Single Image Smoke Detection

Hongda Tian[✉], Wanqing Li, Philip Ogunbona, and Lei Wang

Advanced Multimedia Research Lab, ICT Research Institute,
School of Computer Science and Software Engineering,
University of Wollongong, Wollongong, Australia
{ht615,wanqing,philipo,leiw}@uow.edu.au

Abstract. Despite the recent advances in smoke detection from video, detection of smoke from single images is still a challenging problem with both practical and theoretical implications. However, there is hardly any reported research on this topic in the literature. This paper addresses this problem by proposing a novel feature to detect smoke in a single image. An image formation model that expresses an image as a linear combination of smoke and non-smoke (background) components is derived based on the atmospheric scattering models. The separation of the smoke and non-smoke components is formulated as convex optimization that solves a sparse representation problem. Using the separated quasi-smoke and quasi-background components, the feature is constructed as a concatenation of the respective sparse coefficients. Extensive experiments were conducted and the results have shown that the proposed feature significantly outperforms the existing features for smoke detection.

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 and providing early detection with information on the location and intensity [1–5]. Despite the recent advances [4,5], almost all existing detection algorithms are video-based and the video is assumed to be captured by stationary cameras in order to facilitate the motion detection and feature extraction involved in these algorithms. However, such requirement can be hardly met in an open space where cameras are inevitably jittering under severe and dynamic environment, such as wind. Our experiments (see Sect. 5.7) have shown that camera jittering can significantly degrade the performance of video-based smoke detection. If the surveillance is based on battery-powered sensor network, the available power supply, computing resource, or bandwidth is hardly sufficient for video processing and smoke detection. In this case, surveillance images rather than videos are available. Furthermore, when a pan-tilt-zoom (PTZ) camera is used in video-based smoke detection, the unreliable background modeling will cause the failure of most systems. In such circumstances, detection of smoke from single images becomes highly desirable. This desirability comes at a price because image-based detection is much more challenging than video-based systems as it

is no longer possible to estimate the background required for the separation of the smoke component in the state-of-the-art methods recently proposed in [4,5]. To the best of our knowledge, there is little study reported on single image-based smoke detection. This paper presents a novel method to address this problem.

The main contributions of the paper are three-fold: (i) Based on the atmospheric scattering (attenuation and airlight) models [6], an image formation model for smoke is derived. This model explains how smoke scatters the light reflected from the background of the scene and also serves as a source of light through scattering. The model suggests that an image patch covered by smoke can be approximated as a linear combination of two components; one component is contributed by smoke while the other is contributed by the background. The weight of the composition is a function of the thickness and the scattering coefficient of the smoke. (ii) Guided by the image model, dictionary-based sparse representation for the two components is used to separate an image into quasi-smoke and quasi-background components through a convex optimization process. The coefficients of both components are concatenated as a novel feature for detection. The experimental results verified that the proposed feature is reliable and highly discriminative. (iii) A method to differentiate light smoke and heavy smoke and a method to differentiate smoke and fog/haze are presented. Preliminary results on these are reported in the paper.

The remainder of the paper is organized as follows: a brief review of existing video-based smoke detection methods is provided in Sect. 2. Based on the atmospheric scattering models, an image formation model for smoke is derived in Sect. 3. The proposed method based on the image formation model is presented in Sect. 4. Experimental results are shown in Sect. 5 along with discussions. Finally, the paper is concluded with some perspectives on future work in Sect. 6.

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 feature some representative video-based smoke detection methods are reviewed with respect to the features they used. The features have been based on the characteristics of smoke including motion, color, edge and texture.

From motion point of view, an accumulative motion model has been proposed to capture the motion characteristics of smoke in [7]. Other research efforts have extracted motion features of smoke using optical flow [8,9]. However, no motion information is available from a single image. The fact that the color of smoke is usually grayish provides a clue for the extraction of color features [1–3,10–13]. However, this paper focuses on detection of smoke from single gray-scale images.

Given the video of a scene, blurred edges could be observed in smoke-covered areas and the consequent decrease in high frequency has been used as cue to perform smoke detection [1,2,11]. 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. Owing to the dispersive distribution
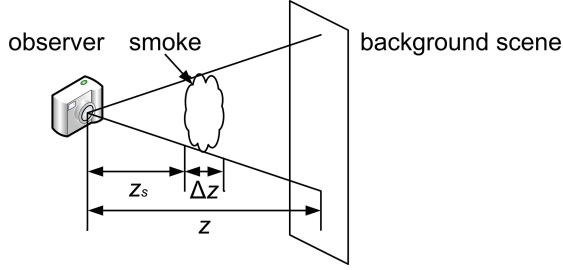
Fig. 1. Smoke usually appears at a certain distance from the observer with limited thickness along the line of sight.

of smoke, texture features have been extracted for smoke detection [3,10,13,14]. Additionally, it is also noted that the transmission [15], fractal [16] and histograms of oriented gradient (HOG) [9] have been employed to detect smoke.

