Detection and Separation of Smoke From Video using novel method based on finite element method (FEM)

Rajeswary .B

Assistant Professor, Electronics & Communication Engineering Department, Misrimal Navajee Munoth Jain Engineering College, Chennai-97, India

Abstract— This paper proposes novel methods based on finite element method (FEM) for detecting and separating smoke from a single image frame of a video. Specifically, an image formation model is derived based on the atmospheric scattering models. The video is divided into number of frames and then apply the novel methods for detecting and separating smoke from a single image frame. The separation of a frame into quasi-smoke and quasi-background components is formulated as convex optimization that solves a sparse representation problem using dual dictionaries for the smoke and background components, respectively. A novel feature is constructed as a concatenation of the respective sparse coefficients for detection. In addition, a method based on the concept of deep image matting is developed to separate the true smoke and background components from the smoke detection results. Previous algorithms have poor performance when an image has similar foreground and background colors or complicated textures. Extensive experiments on detection were conducted and the results showed that the proposed feature significantly outperforms existing features for smoke detection. In particular, the proposed method is able to differentiate smoke from other challenging objects (e.g. fog/haze, cloud, and so on) with similar visual appearance in a gray-scale frame. Experiments on smoke separation also demonstrated that the proposed separation method can effectively estimate/separate the true smoke and background components.

Index Terms— Smoke detection, smoke separation, finite element method (FEM), deep image matting, surveillance, sparse representation.

1. INTRODUCTION

VISION-based smoke detection has many advantages over the traditional photoelectric- or ionization-based smoke detectors, including being suitable for both closed and open spaces. Furthermore, it can provide early detection with information on the location and intensity [1]–[5] of smoke.

To the best of our knowledge, almost all existing detection algorithms are video-based and the video is assumed to be cap-tured by a stationary camera in order to facilitate the motion detection and feature extraction required by these algorithms. However, such requirement can hardly be met in practical scenarios. First, cameras in an open space inevitably suffer jitter under severe and dynamic environment such as heavy rain and wind. The detection accuracy of the state-of-the-art video-based method [5] dropped from 95.5% on video from a stationary camera to 54.5% on video from a non-stationary camera with only a few pixels of jitter. Non-stationary or moving cameras are typically installed on commonly used unmanned aerial vehicle (UAV)-based and recently drone-based bushfire surveillance. Second, surveillance images are often streamed in a resource-limited infrastructure such as sensor networks. In these applications, detection of smoke from single frames becomes imperative. However, there is little study so far on single frame-based smoke detection. This article presents a novel method to address this problem.

Conventional vision-based smoke detection techniques [1], [2], [6] usually follow an object detection framework in which each frame of the input video is divided into blocks. For each block, salient features are extracted and employed to classify the block into smoke and non-smoke. The success of these techniques depends on identification of robust visual features that can characterize smoke and reliable extraction of these features. Among the commonly used visual features, texture is the most reliable because motion is highly subject to the environment and weather conditions and color is subject to a number of factors including the type of material that is burning and illumination conditions. For example cameras often switch to gray-scale under low illumination during evening or at night. In addition, due to the deformable and transparent nature of smoke, extracted visual features are often corrupted by the non-smoke background, especially in cases where the smoke is light or at its early stage. To address this issue, Tian *et al.* [4], [5] recently proposed to separate the smoke component from the background and to extract features from the smoke component. While this approach has significantly improved the detection rate as reported in [4] and [5], it requires a video captured by a stationary camera to estimate the background through background modeling. For a single frame, it is no longer possible to estimate the required background.

The proposed method follows the concept of image separation for smoke detection as presented in [5] and takes a further step to separate the smoke component from a single gray-scale frame. The weight of the

composition is a function of the thickness and the scattering coefficient of the smoke. A dual over complete dictionary approach is then proposed to model the background and smoke components. The separation of the two components is formulated as a sparse representation and convex optimization problem. This dual-dictionary model provides us with the sparse coefficients of the smoke and background components as effective features for detection. However, the sparse representation can only allow us to estimate a scaled version of the smoke or background components; herein referred to as quasi-smoke and quasi-background respectively. A method based on the concept of deep image matting is developed to further estimate the true smoke and background components.

This paper is a significant extension of the work presented at the 2014 Asian Conference on Computer Vision [8]. The extension is five-fold: a) a method based on cascaded multi-ple classifiers to differentiate smoke from other objects that share similar visual appearance to smoke in a gray-scale image frame. Examples of such objects include fog/haze, cloud, steam, sky, shadow, glass, etc.; b) a novel image-based method to estimate the true smoke and background components from the detected smoke blocks, experimental evaluation of the method and its comparison to the Guided Image Filtering (GIF) method; c) detailed analysis of the image formation model and its difference from the model used in haze/fog removal in the literature; d) formalization of the dual-dictionary approach to separating an image into quasi-smoke and quasi-background components and a method of using the estimated sparse coefficients for detection; and
e) systematic and extensive empirical evaluation of the separation and detection methods.

The remainder of the article is organized as follows: a review of some representative video-based smoke detection methods is provided in Section II. An image formation model for smoke is derived in Section III-A. The proposed method for single frame smoke detection based on the image formation model is presented in Section III-B.

On the basis of the detection result, a method for single frame smoke separation is proposed in Section IV. Experimental results are presented in Section V along with discussions. The paper is concluded with some perspectives on future work in Section VI.

2. RELATED WORK

The success of existing video-based smoke detection methods lies in identifying robust visual features to characterize smoke. To motivate the rationale for the proposed methods some representative video-based smoke detection methods are reviewed with respect to the features they used, viz. motion, color, edge and texture. Review of other related works such as image separation and deep image matting from which the separation of true smoke component is developed can be found in [5], [9], and [10], respectively.

From motion point of view, an accumulative motion model has been proposed to capture the motion characteristics of smoke in [11] and [12]. However, the assumption that smoke usually drifts upwards continually through hot airflow makes this model ineffective in a scenario where wind is present. Other research efforts have extracted motion features of smoke using optical flow [13]. For instance, optimal mass transport optical flow, which introduces the concept of mass transport in the sense of minimal transportation cost to optical flow computation, was used as a low-dimensional descriptor of smoke's motion [13]. In addition, a histogram of oriented optical flow was extracted as a temporal feature based on the fact that the diffusion direction of smoke is upward owing to thermal convection. As can be inferred, this temporal feature will not be effective in the presence of strong draught or wind. Moreover, no motion information is available from a single image.

The fact that the color of smoke is usually grayish provides a cue for the extraction of color features [1], [12]. Despite the differences in methods reported in the literature, a commonality amongst them is that the three intensity values from the RGB channels of a smoke pixel are equal or similar. For example, a reference color model for smoke was selected in the RGB color space to measure the deviation of the current pixel color from the model [1]. An analysis in CIELAB color space was performed in and a better clustering of the chromatic features of smoke was reported.

Edge-based features used for smoke detection can be divided into two groups. One group of features is extracted to characterize the boundary of smoke while the other group of features describes the effects of smoke on the edges of objects covered by smoke. In the first group, histogram of oriented gradient (HOG) descriptors were extracted for smoke detection. The underlying assumption in these methods is that the direction of smoke diffusion is upward owing to thermal convection such that the gradient distribution of a smoke boundary has a distinguishable pattern. The phenomenon that blurred edges could be observed in smoke-covered areas of a scene image has motivated the second group of features. As edges contribute to high frequency information, the consequent decrease in high frequency has been used as a cue to perform smoke detection in [2] and [12], where wavelets were adopted to extract horizontal, vertical, and diagonal high frequency information of images. However, this decrease in high frequency is not unique to smoke coverage and is hard to measure its extent from a single image due to the lack of background information.

As a result of the dispersive distribution of smoke, texture features have been extracted for smoke detection [3], [6], [12]. For instance, local binary pattern (LBP), a well-known descriptor for texture classification was applied to smoke detection [3]. In a similar vein a modified local ternary pattern, namely center symmetric local ternary pattern (CS-LTP) was proposed as a texture descriptor for smoke. It was designed to improve the robustness of handling noise and reducing the number of bits in feature description, and thus speeding up its histogram comparison process.

Additionally, smoke has been considered as a self-affine fractal and its Hurst exponent was employed for smoke detection. Inspired by the air light-albedo ambiguity model, the concept of transmission was introduced as a feature of

smoke. Since no single feature can perfectly characterize smoke, feature fusion techniques have been proposed. Representative works are found in [12]. Recently, to reduce the level of noise introduced into the extracted features by the background, an image separation approach has been proposed for smoke detection [4], [5]. It actively separates the smoke component, if any, from the background. Texture features are then extracted from the separated smoke component for detection.

