2015

Vision-based smoke detection

Hongda Tian
University of Wollongong

Vision-Based Smoke Detection

Hongda Tian

Supervisor:
Associate Professor Wanqing Li
Co-supervisors:
Professor Philip Ogunbona, Associate Professor Lei Wang

This thesis is presented as part of the requirements for the conferral of the degree:

Doctor of Philosophy

The University of Wollongong
School of Computer Science and Software Engineering

February 12, 2015

Declaration

I, Hongda Tian, declare that this thesis submitted in partial fulfilment of the requirements for the conferral of the degree Doctor of Philosophy, from the University of Wollongong, is wholly my own work unless otherwise referenced or acknowledged. This document has not been submitted for qualifications at any other academic institution.

Hongda Tian

February 12, 2015

Abstract

To prevent property damage and loss of human lives due to fire accidents, vision-based smoke detection is deemed to be a promising approach to early detection and prediction of fire disasters. Despite advances, challenges still exist for developing robust vision-based smoke detection algorithms, which include reliably quantifying the visual characteristics of smoke and overcoming the limitations of devices in real applications. This thesis has established computational models to address these challenges and developed novel methods based on the models for vision-based smoke detection. Specifically, an image formation model for smoke is derived based on atmospheric scattering models. To capture the texture feature of smoke, non-redundant local binary pattern is adopted to deal with the relative intensity between the bright/dark smoke and the background. In addition, algorithms are developed to separate the smoke component from the background and robust features are extracted from the smoke component for reliable detection. Specifically, three models, namely local smoothness, principal component and sparse representation, are proposed to model the smoke component for effective component separation. Detection performance is significantly improved compared with the traditional video-based methods. For the scenarios where video is not available, novel algorithms are developed for detecting smoke in single images. Specifically, sparse representation based on dual-dictionary modeling is proposed for both the smoke and background components and the sparse coefficients are employed as a feature for smoke detection. The effectiveness of the proposed methods is validated by extensive experiments.

Acknowledgments

The PhD study is a long but rewarding journey for me. Thanks to the help and support of many people, I can overcome so many difficulties and be where I am today.

I would like to express my sincere gratitude to my supervisor Associate Professor Wanqing Li. Without his invaluable guidance, suggestion and encouragement, this work would not have been possible. Heartfelt appreciation also goes to my co-supervisors Professor Philip Ogunbona and Associate Professor Lei Wang for their insightful advice and comments on my research. It is a great honor for me being able to work with all my supervisors. What I have learned from them is far beyond research and they are more of life mentors than supervisors.

I would like to take this opportunity to thank the University of Wollongong and SNS Unicorp Pty Ltd for providing me with scholarships, namely International Postgraduate Tuition Award and Matching Scholarship. In addition, my acknowledgments are extended to the School of Computer Science and Software Engineering and Information and Communication Technology Research Institute for the financial support for academic conference attendance.

I wish to thank my colleagues Ce Zhan, Duc Thanh Nguyen, Yuyao Zhang, Song Liu, Zhimin Zong, and Lingqiao Liu for their constructive ideas on my research which have inspired me a lot. I am also grateful to my friends Xue Wei, Fuchun Guo, and Xiguang Zheng for their help in both of my study and life.

I am deeply indebted to my parents for their everlasting and unconditional love, care, and support throughout my life.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

Fire accidents often cause substantial economic and ecological damage as well as endanger the lives of people. Take Australia as an example, there are approximately 100 fatalities and 3000 injuries each year due to structure fires [7]. According to incomplete statistics, bush fires have accounted for over 800 deaths since 1851 and the total cost is about 1.6 billion dollars. Therefore, both industrial and academic research communities are committed to develop effective methods and equipment that can provide early detection and prediction of fire disasters. Fire detection is a process of issuing fire alarms by sensing different signatures (e.g. smoke and flame) of combustion process [132]. Although flame-based fire detection has been reported in the literature [88], [152], [153], smoke-based fire detection is more meaningful and can serve as an early warning for fire incidences, as the presence of smoke often signals the onset of a fire event.

Conventional smoke detectors [126], which include photoelectric, ionization and air-sampling detectors, mainly detect the presence of certain particles generated by smoke and fire. Photoelectric and air-sampling 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. All these methods depend on the transportation of the smoke towards the detector and sufficient concentration of the molecules or particles being present. Consequently, there is a delay inherent in the transportation and

this is exacerbated in outdoor scenarios 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 of the intensity and scale of smoke or fire.

Compared to the traditional point-based smoke detection 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 a process of sensing smoke by analyzing the visual information captured by cameras. It 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. Due to these advantages, there have been increasing research activities on vision-based smoke detection in the past few years. Most techniques adopt a pattern recognition paradigm in which the input image or video is preprocessed and divided into image blocks. For each block, 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 as reported in the literature [155], [186], [16], [185].

However, there are many challenges in developing robust vision-based smoke detection methods. One of the challenges is reliably quantifying the visual characteristics of smoke. Some of the visual characteristics of smoke include irregular and deformable shape, irregular motion, diverse color, 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, is also hard to quantify. Thus, well-known motion descriptors such as optical flow [61] have not been very successful as features. In addition, the colors of smoke range from white-bluish to white when the temperature of smoke is low and from black-grayish to black when the temperature rises until it catches fire, which means it is hard to extract reliable color-related features. The possibility of using texture as a feature for

smoke is promising because of its dispersive distribution. However, one needs to account for the varying degree of transparency of smoke.

One of the reasons for unreliable characterization of smoke is the lack of understanding and modeling of the imaging process of smoke. Unlike general opaque objects, smoke has its unique imaging properties. An exploration on illumination and reflectance modeling related to smoke imaging can provide some insights on smoke formation and, hence, how to quantify the visual characteristics of smoke.

Another challenge for vision-based smoke detection arises from the limitations of devices in real applications. Almost all existing smoke 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 moving under severe and dynamic environment, such as wind, or in cases where the surveillance is based on a battery-powered sensor network in which neither the power nor bandwidth is sufficient to stream video. In such circumstances, detection of smoke from single images, often known as surveillance images, becomes highly desirable. This desirability comes at a price because image-based detection is much more challenging than video-based systems.

1.2 Contributions

This thesis focuses on establishing computational models to deal with the challenges as presented before and developing novel methods based on the models for vision-based smoke detection. Specifically, the thesis has made the following contributions to the field of study.

- An image formation model is derived based on the atmospheric scattering models and forms the base for vision-based smoke detection. It expresses an image as a linear combination of smoke and non-smoke (background) components.

- Non-redundant local binary pattern is adopted for characterizing texture feature of

smoke and used for smoke detection.

- According to the derived image model, a framework for video-based smoke detection based on image separation is developed. It separates the smoke component from the background and features extracted only from the smoke component are used for smoke detection.

- Local smoothness, principal component and sparse representation are proposed to model the smoke component for effective component separation.

