_\mathrm{x} = \frac{\Delta\mathrm{x}}{\Delta\mathrm{t}}$, $V_\mathrm{y} = \frac{\Delta\mathrm{y}}{\Delta\mathrm{t}}$ and $e = -\frac{\varepsilon}{\Delta\mathrm{t}}$. The constrained Horn-Schunk optimization method minimizes the regularization term

$$\mathcal{E}(\mathrm{x}, \mathrm{y}, \mathrm{t}) = \left(\frac{\partial V_\mathrm{x}}{\partial \mathrm{x}}\right)^2 + \left(\frac{\partial V_\mathrm{x}}{\partial \mathrm{y}}\right)^2 + \left(\frac{\partial V_\mathrm{y}}{\partial \mathrm{x}}\right)^2 + \left(\frac{\partial V_\mathrm{y}}{\partial \mathrm{y}}\right)^2 \tag{4}$$

subject to the constraint in (3). Minimizing this term is equivalent to minimize the variation of the velocities across space. The details of this derivation can be found in [42]. Fig. 2 shows the optical vectors on a video frame with fire and smoke. In general, it can be appreciated that the edges of the fire have a higher velocity field, and the smoke has lower values. Also, it can be observed that the core of the fire does not show any velocity since the temperature is uniform and constant across frames.

2) DIVERGENCE

The next step of our algorithm is the computation of the divergence of the velocity field. The divergence is defined as the net amount of flux (represented by the velocity field) entering or leaving a point. It is defined as positive when there is a net quantity of flux leaving the point, otherwise it is negative or zero. Given a vector field $\vec{D} = V_\mathrm{x}\vec{i} + V_\mathrm{y}\vec{j}$, the divergence is computed as:

$$div\vec{D} = \nabla \cdot \vec{D} = \frac{\partial V_\mathrm{x}}{\partial \mathrm{x}} + \frac{\partial V_\mathrm{y}}{\partial \mathrm{y}} \tag{5}$$

Hence the measure of expansion or compression of an object in the field is given by divergence. By applying divergence, a clear contrast between widening and narrowing of flow vectors can be visualized. As an example, Fig. 3 shows the divergence operator applied to the velocity field of the right panel of Fig. 2. The left panel of Fig. 3 shows a high divergence in the edges of the fire and a lower divergence in the areas with smoke. Since the surface of the fire is constant across frames, the velocity field and hence the divergence, are zero. Thus, both velocity and divergence are significant of smoke and fire edges, but the inclusion of the intensity of pixels needs to be added to detect the fire inside its edges.

3) SIFT FLOW

Besides the dynamic features computed above, SIFT features are added to them to test whether they improve the classification by adding information to the feature vector. This has been added because, in the optical flow feature selection described above, the main assumptions include brightness constancy and velocity smoothness constraint.

Nevertheless, the pixel displacements in images of distinct scenes can be larger than the magnitude of the motion vectors. Thus, the assumptions used in classical optical flow may not be strong enough. These issues are addressed using the SIFT flow algorithm [43]. Primarily the SIFT descriptors are extracted from each pixel location and these descriptors are constant for the pixel displacement field. The SIFT descriptors are brightness independent and view-invariant image structures. Hence when there is significantly different image content, matching these SIFT descriptors helps to establish meaningful correspondences across the images. These descriptors can be used even when the pixel displacements are as large as the image itself. But the smoothness of the pixel displacement across images is still assumed since close

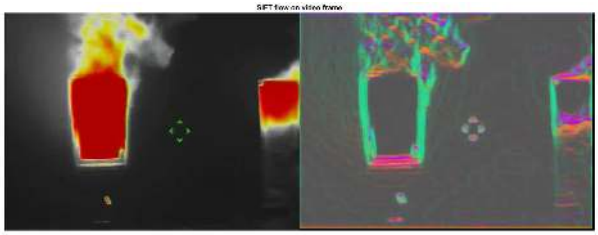

FIGURE 4. Left: Raw image. Right: Visualization of the SIFT image by projecting the SIFT vector into the 3D RGB space. The color-coding scheme uses the first 3 principal components of the SIFT descriptor to map into the principal components of the RGB space. Here the first component is transformed into R+G+B, next into R-G and the last is mapped into R/2+G/2−B. Similar pixel displacements are visualized using similar colors [43].

by pixels tend to have similar displacements. Thus, the search of the correlated SIFT descriptors across the images is formulated as an optimization problem with a cost function as follows:

$$\begin{aligned} E(w) = &\sum_{p} min\Big(\|s_1(p) - s_2(p + w(p))\|_1, t\Big) \\ &+ \sum_{p} \eta\Big(|u(p)|+|v(p)|\Big) \\ &+ \sum_{(p,q)\in\epsilon} min\Big(\alpha|u(p) - u(q)|, d\Big) \\ &+ min\Big(\alpha|v(p) - v(q)|, d\Big) \end{aligned} \tag{6}$$

The above function consists of a data term, displacement term, and a smoothness term. The displacement vector at pixel location $p = (x, y)$ is given by $w(p) = (u(p), v(p))$, ϵ corresponds to the spatial neighborhood of a pixel and $s_i(p)$ denotes the SIFT descriptor extracted at location p in image i. The first term in the above objective function has an L1 norm calculation to account for outliers in SIFT matching whereas a thresholded L1 norm is used in the third term along with the regularization parameter α to model discontinuities in the pixel displacement field. Further, the optimization is done using a dual-layer loopy belief propagation algorithm. Here the smoothness term is decoupled and hence allows to separate u and v during message passing [44]. Thus at one iteration of the message passing the complexity is reduced from $\mathcal{O}(n^4)$ to $\mathcal{O}(n^2)$. The distance transforms [45] is used further to reduce the complexity since the functional form of the objective function has truncated L1 norms.

B. SEGMENTATION

As stated before, the segmentation is performed by the use of a clustering procedure. Once the clusters are estimated, a procedure is applied to determine what are the most likely cluster for each sample. Clustering algorithms are based on a conditional probabilistic model of the observable data x_i, where the condition is its class. Thus each pattern x_i has an associate latent variable $z_i \in \{1, \cdots K\}$. K-means abnd GMM assume that the latent variables are independent and that the observable data are conditionally independent, this is $\forall i, j,\ p(x_i|x_j, z_i = k) = p(x_i|z_i = k)$.

The models for K-means and GMM are usually represented by Gaussian functions. In K-means, the K class conditionals are identical and isotropic. Thus, the observation log-likelihood is proportional to the Euclidean distance of the samples to the K means of the distributions. The posteriors are simply approximated to 1 for the distribution with the closest mean, 0 for the rest. GMM assumes variable covariance matrices, which gives more flexibility to the model, and the posteriors are computed using the Bayes theorem through the data likelihood and the latent variable priors.

The MRF model [46]–[48] uses a likelihood for the data identical to the GMM one, but it assumes that there is a relationship between a pixel and its neighbors, so the latent variables in the same image are remodeled using an undirected graph. The algorithm is usually updated using the Iterated Conditional Modes (ICM) method. The GMRF also models the likelihood similarly and here the random variable associated with a pixel is considered to be jointly Gaussian. The methods are summarized below.