In summary, video-based methods require a stationary camera. Motion and color features are subject to environment and the nature of materials being burned respectively. The proposed method for single frame smoke detection is based on a novel feature derived from the physics of smoke formation and is able to encode reliable information for detection.

3. SMOKE DETECTION BY DUAL-DICTIONARY MODELING

3.1 Physics-Based Image Formation Model

Based on the dichromatic atmospheric scattering model [7] developed for adverse weather conditions (e.g. fog/haze), smoke will act as the scattering medium like fog/haze. However, unlike fog/haze, smoke usually does not occupy the entire space of the scene. Assume that smoke appears at distance z_S from a camera and its thickness along the line of sight is z. Since there are no point sources of light, the irradiance at each background scene point is dominated by the ambient radiance, and the irradiance due to other scene points is not significant. By ignoring the multiple scattering, the observed intensity at pixel x [8] can be expressed as

$$F(x) = (1 - \varepsilon(z, x))\, B(x) + \varepsilon(z, x) S(z_S, x), \quad (1)$$

if the pixel is covered by smoke. In Equation (1) $B(x) = \frac{g}{z^2} L_\infty \rho$ accounts for the background under clear air when there is no smoke (referred to as the background component or non-smoke component interchangeably); $\varepsilon(z, x) = 1 - e^{-\beta z}$ is in the range [0, 1] and referred to as blending parameter; $S(z_S, x) = g e^{-\beta z_S} \mathbf{L}_\infty$ represents the pure smoke at distance z_S from the observer and referred to as the smoke component; z is the scene depth at pixel x ; g is a constant that accounts for the optical settings of the imaging system;

β is the scattering coefficient; L_∞ is the radiance of the horizon ($z = \infty$); and ρ represents the reflectance properties and aperture of the scene point. When the pixel is not covered by smoke, Equation (1) reduces to $F(x) = B(x)$.

Notice that Equation (1) appears similar to the haze image model [24]. But they are in fact different. First, haze occupies the entire space. This leads to a constant global atmospheric light A in the haze model. Most de-hazing methods are derived based on a constant A. In contrast, smoke model assumes that a smoke appears at location (z_S, x) and with thickness, z, which leads to non-constant atmospheric light term, $S(z_S, x)$, for smoke. Second, the attenuation of radiance in haze always exists for all pixels and depends on the scene depth. However, attenuation $\varepsilon(z, x)$ only occurs if the light goes through the smoke and it depends on the thickness of the smoke.

In addition, the purpose of smoke detection and de-hazing is different. In de-hazing, the focus of interest is the back-ground image and the objective of de-hazing is to obtain a haze-free or haze-reduced background image. It is of no concern whether the haze component can be separated or how accurately it can be separated. However, in smoke detection, the focus of interest is the smoke component. The objective is to accurately judge whether there is smoke somewhere in the image or not. If there is, we need to separate the smoke component as accurately as possible to assist the assessment of the intensity and extent of the smoke. Hence, most de-hazing methods are not designed to serve as smoke detectors.

3.2 Smoke Detection on Block Level

Because B, and S vary across the scene (frame) and only F is observed, it is hard to solve for and S globally. This article adopts block-based approach and assumes that the thickness z of the smoke within a block is constant so that

is constant within a block (not over the entire image!). However, we do not assume S is constant even within a block so as to accommodate the local spatial heterogeneity. Sparse representation is adopted to represent the smoke within each block.

Let $\mathbf{f} \in \mathrm{R}^N$ be a given image block with N pixels, $\mathbf{b} \in \mathrm{R}^N$ and $\mathbf{s} \in \mathrm{R}^N$ be the corresponding background and smoke components. Then the image formation model described by Equation (1) can be written as

$$\mathbf{f} = (1 - \omega)\mathbf{b} + \omega\mathbf{s} + \mathbf{n}, \quad (2)$$

where $\mathbf{n} \in \mathrm{R}^N$ represents modeling noise and ω is the blending parameter within the block f . Guided by the image formation model and in order to extract reliable features for smoke detection from a single image block $\mathbf{f}$, the background component $\mathbf{b}$ should be separated from the smoke component $\mathbf{s}$. Intuitively, the problem can be formulated as the minimization of the power of the residual noise:

$$\min_{\omega, \mathbf{b}, \mathbf{s}} \|\mathbf{f} - \omega\mathbf{s} - (1 - \omega)\mathbf{b}\|_2^2 \quad . \omega \in [0, 1]. \quad (3)$$

Given only a single input image block $\mathbf{f}$, Equation (3) is under-determined because there are N equations but $2N + 1$ free variables for a gray-scale image. Further constraints are required to obtain a unique and reliable solution to Equation (3). A good estimation of $\mathbf{b}$, $\mathbf{s}$, and ω is expected if both $\mathbf{b}$ and $\mathbf{s}$ could be well modeled according to the visual property of non-smoke and pure smoke.

Research on image statistics suggests that image patches can be well-represented as a sparse linear combination of elements from an appropriately chosen over-complete dictionary [25]. If each image block is considered as a point in an N-dimensional space, pure smoke images are also likely to lie in multiple low-dimensional subspaces as validated in [5] and [8]. Driven by the progress of sparse representation methods [26] in recent years, if sample smoke images can be collected or generated to capture the distribution of pure smoke in the space, it is expected that any specific pure smoke image would have a sparse representation with respect to the samples. Similar argument can be made for samples of non-smoke

images. Such a collection of samples represents a dictionary and each sample in the dictionary is typically referred to as an atom. Both dictionaries, one for pure smoke and the other for non-smoke, are designed such that they lead to sparse representations over only one type of image content (either pure smoke or non-smoke). In addition, each dictionary should be trained using real samples to adapt to the specific type of image content. In this article, the two dictionaries are trained independently. More sophisticated schemes such as supervised dictionary learning can be adopted to enhance the discriminative power of the dictionaries.

Let $\mathbf{D_S} \in \mathrm{R}^{N \times J}$ (N J) be a dictionary for pure smoke and each column of $\mathbf{D_S}$ be an atom. Then a pure smoke image $\mathbf{s}$ is expected to be sparse in $\mathbf{D_S}$:

$$\mathbf{s} = \mathbf{D_S}\mathbf{x_S} \quad s.t. \quad \mathbf{x_S}_{\,0} \le M_S, \tag{4}$$

where $\mathbf{x_S}_{\,0}$ counts the number of non-zero entries in $\mathbf{x_S}$. Similarly a non-smoke image $\mathbf{b}$ is expected to be sparse in

$$\mathbf{b} = \mathbf{D_b}\mathbf{x_b} \quad s.t. \quad \mathbf{x_b}_{\,0} \le M_b. \tag{5}$$

Here M_S and M_b are the upper bounds for the number of non-zero entries in the sparse coefficients $\mathbf{x_S}$ and $\mathbf{x_b}$ respectively. Considering Equation (4) and (5) as the models for pure smoke and non-smoke respectively, Equation (3) can be rewritten as follows:

$$\min_{\omega,\mathbf{x_b},\mathbf{x_s}} \{\|\mathbf{f} - \omega \mathbf{D_s}\mathbf{x_s} - (1-\omega)\mathbf{D_b}\mathbf{x_b}\|_2^2 + \eta\|\mathbf{x_b}\|_0 + \gamma\|\mathbf{x_s}\|_0\}$$

$$s.t.\ \omega \in [0, 1], \tag{6}$$

where η and γ are regularization parameters. Due to the non-convexity of the 0-norm, a convex approximation is obtained by replacing it with the 1-norm:

$$\min_{\omega,\mathbf{x_b},\mathbf{x_s}} \{\|\mathbf{f} - \omega \mathbf{D_s}\mathbf{x_s} - (1-\omega)\mathbf{D_b}\mathbf{x_b}\|_2^2 + \eta\|\mathbf{x_b}\|_1 + \gamma\|\mathbf{x_s}\|_1\}$$
$$s.t.\ \omega \in [0, 1]. \tag{7}$$

The optimization problem expressed by Equation (7) is convex with respect to one of $\mathbf{x_b}$, $\mathbf{x_S}$, and ω when fixing the other two. One may propose to optimize the three terms alternately.