- Single image based smoke detection is formulated as an image separation problem. Dictionary-based sparse representation is proposed for component separation. The sparse coefficients for the separated components are employed for smoke detection.

- Based on the results of single image smoke detection, single image smoke separation is formulated as an image matting problem. A new method is developed for smoke separation.

1.3 Publication list

1.3.1 Published papers

1. Hongda Tian, Wanqing Li, Lei Wang, and Philip Ogunbona. Smoke Detection in Video: An Image Separation Approach. *International Journal of Computer Vision*, vol. 106, no. 2, pp. 192-209, Jan. 2014.

2. Hongda Tian, Wanqing Li, Philip Ogunbona, and Lei Wang. Single Image Smoke Detection. *Proceedings of 12th Asian Conference on Computer Vision*, 2014.

3. Hongda Tian, Wanqing Li, Lei Wang, and Philip Ogunbona. A Novel Video-Based Smoke Detection Method Using Image Separation. *Proceedings of IEEE International Conference on Multimedia and Expo*, 2012, pp. 532-537.

4. Hongda Tian, Wanqing Li, Philip Ogunbona, Duc Thanh Nguyen, and Ce Zhan. Smoke Detection in Videos Using Non-Redundant Local Binary Pattern-Based Features. *Proceedings of IEEE International Workshop on Multimedia Signal Processing*, 2011, pp. 1-4.

1.3.2 Papers in preparation

1. Hongda Tian, Wanqing Li, Philip Ogunbona, and Lei Wang. Single Image Smoke Detection and Separation. (in preparation, to be submitted to *IEEE Transactions on Image Processing*)

1.4 Organization of the thesis

The rest of this thesis is organized as follows:

Chapter 2 provides a critical review of the state-of-the-art of vision-based smoke detection and its related topics. Firstly, the techniques of traditional smoke detectors are presented. Secondly, the existing vision-based smoke detection methods are summarized. Thirdly, some representative image separation techniques are reviewed. Fourthly, a discussion about the approaches for image matting is provided. Finally, methods for fog/haze detection and removal are included here to provide a context for understanding the separation and detection of smoke.

Chapter 3 presents the process of smoke imaging. Based on the atmospheric scattering models, it presents the derivation of a physics-based image formation model for smoke detection.

Chapter 4 describes the adopted texture feature–Non-redundant local binary pattern (NRLBP) for smoke characterization. After a brief introduction of local binary pattern, NRLBP is defined and its effectiveness for smoke detection is evaluated using real videos.

Chapter 5 and Chapter 6 propose video-based smoke detection methods based on the derived model using different image separation techniques. In Chapter 5, the structure of inter-pixel correlation within a smoke image is employed and a local smoothness con-

straint is imposed for smoke separation. In Chapter 6, principal component and sparse representation are considered as constraints for effective smoke separation.

Chapter 7 depicts the scenario of single image smoke detection. Based on the derived image model, sparse representation constraints are imposed on both the smoke and non-smoke components. The detection performance using sparse coefficients is verified experimentally. The details of GMM-based matting for smoke separation are presented.

Chapter 8 concludes the thesis and suggests possible future works.

Chapter 2

Literature review

This chapter provides a critical review of the literature related to the topics covered in this thesis. Firstly, to better understand the motivation of vision-based smoke detection, the main techniques of traditional smoke detection are presented. Secondly, existing vision-based smoke detection methods are summarized in terms of the visual features used for smoke characterization. Thirdly, as the main contribution of this thesis to smoke detection is related to smoke separation, state-of-the-art methods for image separation are reviewed. Fourthly, due to the analogy between the image formation model in smoke separation and the compositing model in image matting, the methods for image matting are discussed. Finally, due to the visual similarity between smoke and fog/haze, representative methods for fog/haze detection and removal are discussed to provide references for smoke separation and detection.

2.1 Traditional smoke detection

In terms of working principles, conventional smoke detectors [126], [24] can be classified into three categories: photoelectric-based, ionization-based, and air sampling-based, as shown in Figure 2.1. Both the techniques and limitations of the three types of traditional smoke detectors are discussed in this section.

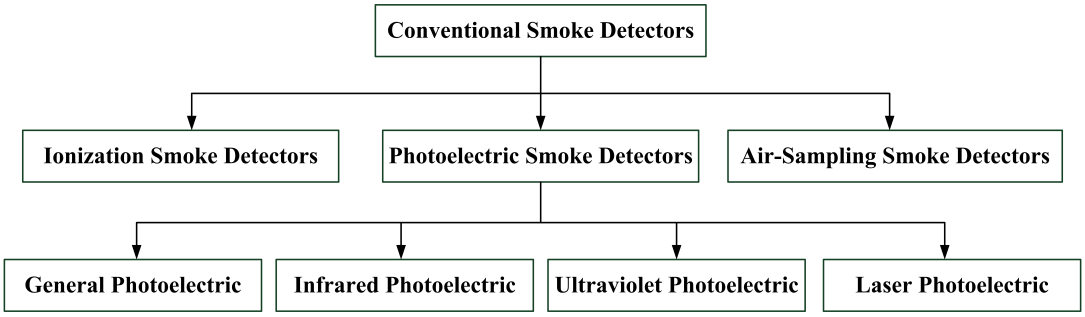


Figure 2.1: Different types of traditional smoke detectors.

2.1.1 Photoelectric (Optical) smoke detectors

Photoelectric smoke detectors [126], [119] are developed based on the principle of light scattering and the most common detectors found in fire alarms. According to different types of light sources, they can be categorized into four groups: general photoelectric, infrared photoelectric, ultraviolet photoelectric, and laser photoelectric. Regardless of different types of light sources, photoelectric smoke detectors utilize a light-sensitive cell in either of two ways. In one type, a light source (e.g. a small spotlight) causes a photoelectric cell to generate electric current that keeps an alarm circuit open. Once the particles of smoke interrupt the ray of light, the circuit is broken and the alarm is set off. The other type of photoelectric detectors are widely used in private dwellings. They employ a detection chamber, which is designed so that the light source and the photoelectric cell are installed at acute angles to each other in separate sections of the chamber. Under normal conditions, the ray from the light source travels along the chamber in a straight line and does not enter the section where the photoelectric cell is located. When smoke enters the chamber, the particles of smoke scatter the light, which causes certain light to travel towards the photoelectric cell. An alarm can be triggered at the required sensitivity by measuring the amount of light reaching the photoelectric cell.

Generally speaking, photoelectric smoke detectors are more sensitive to larger particles of smoke such as that generated by dense smoke or smoldering fires but are less sensitive to smaller smoke particles which can be found in rapidly burning fires. Due to the growing use of flame retardant materials in modern building construction, photoelectric smoke detectors are more widely applied in indoor environments.