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 these methods require a video sequence captured by a stationary camera and in the case of color feature, a color camera. Hence in general, they cannot deal with smoke detection from a single gray-scale image. The proposed feature for single image smoke detection is based on the physics of smoke formation and is able to encode reliable information for detection.

3   Physics-Based Image Formation Model

To develop computer vision systems that are able to operate in adverse weather conditions (e.g. fog/haze), the dichromatic atmospheric scattering model was proposed in [6]. The model accounts for the presence of scattering medium (e.g. fog/haze) in the entire space and expresses the final spectral irradiance $\mathbf{F}(z,\lambda)$ received by the observer (e.g. camera) as the sum of the irradiance $\mathbf{T}(z,\lambda)$ of directly transmitted light and the irradiance $\mathbf{A}(z,\lambda)$ of airlight:

$$\mathbf{F}(z,\lambda) = \mathbf{T}(z,\lambda) + \mathbf{A}(z,\lambda), \tag{1}$$

where z is the distance between the scene and observer, and λ refers to the wavelength of light. Specifically, $\mathbf{T}(z,\lambda)$ is related to the attenuation of a beam of light as it travels through the scattering medium. $\mathbf{A}(z,\lambda)$ is related to the phenomenon whereby the medium behaves like a source of light, which is caused by the scattering of environmental illumination by particles of the medium.

In the case of smoke, the 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 Δz, as shown in Fig. 1. There are no point sources of light, that 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, a formation model for smoke can be derived using the reasoning similar to that in [6] as follows:

$$\mathbf{T}(z,\lambda) = g\frac{e^{-\beta(\lambda)\Delta z}}{z^2}\mathbf{L}_\infty(\lambda)\rho(\lambda); \tag{2}$$

$$\mathbf{A}(z,\lambda) = g\int_{z_s}^{z_s+\Delta z}\mathbf{L}_\infty(\lambda)\beta(\lambda)e^{-\beta(\lambda)z}dz$$
$$= ge^{-\beta(\lambda)z_s}(1 - e^{-\beta(\lambda)\Delta z})\mathbf{L}_\infty(\lambda), \tag{3}$$

where g is a constant that accounts for the optical settings of the imaging system, $\beta(\lambda)$ is the scattering coefficient, $\mathbf{L}_\infty(\lambda)$ is the radiance of the horizon ($z = \infty$) at wavelength λ, and $\rho(\lambda)$ represents the reflectance properties and aperture of the scene point. Substituting Eqs. (2) and (3) into Eq. (1) yields

$$\mathbf{F}(z,\lambda) = (1 - \Omega(\Delta z,\lambda))\mathbf{B}(z,\lambda) + \Omega(\Delta z,\lambda)\mathbf{S}(z_s,\lambda), \tag{4}$$

where

$$\Omega(\Delta z,\lambda) = 1 - e^{-\beta(\lambda)\Delta z};$$
$$\mathbf{B}(z,\lambda) = \frac{g}{z^2}\mathbf{L}_\infty(\lambda)\rho(\lambda); \tag{5}$$
$$\mathbf{S}(z_s,\lambda) = ge^{-\beta(\lambda)z_s}\mathbf{L}_\infty(\lambda).$$

Equation (4) is the image formation model for smoke. $\mathbf{B}(z,\lambda)$ accounts for the background under clear air when there is no smoke. In the rest of the paper, it is referred to as the background component or non-smoke component interchangeably. $\mathbf{S}(z_s,\lambda)$ represents the pure smoke at distance z_s from the observer, which is referred to as the smoke component. The parameter $\Omega(\Delta z,\lambda) \in [0,1]$ depends on the thickness Δz of the smoke. It can be assumed constant within a small area where Δz would not vary much. In the rest of the paper it is referred to as the blending parameter. Note the derived model Eq. (4) indicates an additive relationship between smoke and non-smoke components.

4   Proposed Method

This paper adopts block-based detection scheme in order to achieve early detection (smoke will usually cover a very small area at the early stage) and localization of the smoke.