However, ω and $(1-\omega)$ are coupled with $\mathbf{x_S}$ and $\mathbf{x_b}$ respectively by multiplication, which indicates that $\mathbf{x_b}$, $\mathbf{x_s}$, and ω may not be well estimated to reflect their true values, if no other constraints are imposed. Noting that the optimal ω is a scalar, we can always absorb ω into $\mathbf{x_S}$ and $(1-\omega)$ into $\mathbf{x_b}$ in Equation (7), and solve for $\omega\mathbf{x_S}$ and $(1-\omega)\mathbf{x_b}$. The only changes are to scale down γ and η by ω and $(1-\omega)$ respectively. This does not significantly change the essence of optimization, but helps to reduce one unknown ω. Based on this consideration, the following variables are defined

$$\mathbf{y_b} = (1-\omega)\mathbf{x_b}; \quad \mathbf{y_S} = \omega\mathbf{x_S}. \tag{8}$$

Then Equation (7) can be written as

$$\min_{\mathbf{y_b},\mathbf{y_s}} \|\mathbf{f} - \mathbf{D_s}\mathbf{y_s} - \mathbf{D_b}\mathbf{y_b}\|_2^2 + \eta\|\mathbf{y_b}\|_1 + \gamma\|\mathbf{y_s}\|_1. \tag{9}$$

Algorithm 1 Single Image Frame Smoke Detection

Input:
block image $\mathbf{f}$
dictionary $\mathbf{D_b}$ for non-smoke
dictionary $\mathbf{D_s}$ for pure smoke
regularization parameters η' and γ'
threshold Th for convergence checking
initial values obj_0 and obj_1 of the objective function (Equation (9)) such that $|obj_0 - obj_1| > Th$

Output:
class label for $\mathbf{f}$

initialize $\mathbf{y_s}$ by solving $\min_{\mathbf{y_s}} \|\mathbf{f} - \mathbf{D_s}\mathbf{y_s}\|_2^2 + \gamma'\|\mathbf{y_s}\|_1$;
initialize $\mathbf{y_b}$ by solving $\min_{\mathbf{y_b}} \|\mathbf{f} - \mathbf{D_b}\mathbf{y_b}\|_2^2 + \eta'\|\mathbf{y_b}\|_1$;
while $|obj_0 - obj_1| > Th$ **do**
  calculate $\mathbf{y_b}$ by solving Equation (9) with $\mathbf{y_s}$ fixed;
  $\mathbf{y_b} \Leftarrow \mathbf{y_b}$;
  calculate $\mathbf{y_s}$ by solving Equation (9) with $\mathbf{y_b}$ fixed;
  $\mathbf{y_s} \Leftarrow \mathbf{y_s}$; $obj_0 \Leftarrow obj_1$;
  calculate the current value obj_1 of the objective function (Equation (9));
  $obj_1 \Leftarrow obj_1$;
end while
concatenate $\mathbf{y_s}$ and $\mathbf{y_b}$ and input them to an SVM;
return the classification label.

$\mathbf{D_b}\mathbf{y_b}$ and $\mathbf{D_S}\mathbf{y_S}$ can be quasi-background and quasi-smoke component respectively hereinafter. Given $\mathbf{f}$, $\mathbf{D_b}$, and $\mathbf{D_s}$, Equation (9) can be solved through alternate optimization with regard to $\mathbf{y_b}$ and $\mathbf{y_s}$ respectively by using sparse coding algorithms such as the feature-sign search algorithm. Each is a convex problem and the convergence of the optimization is guaranteed. Once the difference between the objective function (Equation (9)) values in two consecutive iterations is less than a predefined threshold, the optimal $\mathbf{y_b}$ and $\mathbf{y_s}$ can be obtained.

For any input image block $\mathbf{f}$ irrespective of whether it contains smoke, the separated quasi-smoke component tends to be more like smoke. Thus features extracted from quasi-smoke component only may not be sufficient for deciding whether there is smoke or not in $\mathbf{f}$. This is particularly true when $\mathbf{f}$ is not covered by any smoke. Similar argument can be made for the quasi-background component. However, features extracted from both components jointly characterize $\mathbf{f}$ and are expected to have a good discriminative power. One may propose to extract LBP from both components and concatenate them as a feature to characterize $\mathbf{f}$. Actually the coefficient vectors $\mathbf{y_b}$ and $\mathbf{y_s}$ indicate which atoms and the proportion thereof that contribute to the construction of the quasi-background and quasi-smoke components. Thus, $\mathbf{y_b}$ and $\mathbf{y_s}$ are expected to encode more discriminative information. Based on this insight, $\mathbf{y_b}$ and $\mathbf{y_s}$ are concatenated as a novel feature to characterize $\mathbf{f}$. The extracted feature is input to an SVM classifier. A decision is made on whether there is smoke or not in $\mathbf{f}$.

3.3 Discussions

In the proposed detection method, the two dictionaries are learned offline and, once learned, they can be generally applied to any image block for smoke detection. Therefore, most computation time in detection is spent in the step of feature extraction; that is, obtaining the sparse coefficients to represent the quasi-smoke and quasi-background components by solving Equation (9). In this step, the sparse coefficients $\mathbf{y_b}$ and $\mathbf{y_s}$ are alternately calculated using the feature-sign search algorithm. The complexity of this step is $O(K_1 K_2(K_3^3 + K_4^3))$, where K_1 is the number of iterations within the feature-sign search algorithm, K_2 is the number of alternations, K_3 is the number of non-zero entries in $\mathbf{y_b}$, and K_4 is the number of non-zero entries in $\mathbf{y_s}$. Typical values of K_1, K_2, K_3 and K_4 for our experiments are 5, 15, 30 and 20 respectively.

It is noted that an image formation model similar to Equation (2) was also used for video-based smoke detection in [4] and [5]. In [4] and [5], background modeling based on the information of previous video frames is a strict prerequisite for image separation. In this article a different separation method is proposed for single frame smoke detection. Our work differs from it in two key aspects. First, the separation problem was for a mixture of texture and piece-wise smooth components. Second, the dictionaries used in that work were restricted to well-known transforms such as the curve let and discrete cosine transforms. As shown later in the article, the dictionaries $\mathbf{D_s}$ and $\mathbf{D_b}$ are learned from real samples so as to adapt to the smoke and non-smoke classes. A dictionary was trained to represent the high frequency part of an image and the dictionary was divided into two parts, one represents rain component while the other represents non-rain component.

4. SEPARATION OF THE SMOKE COMPONENT

While the separated quasi-smoke and quasi-background components for each image block are sufficient for detection of smoke, the separation of the true smoke component, if any, from the background is required to assess the intensity of the smoke. However, this cannot be achieved at the block level through the dual-dictionary modeling. The reason is that the dual-dictionary modeling only imposes constraints on the smoke and background component without constraining the blending parameter, which is coupled with both components. A reliable estimation of the blending parameter requires appropriate constraints on it.

Considering each pixel has a blending parameter, it is necessary to impose certain constraints based on the intuition that the pixel-wise blending parameters should be smooth across the entire image. Given one single image, block-based smoke detection can be performed using Algorithm 1 in a sliding window manner. Every pixel in the image can be classified as smoke or non-smoke according to the block-based detection results. Furthermore, by setting thresholds for the output score of SVM classifier, each pixel can be further classified into one of three categories: reliable smoke, reliable non-smoke, and uncertain. For both reliable smoke and non-smoke pixels, their blending parameters are close to 1 and 0 respectively. Thus the problem reduces to the estimation of blending parameters for uncertain pixels. Based on this argument, the blending parameter may be estimated in a propagating fashion from the known pixels to the unknown ones.

Such a formulation of the problem to estimate the blending parameter shares some common properties with image matting [9]. However, the separation of smoke component differs from deep image matting in three aspects. First, a trimap that serves as the constraints for deep image matting is accurate or error-free as it is usually manually constructed. In our case, a similar map is to be obtained through automatic detection; and this will be referred to as detection map hereinafter. Obviously, it includes detection errors. Second, almost all the deep image matting algorithms are developed to deal with cases where a substantial area of foreground objects to be separated is opaque. However, much of the area covered by smoke may be semi-transparent. Third, most existing matting methods only focus on alpha extraction but not foreground estimation. In our case, smoke is also expected to be extracted from single gray-scale images in addition to the blending parameter. For these differences, it has been found through experiments that the existing deep image matting algorithms do not work satisfactorily in our case. Therefore, a new method is developed for separating smoke component in a single image.

For each uncertain pixel based on the detection results, the corresponding smoke component, background component, and blending parameter are all unknown. Thus some assumptions are required to solve this under-determined problem. First, based on the intuition that pixels that are spatially close and visually similar should have similar blending parameters, the following assumption is made: the blending parameter of a pixel can be approximated by a linear combination of the parameters of similar pixels with the weights being given by a similarity function $K(i, j)$ measuring the similarity between pixels. Such an assumption can be expressed mathematically as follows:

$$\omega_i \approx \omega_j \sum_j K(i, j)/D_i \quad \text{or} \quad D_i \omega_i \approx \mathbf{K}_i \omega, \tag{10}$$

where $\omega \in R^n$ is a vector including ω values for all the n pixels in the input image. D_i and $\mathbf{K}_i$ are defined as follows:

$$D_i = \sum_j K(i, j), \tag{11}$$

$$\mathbf{K}_i = [K(i, j), \cdots, K(i, n)] \tag{12}$$

Under this assumption, two key issues need to be addressed for the estimation of ω. First is the question of how to gather similar pixels j, given a pixel i of interest. Second is how to construct the similarity function, $K(i, j)$.