1) K-MEANS

K-means clustering [49] is an unsupervised learning algorithm. It partitions the data points into K clusters and each of the data points belongs to the cluster with the closest mean value. Based on the feature similarity, the algorithm works iteratively to assign each data point to one of K clusters. The algorithm inputs the data set and the number of clusters K. The initial estimates for K centroids are either generated randomly or selected from the data set. Then it iterates between two steps:

- Data assignment step: Here each centroid defines one of the clusters. Based on the squared Euclidean distance, each data point is assigned to its nearest centroid. Let c_i be the collection of centroids in set C, then each data point x is assigned to a cluster K based on the following:

$$\arg\min_{c_i \in C} dist(c_i, x)^2 \tag{7}$$

- Centroid update step: The mean is calculated for all data points assigned to a centroid's cluster and thereby the centroids are recomputed.

$$c_i = \frac{1}{|S_i|} \sum_{x_i \in S_i} x_i \tag{8}$$

where S_i be the set of data point assignments for each i^{th} cluster centroid.

The algorithm iterates between these steps and it converge when none of the data points change clusters and the sum of the distances is minimized.

2) GAUSSIAN MIXTURE MODELS

A probabilistic model can be used in the representation of normally distributed subpopulation in a dataset. Gaussian mixture models [16] are such models that learn about

the subpopulation without knowing which subpopulation a data point belongs to. This constitutes a form of unsupervised learning since the assignment of the subpopulation is unknown. The mixture of Gaussians is represented as follows:

$$p(x_i|\mu_k, \Sigma_k) = \sum_{k=1}^{K} \pi_k \mathcal{N}(x_i|\mu_k, \Sigma_k) \tag{9}$$

$$\mathcal{N}(x_i|\mu_k, \Sigma_k) = \frac{1}{\sqrt{(2\pi)^K |\Sigma_k|}} exp(-\frac{1}{2}(x_i-\mu_k)^T \Sigma_k^{-1}(x_i-\mu_k)) \tag{10}$$

where x_i denotes the observed variables. The mixture component weights and the component mean and covariances characterizes a Gaussian mixture model. In the multivariate case μ_k denotes the mean whereas Σ_k corresponds to the covariance matrix. For each latent variable z_k, we define prior probabilities π_k. The total probability distribution normalizes to 1 with the constraint that $\sum_{k=1}^{K} \pi_k = 1$.

When the number of components K is known, expectation-maximization is employed to estimate the parameters of the mixture model. It is a numerical technique used in the estimation of maximum likelihood. It is an iterative technique with the property that with each subsequent iteration the maximum likelihood of the data increases strictly. Hence it reaches a local maximum at the end of the procedure. The expectation-maximization consists of two steps. In the Expectation step, the posterior probability γ_{ik} that, each data point belongs to each cluster is calculated using the current estimated mean vectors and covariance matrices. While in the Maximization step, the cluster means and covariances are recalculated based on the probabilities calculated in the expectation step. The steps are repeated until the algorithm converges, providing a maximum likelihood estimate. Thus, the main algorithm is as follows:

- Evaluation of the log likelihood after initializing the means, covariances and the mixture component weights.
- E-step: Calculation of the posterior probability that the data point x_i belongs to component z_k. Thus $\gamma_{ik} = p(z_k|x_i, \pi, \mu, \Sigma)$

$$\gamma_{ik} = \frac{\pi_k \mathcal{N}(x_i|\mu_k, \Sigma_k)}{\sum_{k=1}^{K} \pi_k \mathcal{N}(x_i|\mu_k, \Sigma_k)} \tag{11}$$

- M-step: Re-estimate the new parameter values using the γ_{ik} calculated in the E-step.

$$\pi_k = \sum_{i=1}^{N} \frac{\gamma_{ik}}{N}$$

$$\mu_k = \frac{\sum_{i=1}^{N} \gamma_{ik} x_i}{\sum_{i=1}^{N} \gamma_{ik}}$$

$$\Sigma_k = \frac{\sum_{i=1}^{N} \gamma_{ik}(x_i - \mu_k)^2}{\sum_{i=1}^{N} \gamma_{ik}} \tag{12}$$

- Evaluation of the log likelihood function using the new values of mean, covariance and mixture component weights.

$$\ln p(X|\mu, \Sigma, \pi) = \sum_{i=1}^{N} \ln(\sum_{k=1}^{K} \pi_k \mathcal{N}(x_i|\mu_k, \Sigma_k)) \tag{13}$$

If there is no convergence,the E step is repeated and finally using the fitted model density estimation and clustering is done.

3) MARKOV RANDOM FIELD

In images, neighboring pixels exhibit similar properties such as intensity, texture and color information. The Markov random field (MRF) [50] is an undirected graphical model that makes use of this contextual information and represent them in probabilistic terms. Based on the Markov random field theory, any digital image consists of a discrete set of pixels that can be modeled using a set of random variables. The site is a term that is used to denote every pixel in an image and each site is given a label y which represents the intensity value of a pixel. Let an $M \times N$ digital image be described as $S = \{(i, j)|1 \leq i \leq m, 1 \leq j \leq n\}$ where S is a rectangular grid. The relations between the sites in S are defined using a neighborhood system and a set of sites in S is said to be a clique C if every pair of sites in C is neighbors to each other. Hence there exist two random fields; the label random field $y = \{y_i|s_i \in S\}$ and the observable random field $x = \{x_i|s_i \in S\}$.

The main goal of segmentation [51] is to find the optimum estimation of hidden field y from observed field x i.e., to estimate the correct classification for each pixel. The MRF uses the maximum a posterior probability estimation to minimize the probability of misclassification.

$$\hat{y} = \arg\max_{y} P(y|x) \tag{14}$$

The Hammersley-Clifford theorem [52] states that any MRF can be described by a probability distribution P(y) which follows Gibbs form.

$$P(y) = \frac{1}{Z} e^{-\frac{U(y)}{T}} \tag{15}$$

where P(y), Z and T denote the prior probability, normalization constant and temperature parameter respectively. The energy function U(y) can also be represented as follows.

$$U(y) = \sum_{c \in C} V_c(y) \tag{16}$$

where $V_c(y)$ denotes the potential function. Here U(y) is the sum of clique potentials $V_c(y)$ over all possible cliques C. It is also assumed that one pixel has at most 4 neighbors.

Therefore the clique potential can be either singleton, doubleton and other higher orders depending on the number of neighbors. Thus a clique consisting of two neighboring pixels is given as follows:

$$V_c(y_i, y_j) = \beta\delta(y_i, y_j) \tag{17}$$

where β is the coupling coefficient and when it increases the regions becomes more homogenous.

The segmentation problem is solved using one of MRF's pixel labeling algorithm named Iterated Conditional Modes (ICM) [53]. This algorithm iteratively optimizes a statistical criterion by approximating the Maximum A-Posteriori (MAP) estimate. In the MAP approach, a posterior probability measure $P(y|x)$ and we try to find optimal labeling $\hat{x}$ which maximizes this probability. It is also similar to minimizing the posterior energy function $U(y|x)$. ICM is thereby a greedy algorithm which tries to find a local minimum. For each pixel, the algorithm initially provides an estimate of the labeling and it chooses the label giving the largest decrease of the energy function. The posterior energy $U(y|x)$ is given by the sum of the likelihood energy function and the prior energy function as follows:

$$U(y|x) = U(x|y) + U(y) \tag{18}$$

ICM, when compared with other approaches such as simulated annealing, doesn't allow the temporary increase in the potential function to obtain minimum potential. The ICM algorithm can be summarized using the following steps.

- Initialize by assigning an arbitrary labeling y at step n = 0.
- At step n, we find,

$$y^{n+1} = \arg\min_{y} U(y|x) \tag{19}$$

- Repeat the above step until convergence is obtained.