4.1   Smoke Detection on Block Level

Let $\mathbf{f} \in \mathbb{R}^N$ be a given image block with N pixels, $\mathbf{b} \in \mathbb{R}^N$ and $\mathbf{s} \in \mathbb{R}^N$ be the corresponding background and smoke components. Then the image formation model described by Eq. (4) can be written as

$$\mathbf{f} = (1 - \omega)\mathbf{b} + \omega\mathbf{s} + \mathbf{n}, \tag{6}$$

where $\mathbf{n} \in \mathbb{R}^N$ represents modeling noise. From Eq. (5), it is apparent that the blending parameter $\Omega(\Delta z, \lambda)$ depends on the scattering coefficient $\beta(\lambda)$ of smoke and the thickness Δz of the smoke along the line of sight. Assuming that the scattering coefficient of smoke does not change appreciably within the visible wavelength and the thickness of smoke is constant within a small image block, $\Omega(\Delta z, \lambda)$ is a constant within the image block, and the quantity is referred to as the blending parameter, ω, on block level. 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:

$$\min_{\omega, \mathbf{b}, \mathbf{s}} \|\mathbf{f} - \omega\mathbf{s} - (1 - \omega)\mathbf{b}\|_2^2 \qquad s.t. \quad \omega \in [0, 1]. \tag{7}$$

Given only a single input image block $\mathbf{f}$, further constraints are required to obtain an unique and reliable solution to Eq. (7). 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. If each image block is considered as a point in an N-dimensional space, pure smoke images are likely to lie in multiple low-dimensional subspaces. Driven by the progress of sparse representation [17] 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 a basis. 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). To fix these ideas let $\mathbf{D_s} \in \mathbb{R}^{N \times J}(N \ll J)$ be a dictionary for pure smoke and each column of $\mathbf{D_s}$ be a basis. Then a pure smoke image $\mathbf{s}$ is expected to be sparse in $\mathbf{D_s}$:

$$\mathbf{s} = \mathbf{D_s}\mathbf{x_s} \qquad s.t. \quad \|\mathbf{x_s}\|_0 \leq M_s, \tag{8}$$

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 a dictionary $\mathbf{D_b} \in \mathbb{R}^{N \times L}(N \ll L)$ for non-smoke:

$$\mathbf{b} = \mathbf{D_b}\mathbf{x_b} \qquad s.t. \quad \|\mathbf{x_b}\|_0 \leq M_b. \tag{9}$$

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 Eqs. (8) and (9) as the models for pure smoke and non-smoke respectively, Eq. (7) can be rewritten as follows:

$$\min_{\omega, \mathbf{x_b}, \mathbf{x_s}} \{\|\mathbf{f} - \omega\mathbf{D_s}\mathbf{x_s} - (1 - \omega)\mathbf{D_b}\mathbf{x_b}\|_2^2 + \eta\|\mathbf{x_b}\|_0 + \gamma\|\mathbf{x_s}\|_0\} \qquad s.t. \qquad \omega \in [0, 1], \tag{10}$$

where η and γ are regularization parameters. Due to the non-convexity of the ℓ_0-norm, it is replaced with the ℓ_1-norm, which is the common practice in the literature:

$$\min_{\omega,\mathbf{x_b},\mathbf{x_s}} \{\|\mathbf{f} - \omega\mathbf{D_s}\mathbf{x_s} - (1-\omega)\mathbf{D_b}\mathbf{x_b}\|_2^2 + \eta\|\mathbf{x_b}\|_1 + \gamma\|\mathbf{x_s}\|_1\} \quad s.t. \quad \omega \in [0,1]. \tag{11}$$

The optimization problem expressed by Eq. (11) 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 Eq. (11), 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

$$\mathbf{y_b} = (1-\omega)\mathbf{x_b}; \qquad \mathbf{y_s} = \omega\mathbf{x_s}. \tag{12}$$

Then Eq. (11) can be written as

$$\min_{\mathbf{y_b},\mathbf{y_s}} \|\mathbf{f} - \mathbf{D_s}\mathbf{y_s} - \mathbf{D_b}\mathbf{y_b}\|_2^2 + \eta'\|\mathbf{y_b}\|_1 + \gamma'\|\mathbf{y_s}\|_1. \tag{13}$$