2.1.2 Ionization smoke detectors

Ionization smoke detectors [126], [138] make use of radioactive material to ionize the air molecules between a pair of electrodes in a detection chamber such that a small electric current can be conducted by the ionized air. When smoke enters the chamber, the particles of smoke attach themselves to ions and reduce the flow of the electric current. An alarm can be issued at the required sensitivity by measuring the reduction in electric current.

Ionization smoke detectors generally have better performance on detecting smaller particles of smoke such as that produced by rapidly burning fires but are less sensitive to larger smoke particles generated by smoldering fires or dense smoke. In addition, ionization detectors are less effective in environments where strong air flow exists. As a result, photoelectric smoke detectors are more reliable for detecting smoke at both the smoldering and flaming stages of a fire.

2.1.3 Air-sampling smoke detectors

Air-sampling smoke detectors, most of which are aspirating smoke detectors, are good at detecting microscopic particles of smoke. They actively draw air through a network of pipes arranged above or below a ceiling in parallel covering a monitored area. Small holes are drilled into each pipe to form a matrix of holes, which provides an uniform distribution across the pipe network. The small holes can be considered as sampling points and sampled air is passed through a sensitive optical device (e.g. a solid-state laser) tuned to detect the extremely small particles of combustion.

Most air-sampling smoke detectors have a higher sensitivity than point-based smoke detectors and can provide multiple levels of alarm threshold, such as Alert, Action, Fire 1 and Fire 2. This makes it possible to detect a developing fire earlier than point-based smoke detection, which allows manual intervention or activation of automatic suppression systems before a fire has developed beyond the smoldering stage, and thus increases the time available for evacuation and minimizing the damage due to fire.

2.1.4 Discussions

According to the working principles, traditional smoke detectors 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 (see Figure 2.2) 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.

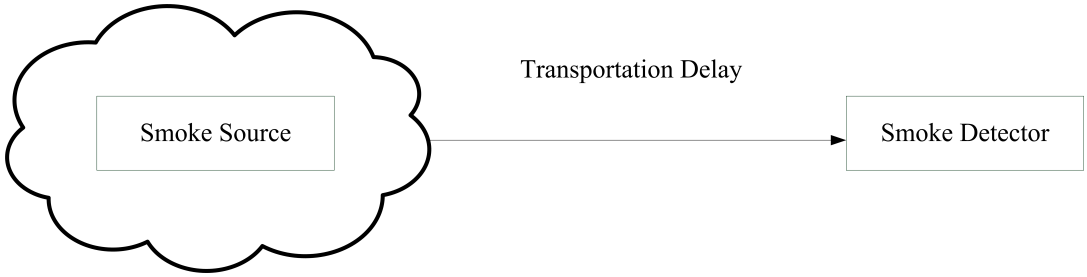


Figure 2.2: Schematic illustration of the transportation delay problem for traditional smoke detectors.

In terms of different applications, traditional smoke detectors can be categorized into two groups: stand alone and network smoke detectors. Once smoke is detected, stand alone smoke detectors issue acousto-optic alarm on the spot only, which are usually used in private dwellings, dormitories, shops and so on. No matter whether there is acousto-optic alarm on the spot when smoke is detected, network smoke detectors send an alarm signal to the control center to take further measures. They are more applicable in factories, hotels, residence communities, communication base stations and so on.

2.2 Vision-based smoke detection

Almost all existing vision-based smoke detection methods 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. They mainly follow the traditional pattern recognition paradigm, as shown in Figure 2.3. Regardless of training or testing stages, image sequences are preprocessed first. At this step, image frames are usually

divided into image blocks (either overlapped or non-overlapped). The purpose of this procedure is to serve early smoke detection as smoke only covers a small area at the very beginning of fire events. Consequently, the problem is reduced to that of determining if an image block is covered by smoke or not. Meanwhile, background modeling and motion detection are performed to select candidate image blocks for detection. Then features are extracted from the candidate image blocks. The features extracted from training set are used to train a classifier (e.g. support vector machine (SVM) [175]), which considers the features extracted from testing data as input and decides whether there is smoke or not in the testing data. Under the detection framework depicted by Figure 2.3, the success of smoke detection algorithms lies in identifying robust visual features to characterize smoke [75] and most research activities focus on extracting different visual features of smoke regions such as motion, color, edge, texture and so on. This section provides a review of the features that have been studied for smoke detection.

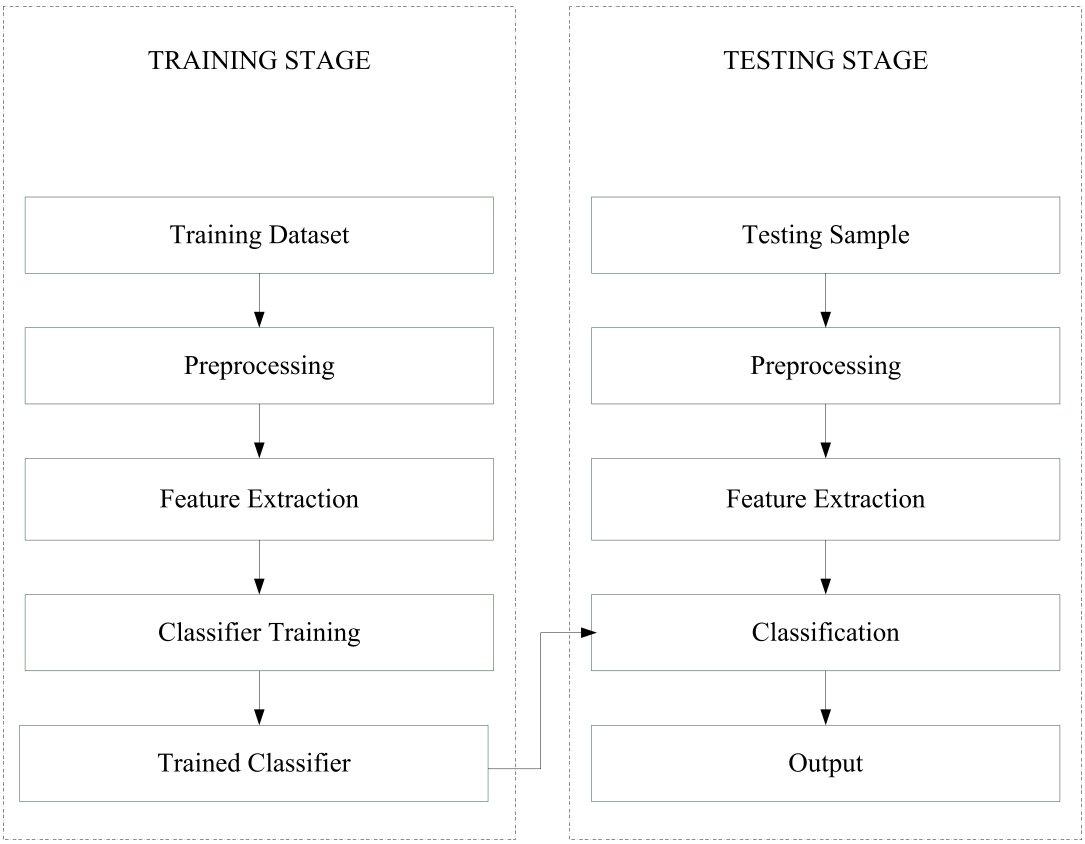


Figure 2.3: Existing vision-based smoke detection framework.