4) GAUSSIAN MARKOV RANDOM FIELD

A Gaussian Markov random field (GMRF) is an undirected gaussian graphical model with values of the random field at the nodes to be jointly Gaussian [54], [55]. GMRFs fit nicely into a Bayesian framework since they are analytically tractable. It is a continuously-valued random vector having a multivariate Gaussian distribution of the following form:

$$p(y) \propto exp(-\frac{1}{2}(y-\mu)^T \Sigma^{-1}(y-\mu)) \tag{20}$$

where $\Sigma^{-1} = \Lambda$ is the the inverse covariance matrix. The quadratic form of the exponent is given as follows:

$$y^T \Lambda y = \sum_i \sum_j y_i y_j \Lambda_{i,j} \tag{21}$$

There does not exist an edge between y_i and y_j in the model when $\Lambda_{i,j} = 0$ and hence the neighborhood system is determined by the matrix Λ. The nonzero pattern of Λ helps to determine whether two nodes are conditionally independent. Here Λ is sparse, that is $\Lambda_{i,j} = 0$ if and only if y_i andy_j are conditionally independent. In practice, the GMRFs are defined using the quadratic energy function [56] given by:

$$U(y) = \frac{1}{2} y^T \Lambda y - y^T b \tag{22}$$

where $b \in \mathbb{R}$. In the application of Bayesian image processing [57], consider the image to have a similar $M \times N$ rectangular lattice structure as of MRFs. When a suitable prior p(y) is chosen, the maximum a posteriori (MAP) is estimated to find the optimal labels for segmentation using the ICM labeling algorithm.

IV. EXPERIMENTS AND RESULTS

The proposed experimental framework reads a captured video and extracts the frame. The dataset consists of 10 IR videos with a total of 10754 frames. The data is collected from the Kill the Flashover project conducted by the Fire Research Division of NIST (National Institute of Standards and Technology) [58]. It is focused on demonstrating the dynamics of fire behavior. All the videos are captured using an ISG E380 thermal imager and have both fire and smoke in real-time. The videos consist of both interior and exterior environments and hence serves as a generalized dataset. Each of the videos is shot in a time duration varying between 1 minute and 8 minutes. The thermal footages have a frame rate of 30 fps. There are 3 videos of size 960×720, 4 videos which have a resolution of 480×360 and the other 3 are 540×360, 640×358, and 1280×720 respectively.

Here the information extracted from the first 10 frames will act as prior knowledge for the test images. Initially, the training is done using images with fire and smoke and the primary step during this phase is feature extraction. The main features used for experimentation are the intensity values of the image, magnitude of motion vectors, SIFT flow features and the divergence of the image. The intensity values are taken into consideration since fire will be having a higher intensity value compared to the smoke and background. The divergence of the image is another feature that gives the amount of flow passing through a surface surrounding a pixel. Additionally, the SIFT flow features preserve spatial discontinuities and it helps to compute pixel-wise SIFT features between two images. The motion features from the frames are captured by applying the Horn Schunck optical flow algorithm [59]. Let the intensity value at pixel of discrete coordinates x, y and at time t be denoted as I(x, y, t). To calculate the optical flow features the brightness constancy is assumed i.e. $\frac{dI}{dt} = 0$. Further, by computation, a solution is obtained by minimizing the regularization term in (4). The average velocity vectors u and v are computed for each pixel in the image of size $M \times N$. After calculating the optical flow, we obtain its magnitude as a feature from the first P frames, the feature vector becomes $M \times N \times P$. The direction turns out not to be informative because the fire is not always traveling up. Indeed, horizontal velocity is often observed close to ceilings and turbulences cause velocities in direction to the floor. The divergence of

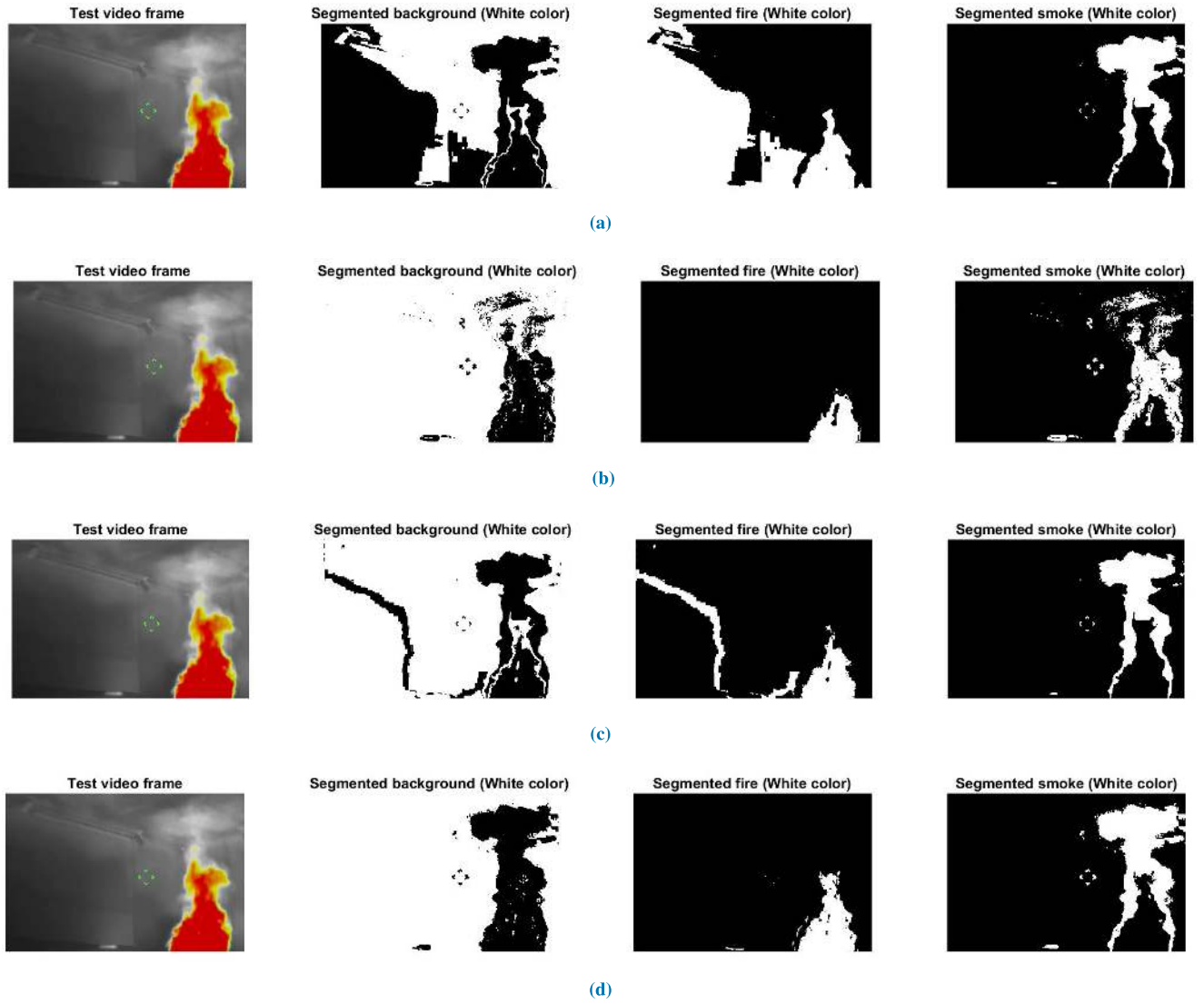

FIGURE 5. Sample segmentation using (a) K-Means (b) GMM (c) GMRF (d) MRF.

these P frames is calculated using (5), hence the size of each feature vector becomes $M \times N \times P$. The intensity values of the pixels are themselves used as a feature with the same dimensions. Then these feature vectors are reshaped into column vectors. Assuming orthogonality of the feature vectors, the column vectors corresponding to intensity, optical flow and divergence are concatenated to obtain the final feature matrix.