In this case, $\mathbf{D_b}\mathbf{y_b}$ and $\mathbf{D_s}\mathbf{y_s}$ can be regarded as the scaled version of the background and smoke component respectively; and they will be referred to as quasi-background and quasi-smoke component respectively in the rest of the paper. Given $\mathbf{f}$, $\mathbf{D_b}$, and $\mathbf{D_s}$, Eq. (13) 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 [18]. Each is a convex problem and the convergence of the optimization is guaranteed [19]. Once the difference between the objective function (Eq. (13)) 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 $\mathbf{f}$ and irrespective of whether it contains smoke, $\mathbf{y_b}$ and $\mathbf{y_s}$ are estimated to model the quasi-background and quasi-smoke component respectively. Both $\mathbf{y_b}$ and $\mathbf{y_s}$ are expected to encode useful information of the input image block $\mathbf{f}$. As a result, they are concatenated as a novel feature to characterize $\mathbf{f}$. The extracted feature is input to a support vector machine (SVM) classifier. A decision is made on whether there is smoke or not in $\mathbf{f}$.

4.2　Discussions

It is noted that an image formation model similar to Eq. (6) was also used for video-based smoke detection in [4,5], image matting in [20,21], and single image haze removal in [22]. In [4,5], background modeling based on the information of previous video frames is a strict prerequisite for image separation. In this paper

a different separation method is proposed for single image smoke detection. User interactions are usually required for image matting. Our image model was derived from the atmospheric scattering models and the proposed method for smoke detection is fully automatic. A dark channel prior was assumed for outdoor haze-free images for restoring high quality haze-free images in [22]. The removal of haze does not require a good separation of haze and the input image in [22] must be a color image to employ the dark channel prior. In this paper, given a single gray-scale image, quasi-smoke is separated from quasi-background to extract reliable features for smoke detection. A somewhat related work was reported in [23] but our work differs from it in two key aspects. First, the separation problem in [23] was for a mixture of texture and piece-wise smooth components. Second, the dictionaries used in that work were restricted from well known transforms such as the curvelet and discrete cosine transforms. As shown later in the paper, the dictionaries $\mathbf{D_b}$ and $\mathbf{D_s}$ are learned from real samples so as to adapt to the smoke and non-smoke classes.

5   Experimental Results

In this section, some preparations for the experiments including the data sets used are described in Sect. 5.1. Some separated quasi-smoke and quasi-background components are shown in Sect. 5.2. To explore the separability between smoke and general non-smoke classes based on the proposed feature, experiments with a binary classification task are performed in Sect. 5.3. To explore the separability among the classes of heavy smoke, light smoke, and general non-smoke based on the proposed feature, results of a ternary classification task are reported in Sect. 5.4. As fog/haze share similar visual appearance with smoke, they may pose a challenge for single image smoke detection. Thus, it is instructive to test whether smoke and fog/haze could be differentiated using the proposed feature; and this is studied in Sect. 5.5. The effectiveness of the proposed feature for smoke detection in real applications is validated in Sect. 5.5 as well. Furthermore, the computational complexity of the proposed algorithm is analyzed in Sect. 5.6. Finally, to make a comparison between video-based smoke detection and image-based smoke detection under the situation that cameras are jittering, experiments are conducted in Sect. 5.7.

5.1   Data Sets and Experimental Setup

Smoke and non-smoke images with the size of 16×16 pixels were collected. These block images were then divided into two parts, one for training the smoke and non-smoke dictionaries and the other for training and testing the classifiers for smoke detection. Notice that the images for learning dictionaries were strictly excluded for classifier training/testing.

Given an input image block $\mathbf{f}$, two over-complete dictionaries $\mathbf{D_b}$ and $\mathbf{D_s}$ are required to solve Eq. (13). To adapt smoke and non-smoke classes, K-SVD [24] was adopted to train $\mathbf{D_b}$ and $\mathbf{D_s}$ from the training samples. Specifically,

Fig. 2. Examples of the bases from the learnt dictionary $\mathbf{D_s}$ for smoke.
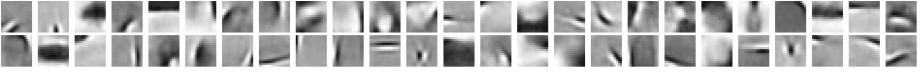

Fig. 3. Examples of the bases from the learnt dictionary $\mathbf{D_b}$ for non-smoke.