Both the reliable smoke and reliable non-smoke pixels in a detection map are modeled by a Gaussian mixture model (GMM). The reason that GMM is adopted to cluster the known pixels is as follows. The detection map constructed from smoke detection includes some errors, which means that there are some misclassifications among reliable smoke, reliable non-smoke, and unknown pixels. Given a pixel of interest, some misclassified pixels may be selected as similar pixels to

the one of interest, which will lead to an inaccurate estimation of ω. In order to reduce or eliminate the adverse effects of misclassified pixels on similar pixel selection process, we propose to cluster the reliable smoke and reliable non-smoke pixels using GMM. In addition, the similarity measurements between pixels are deemed to include both photometric and geometric information. Specifically, a local patch centered at each known pixel is selected and the gray-scale values of all the pixels in the local patch are used to capture the photometric information of the centering pixel. The spatial coordinates of each known pixel are regarded as the geometric information of the pixel. The photometric and geometric information form a feature vector and the clustering using GMM will be performed in the feature space. As an illustration, for the mth Gaussian, the corresponding Gaussian probability density function (PDF) is expressed as:

$$f_m[\mathbf{p}(i)] = \frac{1}{(2\pi)^{M/2}|\Sigma_m|^{1/2}} \times \exp\{-\frac{1}{2}[\mathbf{p}(i) - \mu_m]^T \Sigma_m^{-1}[\mathbf{p}(i) - \mu_m]\}, \quad (13)$$

where μ_m and Σ_m are the mean vector and covariance matrix of the mth Gaussian, $\mathbf{p}(i) \in \mathrm{R}^M$ is the M-dimensional feature vector for the i th pixel. Given an unknown pixel i of interest, the PDFs for the pixel based on all the Gaussians can be calculated. The Gaussian that achieves the highest PDF value will be selected and the corresponding known pixels used to learn this Gaussian will be regarded as similar pixels as the i th pixel.

The similarity function should be constructed in such a way that it has a larger value if the distance between two pixels in the feature space is smaller, and vice versa. Here Manhattan distance is adopted and the similarity function is defined as [33]:

$$K(i, j) = 1 - \frac{\|\mathbf{p}(i) - \mathbf{p}(j)\|_1}{C}, \quad (14)$$

where $\mathbf{p}(i)$ and $\mathbf{p}(j)$ are the feature vectors for the i th and j th pixel respectively and C is a normalization constant.

After addressing the aforementioned two issues, we further define:

$$\mathbf{A} = \begin{bmatrix} K(1,1) & \cdots & K(1,n) \\ \vdots & \ddots & \vdots \\ K(n,1) & \cdots & K(n,n) \end{bmatrix}, \quad (15)$$

and

$$\mathbf{D} = \begin{bmatrix} D_1 & 0 & \cdots & 0 \\ 0 & D_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & D_n \end{bmatrix}, \quad (16)$$

$\mathbf{A} \in \mathrm{R}^{n\times n}$ is often referred to as affinity matrix and $\mathbf{D} \in \mathrm{R}^{n\times n}$ is a diagonal matrix. According to Equation (10), we have:

$$\mathbf{D}\omega \approx \mathbf{A}\omega. \quad (17)$$

Thus,

$$(\mathbf{D} - \mathbf{A})\omega \approx \mathbf{0} \quad or \quad \omega^T \mathbf{L}_C \omega \approx 0, \quad (18)$$

where $\mathbf{L}_C = (\mathbf{D} - \mathbf{A})^T (\mathbf{D} - \mathbf{A})$ is referred to as the clustering Laplacian. In consideration of the constraints from the detection map, the blending parameter can be estimated by solving the following optimization problem:

$$\omega^* = \text{argmin}\ \omega^T \mathbf{L}_C \omega + \lambda(\omega^T - \mathbf{v}^T)\mathbf{V}(\omega - \mathbf{v}), \quad (19)$$

where λ is a regularization parameter, $\mathbf{V} \in \mathrm{R}^{n\times n}$ is a diagonal matrix whose diagonal elements are one for known pixels (i.e. either reliable smoke or non-smoke) and zero for all the unknown pixels, and $\mathbf{v} \in \mathrm{R}^n$ is a vector containing the specified ω values for the known pixels and zero for all the other pixels. Intuitively, the i th entry v_i of $\mathbf{v}$ can be expressed as follows:

$$v_i = \begin{cases} 1 & \text{if pixel } i \text{ is reliable smoke} \\ 0 & \text{otherwise.} \end{cases} \quad (20)$$

The solution to Equation (19), which is a quadratic function of ω, can be obtained by differentiating it with respect to ω and setting the derivatives to zero. Thus the optimal blending parameter is:

$$\omega^* = (\mathbf{L}_C + \lambda\mathbf{V})^{-1}(\lambda\mathbf{v}). \quad (21)$$

For the smoke component, a similar assumption can be made: the smoke component of a pixel can be approximated by a linear combination of the smoke components of the similar pixels with the weights being given by a similarity function. A similar analysis to that of the blending parameter leads to the following optimization problem:

$$\mathbf{s}^* = \text{argmin}\ \mathbf{s}^T \mathbf{L}_C \mathbf{s} + \lambda(\mathbf{s}^T - \mathbf{u}^T)\mathbf{V}(\mathbf{s} - \mathbf{u}), \quad (22)$$

where $\mathbf{u} \in \mathrm{R}^n$ is a vector containing the pixel values from the input image $\mathbf{f}$ for reliable smoke pixels and zero for all the other pixels. Intuitively, the i th entry u_i of $\mathbf{u}$ can be expressed as follows:

$$u_i = \begin{cases} f_i & \text{if pixel } i \text{ is reliable smoke} \\ 0 & \text{otherwise.} \end{cases} \quad (23)$$

The optimal smoke component can be expressed as:

$$\mathbf{s}^* = (\mathbf{L}_C + \lambda\mathbf{V})^{-1}(\lambda\mathbf{u}). \quad (24)$$

Once the blending parameter and smoke component are estimated through solving Equations (21) and (24), they are substituted into the image formation model and the background component can be calculated.

5. EXPERIMENTAL RESULTS

To verify the methods proposed in this article, the following experiments were conducted and the related results are reported in this section:

(1) Visual assessment of the quasi-smoke and quasi-background components.

(2) Binary classification of smoke and non-smoke image blocks to evaluate the discriminative power of the proposed feature.

(3) Classification of heavy smoke, light smoke, and non-smoke based on the proposed feature.

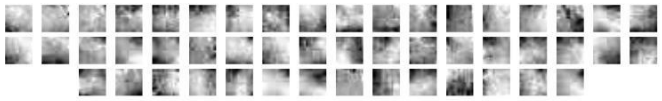

Fig. 1. Examples of the bases from the learned smoke dictionary $\mathbf{D_S}$.

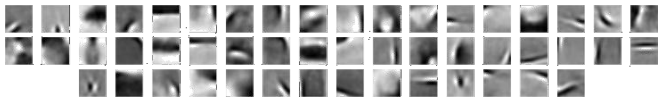

Fig. 2. Examples of the bases from the learned non-smoke dictionary $\mathbf{D_b}$.

(4) Evaluation of the proposed feature on challenging images, such as fog, that have similar visual appearance as smoke.
(5) Validation of the effectiveness of the proposed feature on real data.
(6) Experiments of separating smoke from background on real images.

5.1 Data Sets

Five data sets were constructed for the above experiments.

1) DS1: The data set consists of 1000 pure smoke images with the size of 16 × 16 pixels and was used to learn $\mathbf{D_S}$. The widely-used K-SVD was adopted to train $\mathbf{D_S}$ which has the size of 256 × 500. Some basis samples from $\mathbf{D_S}$ are shown in Fig. 1.

2) DS2: This data set consists of non-smoke images with the size of 16 × 16 pixels and was used to learn $\mathbf{D_b}$. To ensure that $\mathbf{D_b}$ has good generalization ability, it was constructed by randomly cropping 60000 non-smoke images with the size of 16×16 pixels from the images in the CIFAR-100 data set. Similarly, K-SVD was used to train $\mathbf{D_b}$ which has the same size as $\mathbf{D_S}$. Some basis samples from $\mathbf{D_b}$ are shown in Fig. 2.

