Int J Comput Vis (2014) 106:192–209
DOI 10.1007/s11263-013-0656-6

Smoke Detection in Video: An Image Separation Approach

Hongda Tian · Wanqing Li · Lei Wang · Philip Ogunbona

Received: 28 December 2012 / Accepted: 22 August 2013 / Published online: 8 September 2013
© Springer Science+Business Media New York 2013

Abstract Existing video-based smoke detection methods often rely on the visual features extracted directly from the original frames. In the case of light smoke, the background is still visible and it deteriorates the quality of the features. This paper presents an approach to separating the smoke component from the background such that visual features can be extracted from the smoke component for reliable smoke detection. Specifically, an image is assumed to be a linear blending of a smoke component and a background image. Given a video frame and its background, the estimation of the blending parameter and the actual smoke component can be formulated as an optimization problem. Three methods based on different models for the smoke component are proposed to solve the optimization problem. Experimental results on synthesized and real video data have shown that the proposed approach can effectively separate the smoke component and the smoke detection performance is significantly improved by using the visual features extracted from the smoke component.

Keywords Smoke detection · Smoke texture · Image separation · Sparse representation

H. Tian · W. Li (✉) · L. Wang · P. Ogunbona
School of Computer Science and Software Engineering, University of Wollongong, Northfield Avenue, Wollongong, NSW 2522, Australia
e-mail: wanqing@uow.edu.au

H. Tian
e-mail: ht615@uow.edu.au

L. Wang
e-mail: leiw@uow.edu.au

P. Ogunbona
e-mail: philipo@uow.edu.au

1 Introduction

The presence of smoke often signals the onset or possibility of a fire event. Thus, early detection of smoke can serve as a warning for incidence of fire. Conventional point smoke detectors, which include photoelectric and ionization detectors, mainly detect the presence of certain particles generated by smoke and fire. Photoelectric detectors make use of photometry to detect the presence of these particles. Ionization detectors achieve this by means of monitoring the reduced quantities of ionized air molecules. Both methods depend on the transportation of the smoke towards the detector and sufficient concentration of the molecules or particles being present. There is a delay inherent in the transportation and this is exacerbated in outdoor scenario where there could be draught or wind. Smoke detectors used in open areas require to be in close proximity of the source in order to be effective. Apart from the limitation of proximity these detectors do not provide information about the location of the fire, its burning rate or other key indicators.

Compared to the methods described above, visual inspection and detection do not suffer similar drawbacks. Recent advances in real-time video-based surveillance techniques have made vision-based smoke detection a promising approach to early detection of fire. Vision-based smoke detection is suitable in both enclosed and open spaces and there is an additional benefit of being able to specify the location of the fire, its scale and intensity.

Most vision-based smoke detection techniques adopt a pattern recognition paradigm in which the input image or video is preprocessed and 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 identifying robust visual features that can characterize smoke and to this end, there are reported studies

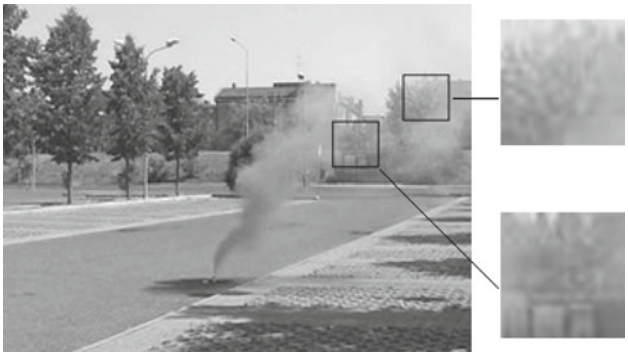

Fig. 1 Two image blocks covered by light smoke

in the literature (Toreyin et al. 2005; Yuan 2008; Calderara et al. 2011). Despite these efforts, vision-based smoke detection is still a challenge because of the difficulty in quantifying the visual characteristics of smoke. Some of the visual characteristics of smoke include irregular and deformable shape, irregular motion and varying degree of transparency. Unlike rigid objects, the irregular and deformable nature of smoke makes it difficult to extract geometrical or edge-based features. The motion of smoke, especially in the presence of draught or wind, can be irregular and somewhat random. Thus, well-known motion descriptors such as optical flow have not been very successful as features. The possibility of using texture as a feature descriptor for smoke is promising because of its dispersive distribution. However, one needs to account for the varying degree of transparency as the background covered by smoke may be included in the feature being extracted.

The potential offered by texture as feature to characterize smoke serves as a key motivation to pursue methods of dealing with the varying degree of transparency associated with smoke.Opaque objects, such as human faces and vehicles, act as solid objects and will obscure other objects lying directly behind them. However, smoke can be transparent and the level of transparency changes with the density of smoke. When smoke is heavy enough, it behaves as an opaque object and no visual information of the covered scene (background) is available. Light smoke, on the other hand, has a degree of transparency and the background is partially visible regardless of the color of the smoke. Texture features extracted from an image with light smoke will capture the visual information of both smoke and background, which renders such features ineffective in describing smoke. The images shown in Fig. 1 illustrate the possible varied effects different background images can have on the features extracted from light smoke.

We therefore seek to remove the effect of background image and extract textural features from the smoke component. A straightforward background subtraction method is ineffective in this case because of the gradual and slow temporal change in the density of smoke. When we consider the state-of-the-art background subtraction methods, most of the efforts are directed at background modeling. Given a video frame, each pixel is classified as either background or foreground pixel based on the background model. However, foreground pixels with light smoke certainly include background information. Figure 2 illustrates the presence of residual background information when a simple background subtraction method is used.

This paper presents a novel approach to separating the smoke component from the background such that visual features can be extracted from the smoke component for reliable smoke detection. We proceed by assuming that the image under consideration is a linear blending of a smoke component and a background component, also referred to as pure smoke and background respectively hereafter. We capture the characteristics of the smoke component using three variants of the model. In the first model we take advantage of the structure of inter-pixel correlation within a smoke image and impose a "local smoothness" constraint. Algebraically, the smoke component may be a manifold embedded in a higher dimensional space. Hence in the second model the smoke component is restricted to be a linear combination of the bases obtained by principal component analysis. Using similar algebraic argument, we construct the third model such that the smoke component is sparsely represented using an overcomplete dictionary that can be learned from pure smoke samples. Given a video frame and its background, the estimation of the blending parameter and the actual smoke component is formulated and solved as an optimization problem.

Following this introduction, we present a review of existing smoke detection and image separation methods in Sect. 2. The image formation model and the formulation of the smoke component separation models are presented in Sect. 3. The solutions of the three models are also provided. A smoke detection framework based on the estimated blending parameter of the image formation model and the smoke component is described in Sect. 4. We generated synthesized smoke images and also used real smoke video to validate the proposed models. Experimental results based on both real and synthetic data are presented in Sect. 5 along with discussions. Some conclusions and perspectives on vision-based smoke detection research are presented in Sect. 6.

2 Related Work

First, we review existing works on vision-based smoke detection and highlight the novelty of the proposed approach. In order to provide further context for the approach we have taken in this paper we also review some representative image separation techniques.

Fig. 2 Background subtraction (**c**) between a smoke image (**b**) and its background (**a**). The background information is clearly noticed in the difference image

2.1 Methods for Vision-Based Smoke Detection

Existing smoke detection methods mainly follow the traditional pattern recognition paradigm and characterize smoke by using visual signatures of smoke regions such as motion, color, edge and texture. For instance, an accumulative motion model has been proposed to capture the motion characteristics of smoke (Yuan 2008). However, the assumption that smoke usually drifts upwards makes this model ineffective in a scenario where strong wind is present. Other research efforts have extracted motion features of smoke using optical flow (Kolesov et al. 2010; Yu et al. 2010); the results have been mixed. Recognizing the fact that the color of smoke is usually grayish, Chen et al (2006) extracted chromatic features of smoke according to a set of decision rules. In a similar vein Calderara et al. (2011) selected a reference color model in the RGB color space to represent smoke. Given the image 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 (Toreyin et al. 2005). However, this decrease in high frequency is not unique to smoke covering as occlusion from other opaque objects could also lead to a decrease in high frequency. Local binary pattern (LBP) has been used successfully to capture texture features and was applied to smoke detection (Tian et al. 2011). Recent progress in smoke detection research points at two emerging trends. First, increased attention is now being directed at characterizing the nature of smoke. For example, the fractal (Maruta et al. 2010) and transmission (Long et al. 2010) property of smoke have been employed to detect smoke. Second, techniques based on feature fusion are being reported. In a representative work (Tung and Kim 2011), motion, surface roughness and area randomness information of smoke were all included in the feature vector.

It is important to note that the different types of features investigated in the literature are usually extracted directly from the original images captured by a stationary camera. A distinguishing point about our proposed approach is that the textural features of smoke are extracted from pure smoke separated from the background in an image sequence.

2.2 Methods for Image Separation

A well-known image separation method is the independent component analysis (ICA) (Bell and Sejnowski 1995) in which a fundamental assumption is that the source images are statistically independent. Some representative applications of ICA-based separation include: separating lighting and reflections (Farid and Adelson 1999), separating artefact in astrophysical images (Funaro et al 2003), separating real-life, nonlinear mixtures of documents acquired through scanning (Almeida 2005), separating reflective and fluorescent components in images (Zhang and Sato 2011). Recently, Minh and Wiskott (2011) relaxed the statistical independence assumption by using so-called slow feature analysis and decorrelation filtering. Generally, these separation methods require multiple mixture observations of source images (e.g. multiple mixtures of the same source images under different illuminations). In video-based smoke detection, the aim is to determine whether there is smoke in the current image frame. Even if smoke does exist in the current image frame, only one mixture observation of the background and pure smoke is available. Hence these methods are of limited use in our case.

We present a quick survey of other representative methods of image separation below. The decomposition of images into cartoon (i.e. piece-wise smooth) and texture layers was addressed by Meyer et al. (2002), Osher et al. (2003), and Starck et al. (2005). Specifically, the wavelet-packet transform (Meyer et al. 2002), total variation minimization (Osher et al. 2003), and both the variational and the sparsity mechanisms (Starck et al. 2005) were employed respectively to solve this problem. The structure of the problem addressed and the solution are not amenable to the current problem where pure smoke is to be separated from background. The background can be either piece-wise smooth or textural.

Tonazzini et al. (2006) formulated blind image separation as a Bayesian estimation problem and proposed an expectation-maximization algorithm with the mean field approximation to solve the problem. Guo and Garland (2006) proposed to use a direct minimization of an entropy-like function to solve the separation problem. In addition, Guidara et al. (2009) proposed a maximum likelihood approach to blind source separation using non-symmetrical half-plane Markov random fields. Based on sparse representation of signals, morphological component analysis was proposed to perform image separation by Fadili et al. (2010). However, these methods inherently need multiple observations of image mixtures which are not available in our case.