1000 pure smoke images with the size of 16×16 pixels were used to learn $\mathbf{D_s}$. To make $\mathbf{D_b}$ have good generalization ability, 60000 non-smoke images with the size of 16×16 pixels that were randomly cropped from the images in the CIFAR-100 data set [25], were used to learn $\mathbf{D_b}$. In the experiments, both $\mathbf{D_b}$ and $\mathbf{D_s}$ have the size of 256×500. Some basis samples from $\mathbf{D_s}$ and $\mathbf{D_b}$ are shown in Figs. 2 and 3 respectively.

To construct a data set of smoke for training and testing the classifier, 5000 images with the size of 16×16 pixels 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. Furthermore, half of the 5000 block images are heavy smoke and the rest are light smoke.

To construct a data set of general non-smoke for training and testing the classifier, which cover a large variety of real life image patches, 5000 images with the size of 16×16 pixels were randomly cropped from the images in the 15-scene data set [26].

To construct a data set of fog/haze image patches for training and testing the classifier, 10 fog/haze images were collected from [22,27–29]. 2500 images with the size of 16×16 pixels were cropped from the fog/haze regions in those images; there are 250 block images in each collected image.

In addition, four video clips that were captured by unstable cameras were chosen. 1000 images with the size of 16×16 pixels were manually cropped from the videos. Half of these block images are smoke (either heavy or light) and the rest are non-smoke foreground objects. Notice the 1000 cropped block images are associated with 1000 background block images that were estimated through video-based background modeling [30].

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 Eq. (13). 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 validate the separation performance, the collage in Fig. 4 shows some separated quasi-smoke and quasi-background components.
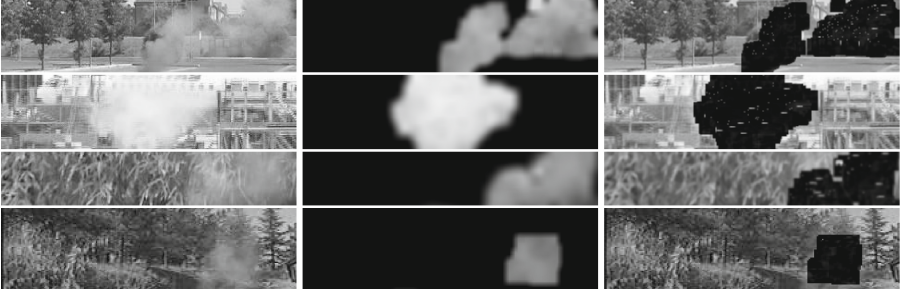
Fig. 4. Quasi-smoke and quasi-background separation (column 1: the test images, column 2: the separated quasi-smoke components, column 3: the separated quasi-background components).

5.3   Binary Classification with the Proposed Feature

Given 5000 smoke image blocks and 5000 general non-smoke image blocks, the separability between them based on the proposed feature was studied. Specifically, each of the 10000 block images was considered as $\mathbf{f}$. 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 Eq. (13). The concatenated $\mathbf{y_b}$ and $\mathbf{y_s}$ was considered as a novel feature to characterize the test image block and as input to SVM classifier to determine whether it contains smoke. In the rest of the paper, the proposed feature will be referred to as SC.

The visual features based on motion, color, and edge are not suitable for smoke detection from a single gray-scale image. Thus texture feature was adopted in this paper. As local binary pattern (LBP) [31] has been successfully used in texture classification tasks and was applied to video-based smoke detection in [3–5], it was adopted for comparison in our experiments. As shown in [4,5], the texture feature extracted from the separated smoke component is 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 obtained, quasi-background component $\mathbf{D_b y_b}$ and quasi-smoke component $\mathbf{D_s y_s}$ could be estimated. Similar to the trick used in [4,5], LBP simply extracted from $\mathbf{D_s y_s}$ was considered as a feature for smoke detection, and will be referred to as LBP_S in the rest of the paper. 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 in the rest of the paper. For completeness, LBP that was extracted from the original image block $\mathbf{f}$ without performing separation was also tested; and this will be referred to as LBP in the rest of the paper.

Both linear and radial basis function (RBF) kernel SVM were tested and 5-fold cross validation was performed in our experiments in the rest of the paper, unless otherwise specified. The classification accuracies are reported in Table 1. 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 general

Table 1. Accuracies for binary classification of smoke and general 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).