2.2.1 Motion features

Based on the fact that humans can easily recognize the presence or absence of smoke using the motion of smoke, some researchers decide to utilize motion characteristics for smoke detection. For instance, an accumulative motion model was proposed to capture the motion characteristics of smoke in [186]. However, the assumption that smoke usually drifts upwards continually by hot airflow makes this model ineffective in a scenario where wind is present. In addition, the velocity is a vector composed of a magnitude and an orientation. However in this paper, only the orientation of motion was calculated from sequential images, which results in inaccurate motion estimation. Based on the observation that various objects may cause local dynamic envelopes of pixels, which can be considered as sets of temporal intensity variations of spatially close pixels, Guillemant and Vicente proposed to discriminate smoke from other objects by estimating the various motions within the given envelopes [45]. The principle of this method is to extract local motions by cluster analysis in a multidimensional temporal embedding space.

Other researchers have extracted motion features of smoke using optical flow [61]. Rather than directly employ optical flow to extract the motion characteristics of smoke [181], Kopilovic et al. proposed to detect the irregularities in the optical velocity field due to smoke motion [72]. Optical flow techniques and wavelet analysis were combined in [127] to better capture the motion property of smoke. In this way the optical flow helps to obtain an approximation of the smoke motion and the wavelets help to obtain this motion at different levels of resolution. Optimal mass transport optical flow, which introduces the concept of mass transport in the sense of minimal transportation cost to optical flow computation, was used as a low-dimensional descriptor of smoke motion [69]. Based on the understanding that brightness constancy assumption holds well for rigid objects with a Lambertian surface but is less appropriate for fluid and gaseous materials (e.g. smoke) [34], optical flow residual was proposed for extracting dense motion features of smoke in [197]. In addition, a histogram of oriented optical flow was extracted as a temporal feature based on the fact that the diffusion direction of smoke is upward owing to thermal convection [67], [120]. As can be inferred, this temporal feature will not be effective in

the case of strong draught or wind.

2.2.2 Color features

Based on the observation that smoke is usually grayish, chromatic features have been used for smoke characterization. Despite the differences in methods reported in the literature, one common base is that the RGB values of a smoke pixel are equal or similar. In [21] and [162], such a grayish color was classified into two gray-level regions (i.e. light gray and dark gray). A set of decision rules were designed to extract color features. Thresholds for the intensity values coming from RGB channels of a smoke pixel were empirically set in [47], [48], [79]. A color model for describing smoke was defined using a statistical analysis of samples extracted from different types of video sequences and images [17]. A Gaussian mixture model was used in RGB color space to build a color model for smoke in [92]. Then color blending coefficients defined based on the model served in smoke detection. In a similar vein a reference color model for smoke was selected in the RGB color space to measure the deviation of the current pixel color from the model [160], [123], [16]. Recently with the definition of between-class scatter matrices and within-class scatter matrices of Fisher linear discriminant, the color model for smoke was estimated by maximizing the ratio of these two scatter matrices [189].

Considering a pixel covered by smoke, although smoke lightens or darkens each component in RGB color space, it does not severely change the intensity values of each component within a certain period of time. Motivated by this understanding, the temporal high frequency information of the three channels in the RGB color system for smoke-covered area, which was obtained by means of discrete wavelet transform in time domain, was employed to characterize the chromatic property of smoke [80]. An analysis in CIELAB color space was performed recently and a better clustering of the chromatic features of smoke was noticed in [106].

2.2.3 Edge features

From edge point of view, the features used for smoke detection can be further divided into two groups. One group of features is extracted to characterize the boundary of smoke while the other group of features describes the effects of smoke on the edges of objects covered by smoke.

In the first group, histogram of oriented gradient (HOG) [27] descriptors were extracted for smoke detection [68], [67], [120]. The underlying assumption of these methods is that the direction of smoke diffusion is upward owing to thermal convection such that the gradient distribution of a smoke boundary has a distinguishable pattern. Other works relating to smoke characterization based on its contour were presented in [194], [64], [65].

The phenomenon that blurred edges could be observed in smoke-covered areas given the image of a scene motivates the second group of features. As edges contribute to high frequency information, the consequent decrease in high frequency has been used as a cue to perform smoke detection in [155], [154], [188] where wavelets were adopted to extract horizontal, vertical, and diagonal high frequency information of images. However, this decrease in high frequency is not unique to smoke covering as occlusion from some pure-color objects could also lead to a decrease in high frequency. In addition to this ambiguity, smoke also produces high frequency signal because of the rich edge information of its contour. To overcome these problems, wavelet transform was implemented based on the RGB Contrast-image, which presents the contrast of RGB channels in a neighboring area [18]. In another work, a mixture of Gaussians of the energy variation in the wavelet domain was considered [15]. Rather than perform wavelet transform only at one scale, characterization of smoke was carried out by extracting wavelet features from approximate coefficients and three levels of detailed coefficients in [43]. To better distinguish smoke from cloud in the scenario of forest fire smoke detection, a correlation analysis on high frequency components of two consecutive image frames was performed after wavelet transform [171].

2.2.4 Texture features

When the object to be analyzed is distributed dispersively, texture analysis has been proved to be effective. Due to the dispersive distribution of smoke, researchers explore different types of texture features to characterize smoke. Gray level co-occurrence matrices (GLCM), which represents second-order statistics features and provides information related to both the gray level distribution of the image and the relative positions of the various gray levels within the image, have been used for smoke detection [26], [95], [180], [113], [141]. Local binary pattern (LBP), a well-known descriptor for texture classification [116], was applied to smoke detection [187], [185], [97]. To well characterize the anisotropic deformed shape of smoke, an anisotropic LBP descriptor was defined in [96], which is a texture operator computed using anisotropic neighborhood pixel values. It is also noted that both GLCM and LBP were considered in one piece of work of smoke detection [12]. A modified local ternary pattern, namely center symmetric local ternary pattern (CS-LTP) was proposed as the texture descriptor for smoke [167]. It is designed for improving the robustness for handling the noise and reducing the number of bits in feature description, and thus speeding up its histogram comparison processing.