There are reported results on efforts to separate transparent layers due to reflection in images. However, transparency induced by reflection does not share common properties with pure smoke which is largely textured. We review some of the results here. Szeliski et al. (2000) used constrained least squares to recover the layer images. However, this method has a limitation that each layer must have a fixed transparency. Schechner et al. (2000) exploited focus difference between the background scene and reflected scene as a cue for separating each layer. Such focus difference between background and pure smoke does not exist in the current problem. Levin et al. (2004) proposed a separation method wherein the total amount of edges and corners is minimized. However, this method may not work well on the images including many intersections of edges from different layers. Sarel and Irani (2005) employed a global-to-local space-time alignment approach to detecting and aligning the repetitive behavior which was assumed as one of two transparent layers. After a median operator was applied to space-time derivatives, two transparent layers could be separated. Levin and Weiss (2007) incorporated user input into the sparsity prior of image gradients to separate reflection from a single image. In the smoke detection problem, an automatic operation is an essential requirement; methods requiring user inputs are unsuitable. The separation of multiple layers with unknown spatial shifts and varying mixing coefficients was addressed by Gai et al. (2008). However, this method is limited to uniform translations. Kong et al. (2011) regarded the exclusiveness of the image gradients of the background layer and the reflection layer as a cue to separate reflection. In our case, this assumption about the image gradients may not always be true as many background scenes share similar image gradients with pure smoke (e.g. uniform walls and homogeneous smoke).

3 Separation of the Smoke Component

In this section, we first present a formulation of the problem by assuming an image is a linear combination of background and a smoke component. Three different models that constrain the smoke component are then introduced. Solutions to separate the smoke component from a given video frame (that may or may not be covered by smoke) and its estimated background are developed under the three constraints, respectively.

3.1 Problem Formulation

In the early stage of fire in a given monitored area, smoke will generally cover a very small area. An important problem is how to achieve early vision-based detection. Furthermore, the utility of the detection increases significantly if the source of the smoke can also be localised. To achieve early detection and localization of smoke, a video frame is divided into overlapped or non-overlapped small-sized image blocks. Conceptually, the problem reduces to that of determining if an image block is covered by smoke. This paper focuses on this problem. In the following, we discuss the physics of the linear image formation model adopted in the paper and provide an analytical formulation.

Let $\mathbf{f}_t \in \mathbb{R}^N$ be an image block with N pixels at time t. According to the two fundamental atmospheric scattering models (the attenuation and airlight models) (Narasimhan and Nayar 2002), smoke, if existent, would serve as a medium to attenuate the light reflected from the background before it reaches the camera due to scattering. Meanwhile, the smoke will generate airlight, a mechanism that causes the atmosphere (i.e. smoke here) to behave like a source of light through scattering as well. Therefore, $\mathbf{f_t}$ is determined by the attenuation model and the airlight model. Assuming that there are no specific point sources of light and the scattering coefficient of the smoke does not change appreciably within the visible wavelength, $\mathbf{f}_t$ can be modeled as a linear blending of $\mathbf{s}_t$ and $\mathbf{b}_t$,[1] such that,

$$\mathbf{f}_t = \alpha_t \mathbf{s}_t + (1 - \alpha_t)\mathbf{b}_t + \mathbf{n}_t \tag{1}$$

where $\mathbf{n}_t \in \mathbb{R}^N$ represents modeling noise, $\mathbf{b}_t \in \mathbb{R}^N$ represents the background under clear air (or as if no smoke exists), and $\mathbf{s}_t \in \mathbb{R}^N$ is the airlight (scattering) component by the smoke of infinite thickness. $\alpha_t \in [0, 1]$ is the blending weight at time t. It depends on the smoke scattering coefficient and the thickness of the smoke along the line of sight. Within a small image block, it is assumed that the thickness of smoke is constant. Therefore, α_t is a constant as well within the image block. However, it may vary from block to block.

[1] Notice that in Narasimhan and Nayar (2002), the attenuating medium, such as fog, is assumed to occupy the entire space between the scene and the camera. For smoke, this is usually not the case. Smoke often appears at a certain distance and is of limited thickness along the line of sight.

For brevity and without loss of generality, the subscript t will be dropped in the rest of the paper.

Assuming $\mathbf{f}$ is acquired by a stationary camera, background modeling techniques such as Gaussian mixture model (GMM) (Stauffer and Grimson 2000) can be adopted to obtain an approximation of $\mathbf{b}$. Thus, the problem can be formulated as the estimation of α and $\mathbf{s}$ given $\mathbf{f}$ and $\mathbf{b}$ by minimizing the residual noise:

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

Equation (2) is under-determined because there are N equations but $N+1$ free variables and will have infinite number of solutions. There is hope to obtain a unique solution by constraining either $\mathbf{s}$ or $\mathbf{b}$ or both. Considering that the background $\mathbf{b}$ can vary significantly from one application to another and smoke has relatively consistent visual characteristics, we introduce three different models for the smoke component $\mathbf{s}$, namely, the local smoothness, principal component and sparse representation. We then present the corresponding solutions to Eq. (2).

3.2 Local Smoothness Model

The local smoothness model is based on the observation that a small-sized pure smoke image bock $\mathbf{s}$ is generally smooth. In other words, a pixel in a pure smoke image block is likely to have a similar intensity or color to its neighboring pixels. This observation naturally leads to the following optimization problem:

$$\min_{\alpha,\mathbf{s}} \|\mathbf{f} - \alpha\mathbf{s} - (1-\alpha)\mathbf{b}\|_2^2 + \lambda \sum_{i=1}^{N} \sum_{j\in\Omega_i} (s_i - s_j)^2$$
$$s.t. \quad \alpha \in [0,1] \tag{3}$$

where s_i is the ith element of $\mathbf{s}$, Ω_i is a small neighborhood centered at the ith pixel, and λ is a weighting parameter which trades off the residual error and the smoothing constraint expressed by the second term. By defining a matrix $\mathbf{T} \in \{-1, 0, 1\}^{M\times N}$, we can rewrite the second term as:

$$\sum_{i=1}^{N} \sum_{j\in\Omega_i} (s_i - s_j)^2 = \|\mathbf{T}\mathbf{s}\|_2^2 = \mathbf{s}^T\mathbf{T}^T\mathbf{T}\mathbf{s} = \mathbf{s}^T\mathbf{A}\mathbf{s} \tag{4}$$

where $\mathbf{A}$ is defined as $\mathbf{T}^T\mathbf{T}$. Also, the number, M, of rows in $\mathbf{T}$ is determined by the size of neighborhood Ω. Substituting Eq. (4) into Eq. (3) yields

$$\min_{\alpha,\mathbf{s}} \|\mathbf{f} - \alpha\mathbf{s} - (1-\alpha)\mathbf{b}\|_2^2 + \lambda\mathbf{s}^T\mathbf{A}\mathbf{s} \quad s.t. \quad \alpha \in [0,1] \tag{5}$$

For Eq. (5), we propose to solve for α and $\mathbf{s}$ alternately. First, we solve for $\mathbf{s}$ by fixing α. In this case, Eq. (5) is a quadratic function of $\mathbf{s}$. An analytical solution $\hat{\mathbf{s}}$ with respect to α can be derived. That is,

$$\hat{\mathbf{s}} = (\alpha^2\mathbf{I} + \lambda\mathbf{A})^{-1}\alpha(\mathbf{f} - \mathbf{b} + \alpha\mathbf{b}) \tag{6}$$

where $\mathbf{I} \in \mathbb{R}^{N\times N}$ denotes the identity matrix.

Next, we solve for α by fixing $\mathbf{s}$. In this case, Eq. (5) is a quadratic function of α. The minimizer of the quadratic function of α is:

$$\alpha^* = \frac{(\mathbf{b}-\mathbf{s})^T(\mathbf{f}-\mathbf{b})}{(\mathbf{b}-\mathbf{s})^T(\mathbf{s}-\mathbf{b})} \tag{7}$$

Considering the constraint $\alpha \in [0,1]$, the current solution $\hat{\alpha}$ with respect to $\mathbf{s}$ can be obtained as follows:

$$\hat{\alpha} = \begin{cases} 0 & \text{if } \alpha^* \le 0 \\ \alpha^* & \text{if } 0 < \alpha^* < 1 \\ 1 & \text{if } \alpha^* \ge 1. \end{cases} \tag{8}$$

The alternating optimization process is carried out until $\mathbf{s}$ and α converge or the number of iterations reaches a predefined value.

3.3 Principal Component Model

The local smoothness model would fail to distinguish smoke from other objects whose surfaces bear similar property of smoothness. Considering each image block with N pixels as a point in an N-dimensional space, pure smoke images, being similar in overall textural configuration, are likely to lie in a low-dimensional subspace. If this subspace is located, it could well describe pure smoke images. In this paper, the widely used principal component analysis (PCA) (Turk and Pentland 1991) is employed to locate the subspace of pure smoke images. Specifically, given a set of pure smoke images, an $N \times N$ covariance matrix is computed, and its eigenvectors and eigenvalues are obtained. The eigenvalue represents the variance of pure smoke images along the corresponding eigenvector. If the eigenvectors are ranked according to the magnitudes of their corresponding eigenvalues, a subset of the eigenvectors with large eigenvalues can be selected to form the subspace of pure smoke.

Let $\mathbf{P} \in \mathbb{R}^{N\times L}(L < N)$ be a matrix, where L is the dimension of the obtained subspace. Each column of $\mathbf{P}$ is an eigenvector chosen according to the aforementioned criterion. Then a pure smoke image $\mathbf{s}$ can be expressed as

$$\mathbf{s} = \mathbf{P}\mathbf{y} \tag{9}$$

where $\mathbf{y} \in \mathbb{R}^L$ is the coefficient vector of projecting $\mathbf{s}$ onto the subspace $\mathbf{P}$ of pure smoke. Substituting Eq. (9) into Eq. (2) yields

$$\min_{\alpha,\mathbf{y}} \|\mathbf{f} - \alpha\mathbf{P}\mathbf{y} - (1-\alpha)\mathbf{b}\|_2^2 \quad s.t. \quad \alpha \in [0,1] \tag{10}$$

Notice that Eq. (10) is a quadratic function of α (or $\mathbf{y}$) when $\mathbf{y}$ (or α) is fixed. We can solve for α and $\mathbf{y}$ alternately as well, and then reconstruct $\mathbf{s}$ through Eq. (9).

Specifically, let $\hat{\mathbf{s}}$ be the current solution for $\mathbf{s}$ when α is fixed, then we have:

$$\hat{\mathbf{s}} = \mathbf{P}(\alpha \mathbf{P}^T \mathbf{P})^{-1} \mathbf{P}^T (\mathbf{f} - \mathbf{b} + \alpha \mathbf{b}) \tag{11}$$

By fixing $\mathbf{y}$, the current solution $\hat{\alpha}$ for α can be expressed as Eq. (8) as well, where α^* is the minimizer of the quadratic function of α and can be obtained as

$$\alpha^* = \frac{(\mathbf{b} - \mathbf{P}\mathbf{y})^T (\mathbf{f} - \mathbf{b})}{(\mathbf{b} - \mathbf{P}\mathbf{y})^T (\mathbf{P}\mathbf{y} - \mathbf{b})}. \tag{12}$$

3.4 Sparse Representation Model

A single linear subspace obtained by PCA may not be sufficient to describe all possible variations of smoke since they may lie in multiple low-dimensional subspaces. According to the theory of sparse representation (Wright et al. 2010), if sample smoke images can be collected or generated to capture the distribution of pure smoke images, it is expected that any specific pure smoke image would have a sparse representation with respect to these samples. Such collection of samples represent a dictionary and each sample in the dictionary is typically referred to as a basis. Following this intuition, we propose a sparse model which is expected to offer more robust representation of smoke.