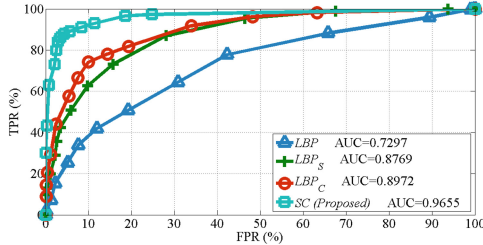| Feature	LBP	LBP_S	LBP_C	**SC (Proposed)**
Accuracy (%)	68.96	80.49	85.58	**94.9**

Fig. 5. ROC curves for binary classification of smoke and general 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).

non-smoke. As expected, the texture feature *LBP* extracted without component separation 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. Furthermore, the receiver operating characteristics (ROC) curves are adopted as performance measurement. They are shown in Fig. 5 along with area under the curve (AUC) values. It is evident that the proposed feature *SC* outperforms all the other three features.

The optimum SVM parameters obtained after tuning (5-fold cross validation on 10000 image blocks) were used to train a SVM classifier using the proposed feature. Some classification results based on the SVM are shown in Fig. 6. In each scene shown in Fig. 6, 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 classification errors on block level, the selected regions indicated by blue rectangle will not be misclassified if simple majority voting is employed.

5.4   Ternary Classification with the Proposed Feature

Generally at the onset, smoke starts out lightly colored in a video surveillance scene. In order to be useful for early smoke detection, the algorithm should be able to differentiate amongst heavy smoke, light smoke, and non-smoke. Furthermore,
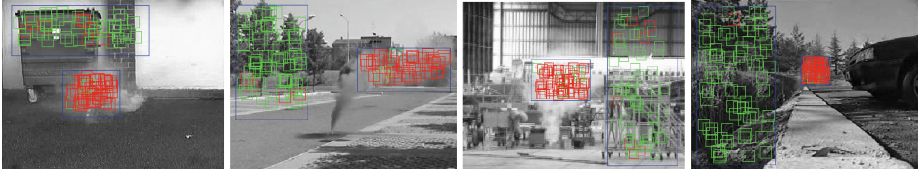
Fig. 6. Illustrative classification results (blue rectangle: the selected region; red block: classified as smoke; green block: classified as non-smoke) (Color figure online).

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 explore the separability among the classes of heavy smoke, light smoke, and general non-smoke based on the proposed feature. For this, a ternary classification task was conducted, which has not been reported in the literature.

Specifically, 2500 block images were randomly selected from the data set of general non-smoke. Given these 2500 general non-smoke, 2500 heavy smoke, and 2500 light smoke image blocks, separation experiments were performed and the proposed feature SC was extracted. For our comparative evaluation, LBP, LBP_S and LBP_C were also extracted as texture feature. The classification accuracies are reported in Table 2. Similar to the binary classification case, among all the four features the highest accuracy is observed when using the proposed feature SC. It is also noted that, for ternary classification of heavy smoke, light smoke, and general non-smoke, the features LBP_S, LBP_C and SC extracted based on the separated components still outperform LBP. For clarity, the confusion matrix for ternary classification based on SC is shown in Table 3. As can be noticed, most non-smoke can be differentiated from heavy smoke and light smoke. The main misclassification occurs between heavy smoke and light smoke.

5.5 Smoke Detection: Real Application Considerations

The separability between smoke and general non-smoke classes based on the proposed feature has been validated in Sect. 5.3 and 5.4. As mentioned before, fog/haze may pose a challenge for single image smoke detection. To better understand this challenging case, the separability between smoke and fog/haze classes was explored. Note this is the first time it is being reported in the literature. This consideration is also useful when specifying the classifiers to be used in a real smoke detection application.

Table 2. Accuracies for ternary classification of heavy smoke, light smoke, and general 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 (%)	51.92	62.77	73.61	**84.47**

Table 3. Confusion matrix for ternary classification of heavy smoke, light smoke, and general 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%

2500 block images were randomly selected from the smoke data set. Given these 2500 smoke (including both heavy and light) and 2500 fog/haze block images, separation experiments were conducted and the proposed feature SC was extracted. To make a comparison, LBP_C which has been proved to be the best among LBP features was extracted from quasi-smoke and quasi-background components as texture feature.

A binary classification task on these image blocks yielded classification accuracies of 76.6% and 77.5% when using LBP_C and SC respectively. Note that the study on the differentiation of smoke from fog/haze is preliminary. It can be expected from the above experiments that the proposed feature SC will outperform LBP-based features in a realistic case where smoke, fog/haze and non-smoke (excluding fog/haze) coexist.