2.2.5 Other features

Except the features presented above, features based on other characteristics of smoke are also reported in the literature and they are listed here for completeness. The flickering property of smoke has been employed for smoke detection [154], [172], [173], [59], [60]. For example, the flickering process of smoke was analyzed by means of a Hidden Markov model using temporal wavelet transform coefficients in [154]. Due to its intrinsic property of diffusion, smoke regions will gradually increase at the initial stage. The growth rate of smoke regions was defined for smoke detection [168], [176], [170]. In addition, smoke was considered as a self-affine fractal and its Hurst exponent was employed for smoke detection in [94], [10]. Inspired by the airlight-albedo ambiguity model which describes the relative uncertainty between airlight and albedo, the concept of transmission was introduced as a feature of smoke [91]. To capture the irregularity of smoke spreading, disorder

characteristics were defined in [125]. Any single feature can not perfectly characterize smoke and thus feature fusion techniques are adopted. Some representative work is found in [157], [39], [77].

2.3 Image separation

Given a mixture (or mixtures) of multiple images, image separation involves the recovery of the original source images, with limited prior information about the sources and the mixing model. It plays a vital role in applications such as compression, enhancement, restoration and so on. The existing methods for image separation often require multiple mixture observations of source images (e.g. multiple mixtures of the same source images under different illuminations).

2.3.1 General methods for image separation

According to the assumptions made on the source images and/or mixing parameters, methods for image separation can be divided into four categories. The representative methods in each category are reviewed in this section.

The first and oldest category is the well-known independent component analysis (ICA) [9] 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 [32], separating artefact in astrophysical images [35], separating real-life nonlinear mixtures of document images acquired through scanning [4], [5], separating reflective and fluorescent components in images [193], [192]. The statistical independence assumption was relaxed by using so-called slow feature analysis and decorrelation filtering in [104].

The second category, which is often called sparse component analysis (SCA), is based on the sparsity properties of the source images, i.e. often on the possibility of isolating each source image in a certain domain. For this purpose, some well-known transforms, such as the local cosine transform [102] and the Gabor transform [142], were adopted to

achieve sparsity of periodic textures. Based on the fact that the high-frequency components of images are sparse and are stronger on one side of the paper than on the other one, wavelet transform was used for the separation of nonlinear real-life mixture of document images obtained by scanning [6]. Motivated by sparse representation of signals, morphological component analysis was proposed to perform image separation in [30]. To overcome the difficulty that the sparse domains of the source images may not be well defined by the known transforms, a local dictionary was adaptively learned for each source image along with separation in [1]. In some cases, the source images are not sufficiently sparse to the standard SCA method. As a result, a feedback mechanism based on SCA was proposed for blind image separation [184]. Recently source separation from compressed hyperspectral images was explored by means of sparse models [40].

The third category consists of methods based on non-negative matrix factorization (NMF) which needs the non-negativity of the observations, mixing coefficients and source images [81]. By taking into account the local smoothness properties of source images, Zdunek and Cichocki combined NMF with a Gibbs smoothing prior for blind image separation in [190]. In addition, NMF was extended for the application of spectral unmixing for urban hyperspectral images [100].

The last category corresponds to Bayesian methods, which can use different prior information about the observations, sources or mixing coefficients and usually need to associate probability laws with the involved components. For instance, Tonazzini et al. formulated blind image separation as a Bayesian estimation problem and proposed an expectation-maximization algorithm with the mean field approximation to solve the problem [151]. Particle filtering, which is an advanced Bayesian estimation method that can deal with non-Gaussian nonlinear models and additive space-varying noise, was used for image separation in [25]. Additionally, Guidara et al. proposed a maximum likelihood approach to blind source separation using non-symmetrical half-plane Markov random fields (MRF) [44]. In a similar vein, the natural image prior was modeled via a high-order MRF and was integrated into a Bayesian framework for estimating the source images [195].

Except the aforementioned four main categories of methods, other typical techniques for image separation are summarized here as well. Total variation minimization was employed to separate images into cartoon (i.e. piece-wise smooth) and texture components in [118]. Guo and Garland proposed to use a direct minimization of an entropy-like function to solve the separation problem [46]. The fractional Fourier transform was used to reduce the correlation between the source images to serve image separation [135], [108]. The success of many separation methods depends on the selection of permutations corresponding to the orders of the update operations. To solve such a permutation ambiguity problem, Hara et al. decorrelated the mixtures while searching for the appropriate update permutation using a pruning technique [50]. Instead of minimizing the correlation between structure and texture components, the correlation between structure gradient and texture components was minimized for image separation in [89]; and superior performance is noticed. Unlike other image separation methods which are commonly based on iterative optimization, an non-iterative method based on the analysis allowing calculation of a relative transformation between the mixtures was proposed in [49]. Driven by the advances of RGB-D imaging sensors, depth cues were considered to facilitate intrinsic image separation [19].

2.3.2 Transparent layer separation

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. For completeness, representative methods for transparent layer separation are listed in this section.

Reflection separation was achieved by using simple comparison and subtraction operations on a series of images captured by rotating a polarizing filter in [115]. The layer images were recovered based on imaging through a polarizing filter at two orientations in [130]. Szeliski et al. used constrained least squares to recover the layer images [147]. However, this method has a limitation that each layer must have a fixed transparency. Schechner et al. exploited focus difference between the background scene and reflected

scene as a cue for separating each layer [131]. By searching for chains of polarity-preserving junctions, an algorithm based on the propagation of local junction information was proposed for transparent layer separation [139]. Levin et al. presented a separation method wherein the total amount of edges and corners is minimized [86]. However, this method may not work well on the images including many intersections of edges from different layers. Sarel and Irani 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 [128]. After a median operator was applied to space-time derivatives, two transparent layers could be separated. Based on spatio-temporal analysis using epipolar plane images (EPI), the edge and color information of EPI was used to separate the reflected and transparent layers in [117]. Levin and Weiss incorporated user input into the sparsity prior of image gradients to separate reflection from a single image [84]. The separation of multiple layers with unknown spatial shifts and varying mixing coefficients was addressed in [37]. However, this method is limited to uniform translations. Additionally, an expectation-maximization algorithm that employed the hidden Markov model was proposed to separate smooth layers and the substantially-textured background from a single image [177]. Kong et al. regarded the exclusiveness of the image gradients of the background layer and the reflection layer as a cue to separate reflection [70]. In the case of smoke, this assumption about the image gradients may not be true as many background scenes share similar image gradients with pure smoke (e.g. uniform walls and homogeneous smoke). Based on the statistics of natural images, both the sparsity and joint behavior patterns of image gradients were employed as cues to estimate the transparent layer [36].

2.4 Image matting

Matting refers to the problem of accurate foreground estimation in image and video. It is one of the key techniques in image editing and film production, thus has been extensively studied in the literature. Accurately separating a foreground object from the background involves determining both full and partial pixel coverage, also known as pulling a matte.