Let $\mathbf{D} \in \mathbb{R}^{N \times J} (N \ll J)$ be a dictionary for pure smoke and each column of $\mathbf{D}$ is a basis. Then a smoke image $\mathbf{s}$ is expected to be sparse in $\mathbf{D}$;

$$\mathbf{s} = \mathbf{D}\mathbf{x}, \tag{13}$$

where $\mathbf{x} \in \mathbb{R}^J$ is the sparse coefficient vector and many of its elements are expected to be zero or close to zero (i.e. sparse). The coefficient vector encodes infomation about which bases and the proportion thereof contribute to the construction of $\mathbf{s}$ from $\mathbf{D}$. Based on Eq. (13) and the sparseness conditions, Eq. (2) can be rewritten as follows:

$$\min_{\alpha, \mathbf{x}} \|\mathbf{f} - \alpha \mathbf{D}\mathbf{x} - (1-\alpha)\mathbf{b}\|_2^2 + \eta \|\mathbf{x}\|_0 \quad s.t. \alpha \in [0, 1] \tag{14}$$

where η is a regularization parameter balancing the residual error term and the sparseness of $\mathbf{x}$. Since the ℓ_0-norm is non-convex, we follow the common trick in the literature by replacing it with the ℓ_1-norm to make the optimization problem solvable. Thus,

$$\min_{\alpha, \mathbf{x}} \|\mathbf{f} - \alpha \mathbf{D}\mathbf{x} - (1-\alpha)\mathbf{b}\|_2^2 + \eta \|\mathbf{x}\|_1 \ s.t. \ \alpha \in [0, 1] \tag{15}$$

Similarly to the local smoothness and principal component models, an optimal $\mathbf{x}$ (or α) can be obtained by alternately fixing α (or $\mathbf{x}$). Let $\hat{\alpha}$ be the current solution for α when $\mathbf{x}$ is fixed. Again, it can be expressed as Eq. (8), with α^*

$$\alpha^* = \frac{(\mathbf{b} - \mathbf{D}\mathbf{x})^T (\mathbf{f} - \mathbf{b})}{(\mathbf{b} - \mathbf{D}\mathbf{x})^T (\mathbf{D}\mathbf{x} - \mathbf{b})}. \tag{16}$$

When α is fixed, Eq. (15) becomes an ℓ_1-regularized least squares problem:

$$\min_{\mathbf{x}} \|\mathbf{f} - \alpha \mathbf{D}\mathbf{x} - (1-\alpha)\mathbf{b}\|_2^2 + \eta \|\mathbf{x}\|_1 \tag{17}$$

To solve for $\mathbf{x}$ efficiently, the feature-sign search algorithm (Lee et al. 2007) is adopted in this paper. The main idea of this algorithm is to preserve an active set of potential nonzero entries in $\mathbf{x}$ and their corresponding signs. Specifically, the algorithm proceeds in a series of "feature-sign steps" to search for the optimal active set and coefficient signs.

Suppose $\mathbf{g} \in \mathbb{R}^J$ and its ith entry $g_i \in \{-1, 0, 1\}$ denotes the sign of x_i, which is the ith entry in $\mathbf{x}$. Given the current active set and signs, let $\bar{\mathbf{D}}$ be a submatrix of $\mathbf{D}$ that contains only the columns corresponding to the active set. Meanwhile, let $\bar{\mathbf{x}}$ and $\bar{\mathbf{g}}$ be subvectors of $\mathbf{x}$ and $\mathbf{g}$ corresponding to the active set. With the active set being considered only, Eq. (17) reduces to the following quadratic optimization problem:

$$\min_{\bar{\mathbf{x}}} \|\mathbf{f} - \alpha \bar{\mathbf{D}}\bar{\mathbf{x}} - (1-\alpha)\mathbf{b}\|_2^2 + \eta \bar{\mathbf{g}}^T \bar{\mathbf{x}} \tag{18}$$

A solution to this problem can be analytically obtained as:

$$\hat{\bar{\mathbf{x}}} = (\alpha^2 \bar{\mathbf{D}}^T \bar{\mathbf{D}})^{-1} [\alpha \bar{\mathbf{D}}^T (\mathbf{f} - \mathbf{b} + \alpha \mathbf{b}) - \eta \bar{\mathbf{g}}/2] \tag{19}$$

The objective function value at $\hat{\bar{\mathbf{x}}}$ and all points where any coefficient changes sign are further checked. Then $\bar{\mathbf{x}}$ is updated to the point with the lowest objective function value. Accordingly, the solution to Eq. (17) is updated. Once the updated solution satisfies the following optimality conditions, the optimal $\mathbf{x}$ will be obtained.

$$\begin{aligned} &\frac{\partial \|\mathbf{f} - \alpha \mathbf{D}\mathbf{x} - (1-\alpha)\mathbf{b}\|_2^2}{\partial x_i} + \eta g_i = 0 \ \forall x_i \neq 0 \\ &\left| \frac{\partial \|\mathbf{f} - \alpha \mathbf{D}\mathbf{x} - (1-\alpha)\mathbf{b}\|_2^2}{\partial x_i} \right| \leq \eta \quad \forall x_i = 0 \end{aligned} \tag{20}$$

Using the optimal $\mathbf{x}$, $\mathbf{s}$ can be calculated by Eq. (13).

3.5 Discussions

We note that similar image model described by Eq. (1) was used in the formulation of image matting (Wang and Cohen 2005; Levin et al. 2008; Bai and Sapiro 2009). Typically to deal with the problem of image matting, user interactions are required to generate either a trimap labeling each pixel as foreground, background or unknown, or some scribbles indicating background and foreground pixels. However, our methods do not require any user interaction and automatically extract the smoke component from the mixture of background and pure smoke. In the single-image haze removal problem (He et al. 2011), a model similar to Eq. (1) was also used. A dark channel prior was assumed for outdoor haze-free images and the work aimed to restore high quality haze-free or haze-reduced images. Removal of haze does not

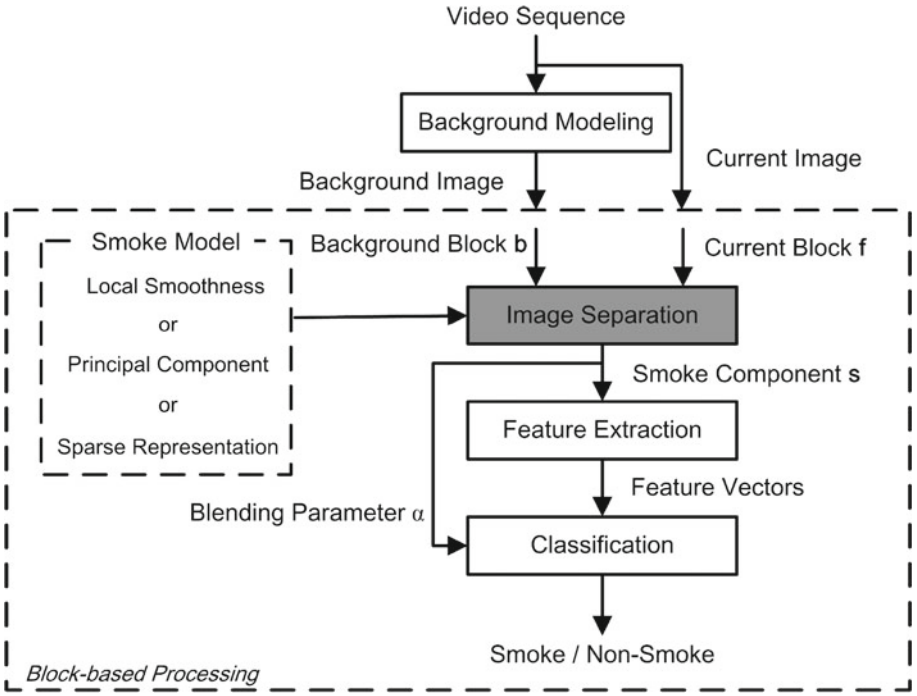

Fig. 3 Proposed framework for video-based smoke detection

necessarily require accurate or reliable separation of the haze component. However, in the present work, the prior knowledge (i.e. local smoothness, principal component, and sparse representation) is assumed and our purpose is to extract the smoke component, if any, as reliably as possible for further classification.

In practical applications, the input image **f** may be covered by nothing, smoke or other opaque objects. When **f** is not covered by anything, α is expected to be extremely small or close to zero. In the case that **f** is covered by heavy smoke or an opaque object, α would be large or close to 1.0. However, the separated "smoke component" is expected to have different nature. When covered by heavy smoke, this represents the actual smoke component; when covered by an opaque object, it would be a low-pass filtered version of the object due to the imposed constraints on **s**. This indicates that further feature extraction and classification are required on **s** for reliable smoke detection, the subject of discussion in the next section.

4 Smoke Detection Framework

Based on the image separation techniques presented in Sect. 3, a novel block-based smoke detection framework is proposed and depicted in Fig. 3. Given a video sequence, background modeling is performed. The current image and its estimated background image are divided into blocks. The division can be either overlapped or non-overlapped. In this paper, the non-overlapped division is adopted. For each frame

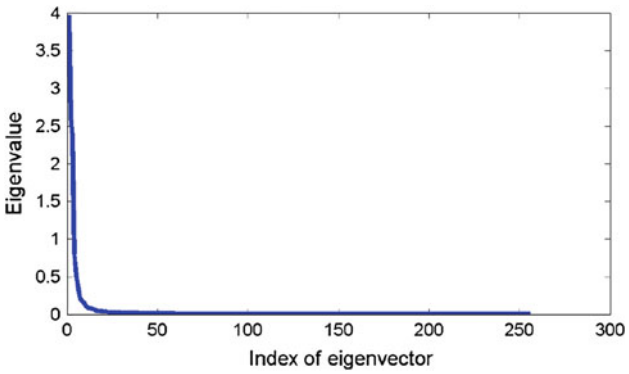

Fig. 4 Principal component analysis on pure smoke images

block **f** and its associated background **b**, the blending parameter α and smoke component **s** are computed using one of the separation techniques presented in Sect. 3. To detect whether **f** is covered by smoke or not, the estimated α value is first checked. If α is less than a very small threshold (or nearly zero), **f** is deemed not to be covered by smoke or anything else and, hence, is classified as non-smoke. In the cases that α is greater than the threshold, the block is considered to be covered by a foreground object, either smoke or non-smoke. The separated component **s** is then passed on to the next step in which LBP feature is extracted from **s** and the extracted feature is input to a binary support vector machine (SVM) classifier. A decision is made on whether **s** is smoke or non-smoke. Notice that accurate detection of smoke at block level is essentially required for early detection since the area that is covered by smoke at early stage is usually very small.