Further various clustering techniques are used to separate the fire and smoke from the background. Thus these computed feature vectors [60] are given as input to different clustering techniques such as K-Means, GMM, MRF and GMRF to segment the smoke and fire from the desired frame. Each of these clustering algorithms [61] provides the indices for the pixel values corresponding to each of the classes. These pixel indices are mapped to the original frames to perform segmentation of the region of interest. The final classification of fire, smoke, and background are done based on the intensity values belonging to those clusters. Hence the labeling of the clusters is done by assigning the cluster with the highest intensity values to fire, intermediate intensity values to smoke and the cluster which has a lesser intensity to the background. Further, the performance evaluation is done by calculating the accuracy values for different algorithms. But the classification accuracy alone is not sufficient to select a model since it hides the details required to better understand the performance of the model. Hence the confusion matrix was computed to overcome the limitations of using only the accuracy as a decision parameter for performance evaluation. The estimation of the confusion matrix involved the manual labeling of the test frames to obtain the ground truth. Here the pixel-wise comparison is done between the segmented and ground truth values to obtain a summary of predictions made by the algorithm for each class.

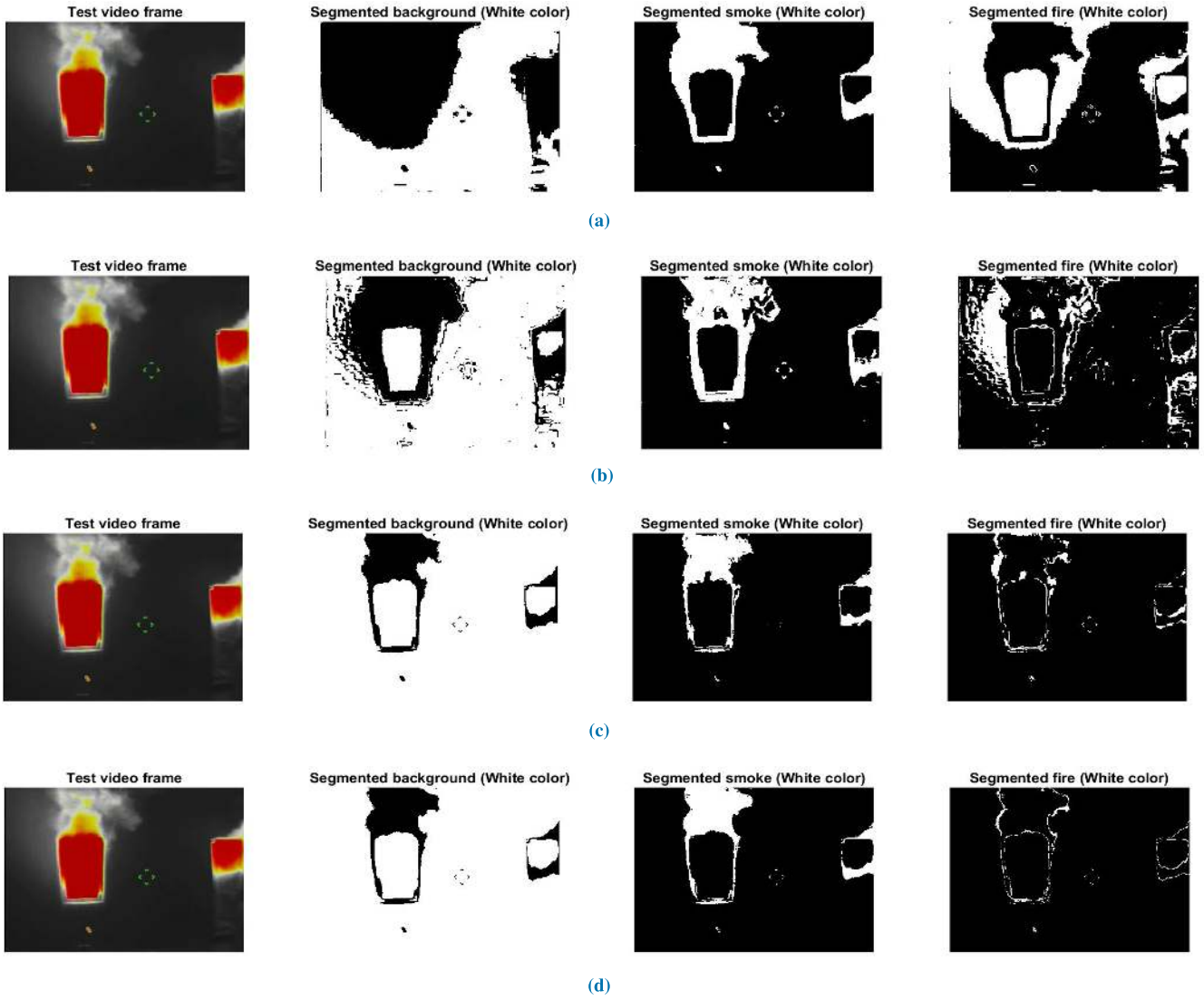

FIGURE 6. MRF segmentation using (a) Intensity feature (b) SIFT Flow feature (c) Optical flow feature (d) Divergence feature.

A. SAMPLE SEGMENTATION COMPARISON USING DIFFERENT ALGORITHMS AND FEATURE VECTORS

Fig. 5 shows the segmentation results using different clustering algorithms. The main features used for this experimentation are the divergence, intensity, and the optical flow values. The first row in Fig. 5 shows the K-means result. The segmentation is poor for two reasons which are not independent. First, the likelihood model assumed by this algorithm is circularly symmetric, which does not seem to be very accurate for this problem. Second, the posterior probabilities are approximated by one for the posterior of the closest cluster using Euclidean distance and zero otherwise. These approximations are too strong for the distribution at hand, as it is stated by the fact that the GMM model, which does not have this restriction, produces covariance matrices that are not diagonal. For the case of the GMRF, the most likely explanation of the poor performance in some background regions is that the Gaussian assumption is not adequate when the Markovian behavior is assumed, which can be seen in the third row of Fig. 5 for the segmentation of background and fire. Hence it was observed that the cascaded system using the feature vectors and MRF performs a more successful segmentation of fire and smoke than the rest of the algorithms. The rest of the segmentation algorithms misses some pixels of inner parts of smoke and fire. The MRF based approach was able to detect the smoke regions which appear blurry and indistinguishable for the human eye. Hence this prior information can be of paramount importance for the first responders during real-time fire-fighting situations. It is also evident that the proposed algorithm can disregard the unwanted artifacts from the frames and only MRF can capture both the static and dynamic properties of the area of interest. The clear definition of the shape of fire and smoke will help in further analysis such as classification of fire as small, medium or large and thereby provide situational awareness.