3) DS3: This data set is a collection of smoke images with the size of 16 × 16 pixels and was used for testing the proposed detection algorithm. Specifically, it includes 5000 image blocks that were manually cropped based on visual observation from 25 publicly available video clips of smoke. These video clips [1]–[3], cover indoor and outdoor, short and long distance surveillance scenes with different illuminations; example frames of which are shown in Fig. 3. Furthermore, half of the 5000 image blocks are heavy smoke and the other half are light smoke.

4) DS4: This is a collection of general non-smoke image blocks used for testing the proposed detection algorithm and covers a large variety of real life image patches. Specifically, 5000 images with the size of 16 × 16 pixels were randomly cropped from the images in the 15-scene data set.

5) DS5: This data set is composed of challenging non-smoke block images used for testing the proposed detection algorithm. In a single image smoke detection problem there are three categories of challenges that could lead to misclassification or false detection. These challenges arise respectively from the physical formation process, transparency and homogeneity properties of smoke. As a result, each image in *DS5* must exhibit at least one of the challenges in the three categories. The first category includes some objects sharing similar physical formation process as smoke, such as haze/fog, cloud, and steam. As shadow and glass have similar transparency property as smoke, they can be considered as the representative samples of the second category. Due to the homogeneous property of some pure smoke, image patches with high homogeneity may pose a challenge to accurate detection. Thus smooth wall, sky, clothes and vehicle body are included in the third category. Water is also considered due to its transparency property and high homogeneity. In summary, to construct *DS5*, images a total of 10 challenging object classes were collected from Internet and 2500 non-smoke images with the size of 16 × 16 pixels were cropped randomly from the related challenging regions, with 250 block images in each class. Samples of challenging non-smoke images and the corresponding block images cropped are shown in Fig. 4.

Fig. 3. Example frames of the video clips of smoke.

TABLE I
DATA SETS USED IN THE RELATED DETECTION EXPERIMENTS

Experiment	Data sets involved
(2)	*DS3*, *DS4*
(3)	*DS3*, *DS4*
(4)	*DS3*, *DS5*
(5)	*DS3*, *DS4*, *DS5*

In a summary, the data sets used in the related detection experiments are listed in Table I.

5.2 Separation of Quasi-Smoke and Quasi-Background

Given a test image block $\mathbf{f}$ and the trained dictionaries $\mathbf{D_b}$ and $\mathbf{D_S}$, the corresponding sparse coefficients $\mathbf{y_b}$ and $\mathbf{y_S}$ are estimated by solving Equation (9). Then quasi-background component $\mathbf{D_b y_b}$ and quasi-smoke component $\mathbf{D_S y_S}$ are calculated. For an image which includes many blocks, the separation can be performed on every block in a sliding window manner. To evaluate the separation performance qualitatively, visual inspection was conducted for both indoor and outdoor scenarios. Some separated quasi-smoke and quasi-background components are shown in Fig. 5. Notice that the pure non-smoke areas in the test images can be well represented by the corresponding quasi-background component. For the smoke-covered regions, most smoke information is included in the corresponding quasi-smoke component while the details

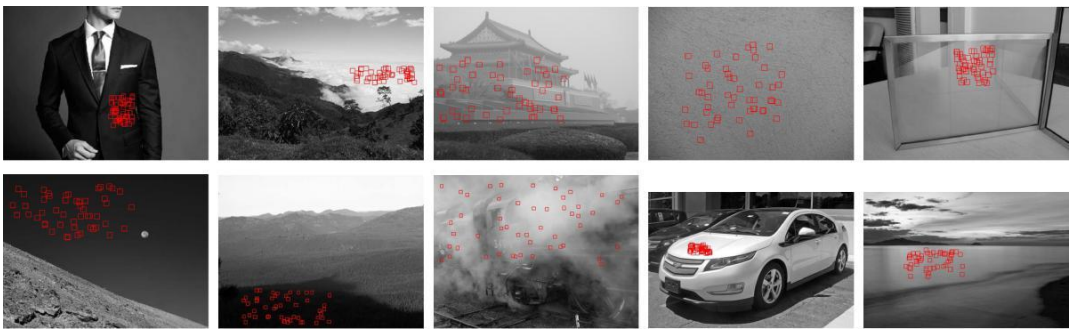

Fig. 4. Sample challenging non-smoke images and the corresponding block images cropped (The block images indicated by red rectangle in each image represent one of the following challenging object classes respectively: clothes, cloud, haze/fog, wall, glass, sky, shadow, steam, vehicle body, and water).

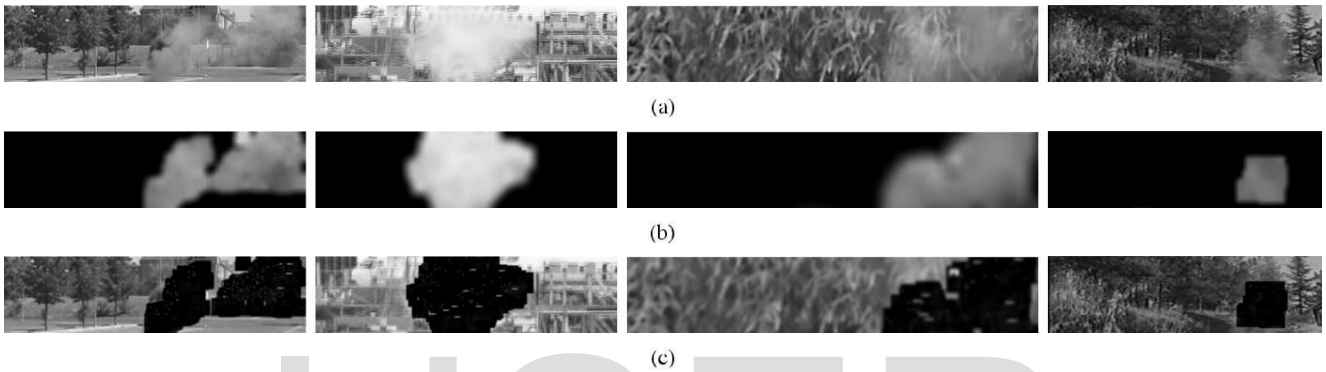

Fig. 5. Quasi-smoke and quasi-background separation ((a) the test images, (b) the separated quasi-smoke components, (c) the separated quasi-background components).

of non-smoke are characterized by the corresponding quasi-background component.

5.3 Classification of Smoke and Non-Smoke

Given 5000 smoke image blocks (*DS3*) and 5000 general non-smoke image blocks (*DS4*), the discriminative power of the proposed feature was studied. Specifically, each of the 10000 image blocks was considered as **f**. Given the trained dictionaries $\mathbf{D_b}$ and $\mathbf{D_s}$, the corresponding sparse coefficients $\mathbf{y_b}$ and $\mathbf{y_s}$ were estimated by solving Equation (9). The novel feature, (i.e. concatenated $\mathbf{y_b}$ and $\mathbf{y_s}$) characterizing the test image block was input to an SVM classifier to determine whether it contains smoke. In the rest of the article, the proposed feature will be referred to as *SC*.

To compare the *SC* feature with conventional image features for smoke detection, the texture feature, LBP, was chosen since it offered the state-of-the-art performance in video-based smoke detection [3]. As shown in [5], the texture feature extracted from the separated smoke component was more reliable than that extracted from the original video frame. In our experiments LBP was extracted from the separated components as well. After $\mathbf{y_b}$ and $\mathbf{y_s}$ were estimated, quasi-background component $\mathbf{D_b y_b}$ and quasi-smoke component $\mathbf{D_s y_s}$ were obtained. Similarly to the trick used in [5], LBP of $\mathbf{D_s y_s}$ was considered as a feature for smoke detection, and will be referred to as LBP_S in the rest of this article. Additionally, the concatenated LBP extracted respectively from $\mathbf{D_b y_b}$ and $\mathbf{D_s y_s}$ may encode discriminative information and was tested as well; and this will be referred to as LBP_C hereafter. For completeness, LBP that was extracted from the original image block **f** without performing separation was also tested; and this will be referred to as LBP in the rest of this article.