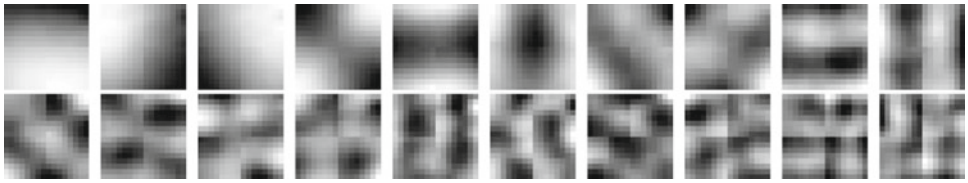

Fig. 5 The eigenvectors corresponding to the top 20 largest eigenvalues shown as 16 × 16 elemental patches

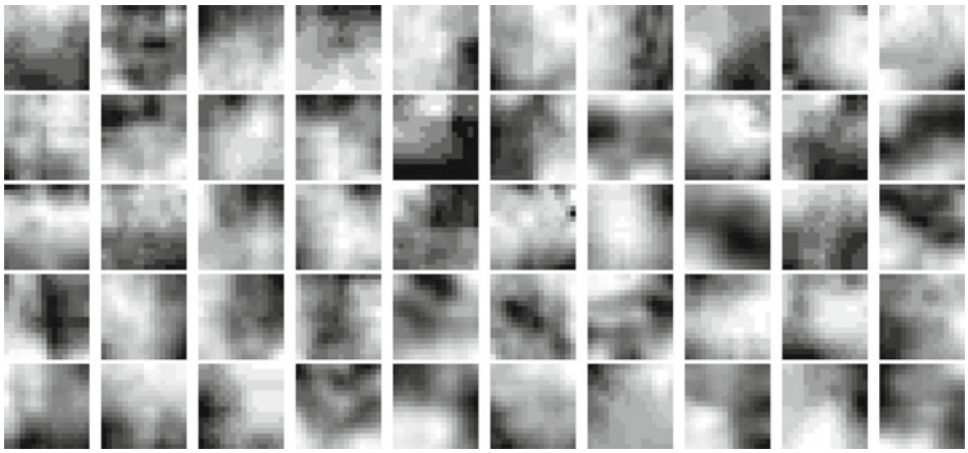

Fig. 6 50 basis samples, shown as 16 × 16 elemental patches, out of the 500 bases in an over-complete dictionary

5 Experimental Results and Discussions

Extensive experiments were conducted using both simulated and real video to evaluate the efficacy of the smoke models and the associated separation algorithms. In this section, results on smoke modeling, smoke separation on both simulated and real video data and smoke detection on real video sequences are presented. Discussions are also provided along with the results.

5.1 Smoke Component Modeling

According to Sect. 3, the principal component model for the smoke component requires the subspace **P** of pure smoke. Similarly, an over-complete dictionary **D** is necessary for the sparse representation model. We present here how **P** and **D** were constructed from training samples.

5.1.1 Learning the Subspace **P** *of Pure Smoke*

In order to separate the smoke component based on PCA, the subspace of pure smoke should be learned first. One thousand (1000) samples of pure smoke were collected to train the subspace of pure smoke. Each sample is 16 × 16 pixels and L (i.e. the number of columns in **P**) was chosen to be much less than 256 ($L \ll 256$).

The eigenvalues obtained from the application of PCA on the training samples were ranked (maximum to minimum) and shown in Fig. 4. A quick analysis indicated that the 20th largest eigenvalue was approximately 1 % of the largest eigenvalue. Thus only the eigenvectors corresponding to the 20 largest eigenvalues contributed to the construction of the subspace **P**. The selected eigenvectors are shown in Fig. 5.

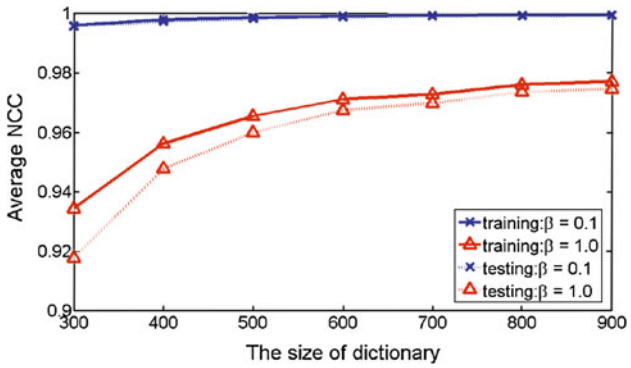

Fig. 7 Average NCC between the reconstructed and actual pure smoke images

5.1.2 Learning Dictionary **D**

The one thousand (1000) pure smoke samples used to learn the subspace of pure smoke were also used for training the dictionary. Additionally, a separate 1000 pure smoke images were collected to test the effectiveness of the learned dictionary. Each sample is 16 × 16 pixels. Therefore, the number of bases in an over-complete dictionary should be much larger than 256.

The widely-used K-SVD (Aharon et al. 2006) was adopted to train an over-complete dictionary of size 500 from the training samples. Fifty (50) bases from the trained dictionary are shown in Fig. 6. The normalized cross correlation (NCC)

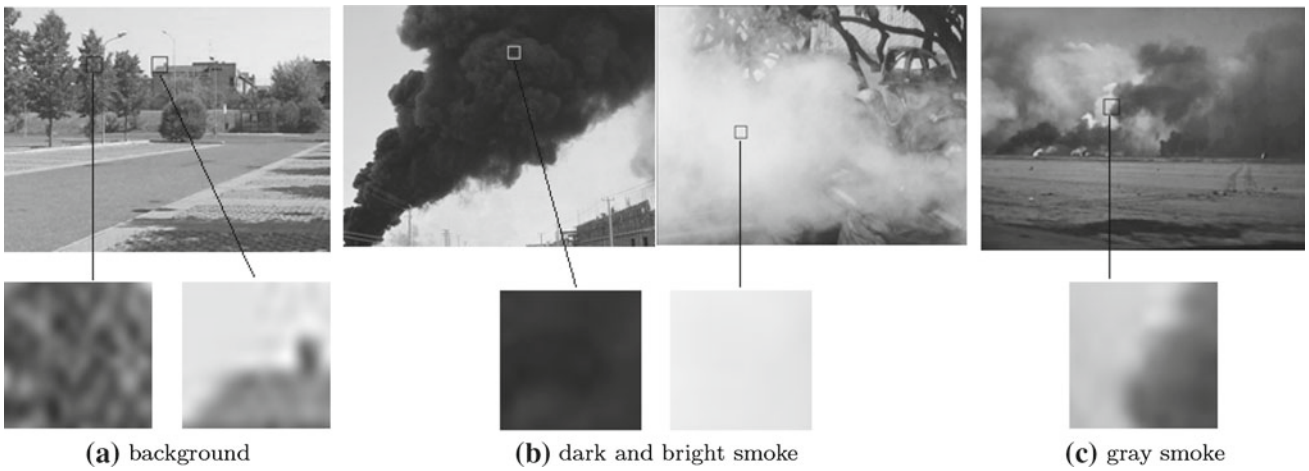

Fig. 8 Background and pure smoke samples used to synthesize smoke images with different blending parameters. **(a)** Texture and piece-wise smooth background images; **(b)** dark and bright smoke with small variation among the pixel values; **(c)** gray smoke with large variation among the pixel values

(Gonzalez and Woods 2007) between the training/test samples (ground truth) and the reconstructed samples from the dictionary was used to measure the effectiveness of the dictionary. Specifically, for a training/test sample $\mathbf{s}_{ori}$, an estimated coefficient vector $\mathbf{x}_{est}$ was obtained by solving the following optimization problem:

$$\mathbf{x}_{est} = \text{argmin}_{\mathbf{x}} \|\mathbf{s}_{ori} - \mathbf{D}\mathbf{x}\|^2 + \beta\|\mathbf{x}\|_1 \quad (21)$$

where β is a constant. The reconstructed version of the smoke image $\mathbf{s}_{ori}$ was obtained as $\mathbf{s}_{rec} = \mathbf{D}\mathbf{x}_{est}$. The graphs shown in Fig. 7 depict the average NCC values for the training and test samples, at different β values, versus the size of dictionary (i.e. the number of bases in the dictionary). As can be seen, both the training and test samples were well reconstructed from the learned dictionary regardless of β values. As expected, better reconstruction was achieved with an increased dictionary size. Increasing the value of β encourages sparseness, but it also sacrifices the reconstruction accuracy.

5.2 Performance of Smoke Separation on Synthesized Images

In this section, the performance of the proposed methods on synthesized images is reported. In particular, the methods were evaluated on noise-free and noisy background **b**, and smoke and non-smoke input images **f**.

5.2.1 Data Sets

In order to synthesize realistic smoke image blocks (sized 16×16), we used two prototypical images each for the background and smoke components; all of them were extracted from real video.The two prototypical background images are respectively, textural and piece-wise smooth. For pure smoke, one type appears dark or bright with small variation among the pixel values and the other is gray with large variation among the pixel values. Figure 8a shows one sample for each type of background images and Fig. 8b and c shows some samples of pure smoke.

Based on these background and pure smoke image samples, 60 images were synthesized using six different combinations of background and pure smoke to construct a test database of smoke images, which is shown in Fig. 9. For each combination, α value ranged from 0.1 to 1.0 with an increment of 0.1.

5.2.2 Using Ground Truth Background as **b**

In the experiments reported in this subsection, each synthesized image in Fig. 9 was considered as an input image **f**, and the corresponding background image in Fig. 8a was considered as **b**. In other words, we have perfect background images. Given the learned subspace **P** or over-complete dictionary **D** for pure smoke if needed, the blending parameter α and smoke component **s** were obtained by solving the optimization problems of Eq. (5), (10), and (15) respectively. To evaluate the accuracy of the estimation of α, the absolute difference between the actual α (ground truth) and the estimated α was adopted as an indicator. For each combination of background and pure smoke, the average absolute differences calculated using different models for the smoke component are reported in Table 1. Note that the average absolute difference between the actual and estimated α is rather small. There is an exception in the case where local smoothness was applied to the separation of gray smoke with large vari-