Based on the results so far obtained, the proposed feature SC, has been validated to effectively separate between the classes of smoke and general non-smoke; and the classes of smoke and fog/haze. In a smoke detection system application it will be preferable to filter out general non-smoke at a first stage of smoke detection. Then smoke and fog/haze are further differentiated at a second stage. Based on this consideration, a tree-structured classifier may have good generalization ability in terms of classification between smoke and non-smoke. To validate this hypothesis, such a classifier was constructed and tested for its effectiveness in detecting smoke. Using the data sets described in Sect. 5.1, two partitions (training and test data) were created. In the training set, there are 1500 block images including either heavy or light smoke, 1500 general non-smoke block images selected randomly, and 1500 fog/haze block images. The test set comprises 3500 smoke block images, 3500 general non-smoke block images, and 1000 fog/haze block images. A SVM was trained using SC on the 1500 smoke block images and 1500 general non-smoke block images; and this classifier is referred to as *Classifier1*. Another SVM was trained using SC on the 1500 smoke block images and 1500 fog/haze block images; and this is referred to as *Classifier2*. For comparison, a SVM was also trained using SC on the 1500 smoke block images and 3000 non-smoke (including both general and fog/haze) block images; and this is referred to as *Classifier3*. *Classifier1* and *Classifier2* were simply concatenated as a tree-structured classifier; and this is referred to as *Classifier4*. Given the 3500 smoke block images and 4500 non-smoke (including 3500 general and 1000 fog/haze) block images in the test set, image separation
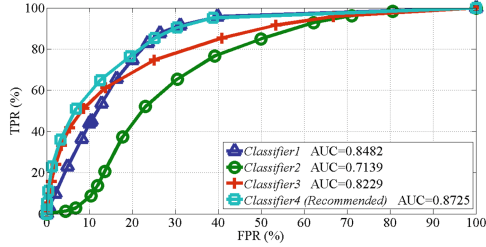
Fig. 7. ROC curves for single image smoke detection based on the four classifiers.

was performed and SC was extracted. The ROC curves for smoke detection based on the four classifiers are shown in Fig. 7, where AUC values are also provided. Overall *Classifier4* outperforms all the other three classifiers. Trained on smoke and fog/haze images only, *Classifier2* gives the worst performance among all the classifiers. The ROC curve based on *Classifier4* also indicates the effectiveness of the proposed feature SC for single image smoke detection.

5.6   Computational Complexity

In the proposed method, most computation time 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 Eq. (13). 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.

5.7   Smoke Detection with Jittering Cameras

When cameras jitter, video-based smoke detection algorithms could lead to poor performance due to the unreliable background modeling and feature extraction. However, single image smoke detection, which does not rely on the information of previous video frames, should perform well. To validate this, experiments using real video data were conducted. Given 1000 block images cropped from the video clips captured by jittering cameras, the proposed single image smoke detection method yielded a classification accuracy of 95.5 %. The state-of-the-art video-based smoke detection algorithm presented in [5] achieved only 54.5 %.

6   Conclusion and Future Work

In this paper, a novel feature, namely the sparse coefficients associated with an over-complete dictionary representation, has been proposed to detect smoke from a single image. The proposed feature arises from two parts; one representing the

smoke component of the input image and the other representing the non-smoke component. We derived a component-based image formation model for smoke using the atmospheric scattering models and formulated an optimization scheme that allowed the separation of quasi-smoke and quasi-background components. The effectiveness of the proposed feature for single image smoke detection was validated by the experimental results. Furthermore, practical consideration for the design of a smoke detection system that could be useful in specifying required classifiers was presented. As an indicator for successful smoke separation, a good estimation of ω is meaningful from the perspective of both theoretical and practical consideration and this will be pursued in our continuing work.