TABLE II
ACCURACIES FOR THE CLASSIFICATION OF SMOKE AND NON-SMOKE (LBP: EXTRACTED FROM THE ORIGINAL IMAGE BLOCK **f**; LBP_S: EXTRACTED FROM THE QUASI-SMOKE COMPONENT $\mathbf{D_s y_s}$ ONLY; LBP_C: EXTRACTED FROM BOTH THE QUASI-SMOKE COMPONENT $\mathbf{D_s y_s}$ AND THE QUASI-BACKGROUND COMPONENT $\mathbf{D_b y_b}$ AND THEN CONCATENATED)

Feature	LBP	LBP_S	LBP_C	***SC* (*Proposed*)**
Accuracy (%)	68.96	80.49	85.58	**94.9**

Both linear and radial basis function (RBF) kernel SVM were tested and 5-fold cross validation was performed in all experiments in this article, unless otherwise specified. The classification accuracies are reported in Table II. As shown in the table, among the four features tested, the proposed feature *SC* achieves the highest accuracy in the binary classification of smoke and non-smoke. As expected, the texture feature LBP extracted from **f** has the worst performance. With the texture information of both quasi-background and quasi-smoke components considered, LBP_C is more discriminative than LBP_S, which only represents the texture feature of quasi-smoke component. For clarity, the confusion matrix based on *SC* is shown in Table III. Furthermore, the ROC curves are adopted as performance measurement. They are shown

TABLE III

CONFUSION MATRIX FOR THE CLASSIFICATION OF SMOKE AND NON-SMOKE BASED ON THE PROPOSED FEATURE *SC*

		Detected	
		Smoke	Non-smoke
Truth	Smoke	95.2%	4.8%
	Non-smoke	5.4%	94.6%

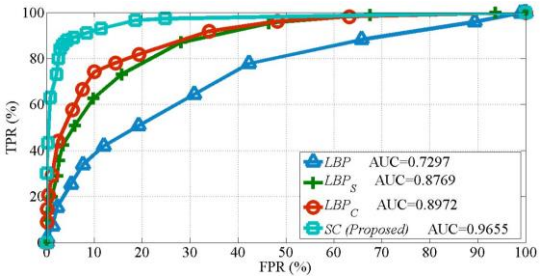

Fig. 6. ROC curves for the classification of smoke and non-smoke (*L B P*: extracted from the original image block **f**; *L B PS*: extracted from the quasi-smoke component $\mathbf{D_{SYS}}$ only; *L B PC* : extracted from both the quasi-smoke component $\mathbf{D_{SYS}}$ and the quasi-background component $\mathbf{D_{BYB}}$ and then concatenated).

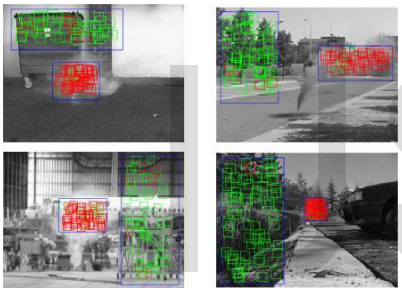

Fig. 7. Illustrative classification results (blue rectangle: the selected region; red block: classified as smoke; green block: classified as non-smoke). Image is best viewed in color.

in Fig. 6 along with area under the curve (AUC) values. It is evident that the proposed feature *SC* outperforms all the other three features.

Some classification results are shown in Fig. 7. In each scene shown in Fig. 7, one smoke region and one non-smoke region were selected manually for illustration purpose; these are indicated using blue rectangle. Then some block images were randomly selected from the two regions as test samples. The smoke and non-smoke blocks classified by using the proposed feature are indicated by red block and green block respectively. Although there are a few misclassifications on block level, the selected regions indicated by blue rectangle will not be misclassified if a simple majority voting is employed.

To compare the proposed detection algorithm with the result reported in [5], a data set was constructed from the data used for smoke detection test in [5]. The constructed data set includes 4000 blocks of images of 16 by 16 pixels, half of which are smoke (either heavy or light) and the rest are non-smoke foreground objects. Notice that all the 4000 image blocks are associated with 4000 background image blocks that

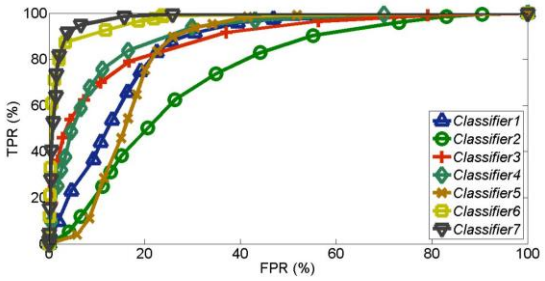

Fig. 8. ROC curves for single image smoke detection based on different classifiers.

were estimated through video-based background modeling. Given this data set, the smoke detection algorithm presented in [5] was implemented on these block images. A 5-fold cross validation was performed and the classification accuracy is 92.8%. The proposed frame-based detection algorithm achieved 86.3% 5-fold cross validation accuracy. Notice that it is expected that the method in [5] would perform better since the video was captured by a stationary camera and it uses temporal information through the background modeling to estimate the background component, However, its accuracy can drop substantially if there is a small jitter of the camera. For the running time, it takes 0.75 second for the method in [5] to process an image. It takes 0.3 second for our frame-based detection method to process an image with the same size. The running time is measured based on a single core MATLAB implementation on a PC with 2.00GHz Intel(R) Core(TM) i7 CPU and 8GB memory.

5.4 Classification of Heavy Smoke, Light Smoke and Non-Smoke

Generally at the onset, smoke starts out lightly in a video surveillance scene. In order to be useful for assessing the stage of smoke, the algorithm should be able to differentiate amongst heavy smoke, light smoke, and non-smoke. Furthermore, the algorithm should not be sensitive to false alarm caused by some objects with high homogeneous appearance such as clothes and vehicle body. This consideration motivates us to conduct a ternary classification of heavy smoke, light smoke, and non-smoke based on the proposed feature. To the best of our knowledge, such a test has not been reported in the literature before.

Specifically, 2500 image blocks were randomly selected from the data set of non-smoke *DS4*. Given these 2500 non-smoke, 2500 heavy smoke, and 2500 light smoke image blocks, quasi-smoke and quasi-background components were separated and the proposed feature *SC* was extracted. For comparative evaluation, *L B P*, *L B PS* and *L B PC* were also extracted as texture feature. The classification accuracies are reported in Table IV. As shown, among all the four features the highest accuracy was obtained when using the proposed feature *SC*. It is also noted that, for the classification of heavy smoke, light smoke, and non-smoke, the features *L B PS* , *L B PC* and *SC* extracted based on the separated components still outperform *L B P*. For clarity, the confusion matrix for

Fig. 9. Examples of image-based detection. Six non-smoke images and three smoke images are sampled from ImageNet and each image is divided into 16 × 16 non-overlapping blocks and each block went through the smoke detector. Smoke blocks are highlighted in red boxes.

TABLE IV

ACCURACIES FOR THE CLASSIFICATION OF HEAVY SMOKE, LIGHT SMOKE, AND NON-SMOKE (LBP : EXTRACTED FROM THE ORIGINAL IMAGE BLOCK **f** ; LBP_S : EXTRACTED FROM THE QUASI-SMOKE COMPONENT $\mathbf{D_{sys}}$ ONLY; LBP_C : EXTRACTED FROM BOTH THE QUASI-SMOKE COMPONENT $\mathbf{D_{sys}}$ AND THE QUASI-BACKGROUND COMPONENT **Dbyb** AND THEN CONCATENATED)

	LBP	LBP_S	LBP_C	***SC* (Proposed)**
Accuracy (%)	51.92	62.77	73.61	**84.47**

TABLE V

CONFUSION MATRIX FOR THE CLASSIFICATION OF HEAVY SMOKE, LIGHT SMOKE, AND NON-SMOKE BASED ON THE PROPOSED FEATURE *SC*

		Detected		
		Heavy	Light	Non-smoke
Truth	Heavy	81.4%	18.2%	0.4%
	Light	13.6%	76.2%	10.2%
	Non-smoke	1.2%	3%	95.8%

ternary classification based on *SC* is shown in Table V. Notice that most non-smoke can be differentiated from heavy smoke and light smoke. The main misclassification occurs between heavy smoke and light smoke. This result is consistent with the classification of smoke and non-smoke.

5.5 Differentiation of Smoke From Challenging Objects

Recall that there are objects with similar visual properties as smoke that may challenge the smoke detector. In this section, experiment on differentiating smoke from some challenging objects is reported. It is noteworthy that this is the first time the result of such experiment was reported.

In this experiment, 2500 smoke (including both heavy and light) image blocks that were randomly selected from the smoke data set *DS3* and 2500 challenging non-smoke image blocks from *DS5* were used. To make a comparison, LBP_C which has been proven to be the best among LBP-based features was extracted from quasi-smoke and quasi-background components as texture feature.

The classification of these image blocks into smoke and non-smoke yielded classification accuracies of 77.16% and 79.2% when using LBP_C and *SC* respectively. Specifically, for each class of the challenging non-smoke objects, the accuracies (%) of being correctly classified as non-smoke are shown in Table VI. Notice that although LBP_C and *SC* lead to similar classification performance on the entire data set, their performance on each class varies significantly. Among all the challenging non-smoke classes considered, when *SC* is used, the test images of sky and steam have the highest and

TABLE VI
ACCURACIES (%) OF EACH CLASS OF OBJECTS IN *DS5* BEING CORRECTLY CLASSIFIED AS NON-SMOKE

Feature	LBP_C	SC
Clothes	75.6	92.4
Cloud	71.2	65.6
Fog/haze	62.4	61.6
Glass	78.4	77.2
Shadow	72.4	83.6
Sky	92.8	96.8
Steam	58.8	54
Vehicle body	97.2	95.2
Wall	83.6	81.2
Water	79.2	84.4

lowest probability of being correctly classified as non-smoke respectively.