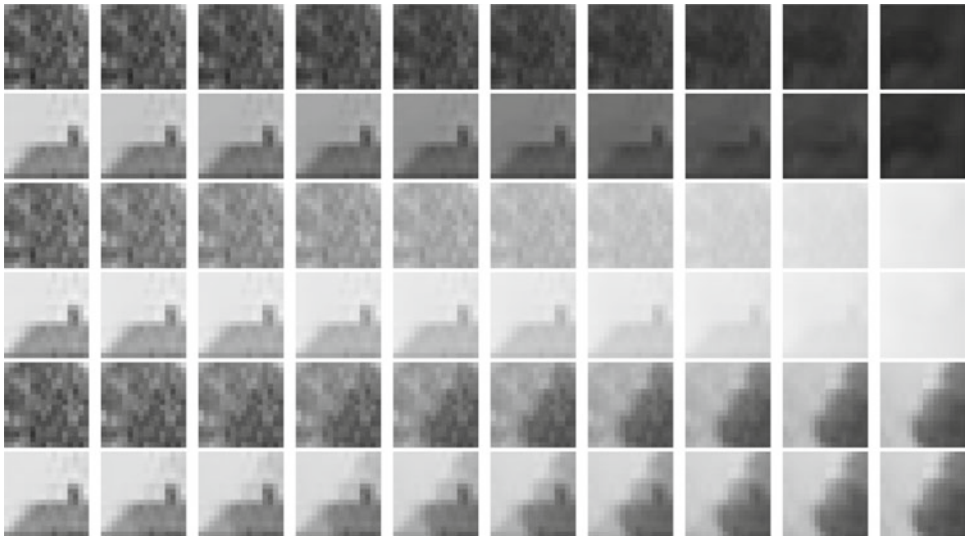

Fig. 9 The test database of synthesized smoke images (From the *top down*, the images in each row were synthesized from (1) texture background and dark smoke, (2) piece-wise smooth background and dark smoke, (3) texture background and bright smoke, (4) piece-wise smooth background and bright smoke, (5) texture background and gray smoke, and (6) piece-wise smooth background and gray smoke respectively)

Table 1 Average absolute differences between the estimated α and the actual α for different combinations of background and pure smoke using different models for the smoke component

	Local smoothness	PCA	Sparse representation
Texture background and dark smoke	0.0032 ± 0.0021	0.0024 ± 0.0015	0.0022 ± 0.0019
Piece-wise smooth background and dark smoke	0.0020 ± 0.0015	0.0011 ± 0.0011	0.0019 ± 0.0015
Texture background and bright smoke	0.0021 ± 0.0016	0.0094 ± 0.0055	0.0023 ± 0.0017
Piece-wise smooth background and bright smoke	0.0022 ± 0.0010	0.0089 ± 0.0042	0.0018 ± 0.0014
Texture background and gray smoke	0.1378 ± 0.0884	0.0027 ± 0.0022	0.0337 ± 0.0266
Piece-wise smooth background and gray smoke	0.0971 ± 0.0511	0.0015 ± 0.0013	0.0062 ± 0.0027

ation among the pixel values. This is likely because the local smoothness assumption does not hold well for smoke with large variation.

To evaluate the performance of the smoke component estimation, NCC between the estimated and true smoke components was computed and plotted against varying blending parameters and for different combinations of the background and smoke component (see Fig. 10). In general a high NCC value, indicative of good separation performance, is achieved from a relatively low α value onwards. Nevertheless, the sparse representation has led to the best separation performance (highest NCC value) among all three models. Compared with the principal component model for the smoke component, the local smoothness is more effective when smoke is dense enough.

Considering different combinations of background and pure smoke with small variation, it can be noted that, regardless of which smoke model is used, when the blending parameter increases, the separated smoke component becomes more similar to the ground truth with increasing and high NCC values. Similar consideration for different combinations of background and pure smoke with large variation indicates that this characteristic is still observed when local smoothness assumption is used in the smoke component model. However, when either the principal component model or the sparse representation model is used, the NCC values are relatively stable with respect to the varying blending parameter. This suggests that the principal component and sparse representation models are relatively insensitive to different blending parameters when separating pure smoke with large variation.

5.2.3 Using Noisy Background as **b**

In real scenarios, background images may be noisy for several reasons including imperfect output from background modeling. We evaluate the proposed methods under a simulated noisy background by adding to the true background white Gaussian noise with zero mean and standard deviation ranging from 0 to 15 [additive white Gaussian noise (AWGN) model]. In this evaluative experiment, we conducted image separation on the synthesized images as shown in Fig. 9 using the noisy background images shown in Fig. 11. The average absolute differences of α and the average NCC values were computed and shown in Figs. 12 and 13.

As expected, the performance of the separation methods deteriorated as the noise level (i.e. standard deviation of the AWGN) increases. However, the deterioration was

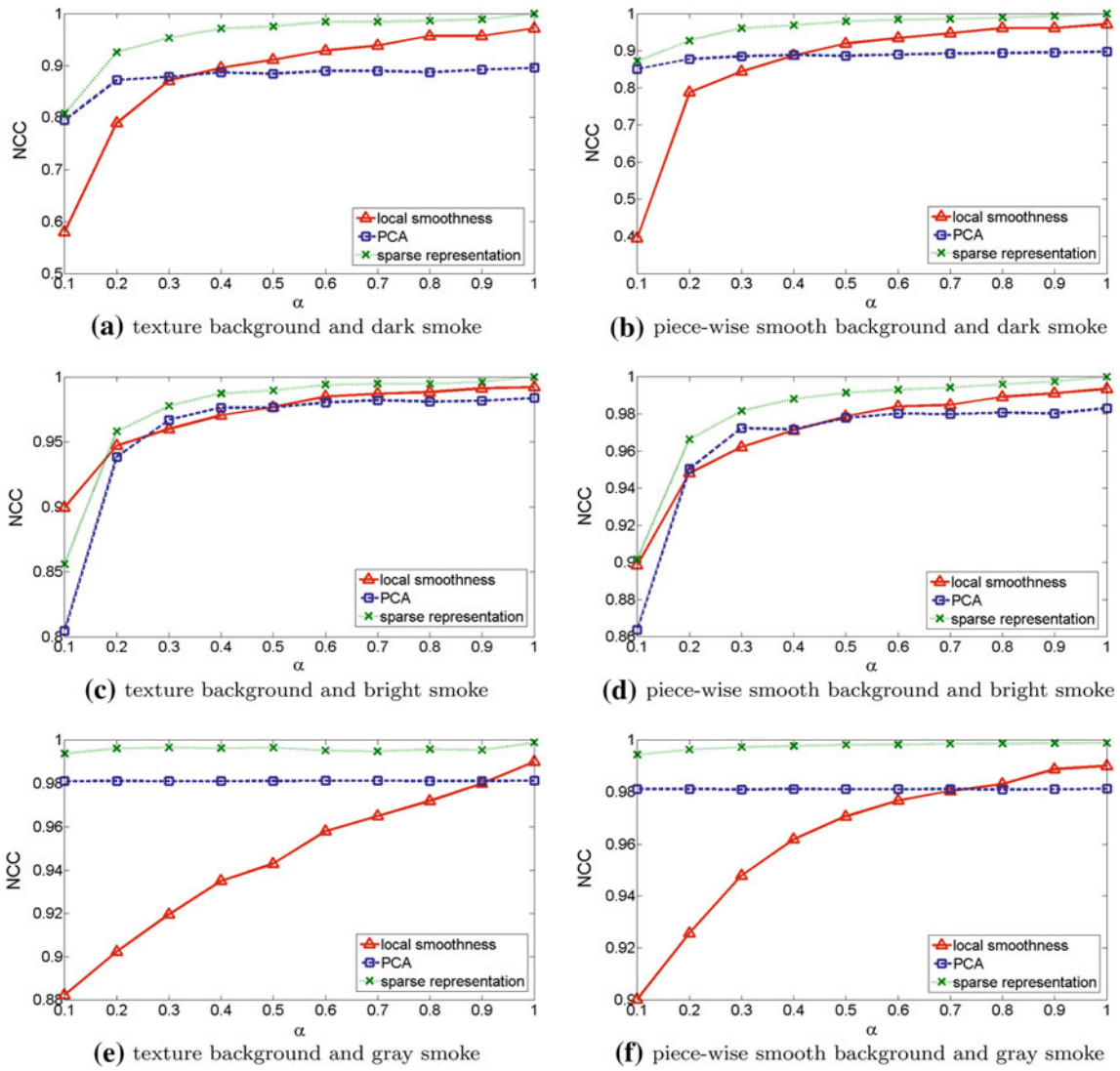

Fig. 10 NCC between the separated smoke component and ground truth for different combinations of background and pure smoke using different models for the smoke component

Fig. 11 Background images with the AWGN (for each type of background, from *left to right*, the standard deviation of the AWGN is 5, 10, and 15 respectively)

slow until the standard deviation reached 5 whereupon there was a marked deterioration. There was an exception in the case where local smoothness model was applied to separating smoke with large variation. The deterioration was large and remained fairly constant (Figs. 12e, f, 13e, f).

We note here, and this will be demonstrated with real videos in Sect. 5.4.2, that the noise level of an estimated background modeled by GMM is usually below a standard deviation of 5. Thus in this case the average difference of α and the average NCC will be relatively constant. It is interesting to note that there is strong visual similarity between two images when their NCC value is larger than 0.3. The trend of the graphs shown in Fig. 12 and 13 suggests that the separation method based on the local smoothness constraint for smoke component has the highest noise tolerance among all the proposed methods.

5.2.4 *Using the Original Background as* **f**

In order to obtain good smoke detection performance, pure smoke must be well separated if indeed smoke exists in the image **f**. However, the question arises as to the nature of the separated "smoke component" when **f** is a non-smoke image;