This problem was mathematically formulated by Porter and Duff [124]. They introduced the alpha channel as the means to control the linear interpolation of foreground and background colors. Mathematically, the observed image $\mathbf{I}_z(z = (x, y))$ is modeled as a convex combination of foreground image $\mathbf{F}_z$ and background image $\mathbf{B}_z$ by using the alpha matte α_z:

$$\mathbf{I}_z = \alpha_z \mathbf{F}_z + (1 - \alpha_z)\mathbf{B}_z \tag{2.1}$$

where $\alpha_z \in [0, 1]$. If α_z is 0 or 1, pixel z is called definite background or definite foreground respectively. Otherwise, pixel z is usually called a mixed pixel. Given only a single input image, all three values α, $\mathbf{F}$ and $\mathbf{B}$ are unknown and need to be determined at every pixel location. The known information we have for a pixel are the three dimensional color vector $\mathbf{I}_z$ (assuming it is represented in some 3D color space), and the unknown variables are the three dimensional color vectors $\mathbf{F}_z$ and $\mathbf{B}_z$, and the scalar alpha value α_z. Matting is thus inherently an under-constrained problem, since 7 unknown variables need to be solved given 3 known values for every pixel. In order to obtain good estimates of the unknown variables, most matting approaches rely on user guidance and prior assumptions on image statistics to constrain the problem. Usually, the user guidance includes either a trimap or scribbles. A trimap is a three-level pixel map where the input image is segmented into three regions: definite background, definite foreground and unknown. Scribbles can be regarded as a coarse trimap, where only a few background and foreground pixels are specified and the majority pixels are considered as unknowns. Generally, existing image matting techniques can be classified into two categories based on their underlying methodologies. Some representative methods in each category are presented in this section.

2.4.1 Sampling-based approach

Statistically, neighboring pixels that have similar colors often have similar alpha values. Sampling-based approach makes use of this local correlation by sampling nearby known

foreground and background colors for each unknown pixel z. According to the local smoothness assumption on the image statistics, the colors of these samples are close to the true foreground and background colors ($\mathbf{F}_z$ and $\mathbf{B}_z$) of z, thus these color samples can be used to estimate $\mathbf{F}_z$ and $\mathbf{B}_z$. Once $\mathbf{F}_z$ and $\mathbf{B}_z$ are determined, α_z can be easily calculated from the composition model expressed by Equation (2.1).

Sampling-based approach can be further divided into parametric and non-parametric methods. Parametric sampling methods fit a parametric statistical model to known foreground and background samples and then estimate alpha by considering the distance of unknown pixels to known foreground and background distributions. For example, Bayesian matting [23] models foreground and background colors as mixtures of Gaussians. It uses a continuously sliding window for neighborhood definition, which marches inward from the foreground and background regions. Some other methods first trains Gaussian Mixture Models (GMMs) on known foreground and background colors globally, and for an unknown pixel, samples are drawn from all the Gaussians to cover all the possibilities that its foreground color could have [8], [163]. Nonparametric methods simply collect a set of known foreground and background samples to estimate alpha matte by finding best samples for unknown pixels. To discover which samples from a large sample set are most valid, an optimized color sampling procedure was proposed in [164]. In some cases that good samples for an unknown pixel can not be found nearby, a global sampling method that used all samples available in the image was proposed [55]. Furthermore, local sampling and global sampling schemes were combined to prevent true foreground or background samples being missed during the sample collection stage [133], [159]. Recently to collect a comprehensive and representative set of samples, the spatial sampling range for pixels was expanded farther from the foreground or background boundary and samples from each color distribution in known regions were ensured to be included in [134].

2.4.2 Affinity-based approach

Affinity-based approach intrinsically models the alpha matte gradient across the image lattice instead of directly estimating the alpha value at each single pixel by defining various affinities between neighboring pixels. Unlike sampling-based approach, most methods focus on first estimating alpha values, and then estimating true foreground colors for unknown pixels based on pre-computed alphas, rather than estimating them jointly for an optimal solution. Additionally, the alpha matte is estimated in a propagation fashion from known pixels to unknown ones, thus small errors could be propagated and accumulated to produce large errors [166].

Poisson matting models the matte gradient by assuming that intensity changes in the foreground and background are locally smooth [145]. In random walk matting [42], given an unknown pixel, its alpha value is set to be the probability that a random walker starting from this location will reach a pixel in the foreground before striking a pixel in the background, when biased to avoid crossing the foreground boundary. A closed form solution for alpha matting was proposed in [85] based on local color line model. A learning-based technique was proposed further to generalize the color line assumption [198]. Although widely adopted, local color line assumption can be easily violated in practice when large kernels are used [57]. As a result, nonlocal principle was introduced for alpha matting in [83]. Recently, the assumption of local color line was further relaxed in KNN matting [20] using nonlocal principle and K nearest neighbors. Until now, spectral matting [87] is the only method that tries to pull out a foreground matte in a completely automatic fashion. However, it is limited to images that consist of a modest number of visually distinct components.

2.5 Fog/haze detection and removal

Fog/haze detection is important for tracking and navigation applications, consumer electronics, and entertainment industries. As fog/haze degrades the perceptual image quality and thus the efficacy of computer vision algorithms based on small features or high fre-

quencies, removal of fog/haze from images as a preprocessing can increase the accuracy of these computer vision algorithms.

Gray level variation in the horizon was used to measure the amount of fog in an image and then estimate the visibility distance in [52]. However, a model needs to be built for measuring distance by using a binocular camera, which means this method is not suitable for fog detection with single camera or classification of fog images. Two clues of fog presence were considered for fog detection in [14]. One is the reduced visibility distance, which was calculated from the camera projection equations; the other is the blurring effect due to fog, which was estimated by measuring the high frequency components in images. Gallen et al. proposed to detect the presence of fog in images captured by in-vehicle multipurpose cameras during nighttime by means of two methods, which are based on the detection of the back-scattered veil induced by the vehicle ego lights and the detection of halos around light sources in the vehicle environment respectively [38]. To distinguish images with fog present from those free of fog, Gabor filters at different frequencies, scales and orientations were adopted as global descriptors for characterizing images [121]. Liu and Lu proposed to calculate the characteristic parameters of the gray-scale histograms and use a series of thresholds to determine which fog level a test image belongs to [90]. A highly effective method for fog region of interest segmentation based on geodesic maps computation was proposed as well as a novel joint fog density and horizon line estimation process in [53].

Early methods for fog and haze removal mainly rely on additional depth information or multiple observations of the same scene. For instance, haze removal was achieved by using two images taken through a polarizer at different angles [129], [137]. By means of a physics-based scattering model, the scene structure was recovered from two or more weather images in [110]. Kopf et al. proposed to dehaze an image by using the scene depth information directly accessible in the georeferenced digital terrain or city models [71]. Besides generating a modified transmission map for defogging video images using multiphase level set formulation, the color distortion between consecutive frames was compensated using the temporal difference ratio of HSV color channels in [178].