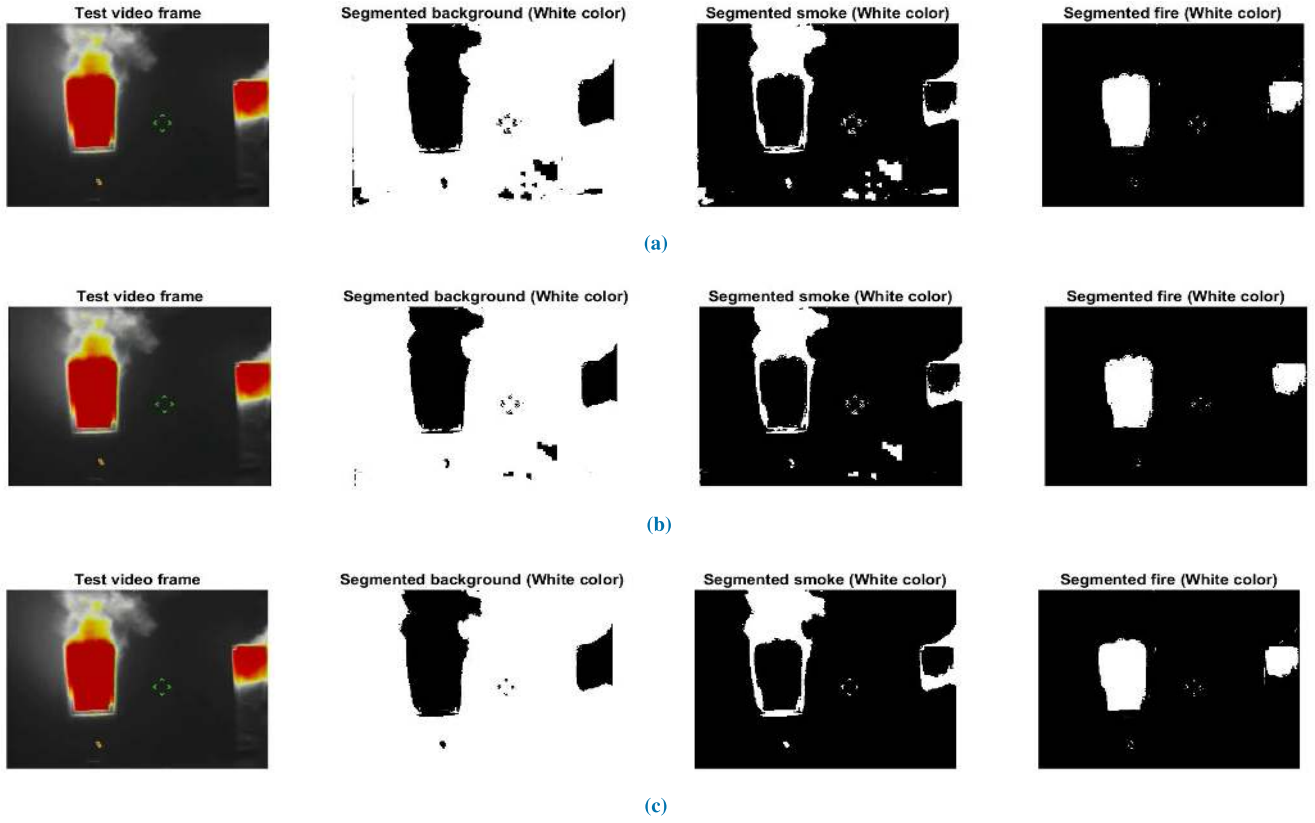

FIGURE 7. MRF segmentation using (a) Intensity and optical flow features (b) Intensity and divergence features (c) Intensity, divergence and optical flow features.

The contribution of different feature vectors in improving the MRF based segmentation is shown in Fig. 6 and Fig. 7. The first row of Fig. 6 illustrates the MRF segmentation using only intensity as a feature vector. Whereas the second row uses SIFT flow, the third row uses optical flow, the final row uses divergence features. In case of Fig. 7 the first row corresponds to MRF segmentation using the fusion of intensity and optical flow features, the second row uses the fusion of intensity and divergence features and the final row uses the fusion of intensity, optical flow, and divergence as the features. When MRF segmentation is applied directly to the image intensity, it was found that their values are not reliable in segmenting the background. Hence different feature extraction methods such as optical flow, SIFT flow and divergence were employed for more accurate segmentation. It was observed that SIFT flow showed the worst performance in segmentation and hence it was discarded as a feature vector. On the other hand, optical flow and divergence used velocity vectors as features, and hence the dynamics of smoke and the moving edges of the fire were extracted properly. The main disadvantage of these features was that the segmented background also includes the static portion of the fire, which has a higher intensity than the background. Hence a feature vector was constructed by concatenating intensity, optical flow, and divergence features. It was observed that when MRF based segmentation was performed on these feature vectors, the smoke, fire, and background were segmented in a more precise manner. The proposed approach is also tested in video frames under different conditions and the segmentation results are shown in Fig. 10.

B. CONFUSION MATRIX

The confusion matrix is an evaluation metric used widely for the analysis of semantic segmentation. It is a square matrix in which each row has instances of the true class and each column has instances of the segmented class. Hence C_{mn} represents the pixels of class m which are classified as class n. Table 1 shows the comparative analysis of the confusion matrix table for different methods. It can be seen that both GMM and MRF based segmentation was able to segment fire in a more accurate way than the other algorithms. In the case of smoke MRF was able to perform the segmentation with 90% accuracy whereas the rest of the methods were unable to distinguish accurately between background and smoke. Further, Fig. 8 shows the final calculated accuracy from the confusion matrix for the proposed methods. It can be seen that the fusion of intensity, divergence, and optical flow features boosts the performance of the segmentation algorithms. In comparison to the other clustering algorithms, MRF based segmentation provides the highest accuracy while using just the intensity features and also when the feature fusion is

TABLE 1. Confusion matrix for K-means, GMM, GMRF and MRF.

K-means

Background	0.54	0.46	0
Fire	0.90	0.10	0
Smoke	0	0.28	0.72
	Background	Fire	Smoke

GMM

Background	0.83	0.01	0.16
Fire	0	1	0
Smoke	0.07	0.17	0.92
	Background	Fire	Smoke

GMRF

Background	0.94	0.06	0
Fire	0.17	0.83	0
Smoke	0.44	0	0.56
	Background	Fire	Smoke

MRF

Background	0.98	0.01	0.01
Fire	0.01	0.98	0
Smoke	0.10	0.0	0.90
	Background	Fire	Smoke

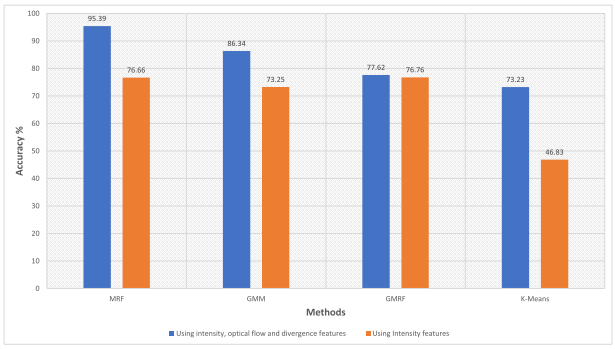

FIGURE 8. Accuracy of the different segmentation algorithms using just intensity features and using fusion of intensity, optical flow and divergence features.

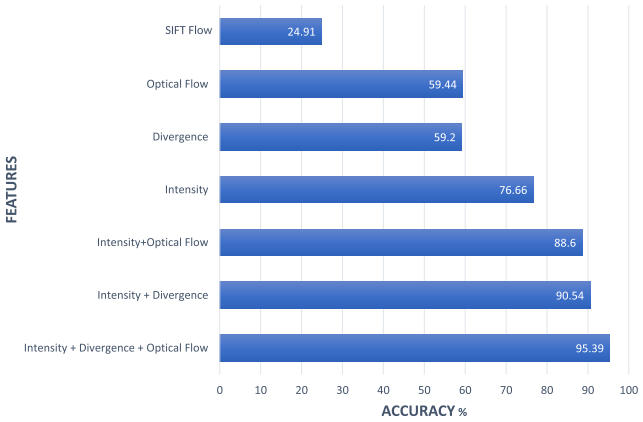

FIGURE 9. MRF segmentation accuracy using different feature extraction methods.

applied. The overall accuracy of the proposed approach was found to be 95.39%.