5.6 Smoke Detection: Real Application Considerations

Based on the results so far obtained, the proposed feature *SC* has been validated to effectively differentiate between the classes of smoke and common non-smoke; and the classes of smoke and challenging non-smoke. However, a classifier which can well differentiate smoke from common non-smoke may not effectively classify smoke from challenging non-smoke. Similar argument can be made for a classifier trained using smoke and challenging non-smoke only. Moreover, in a smoke detection system it will be preferable to filter out common non-smoke at a first stage of smoke detection. Then smoke and challenging non-smoke are further differentiated at a second stage. Based on these considerations, a tree-structured classifier may have good generalization ability in classifying smoke from non-smoke and was constructed. Using the data sets *DS3*, *DS4*, and *DS5* described in Section V-A, two partitions (training and test data) were created. In the training set, there are 1500 image blocks including either heavy or light smoke, 1500 common non-smoke image blocks selected randomly, and 1500 challenging non-smoke block images. The test set comprises 3500 smoke image blocks, 3500 common non-smoke image blocks, and 1000 challenging non-smoke image blocks.

Five SVMs (listed below) were trained using the proposed feature *SC*:

- *Classifier1*: a binary classifier trained on the 1500 smoke image blocks and the 1500 common non-smoke block images
- *Classifier2*: a binary classifier trained on the 1500 smoke block images and the 1500 challenging non-smoke image blocks
- *Classifier3*: a binary classifier trained on the 1500 smoke image blocks and the 3000 non-smoke (including both common and challenging) image blocks
- *Classifier4*: a ternary classifier trained on the 1500 smoke image blocks, the 1500 common non-smoke image blocks, and the 1500 challenging non-smoke image blocks
- *Classifier5*: a binary classifier trained on the 1500 smoke image blocks, the 1500 common non-smoke image blocks, and the 1500 challenging non-smoke image blocks (the smoke image blocks and challenging non-smoke image blocks were considered to be from the same class)

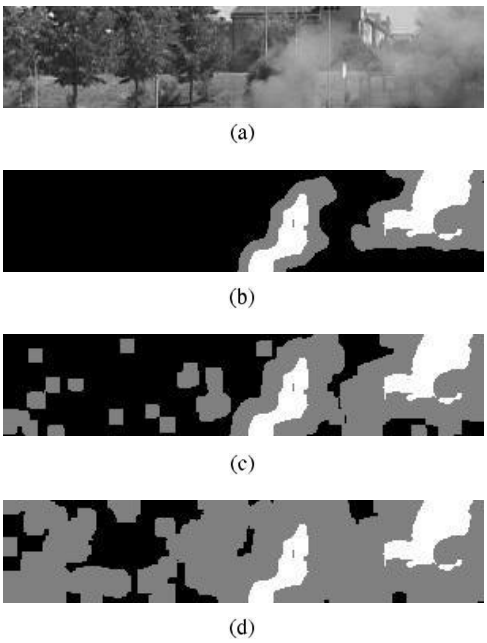

Fig. 10. An input image and its corresponding detection maps constructed based on the results of smoke detection using different thresholds for the output score of SVM classifier. (a) Input image. (b) Detection map-1. (c) Detection map-2. (d) Detection map-3.

Based on the trained five SVMs, two classifiers with tree-structure were constructed. They are:

- *Classifier6*: a concatenated classifier of *Classifier1* and *Classifier2*
- *Classifier7*: a concatenated classifier of *Classifier5* and *Classifier2*

Given the 3500 smoke image blocks and 4500 non-smoke (including 3500 common and 1000 challenging non-smoke) image blocks in the test set, image separation was performed and *SC* was extracted. The ROC curves for smoke detection based on the seven classifiers are shown in Fig. 8. Overall tree-structured classifiers *Classifier6* and *Classifier7* outperform all the other five single SVM classifiers. Furthermore, *Classifier7* leads to the best performance among all the classifiers. Trained on the smoke and challenging non-smoke image blocks only, *Classifier2* gives the worst performance among all the classifiers. The ROC curve based on *Classifier7* also indicates the effectiveness of the proposed feature *SC* for single image smoke detection.

The block-based detector provides a core algorithm for many application systems with specific engineering requirements. In practice, false detections at block-level are expected to be sparse and isolated. They are not difficult to be corrected through a post processing by majority voting, morphological operation or Markov-random field (MRF) modelling on the detection map. If applications permit, these possible false detected smoke patches can also be kept in a queue and given more attention (as suspicious patches) in the following

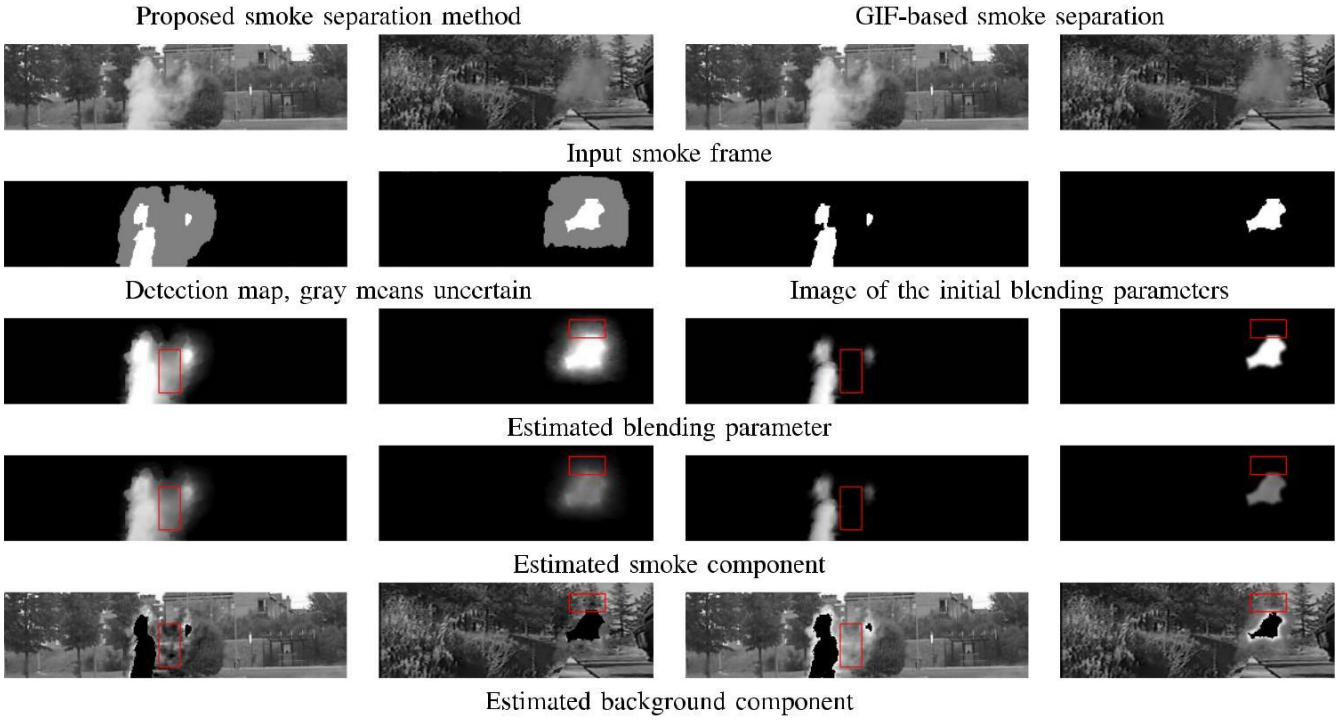

Fig. 11. Results for frame-based smoke separation. The left two columns show the results of the proposed method and the right two columns show the results obtained using GIF.

images taken. This will help to decide whether this is a true positive or a false alarm. However, building a practical system is not the focus of the paper. Figure 9 shows the image-based detection where six non-smoke images and three smoke images are sampled from Image Net. Each image is divided into 16 × 16 non-overlapping blocks and each block went through the smoke detector. If a block is detected as smoke, the block is highlighted in a red box. Notice that there is no false detection in the five of the six non-smoke images and there are only a few isolated false detections in the non-smoke image (d), which is easy to be corrected as discussed above.