Single image fog and haze removal is a more challenging problem, since less information about the scene structure is available. Hautiere et al. first estimated the weather conditions and approximated a 3D geometrical model of the scene, which was inferred a priori and refined during the restoration process [54]. A modified partially overlapped sub-block histogram equalization was applied to enhance contrast in foggy images in [191]. Based on the comparison between the local variance of a foggy image and that of the preliminary enhancement, a method was proposed for enhancing contrast in images captured when fog density varies across the scene [122]. Fattal developed a refined image formation model to account for the surface shading and the scene transmission [33]. Under the assumption that the two functions are locally statistically uncorrelated, a haze image can be broken into regions of constant albedo, from which the scene transmission can be inferred. Tan proposed to enhance the visibility of a haze image by maximizing its local contrast [148]. This method can generate quite compelling results, especially in regions with very dense haze. However, since it is not a physics-based method, the restored image often suffers from distorted colors and significant halos. A cost function based on human visual model was used in luminance image to estimate the fog map, which was then subtracted from the foggy image for restoration purpose [66]. In [73] and [114] an image was modeled as a factorial Markov random field, in which the scene albedo and depth are two statistically independent latent layers. A canonical expectation maximization algorithm was implemented to factorize the image. Tarel and Hautiere estimated a contrast enhanced image of the scene by using combinations of min, max, and median filters to enforce piecewise constancy [149]. To reduce the computational complexity of physics based methods for fog removal, a fuzzy logic based algorithm was proposed in [28]. A fog removal method, which can be less affected by noise, was presented based on contrast limited adaptive histogram equalization [174]. Inspired by the human visual system, Jia et al. constructed a model which accounts for the perceptual sensitivity to noise, compression artifacts, and the texture of image content [62]. Such a model was combined with contrast limited adaptive histogram equalization to adaptively enhance both global and local contrast of foggy images. A nonlinear edge-preserving filter was

proposed to extract transmission map, which was further used to recover haze-free images in [196]. A graph-based method was presented to segment foggy images and the transmission map was refined by using a bilateral filter [31]. Similarly under the assumption that the depth map tends to be smooth except along edges with large depth jumps, fog removal was achieved based on fast bilateral filtering [183], [182]. A dark channel prior, which comes from an observation that most local patches in haze-free images often contain some low intensity pixels, was proposed for single image haze removal [56]. On this basis, lowest level channel prior was presented for fog removal in [22]. This prior is based on an observation that fog-free intensity in a color image is usually the minimum value of trichromatic channels. Unlike other physics-based models that only consider single scattering of light for fog/haze removal, Wang et al. developed a model which can explain multiple scattering effects [165]. Then the blurring effects of multiple scattering was estimated with local minimum filter. Based on the understanding that the transmission map should be smooth for an object except along the edges, Tripathi and Mukhopadhyay proposed to use anisotropic diffusion to recover scene contrast [156]. A wavelength adaptive image formation model was developed to generate the transmission map, which was used to estimate the atmospheric light [179]. According to the modified transmission map and the estimated atmospheric light, fog component was removed in a spatially adaptive manner. From a geometric perspective, boundary constraint on the transmission function was combined with contextual regularization for haze removal in [101]. Based on the understanding that the estimation of the ambient light can have a critical impact on the quality of restored fog-free images, Sun addressed the problem how to remove the interference of white areas as well as calibrate ambient light of degraded images with a large area of sky [146]. Shiau et al. applied a weighted technique that can automatically find the possible atmospheric lights and mix these candidates to refine the atmospheric light [136]. Then a novel prior processing method named difference prior, which can mitigate the halo artifact around the sharp edges, was proposed for haze removal.

2.6 Summary

There is a transportation delay inherently in traditional smoke detection and this is exacerbated in outdoor scenario where there could be draught or wind. Vision-based smoke detection does not suffer similar drawbacks and there is an additional benefit of being able to specify the location of the fire, its scale and intensity. The key for existing vision-based smoke detection is how to achieve reliable characterization of smoke using visual features. To answer this question, a realistic modeling of the imaging process of smoke is expected to provide some insights.

Chapter 3

Physics-based image model

Reliable vision-based smoke detection requires good understanding and modeling of the imaging process of smoke, which defines the topic of this chapter. The chapter starts with an introduction to the principle and properties of scattering by particles of the media in space. Two mechanisms relating to scattering, namely attenuation model and airlight model, are discussed. According to the two models, a smoke imaging model is derived.

3.1 Light scattering

The characteristics of light, notably its intensity, color, spatial distribution and so on, are changed when light travels through atmosphere. The alterations are due to the interactions between the energies carried by the light waves and the media of the atmosphere. As one type of essential interactions, scattering, induced by the media, occurs at all wavelengths in the electromagnetic spectrum.

From a microscopic viewpoint, various media consist of particles with different types and sizes. When a light wave is incident on a type of media, each particle in it scatters the light in a way that depends on not only the properties of the medium but also the shape and size of the particle. The exact pattern of the scattering varies significantly with the size of the particle [105]. Assume λ is the wavelength of light. A small particle with a radius of 0.1λ scatters almost equally in the forward and backward directions with respect to incidence. A particle with medium size (e.g. the radius is 0.25λ) scatters more in the

forward direction. While a large particle the radius of which is larger than λ scatters almost fully in the forward direction.

Particles are adjacent to numerous other particles apparently. However, the average distance between particles is often several times of the particle size. Therefore, particles can be regarded as independent scatterers, which means that the scattering intensities coming from different particles do not interfere with each other. This does not mean that the incident light is merely scattered by a single particle, which is often referred to as single scattering. In fact, each particle is exposed to both the incident light and the light scattered by other particles and this is referred to as multiple scattering.

Assume that a unit volume of scattering medium is illuminated by the spectral irradiance $E(\lambda)$ per unit cross-sectional area. The radiant intensity $\psi(\theta, \lambda)$ of the unit volume in the direction θ with respect to the observer is defined as follows [99]:

$$\psi(\theta, \lambda) = \beta(\theta, \lambda)E(\lambda), \tag{3.1}$$

where $\beta(\theta, \lambda)$ is the angular scattering coefficient, which represents the angular scattering property of a type of media per unit path length of the irradiating light and per steradian centered at the direction of observation. In fact, the irradiance $E(\lambda)$ can be regarded as the flux projecting on the unit volume per unit cross-sectional area. And the radiant intensity $\psi(\theta, \lambda)$ can be interpreted as the flux radiated per unit solid angle and per unit volume of the medium. Hence the omnidirectional flux scattered by this volume is obtained by integrating over the entire sphere:

$$\Psi(\lambda) = \beta(\lambda)E(\lambda), \tag{3.2}$$

where $\beta(\lambda)$ is the total scattering coefficient, which represents the ability of the volume to scatter flux per unit path length for a given wavelength λ in all directions.

Media such as haze, fog, cloud and smoke differ from each other mainly in the types and sizes of the constituent particles and their concentrations in space. A great amount of effort has been made to measure the sizes and concentrations of particles in these media.