C. ACCURACY ANALYSIS FOR FEATURE SELECTION

The comparative analysis of different features for MRF segmentation is now shown in Fig. 9. The best performance is seen when the spatial, temporal and motion features are extracted and used for segmentation. The optical flow, divergence, and intensity features are combined to obtain an overall accuracy of 95.39%. Nevertheless, when intensity is combined with either optical flow or divergence features there is a reduction in accuracy by 5-7%. When all the 3 features are combined it captures the static features of fire and background as well as the randomness of smoke. The main disadvantage of using the features individually is that in the case of optical flow and divergence it extracts only the velocity vectors and hence the background and fire are not segmented properly. Further, when MRF segmentation is performed on the intensity values it results in only 76.66% accuracy due to lack of contrast between the desired regions.

D. RUN TIME ANALYSIS

The experiments were conducted on Intel® Core™ i7-7700HQ CPU @ 2.80Ghz with 16Gb of memory. Using the MRF based approach the overall run time for a single frame in a video is 3.04 seconds. The overall time complexity was calculated by involving feature extraction, feature fusion, model optimization for 10 iterations and unsupervised testing. Few of the other video segmentation algorithms are implemented in videos with much lower resolution and the time complexity is about 355 seconds for 85 frames [62]. Even though some methods have a run time of 170 seconds for an 85-frame video, it is obtained by using GPU with high-end hardware specifications [63]. Hence for one of the videos with 500 frames, it takes about 1500 seconds in a CPU. The proposed method is currently implemented offline, and it can be further optimized for parallel processing especially during feature extraction and training.

TABLE 2. Comparison of time complexities of different approaches.

Methods	Overall run time(s)	Test time (s)
K-Means	1.027	0.23
GMM	3.25	0.11
GMRF	4.74	0.73
MRF	3.04	0.33

Table 2 shows the comparison of overall time complexity and test time for the various algorithms described in this paper. It was found that the K-means method has the least overall computational cost. But the accuracy of this algorithm is poor compared to the rest of the methods. Meanwhile, GMRF takes slightly more time for training and test. Though GMM and MRF have similar time complexities, MRF was

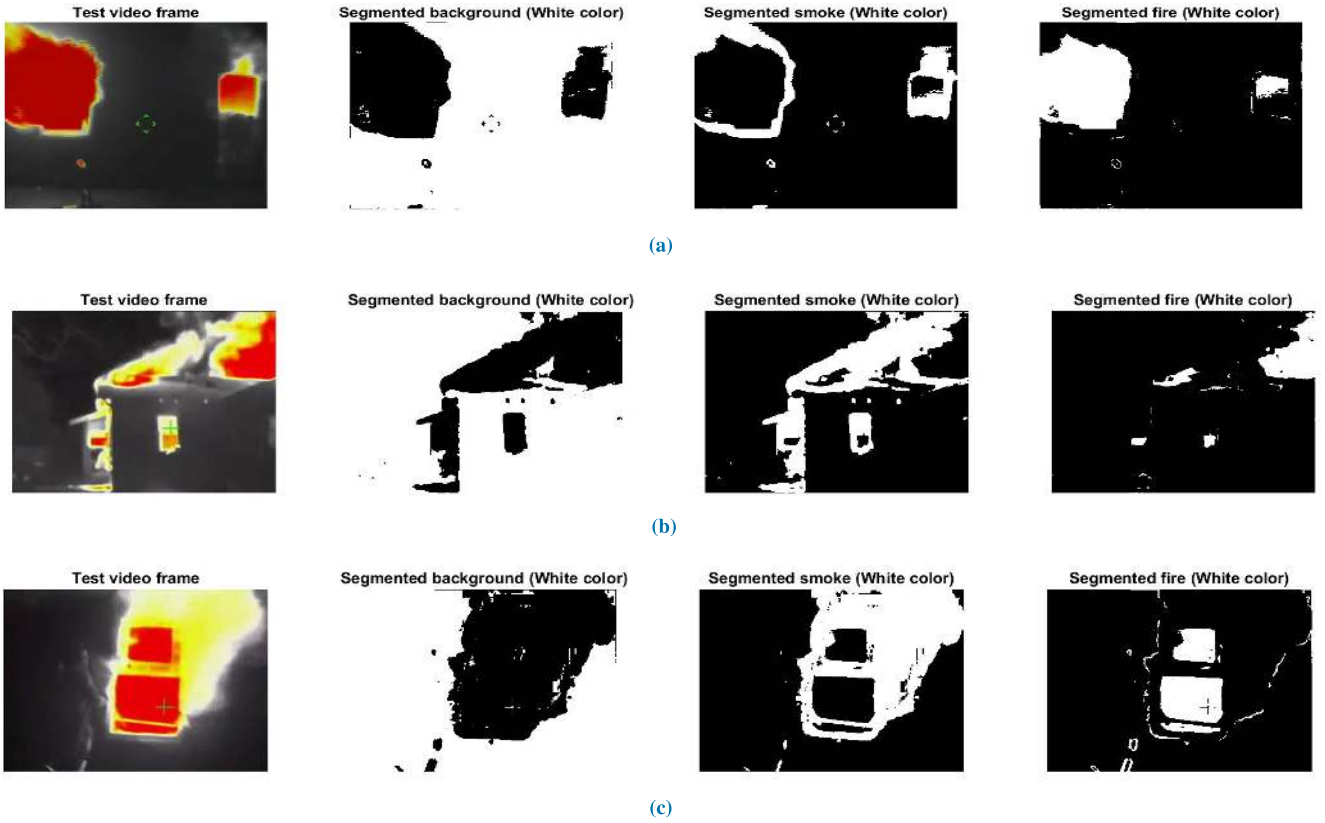

FIGURE 10. Different video frames and corresponding outputs using the proposed MRF based segmentation with intensity, optical flow, and divergence features.

found to be the most efficient amongst the two algorithms with an accuracy of 95.39%. In the case of MRF, a test takes 0.3 seconds, hence a real-time test for our hardware setup is limited to sequences of 10 fps, but an optimized parallel setup will perform the classification in a shorter time, leading to a real-time classification for sequences of 30 fps.

V. CONCLUSION

This paper introduces a novel method for fire and smoke characterization in IR images. This approach can perform unsupervised testing in real-time and it can be trained parallelly in offline mode. The feature extraction methods proposed for this problem are by using optical flow, divergence, and intensity. Even though SIFT flow features were tested but it did not give any significant improvement in segmentation compared to the combination of the above feature extractors. The unsupervised segmentation methods used for the comparative analysis were K-Means, GMM, GMRF, and MRF. It was found that MRF showed better performance in the classification with a higher accuracy of 95.39%. It has been tested visually and quantitatively that MRF was able to distinguish fire, smoke, and background in a more precise manner. Even though GMM was able to segment most of the fire regions but it gave a much lower accuracy of 76% for smoke segmentation. Thus, the fusion of information shows results that outperform single feature-based methods. The future work aims to use multispectral data from UV and RGB sensors to make more accurate predictions for real-time firefighting scenarios. We would also like to extend the experimentation on dynamic and complex fire environments to test the robustness of these approaches.

ACKNOWLEDGMENT

The authors would like to thank the UNM Center for Advanced Research Computing, supported in part by the National Science Foundation, for providing the high performance computing, large-scale storage, visualization resources used in this work.