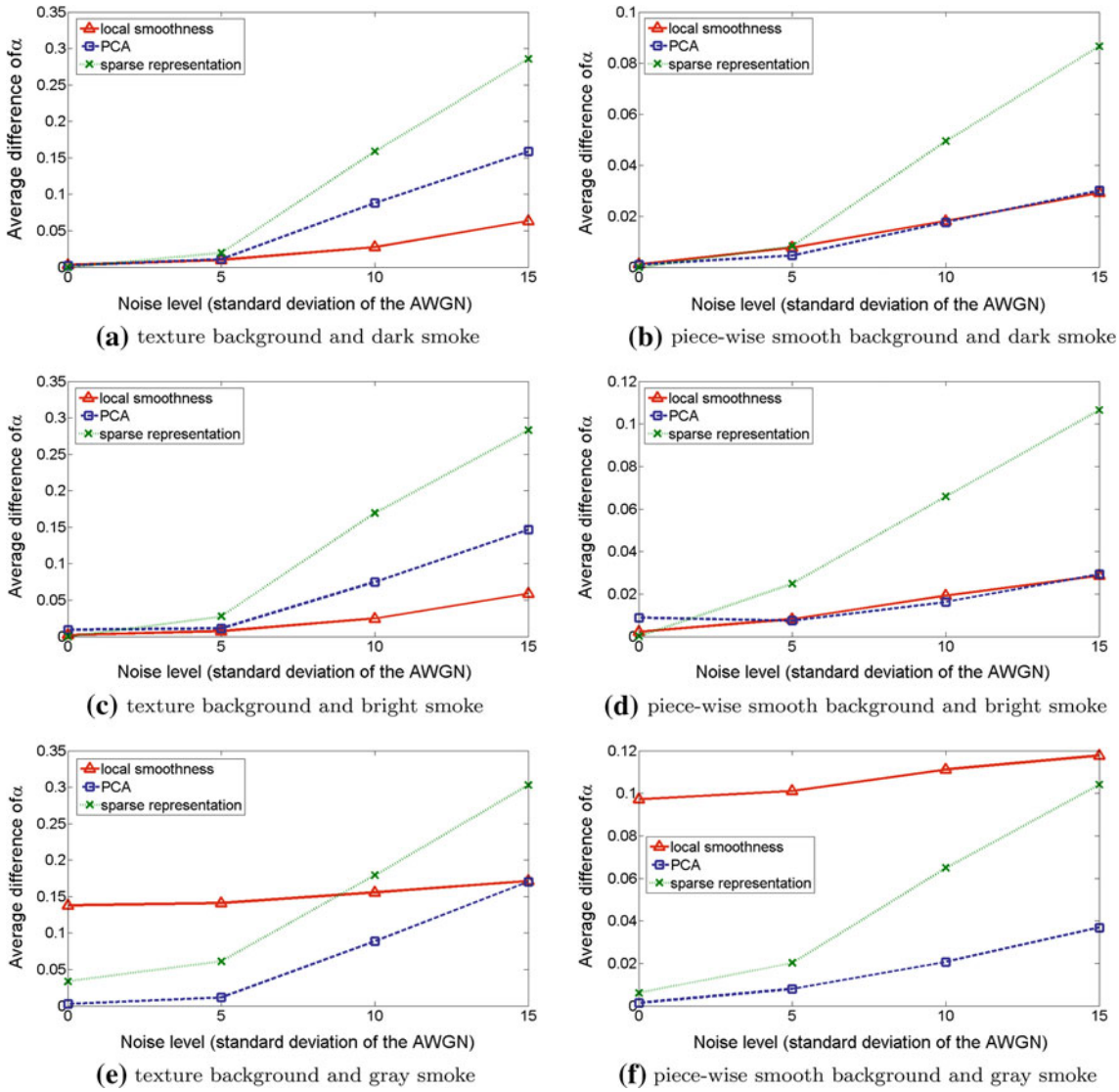

Fig. 12 Average absolute differences of α versus noise levels for different combinations of background and pure smoke using different models for the smoke component

does the separated **s** look like smoke or not? We first proceed by letting **f** be the original background. Hence in these separation experiments, the images in Fig. 8a were considered as both **b** and **f**. The results indicated that, regardless of the different types of background images and separation methods, the estimated α was nearly zero and the separated "smoke component" was almost a homogeneous patch. Although the homogeneous patch is similar to some of the samples of smoke images, it will not be misclassified as smoke because the nearly zero α value will fail the threshold test.

5.2.5 *Using other Foreground Images as* **f**

Next let us assume that **f** represents other foreground images. Here, both textured and piece-wise smooth images are considered and some examples are shown in the second column in Fig. 14. In conjunction with the images shown in the first column in Fig. 14 as **b**, the several separated "smoke components" based on different types of constraints for the smoke component are shown in columns 3–5 in Fig. 14. As expected, regardless of the types of background and foreground images, the separated "smoke component" is similar to the original foreground image if local smoothness or sparse representation is used as the constraint. When local smoothness constraint is used, the separated "smoke component" can be considered as a smooth version of the original foreground. However, when the principal component is employed as the constraint, the separated "smoke component" appears more like pure smoke than the original foreground image. In other words, the use of the principal component constraint

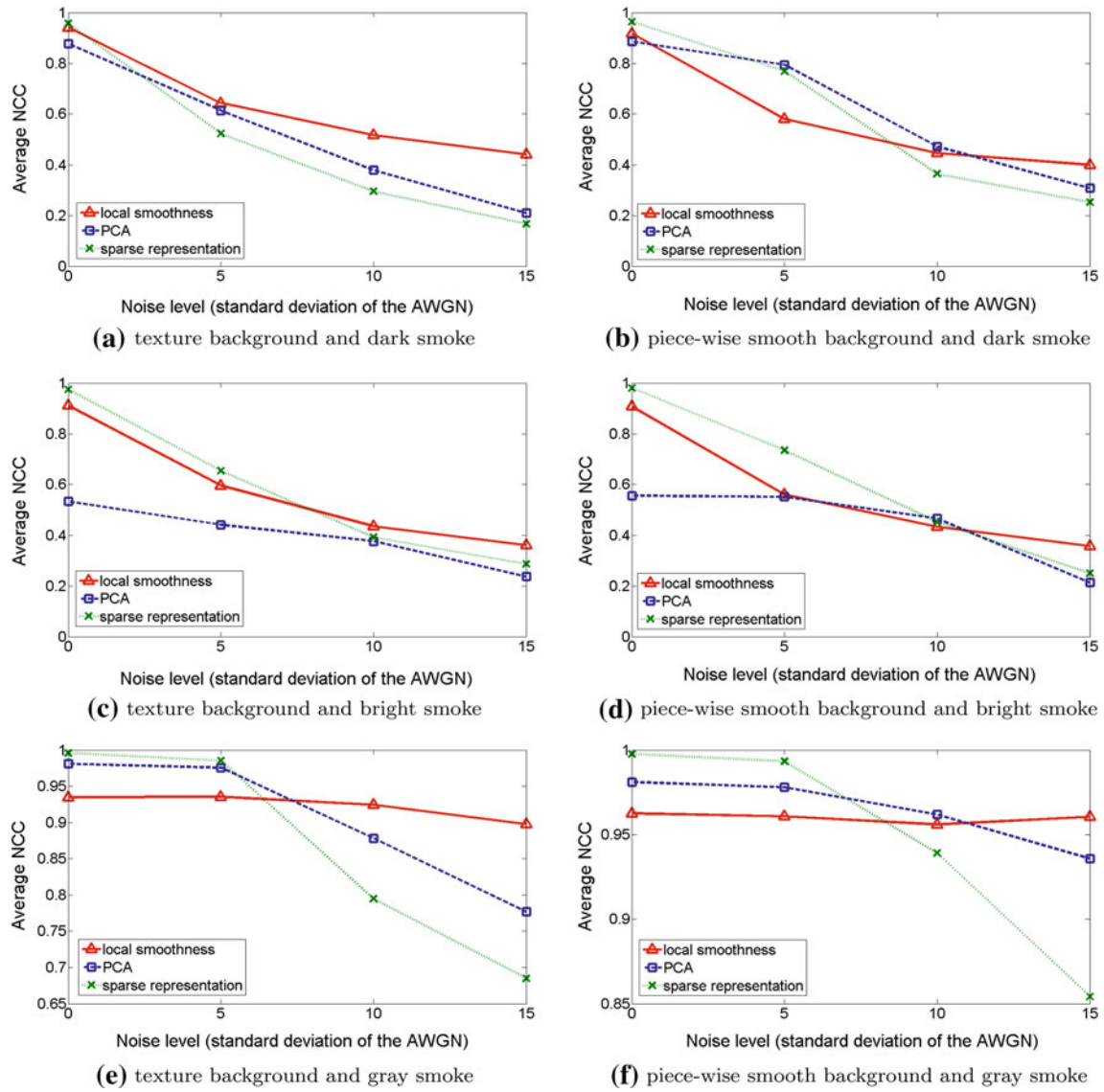

Fig. 13 Average NCC values versus noise levels for different combinations of background and pure smoke using different models for the smoke component

for smoke component results in non-smoke images being modeled as pure smoke. This phenomenon does not accord with our original expectation of the consequence of image separation. It then follows that if the principal component constraint is adopted in the detection task, it will inevitably lead to misclassification. As a result, in the subsequent smoke detection experiments that we conducted and reported, only the local smoothness and sparse representation were used as constraint for smoke component at the stage of image separation. It is also noticed that the estimated α values are all 1.0 in these separation experiments. When a scene is totally opaque due to smoke, the blending parameter will also be estimated as 1.0. Thus, in the smoke detection tasks, simply thresholding the α value will not be sufficient.

5.3 Performance of Smoke Separation on Real Video Frames

We tested the smoke separation methods on real video sequences and some illustrative blocks of the separated smoke components are shown in Fig. 15. The collage in Fig. 16 shows a few scenarios from our test scenes and frames of the separated smoke components. Notice how well the smoke component was separated in indoor and outdoor, long and short distance surveillance scenes. However, situations where the texture of smoke is very similar to that of the covered background and the smoke is very light still present a challenge to our method (see fourth scenario in Fig. 16). Given three image blocks from the video image shown in

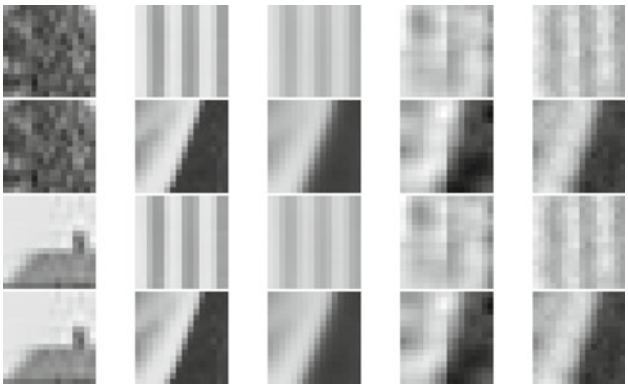

Fig. 14 Foreground object separation (*column 1*: texture and piece-wise smooth background, *column 2*: texture and piece-wise smooth foreground, *columns 3–5*: the separated "smoke component" using the local smoothness, principal component, and sparse representation constraint respectively)

Fig. 17 as the current input images **f**, their estimated α values and separated smoke components are shown in the figure. When the background image has no covering (the most left block in Fig. 17), the estimated α value is close to 0 and the separated "smoke component" looks like a homogeneous patch. When the background image is covered by light or heavy smoke, the separated smoke component hardly includes the background information and the estimated α indicates the degree of heaviness of the smoke.

5.4 Smoke Detection Experiments on Real Video Sequences

In a series of smoke detection experiments we sought to verify how the smoke separation would improve the detection performance compared to the conventional methods that extract features from the original video frames **f** rather than smoke components **s**.

5.4.1 Data Sets

In total 15000 pairs of background images **b** modeled by a GMM (Stauffer and Grimson 2000) and corresponding input images **f** were created from 20 publicly available video clips of smoke.[2] These video clips cover indoor and outdoor, short and long distance surveillance scenes with different illuminations; a few of which are shown in Fig. 16. Specifically, the 15000 input images consist of 10000 smoke images and 5000 non-smoke images. The smoke images are divided into 4 categories, each category having 2500 images:

SI1: images that are fully covered by heavy smoke,
SI2: images that are fully covered by light smoke,
SI3: images that smoke covers more than half of their areas, and
SI4: images that smoke covers less than half of their areas.

The non-smoke input images are divided into two categories, each having 2500 images as well:

NS1: images that are covered by non-smoke opaque objects, and
NS2: images that are not covered by anything.

Notice that each input image **f** in these categories has its corresponding background image **b** obtained using GMM-based background modeling.

5.4.2 The Noise Level of **b** *Modelled by GMM*

Experimental results obtained with synthesized images indicated that when the standard deviation of the AWGN was less than 5, a successful separation was achievable. We estimated the noise level of **b** modelled by GMM to gain insight into the expected performance of the separation methods in real applications. The non-smoke set, NS2, was used to estimate the noise level of the background images. If the GMM background modeling used to obtain the background images had performed perfectly, an image in NS2 would be exactly the same as their corresponding background image and their difference would be zero. Therefore, we simply considered the difference as noise. The means and standard deviations of the differences for all images in NS2 were computed and the average standard deviation value was about 4.7. This value is indicative of a successful separation.