References

1. Calderara, S., Piccinini, P., Cucchiara, R.: Vision based smoke detection system using image energy and color information. Mach. Vis. Appl. **22**, 705–719 (2011)
2. Toreyin, B.U., Dedeoglu, Y., Cetin, A.E.: Wavelet based real-time smoke detection in video. In: EUSIPCO (2005)
3. Yuan, F.: A double mapping framework for extraction of shape-invariant features based on multi-scale partitions with adaboost for video smoke detection. Pattern Recogn. **45**, 4326–4336 (2012)
4. Tian, H., Li, W., Wang, L., Ogunbona, P.: A novel video-based smoke detection method using image separation. In: ICME, pp. 532–537 (2012)
5. Tian, H., Li, W., Wang, L., Ogunbona, P.: Smoke detection in video: an image separation approach. IJCV **106**, 192–209 (2014)
6. Narasimhan, S.G., Nayar, S.K.: Vision and the atmosphere. IJCV **48**, 233–254 (2002)
7. Yuan, F.: A fast accumulative motion orientation model based on integral image for video smoke detection. Pattern Recogn. Lett. **29**, 925–932 (2008)
8. Kolesov, I., Karasev, P., Tannenbaum, A., Haber, E.: Fire and smoke detection in video with optimal mass transport based optical flow and neural networks. In: ICIP (2010)
9. Park, J., Ko, B., Nam, J.Y., Kwak, S.: Wildfire smoke detection using spatiotemporal bag-of-features of smoke. In: WACV, pp. 200–205 (2013)
10. Jakovcevic, T., Stipanicev, D., Krstinic, D.: Visual spatial-context based wildfire smoke sensor. Mach. Vis. Appl. **24**, 707–719 (2013)
11. Labati, R.D., Genovese, A., Piuri, V., Scotti, F.: Wildfire smoke detection using computational intelligence techniques enhanced with synthetic smoke plume generation. IEEE Trans. Syst. Man Cybern.: Syst. **43**, 1003–1012 (2013)
12. Morerio, P., Marcenaro, L., Regazzoni, C.S., Gera, G.: Early fire and smoke detection based on colour features and motion analysis. In: ICIP, pp. 1041–1044 (2012)
13. Wang, Y., Chua, T.W., Chang, R., Pham, N.T.: Real-time smoke detection using texture and color features. In: ICPR, pp. 1727–1730 (2012)
14. Tian, H., Li, W., Ogunbona, P., Nguyen, D.T., Zhan, C.: Smoke detection in videos using non-redundant local binary pattern-based features. In: MMSP, pp. 1–4 (2011)

15. Long, C., et al.: Transmission: a new feature for computer vision based smoke detection. In: Wang, F.L., Deng, H., Gao, Y., Lei, J. (eds.) Artificial Intelligence and Computational Intelligence. LNCS, vol. 6319, pp. 389–396. Springer, Heidelberg (2010)
16. Maruta, H., Nakamura, A., Yamamichi, T., Kurokawa, F.: Image based smoke detection with local hurst exponent. In: ICIP, pp. 4653–4656 (2010)
17. Wright, J., Ma, Y., Mairal, J., Sapiro, G., Huang, T.S., Yan, S.: Sparse representation for computer vision and pattern recognition. Proc. IEEE **98**, 1031–1044 (2010)
18. Lee, H., Battle, A., Raina, R., Ng, A.Y.: Efficient sparse coding algorithms. In: NIPS (2007)
19. Bezdek, J.C., Hathaway, R.J.: Convergence of alternating optimization. Neural Parallel Sci. Computations **11**, 351–368 (2003)
20. Bai, X., Sapiro, G.: Geodesic matting: a framework for fast interactive image and video segmentation and matting. IJCV **82**, 113–132 (2009)
21. Levin, A., Lischinski, D., Weiss, Y.: A closed-form solution to natural image matting. IEEE TPAMI **30**, 228–242 (2008)
22. He, K., Sun, J., Tang, X.: Single image haze removal using dark channel prior. IEEE TPAMI **33**, 2341–2353 (2011)
23. Starck, J.L., Elad, M., Donoho, D.L.: Image decomposition via the combination of sparse representations and a variational approach. IEEE TIP **14**, 1570–1582 (2005)
24. Aharon, M., Elad, M., Bruckstein, A.: K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation. IEEE Trans. Sig. Process. **54**, 4311–4322 (2006)
25. Krizhevsky, A.: Learning multiple layers of features from tiny images. Technical report (2009)
26. Lazebnik, S., Schmid, C., Ponce, J.: Beyond bags of features: spatial pyramid matching for recognizing natural scene categories. In: CVPR, pp. 2169–2178 (2006)
27. Fattal, R.: Single image dehazing. ACM Trans. Graph. **27**, 1–9 (2008)
28. Kopf, J., Neubert, B., Chen, B., Cohen, M., Cohen-Or, D., Deussen, O., Uyttendaele, M., Lischinski, D.: Deep photo: model-based photograph enhancement and viewing. ACM Trans. Graph. **27** (2008)
29. Tan, R.T.: Visibility in bad weather from a single image. In: CVPR (2008)
30. Stauffer, C., Grimson, W.E.L.: Learning patterns of activity using real-time tracking. IEEE Trans. Pattern Anal. Mach. Intell. **22**, 747–757 (2000)
31. Ojala, T., Pietikainen, M., Maenpaa, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE TPAMI **24**, 971–987 (2002)