REFERENCES

[1] B. U. Töreyin, Y. Dedeoğlu, U. Güdükbay, and A. E. Çetin, "Computer vision based method for real-time fire and flame detection," *Pattern Recognit. Lett.*, vol. 27, no. 1, pp. 49–58, 2006. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167865505001819

[2] T.-H. Chen, P.-H. Wu, and Y.-C. Chiou, "An early fire-detection method based on image processing," in *Proc. Int. Conf. Image Process. (ICIP)*, vol. 3, Oct. 2004, pp. 1707–1710.

[3] W. P. Iii, M. Shah, and N. da Vitoria Lobo, "Flame recognition in video," *Pattern Recognit. Lett.*, vol. 23, no. 1, pp. 319–327, 2002. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167865501001350

[4] T. Çelik and H. Demirel, "Fire detection in video sequences using a generic color model," *Fire Saf. J.*, vol. 44, no. 2, pp. 147–158, 2009. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0379711208000568

[5] T. Celik, "Fast and efficient method for fire detection using image processing," *ETRI J.*, vol. 32, pp. 881–890, Dec. 2010.

[6] W.-B. Horng, J.-W. Peng, and C.-Y. Chen, "A new image-based real-time flame detection method using color analysis," in *Proc. IEEE Netw., Sens. Control*, Mar. 2005, pp. 100–105.

[7] T. Çelik and H. Özkaramanli, and H. Demirel, "Fire and smoke detection without sensors: Image processing based approach," in *Proc. 15th Eur. Signal Process. Conf.*, Sep. 2007, pp. 1794–1798.

[8] C.-B. Liu and N. Ahuja, "Vision based fire detection," in *Proc. 17th Int. Conf. Pattern Recognit. (ICPR)*, vol. 4, Aug. 2004, pp. 134–137.

[9] Y. Wei, Y. Chunyu, and Z. Yongming, "Based on wavelet transformation fire smoke detection method," in *Proc. 9th Int. Conf. Electron. Meas. Instrum.*, Aug. 2009, pp. 2-872–2-875.

[10] D. Titterington, A. Smith, and U. Makov, *Statistical Analysis of Finite Mixture Distributions*. New York, NY, USA: Wiley, 1985.

[11] T. Celik, H. Demirel, H. Ozkaramanli, and M. Uyguroglu, "Fire detection using statistical color model in video sequences," *J. Vis. Commun. Image Represent.*, vol. 18, no. 2, pp. 176–185, Apr. 2007. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1047320306000927

[12] Y. Wang, A. Wu, J. Zhang, M. Zhao, W. Li, and N. Dong, "Fire smoke detection based on texture features and optical flow vector of contour," in *Proc. 12th World Congr. Intell. Control Automat. (WCICA)*, Jun. 2016, pp. 2879–2883.

[13] D. Chetverikov, S. Fazekas, and M. Haindl, "Dynamic texture as foreground and background," *Mach. Vis. Appl.*, vol. 22, pp. 741–750, Sep. 2011.

[14] B. U. Töreyin, Y. Dedeoğlu, and A. E. Çetin, "Flame detection in video using hidden Markov models," in *Proc. IEEE Int. Conf. Image Process.*, vol. 2, Sep. 2005, p. II-1230.

[15] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proc. 5th Berkeley Symp. Math. Statist. Probab.*, vol. 1. Berkeley, CA, USA: University of California Press, 1967, pp. 281–297. [Online]. Available: https://projecteuclid.org/euclid.bsmsp/1200512992

[16] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Roy. Statist. Soc. B, Methodol.*, vol. 39, no. 1, pp. 1–38, 1977.

[17] G. R. Cross and A. K. Jain, "Markov random field texture models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-5, no. 1, pp. 25–39, Jan. 1983.

[18] H. Rue and L. Held, *Gaussian Markov Random Fields: Theory and Applications* (Monographs on Statistics and Applied Probability). London, U.K.: Chapman & Hall, 2005.

[19] Y. Liu, W. Qin, K. Liu, F. Zhang, and Z. Xiao, "A dual convolution network using dark channel prior for image smoke classification," *IEEE Access*, vol. PP, p. 1, 05 2019.

[20] Z. Yin, B. Wan, F. Yuan, X. Xia, and J. Shi, "A deep normalization and convolutional neural network for image smoke detection," *IEEE Access*, vol. 5, pp. 18429–18438, 2017.

[21] M. Tehrani, M. Garratt, and S. Anavatti, "Low-altitude horizon-based aircraft attitude estimation using UV-filtered panoramic images and optic flow," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 52, no. 5, pp. 2362–2375, Oct. 2016.

[22] S. Calderara, P. Piccinini, and R. Cucchiara, "Vision based smoke detection system using image energy and color information," *Mach. Vis. Appl.*, vol. 22, no. 4, pp. 705–719, Jul. 2011, doi: 10.1007/s00138-010-0272-1.

[23] H. Tian, W. Li, P. Ogunbona, D. T. Nguyen, and C. Zhan, "Smoke detection in videos using non-redundant local binary pattern-based features," in *Proc. IEEE 13th Int. Workshop Multimedia Signal Process.*, Oct. 2011, pp. 1–4.

[24] H. Maruta, A. Nakamura, T. Yamamichi, and F. Kurokawa, "Image based smoke detection with local hurst exponent," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2010, pp. 4653–4656.

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

[26] L. Rossi and M. Akhloufi, "Dynamic fire 3D modeling using a real-time stereovision system," in *Technological Developments in Education and Automation*, M. Iskander, V. Kapila, and M. A. Karim, Eds. Dordrecht, The Netherlands: Springer, 2010, pp. 33–38.

[27] D. Kim and Y. Wang, "Smoke detection in video," in *Proc. WRI World Congr. Comput. Sci. Inf. Eng.*, vol. 5, Mar. 2009, pp. 759–763.

[28] B. U. Töreyin, Y. Dedeoğlu, and A. E. Çetin, "Contour based smoke detection in video using wavelets," in *Proc. 14th Eur. Signal Process. Conf.*, Sep. 2006, pp. 1–5.

[29] P. Piccinini, S. Calderara, and R. Cucchiara, "Reliable smoke detection in the domains of image energy and color," in *Proc. IEEE 15th Int. Conf. Image Process. (ICIP)*, Oct. 2008, pp. 1376–1379.

[30] B. U. Töreyin, Y. Dedeoğlu, and A. E. Çetin, "Wavelet based real-time smoke detection in video," in *Proc. 13th Eur. Signal Process. Conf.*, Sep. 2005, pp. 1–4.

[31] F. Gomez-Rodriguez, B. C. Arrue, and A. Ollero, "Smoke monitoring and measurement using image processing: Application to forest fires," in *Automatic Target Recognition XIII*, vol. 5094, F. A. Sadjadi, Ed. Bellingham, WA, USA: SPIE, 2003, pp. 404–411, doi: 10.1117/12.487050.

[32] M. Torabnezhad, A. Aghagolzadeh, and H. HadiSeyedarabi, "Visible and IR image fusion algorithm for short range smoke detection," in *Proc. 1st RSI/ISM Int. Conf. Robot. Mechatronics (ICRoM)*, Feb. 2013, pp. 38–42.

[33] J. R. M.-D. Dios and A. Ollero, "A new training-based approach for robust thresholding," in *Proc. World Automat. Congr.*, vol. 18, Jun. 2004, pp. 121–126.