5.4.3 Evaluation Method

The LBP has been successfully used in texture classification tasks because of its ability to describe texture (Ojala et al. 2002). In this work, for each pair of image block **f** and its corresponding background block **b**, the smoke component **s** was separated using either the local smoothness model or the sparse representation model described previously. LBP was extracted from **s** to describe the texture of smoke and input to a Kernel SVM to decide whether **f** is covered by smoke or not. Both Radial Basis Function (RBF) and Histogram Intersection Kernel (HIK) were tested in the experiments and superior results were obtained with the RBF kernel. Thus the results using RBF kernel are presented in this paper.

In the experiments, each of the four categories of smoke images (SI1–SI4 described in Sect. 5.4.1) was combined with the non-smoke image category NS1 to form four test datasets

[2] The related video clips can be downloaded from http://signal.ee.bilkent.edu.tr/VisiFire/Demo/SampleClips and http://imagelab.ing.unimore.it/visor.

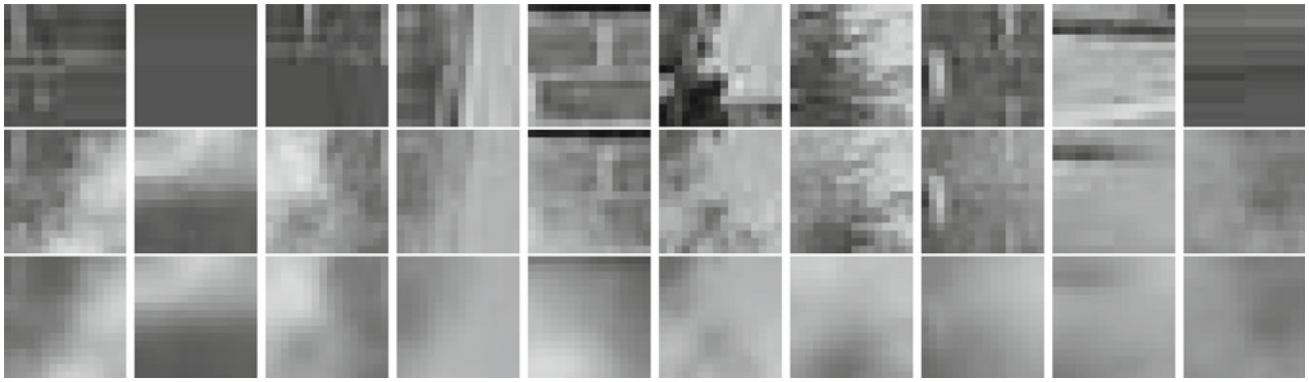

Fig. 15 Some separated pure smoke block images using real video images (*row 1*: the learnt background images **b** using a GMM (Stauffer and Grimson 2000), *row 2*: the current input images **f**, *row 3*: the separated smoke components **s**)

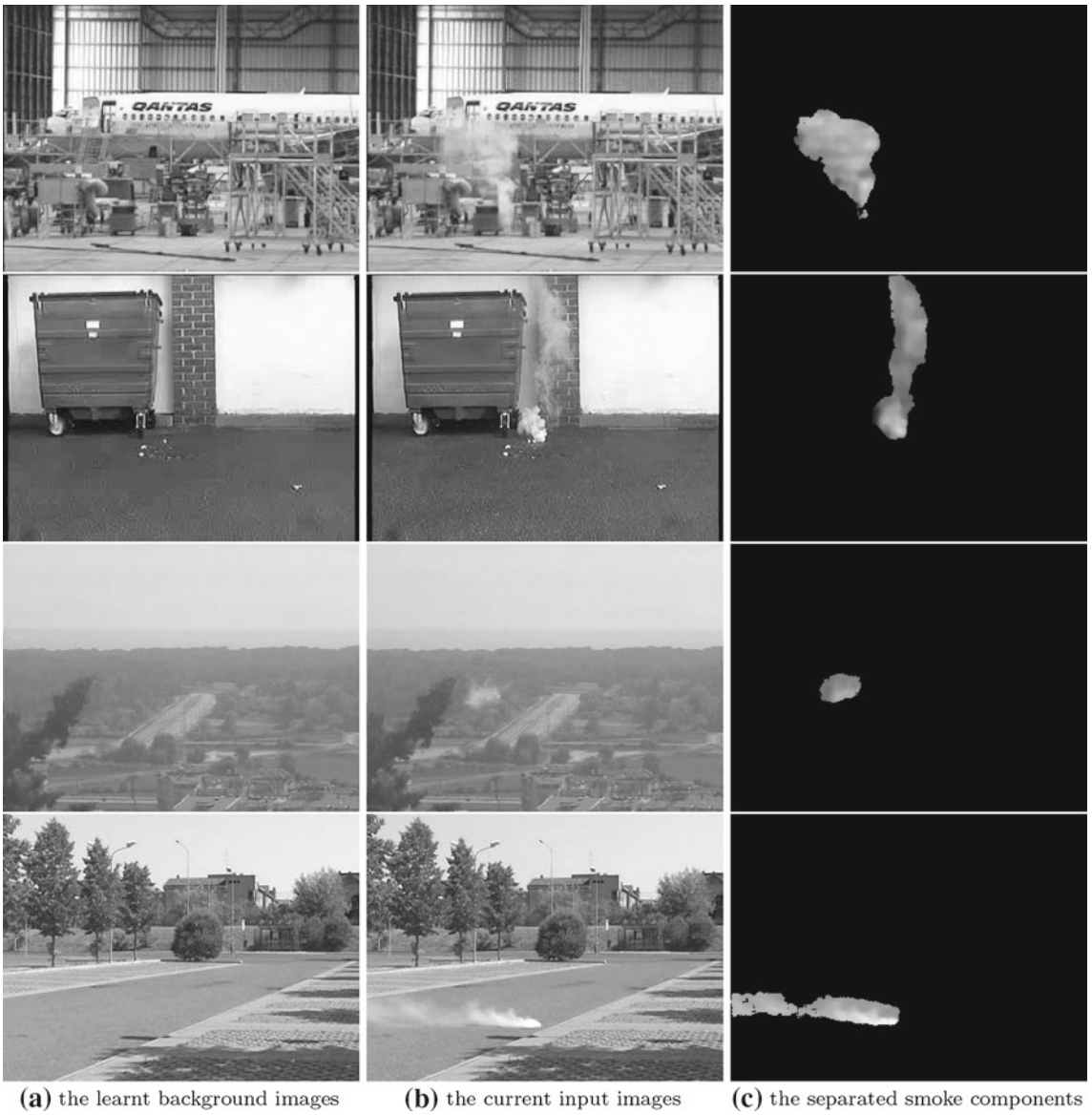

Fig. 16 Some separated pure smoke frames using real video images (*column 1*: the learnt background images **b** using a GMM (Stauffer and Grimson 2000), *column 2*: the current input images **f**, *column 3*: the separated smoke components **s**)

Fig. 17 The α values and separated smoke components for three different image blocks in a real video frame

and ten-fold evaluation was performed on each dataset. The Receiver Operating Characteristics (ROC) curves for the four datasets were generated as performance measurement.

5.4.4 Results

In our evaluation we set the methods proposed by Toreyin et al. (2005) and Tian et al. (2011) as two baselines and compared the results of the proposed methods to the results they published. The two baseline methods extracted wavelet and LBP-based features from the original image **f** to perform smoke detection respectively. Further, as part of our evaluation, we compared our approach to the direct background subtraction in a bid to dispel the notion that it is a simple and appropriate method for smoke separation. The difference image between **f** and **b** was computed and LBP features were extracted for the experiments.

Figure 18 shows the ROC curves of different methods for detecting different types of smoke. As can be seen, irrespective of whether local smoothness or sparse representation is used as the smoke component model, the proposed smoke detection methods based on image separation outperform the baseline methods. Furthermore regardless of whether smoke is heavy or light, fully or partly covering, significant improvements were achieved by using the features extracted from smoke component **s** compared to the case when features were extracted directly from the image **f**. Compared with the case when features were extracted from the difference images, it is noteworthy that the detection performance obtained was much worse than those achieved by the proposed approach; a result that justifies the unsuitability of background subtraction for smoke component separation.

Moreover, in many cases, the difference image could introduce or increase the background information and lead to poor detection performance when compared with the case when features were extracted directly from the image **f** as shown in Fig. 18a–c. In our experiments one exception was observed for the image category SI04 (Fig. 18d) when only

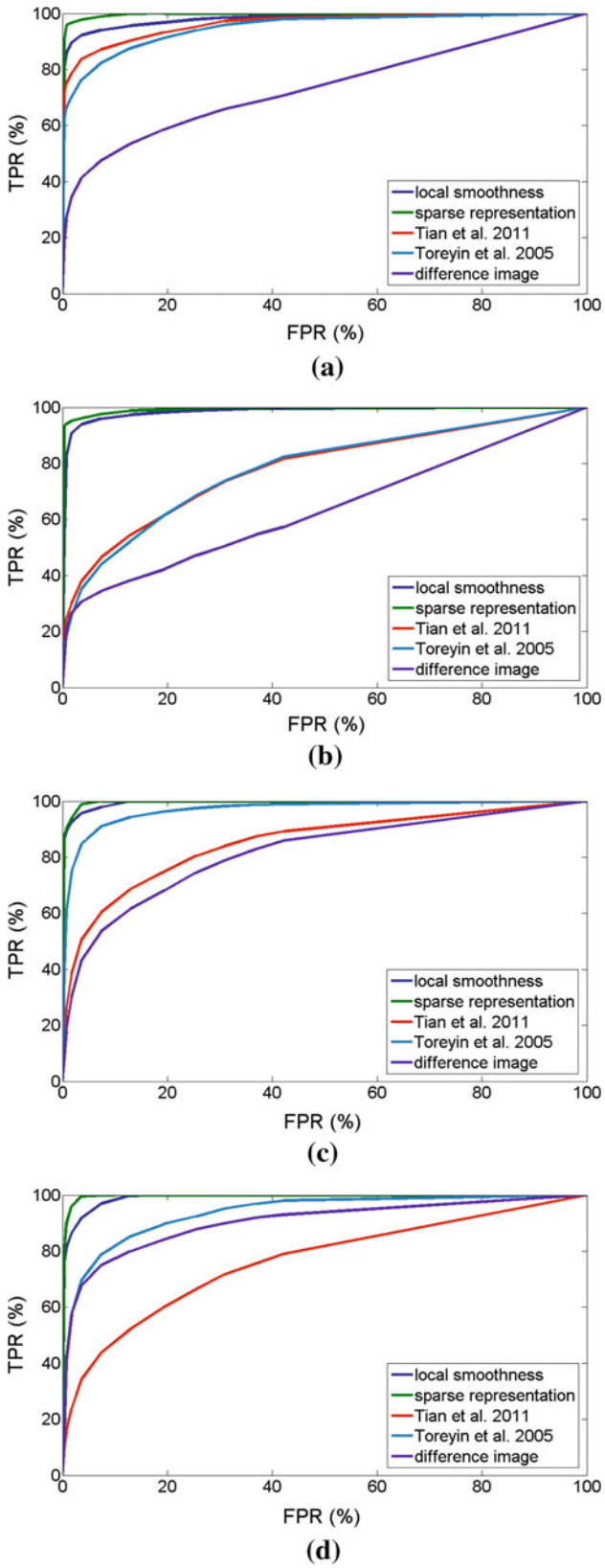

Fig. 18 ROC curves of different methods for detecting different types of smoke (**a** heavy smoke, **b** light smoke, **c** more than half of area is covered by smoke, **d** less than half of area is covered by smoke)

less than half of image area was covered by smoke. The reason adduced for this observation is that in this case background information dominates in both **f** and **b**. Thus the resulting difference image becomes a homogeneous patch.