5.7 Single Image Frame Smoke Separation

To validate the effectiveness of constructing detection map based on the results of smoke detection, an input image and its corresponding detection maps constructed according to different thresholds for the output score of SVM classifier are shown in Fig. 10. Note that reliable smoke, reliable non-smoke, and unknown pixels are indicated with white, black, and gray respectively in the detection maps. Despite the different detection maps, there are some misclassification.

In addition, some results for smoke separation based on the proposed algorithm are shown in Fig. 11 and 12, which cover both outdoor and indoor scenarios with smoke at different stages. As can be noticed from Fig. 11 and the left example of Fig. 12, a small percentage of misclassification leads to a relatively reliable detection map, which further yields good estimation of the blending parameter and smoke component. When misclassification is considerable, as is shown in the right example of Fig. 12, the estimation of the blending parameter and smoke component becomes inaccurate. In terms of background separation, although it is hard to reliably estimate the part that is covered by heavy smoke, certain areas of the background covered by light smoke are recovered.

Experiments were also conduced using the guided image filtering (GIF) to estimate the blending parameter and smoke component of the images shown in Figs. 11 and 12. For a fair comparison, an initial image of blending parameter was obtained by applying the proposed smoke detection algorithm. Then, GIF was applied to refine the blending parameter estimation guided by the original image. For the smoke component, an initial smoke image was extracted by taking initially detected smoke pixels and refining the corresponding smoke image by the GIF, with the original image as the guidance image. The background images were then calculated from the estimated blending parameters and the smoke components. For all the images, parameters of the GIF were tuned to achieve the best quality of both blending parameter and smoke image (judged visually as there is no ground truth for these real images).

Observation of the results shows that (a) GIF-based method tends to fail to estimate the light smoke, which is often semi-transparent. These regions are highlighted in red boxes on the images; (b) it is possible to further tune the GIF parameters, especially to increase the radius, such that the light smoke can be captured to some extent in the smoke component and blending parameter. However, in this case, the background structure would become visible in the smoke component due to fact that the GIF transfers the structure of the guidance image to the filtering output. Computationally, the GIF has advantages. To process a mega-pixel image, GIF takes about 1 second while the proposed smoke separation method takes

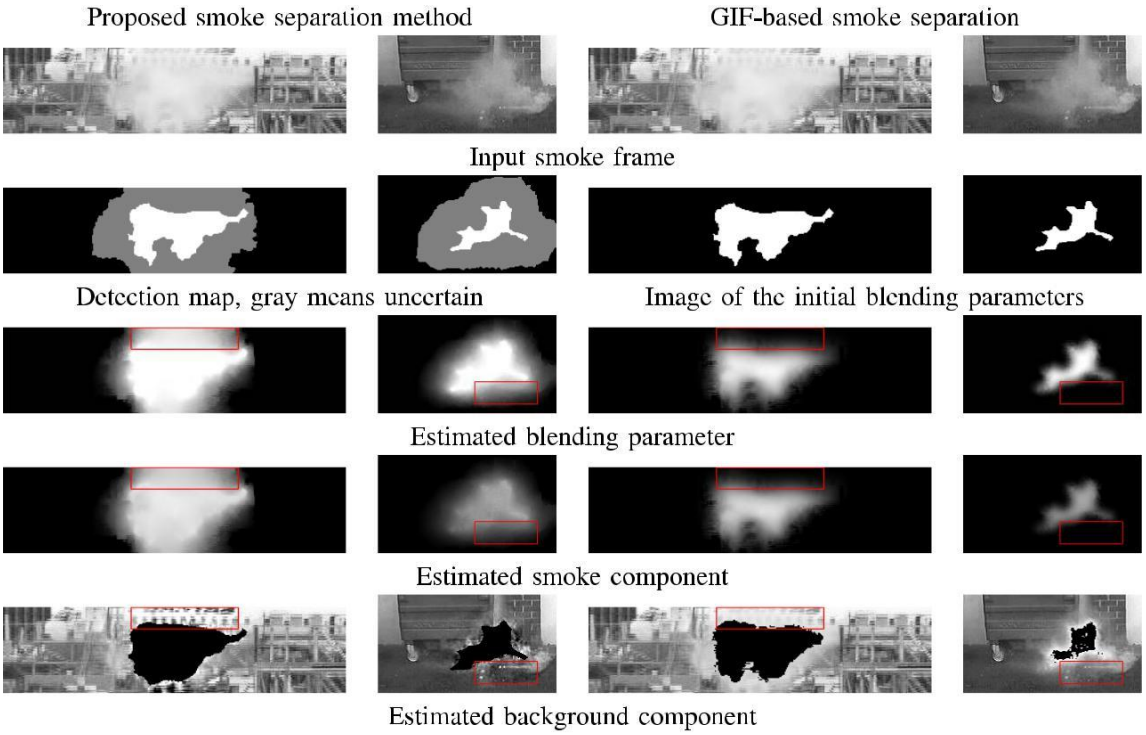

Fig. 12. Results for frame-based smoke separation. The left two columns show the results of the proposed method and the right two columns show the results obtained using GIF (cont.).

40 seconds on the same PC described above. Both algorithms were implemented in MATLAB. It is expected that GPU based implementation could reduce the processing time substantially.

6. CONCLUSION

In this article, novel methods based on finite element method (FEM) have been proposed and verified for the detection and separation of smoke from a single image frame of a video. In particular, based on the imaging model, an optimization scheme allowing the separation of quasi-smoke and quasi-background components was formulated using dual over-complete dictionaries. A novel feature is constructed as a concatenation of the respective sparse coefficients for detection. In addition, a method based on the concept of deep image matting has been developed to separate the true smoke and background components from the automatic detection results. Extensive experiments on detection were conducted and the results indicate that the proposed feature significantly outperforms the existing features for smoke detection. This verifies the efficacy of the detection method. Moreover, the proposed method is able to differentiate smoke from other challenging objects with similar visual appearance in a gray-scale image, viz. fog/haze, cloud, shadow, etc. Experiments on smoke separation have also demonstrated that the proposed separation method can effectively estimate/separate the true smoke and background components. Further improvement may be achieved through nonlinear modeling of the smoke component such as kernel or auto-encoder based modeling.

The proposed framework is suitable and can be extended for detection and separation of transparent or semi-transparent and deformable objects.

REFERENCES

[1] 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.
[2] B. U. Töreyin, Y. Dedeo˘glu, and A. E. Çetin, "Wavelet based real-time smoke detection in video," in *Proc. 13th Eur. Signal Process. Conf.*, 2005, pp. 1–4.
[3] F. Yuan, "A double mapping framework for extraction of shape-invariant features based on multi-scale partitions with adaboost for video smoke detection," *Pattern Recognit.*, vol. 45, no. 12, pp. 4326–4336, Dec. 2012.
[4] H. Tian, W. Li, L. Wang, and P. Ogunbona, "A novel video-based smoke detection method using image separation," in *Proc. IEEE Int. Conf. Multimedia Expo*, Jul. 2012, pp. 532–537.
[5] H. Tian, W. Li, L. Wang, and P. Ogunbona, "Smoke detection in video: An image separation approach," *Int. J. Comput. Vis.*, vol. 106, no. 2, pp. 192–209, Jan. 2014.
[6] 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 Int. Workshop Multimedia Signal Process.*, Oct. 2011, pp. 1–4.
[7] S. G. Narasimhan and S. K. Nayar, "Vision and the atmosphere," *Int. J. Comput. Vis.*, vol. 48, no. 3, pp. 233–254, 2002.
[8] H. Tian, W. Li, P. Ogunbona, and L. Wang, "Single image smoke detection," in *Proc. Asian Conf. Comput. Vis. (ACCV)*, 2014, pp. 87–101.
[9] J. Wang and M. F. Cohen, "Image and video matting: A survey," *Found. Trends Comput. Graph. Vis.*, vol. 3, no. 2, pp. 97–175, 2007.
[10] E. S. Varnousfaderani and D. Rajan, "Weighted color and texture sample selection for image matting," *IEEE Trans. Image Process.*, vol. 22, no. 11, pp. 4260–4270, Nov. 2013.
[11] F. Yuan, "A fast accumulative motion orientation model based on integral image for video smoke detection," *Pattern Recognit. Lett.*, vol. 29, no. 7, pp. 925–932, May 2008.
[12] O. Gunay, B. U. Toreyin, K. Kose, and A. E. Çetin, "Entropy-functional-based online adaptive decision fusion framework with application to wildfire detection in video," *IEEE Trans. Image Process.*, vol. 21, no. 5, pp. 2853–2865, May 2012.
[13] I. Kolesov, P. Karasev, A. Tannenbaum, and E. Haber, "Fire and smoke detection in video with optimal mass transport based optical flow and neural networks," in *Proc. 17th IEEE Int. Conf. Image Process.*, Sep. 2010, pp. 761–764.