[34] J. R. M. de Dios, L. Merino, and A. Ollero, "Fire detection using autonomous aerial vehicles with infrared and visual cameras," *IFAC Proc. Volumes*, vol. 38, no. 1, pp. 660–665, 2005. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S147466701637392X

[35] I. Bosch, A. Serrano, and L. Vergara, "Multisensor network system for wildfire detection using infrared image processing," *TheScientificWorldJournal*, vol. 2013, Jun. 2013, Art. no. 402196.

[36] C. Yuan, Z. Liu, and Y. Zhang, "Fire detection using infrared images for UAV-based forest fire surveillance," in *Proc. Int. Conf. Unmanned Aircr. Syst. (ICUAS)*, Jun. 2017, pp. 567–572.

[37] R. Kaabi, S. Frizzi, M. Bouchouicha, F. Fnaiech, and E. Moreau, "Video smoke detection review: State of the art of smoke detection in visible and ir range," in *Proc. Int. Conf. Smart, Monitored Controlled Cities (SM2C)*, Feb. 2017, pp. 81–86.

[38] S. Verstockt, S. Hoecke, T. Beji, B. Merci, B. Gouverneur, A. E. Çetin, P. De Potter, and R. Van de Walle, "A multi-modal video analysis approach for car park fire detection," *Fire Saf. J.*, vol. 57, pp. 44–57, Aug. 2012.

[39] J. R. Martinez-de Dios, B. C. Arrue, A. Ollero, L. Merino, and F. Gómez-Rodríguez, "Computer vision techniques for forest fire perception," *Image Vis. Comput.*, vol. 26, no. 4, pp. 550–562, 2008.

[40] T. Kohonen, "Self-organized formation of topologically correct feature maps," *Biological Cybern.*, vol. 43, no. 1, pp. 59–69, 1982.

[41] L. Bottou and Y. Bengio, "Convergence properties of the k-means algorithms," in *Proc. Adv. Neural Inf. Process. Syst.*, 1995, pp. 585–592.

[42] B. K. P. Horn and B. G. Schunck, "Determining optical flow," *Artif. Intell.*, vol. 17, nos. 1–3, pp. 185–203, Aug. 1981.

[43] C. Liu, J. Yuen, and A. Torralba, "SIFT flow: Dense correspondence across scenes and its applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 5, pp. 978–994, May 2011.

[44] A. Shekhovtsov, I. Kovtun, and V. Hlavac, "Efficient mrf deformation model for non-rigid image matching," *Comput. Vis. Image Understand.*, vol. 112, pp. 91–99, Oct. 2008.

[45] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient belief propagation for early vision," *Int. J. Comput. Vis.*, vol. 70, no. 1, pp. 41–54, Oct. 2006, doi: 10.1007/s11263-006-7899-4.

[46] X. Descombes, R. D. Morris, J. Zerubia, and M. Berthod, "Estimation of Markov random field prior parameters using Markov chain Monte Carlo maximum likelihood," *IEEE Trans. Image Process.*, vol. 8, no. 7, pp. 954–963, Jul. 1999, doi: 10.1109/83.772239.

[47] S. S. Saquib, C. A. Bouman, and K. Sauer, "ML parameter estimation for Markov random fields with applications to Bayesian tomography," *IEEE Trans. Image Process.*, vol. 7, no. 7, pp. 1029–1044, Jul. 1998, doi: 10.1109/83.701163.

[48] H. Cao and V. Govindaraju, "Preprocessing of low-quality handwritten documents using Markov random fields," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 7, pp. 1184–1194, Jul. 2009, doi: 10.1109/TPAMI.2008.126.

[49] J. Hartigan and M. Wong, "Algorithm AS 136: A k-means clustering algorithm," *Appl. Statist.*, vol. 28, no. 1, pp. 100–108, 1979.

[50] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-6, no. 6, pp. 721–741, Nov. 1984, doi: 10.1109/TPAMI.1984.4767596.

[51] H. Deng and D. A. Clausi, "Unsupervised image segmentation using a simple MRF model with a new implementation scheme," *Pattern Recognit.*, vol. 37, no. 12, pp. 2323–2335, 2004, doi: 10.1016/j.patcog.2004.04.015.

[52] S. Z. Li, *Markov Random Field Modeling in Image Analysis*. Berlin, Germany: Springer-Verlag, 2001.

[53] J. Besag, "On the statistical analysis of dirty pictures," *J. Roy. Stat. Soc. B, Methodol.*, vol. 48, no. 3, pp. 259–279, 1986.
[54] R. Szeliski, "Bayesian modeling of uncertainty in low-level vision," *Int. J. Comput. Vis.*, vol. 5, no. 3, pp. 271–301, Dec. 1990, doi: 10.1007/BF00126502.
[55] M. A. T. Figueiredo and J. M. N. Leitao, "Unsupervised image restoration and edge location using compound gauss-Markov random fields and the MDL principle," *IEEE Trans. Image Process.*, vol. 6, no. 8, pp. 1089–1102, Aug. 1997.
[56] P. Pérez, "Markov random fields and images," *CWI Quart.*, vol. 11, no. 4, pp. 413–437, 1998.
[57] A. Blake, P. Kohli, and C. Rother, *Markov Random Fields for Vision and Image Processing*. Cambridge, MA, USA: MIT Press, 2011.
[58] *Joe Starnes and Kill the Flashover.coms*. [Online]. Available: http://killtheflashover.com/index.html
[59] S. Rinsurongkawong, M. Ekpanyapong, and M. N. Dailey, "Fire detection for early fire alarm based on optical flow video processing," in *Proc. 9th Int. Conf. Electr. Eng./Electron., Comput., Telecommun. Inf. Technol.*, 2012, pp. 1–4.
[60] D. Wu, N. Wang, and H. Yan, "Smoke detection based on multi-feature fusion," in *Proc. 5th Int. Congr. Image Signal Process.*, Oct. 2012, pp. 220–223.
[61] G. Yang, H.-C. Li, and C. Liu, "Unsupervised change detection of remote sensing images using superpixel segmentation and variational Gaussian mixture model," in *Proc. 9th Int. Workshop Anal. Multitemporal Remote Sens. Images (MultiTemp)*, 2017, pp. 1–4.
[62] K. J. F. de Souza, A. de A. Araújo, S. J. F. Guimarães, Z. K. G. do Patrocínio, and M. Cord, "Streaming graph-based hierarchical video segmentation by a simple label propagation," in *Proc. 28th SIBGRAPI Conf. Graph., Patterns Images*, Aug. 2015, pp. 119–125.
[63] C.-P. Yu, H. Le, G. Zelinsky, and D. Samaras, "Efficient video segmentation using parametric graph partitioning," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 3155–3163.

MEENU AJITH received the bachelor's degree in electronics and communication engineering from the Amrita School of Engineering, in 2015, and the master's degree in electrical engineering from The University of New Mexico, in 2017, where she is currently pursuing the Ph.D. degree in electrical engineering. Her research interests are machine learning, computer vision, pattern recognition and image processing.

MANEL MARTÍNEZ-RAMÓN received the degree in telecommunications engineer from the Universitat Politecnica de Catalunya, Spain, in 1996, and the Ph.D. degree in communications technologies form the Universidad Carlos III de Madrid, Spain, in 1999. He is currently a Professor with the ECE Department, The University of New Mexico. His research interests are in machine learning applications to smart antennas, neuroimage, first responders and other cyber-human systems, smart grid and others. His last work is the monographic book *Signal Processing with Kernel Methods*, (Wiley, 2018). He holds the King Felipe VI Endowed Chair of The University of New Mexico, a chair sponsored by the Household of the King of Spain.