In all, it can be concluded that the comparative results presented here verify the effectiveness of the proposed approach.

5.5 Computational Complexity

The proposed smoke detection consists of three major steps: background modeling, separation of smoke component **s** and classification of **s**. Most computation is spent on obtaining the sparse codes to separate the smoke component. In this step, the blending parameter α and the sparse codes to represent the smoke component **s** are alternately solved. The parameter α is analytically calculated using Eq. (16) and the sparse codes for the smoke component are obtained using the feature-sign search algorithm. The complexity of this step is $O(\aleph_1 \aleph_2 K^3)$ where K is the number of non-zero elements in the sparse codes, $\aleph_1$ is the number of iterations within the feature-sign search algorithm, $\aleph_2$ is the number of alternations. Typical values of K, $\aleph_1$ and $\aleph_2$ for our experiments are 30, 4 and 60 respectively. We implemented the detection algorithm in MATLAB on a PC with 2.93GHz Intel(R) Core(TM) i7 CPU and 8GB memory and were able to achieve 1-2 frames per second at early stage of smoke. Notice that the proposed algorithm is highly parallel because each block can be processed independently.

6 Conclusions and Perspectives

The fact that smoke often appears as an overlay on a background image could make the task of smoke detection difficult. In this paper we propose an approach whereby the smoke component, if any, is separated from an image before detection. A mathematical image formation model that linearly blends an amount of smoke and background image formed the basis of the proposed approach. Using several constraints, viz. sparse representation and local smoothness, the blending parameter was estimated and the smoke component was separated. Experimental results have verified the effectiveness of the proposed approach.

In general, this paper presents ideas pioneering a new direction for vision-based smoke detection. However, there are still rooms for improvement. Under this framework, at least three issues are worthy of further study. Firstly, in the image separation, local smoothness and sparse representation constraints on the smoke component have proved effective when the noise level of background modelling is not too high. It is worthwhile to consider constraints based on the nature of smoke in order to achieve better separated smoke component in the presence of high level of noise in the modelled background. Secondly, when imposing the sparse representation constraint on the smoke component, we expected the coefficient vector **x** to encode some information about smoke. However, when it was used directly as a feature vector the detection results were unsatisfactory. Thus a further study is needed to understand how and in what way **x** encodes the characteristics of smoke. Thirdly, once the smoke component has been separated, more features for smoke may be investigated to obtain better characterization although texture has proved to be discriminative in many cases as evidenced by our experiments.

The proposed image separation approach has been inspired by the transparency property of smoke. As a result, this framework may be suitable for detecting other semi-transparent objects, such as fog, haze, water, shadow and steam. An interesting question is whether this framework could be used to distinguish different objects with transparency property, e.g. smoke and fog. From the physical process underlying the formation of smoke and fog, there is a good chance to find some features to distinguish them. This will be investigated in the future.

Acknowledgments This work was partly supported by Beijing Polymer Sensing Technology Co. Ltd.

References

Aharon, M., Elad, M., & Bruckstein, A. (2006). K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on Signal Processing*, *54*(11), 4311–4322.

Almeida, L. B. (2005). Separating a real-life nonlinear image mixture. *Journal of Machine Learning Research*, *6*, 1199–1229.

Bai, X., & Sapiro, G. (2009). Geodesic matting: A framework for fast interactive image and video segmentation and matting. *International Journal of Computer Vision*, *82*(2), 113–132.

Bell, A. J., & Sejnowski, T. J. (1995). An information-maximisation approach to blind separation and blind deconvolution. *Neural Computation*, *7*(6), 1129–1159.

Calderara, S., Piccinini, P., & Cucchiara, R. (2011). Vision based smoke detection system using image energy and color information. *Machine Vision and Applications*, *22*(4), 705–719.

Chen, T., Yin, Y., Huang, S., & Ye, Y. (2006). The smoke detection for early fire-alarming system base on video processing. In *Proceedings of international conference on intelligent information hiding and multimedia signal processing* (pp. 427–430).

Fadili, M. J., Starck, J. L., Bobin, J., & Moudden, Y. (2010). Image decomposition and separation using sparse representations: An overview. *Proceedings of the IEEE*, *98*(6), 983–994.

Farid, H., & Adelson, E. H. (1999). Separating reflections and lighting using independent components analysis. *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *1*, 262–267.

Funaro, M., Oja, E., & Valpola, H. (2003). Independent component analysis for artefact separation in astrophysical images. *Neural Networks*, *16*(3–4), 469–478.

Gai, K., Shi, Z., & Zhang, C. (2008). Blindly separating mixtures of multiple layers with spatial shifts. In *Proceedings of IEEE Computer Society conference on computer vision and pattern recognition* (pp. 15–22).

Gonzalez, R. C., & Woods, R. E. (2007). *Digital image processing* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.

Guidara, R., Hosseini, S., & Deville, Y. (2009). Maximum likelihood blind image separation using nonsymmetrical half-plane Markov random fields. *IEEE Transactions on Image Processing*, *18*(11), 2435–2450.

Guo, L., & Garland, M. (2006). The use of entropy minimization for the solution of blind source separation problems in image analysis. *Pattern Recognition*, *39*(6), 1066–1073.

He, K., Sun, J., & Tang, X. (2011). Single image haze removal using dark channel prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *33*(12), 2341–2353.

Kolesov, I., Karasev, P., Tannenbaum, A., & Haber, E. (2010). Fire and smoke detection in video with optimal mass transport based optical flow and neural networks. In *Proceedings of IEEE international conference on image processing* (pp. 761–764).

Kong, N., Tai, Y. W., & Shin, S. Y. (2011). High-quality reflection separation using polarized images. *IEEE Transactions on Image Processing*, *20*(12), 3393–3405.

Lee, H., Battle, A., Raina, R., & Ng, A. Y. (2007). Efficient sparse coding algorithms. In *Proceedings of advances in neural information processing Systems*.

Levin, A., Lischinski, D., & Weiss, Y. (2008). A closed-form solution to natural image matting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *30*(2), 228–242.

Levin, A., Zomet, A., & Weiss, Y. (2004). Separating reflections from a single image using local features. In *Proceedings of IEEE Computer Society conference on computer vision and pattern recognition* (Vol. 1, pp. 306–313).

Levin, A., & Weiss, Y. (2007). User assisted separation of reflections from a single image using a sparsity prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *29*(9), 1647–1655.

Long, C., Zhao, J., Han, S., Xiong, L., Yuan, Z., Huang, J., et al. (2010). *Transmission: A new feature for computer vision based smoke detection*. Lecture notes in artificial intelligence (Vol. 6319, pp. 389–396). Berlin: Springer.

Maruta, H., Nakamura, A., Yamamichi, T., & Kurokawa, F. (2010). Image based smoke detection with local hurst exponent. In *Proceedings of IEEE international conference on image processing* (pp. 4653–4656).

Meyer, F. G., Averbuch, A. Z., & Coifman, R. R. (2002). Multilayered image representation: Application to image compression. *IEEE Transactions on Image Processing*, *11*(9), 1072–1080.

Minh, H. Q., & Wiskott, L. (2011). Slow feature analysis and decorrelation filtering for separating correlated sources. In *Proceedings of IEEE international conference on computer vision* (pp. 866–873).

Narasimhan, S. G., & Nayar, S. K. (2002). Vision and the atmosphere. *International Journal of Computer Vision*, *48*(3), 233–254.

Ojala, T., Pietikainen, M., & Maenpaa, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *24*(7), 971–987.

Osher, S., Sole, A., & Vese, L. (2003). Image decomposition and restoration using total variation minimization and the H^{-1} norm. *Multiscale Modeling & Simulation*, *1*(3), 349–370.

Sarel, B., & Irani, M. (2005). Separating transparent layers of repetitive dynamic behaviors. In *Proceedings of IEEE international conference on computer vision* (Vol. 1, pp. 26–32).

Schechner, Y. Y., Kiryati, N., & Basri, R. (2000). Separation of transparent layers using focus. *International Journal of Computer Vision*, *39*(1), 25–39.

Starck, J. L., Elad, M., & Donoho, D. L. (2005). Image decomposition via the combination of sparse representations and a variational approach. *IEEE Transactions on Image Processing*, *14*(10), 1570–1582.

Stauffer, C., & Grimson, W. E. L. (2000). Learning patterns of activity using real-time tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *22*(8), 747–757.

Szeliski, R., Avidan, S., & Anandan, P. (2000). Layer extraction from multiple images containing reflections and transparency. In *Proceeding of IEEE Computer Society conference on computer vision and pattern recognition* (Vol. 1, pp. 246–253).

Tian, H., Li, W., Ogunbona, P., Nguyen, D. T., & Zhan, C. (2011). Smoke detection in videos using non-redundant local binary pattern-based features. In *Proceedings of IEEE international workshop on multimedia signal processing* (pp. 1–4).

Tonazzini, A., Bedini, L., & Salerno, E. (2006). A markov model for blind image separation by a mean-field EM algorithm. *IEEE Transactions on Image Processing*, *15*(2), 473–482.

Toreyin, B. U., Dedeoglu, Y., & Cetin, A. E. (2005). Wavelet based real-time smoke detection in video. In *Proceedings of European signal processing conference*.

Tung, T. X., & Kim, J. M. (2011). An effective four-stage smoke-detection algorithm using video images for early fire-alarm systems. *Fire Safety Journal*, *46*(5), 276–282.

Turk, M., & Pentland, A. (1991). Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, *3*(1), 71–86.

Wang, J., & Cohen, M. F. (2005). An iterative optimization approach for unified image segmentation and matting. In *Proceedings of IEEE International Conference on Computer Vision* (pp. 936–943).

Wright, J., Ma, Y., Mairal, J., Sapiro, G., Huang, T. S., & Yan, S. (2010). Sparse representation for computer vision and pattern recognition. *Proceedings of the IEEE*, *98*(6), 1031–1044.

Yu, C., Fang, J., Wang, J., & Zhang, Y. (2010). Video fire smoke detection using motion and color features. *Fire Technology*, *46*(3), 651–663.

Yuan, F. (2008). A fast accumulative motion orientation model based on integral image for video smoke detection. *Pattern Recognition Letters*, *29*(7), 925–932.

Zhang, C., & Sato, I. (2011). Separating reflective and fluorescent components of an image. In *Proceedings of IEEE Computer Society conference on computer vision and pattern recognition* (pp. 185–192).