Some properties of these media are summarized in Table 3.1.

Haze is composed of aerosol which is a dispersion system of small particles suspended in gas [58]. There are various sources for haze including volcanic ashes, combustion products, foliage exudation, and sea salt. The particles produced by these sources respond quickly to the changes in relative humidity and play a role of nuclei (centers) of small water droplets once the humidity is high enough. The particles of haze are smaller than fog droplets but larger than air molecules. Haze tends to produce a distinctive gray or bluish hue and is deemed to affect visibility.

Fog evolves when the relative humidity of air reaches saturation. Under this situation, some of the nuclei turn into water droplets through condensation. Therefore, certain types of haze and fog have similar origins and a sufficient increase in humidity can turn haze into fog. This process is a quite gradual transition and an intermediate status is referred to as mist. Perceptible haze can spread to an altitude of several kilometers, while fog is usually only a few hundred feet thick. One practical difference between fog and haze lies in the dramatically reduced visibility induced by the former. There are many types of fog (eg. radiation fog, advection fog, etc.) which differ from each other in their formation processes [109].

Cloud differs from fog only in existing at higher altitudes rather than locating at the ground level. Most cloud is constituted of water droplets like fog, while some is composed of long ice crystals and ice-coated dust grains. Details on the physics of cloud can be found in [98].

Smoke is a collection of airborne solid and liquid particulates and gases emitted when a material undergoes combustion or pyrolysis, together with the quantity of air that is entrained or otherwise mixed into the mass [107]. Smoke is an aerosol of solid particles and liquid droplets. In the scope of this thesis, the smoke generated at the initial stage of fire is of particular interest. In this case, smoke only covers a small area of the scene. Unlike haze and fog which usually occupy a large part of the scene, smoke only has limited thickness along the line of sight.

Despite of different types of particles, two mechanisms can account for the final spec-

Table 3.1: Different media and associated particle types, sizes and concentrations (adapted from [99])

Medium	Particle type	Radius (μm)	Concentration (cm^{-3})
Air	Molecule	10^{-4}	10^{19}
Haze	Aerosol	$10^{-2} - 1$	$10^{3} - 10$
Fog	Water droplet	$1 - 10$	$100 - 10$
Cloud	Water droplet	$1 - 10$	$300 - 10$
Smoke	Aerosol	$10^{-2} - 1$	$10^{7} - 10^{5}$

tral irradiance received by the observer (e.g. camera), which are often referred to as attenuation model and airlight model in the literature.

3.2 Attenuation model

The first mechanism is the attenuation of a beam of light as it travels through the media, which causes the radiance of a scene point to decrease when its depth from the observer increases. Let a parallel beam of light, which is assumed to have unit cross-sectional area, be projected into one type of media, as shown in Figure 3.1. Consider the beam passing through an infinitesimal length dx along the direction of incidence. According to [99], the fractional change in irradiance due to scattering at location x can be written as:

$$\frac{dE(x,\lambda)}{E(x,\lambda)} = -\beta(\lambda)dx, \tag{3.3}$$

where $E(x,\lambda)$ represents the spectral irradiance at location x.

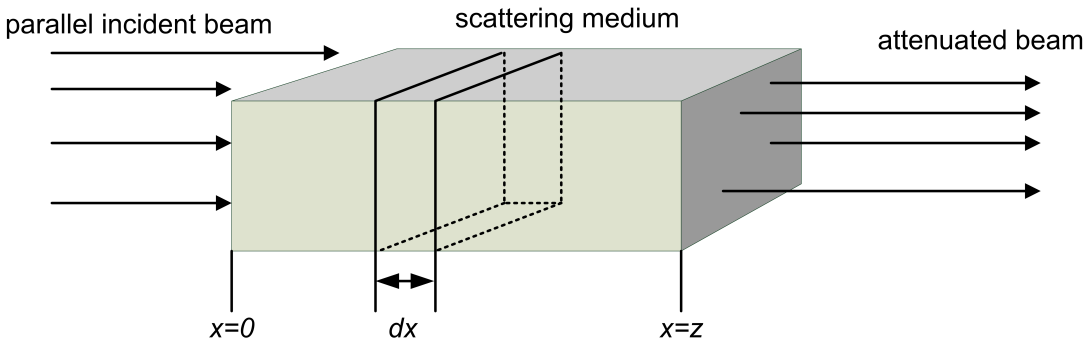


Figure 3.1: Attenuation of a parallel beam of light by particles of media.

By integrating both sides between the limits $x = 0$ and $x = z$, the spectral irradiance $E(z, \lambda)$ at a distance z from the source ($z = 0$) can be expressed as:

$$E(z, \lambda) = E_0(\lambda)e^{-\beta(\lambda)z}, \tag{3.4}$$

where $E_0(\lambda)$ is the irradiance at $z = 0$. Equation (3.4) is Bouguer's exponential law of attenuation [13], which only applies to parallel beams. For diverging beams from point sources, the inverse-square law [51] can be taken into account and incorporated as follows:

$$E(z, \lambda) = \frac{I_0(\lambda)e^{-\beta(\lambda)z}}{z^2}, \tag{3.5}$$

where $I_0(\lambda)$ represents the radiant intensity of the point source. However, Equation (3.5) does not consider the sky illumination and its reflection by scene points. By assuming that the irradiance at each scene point is dominated by the radiance of the sky, and that the irradiance due to other scene points is not significant, Narasimhan and Nayar have shown that the attenuated irradiance at the observer is given by [111]:

$$T(z, \lambda) = g\frac{e^{-\beta(\lambda)z}}{z^2}L_\infty(\lambda)\rho(\lambda), \tag{3.6}$$

where $L_\infty(\lambda)$ represents the radiance of the horizon ($z = \infty$) at wavelength λ, $\rho(\lambda)$ represents the sky aperture (the cone of sky visible to a scene point), and the reflectance of the scene point in the direction of the observer, g is a constant that accounts for the optical settings of the camera (e.g. exposure). It is noted that the flux scattered in the forward direction (i.e. towards the observer) by each particle has been ignored. Fortunately, this component is small in vision applications since the solid angles subtended by the source and the observer with respect to each other are small [103]. It is also noted that multiple scattering of light by particles is not explicitly modeled here, which may make the exponential law not hold in some situations (e.g. heavy fog). In the literature, Equation (3.6) is often referred to as the direct transmission model or the attenuation model.

3.3 Airlight model

A second mechanism referred to as airlight causes the media to behave like a source of light and originates from the scattering of environmental illumination by particles in the media. There can be several sources for environmental illumination, which include direct sunlight, diffuse skylight and light reflected by the ground. Attenuation causes scene radiance to decrease when the length of path from the observer increases, while airlight increases with the length of path. As a result, it causes the apparent brightness of a scene point to increase with the depth from the observer.

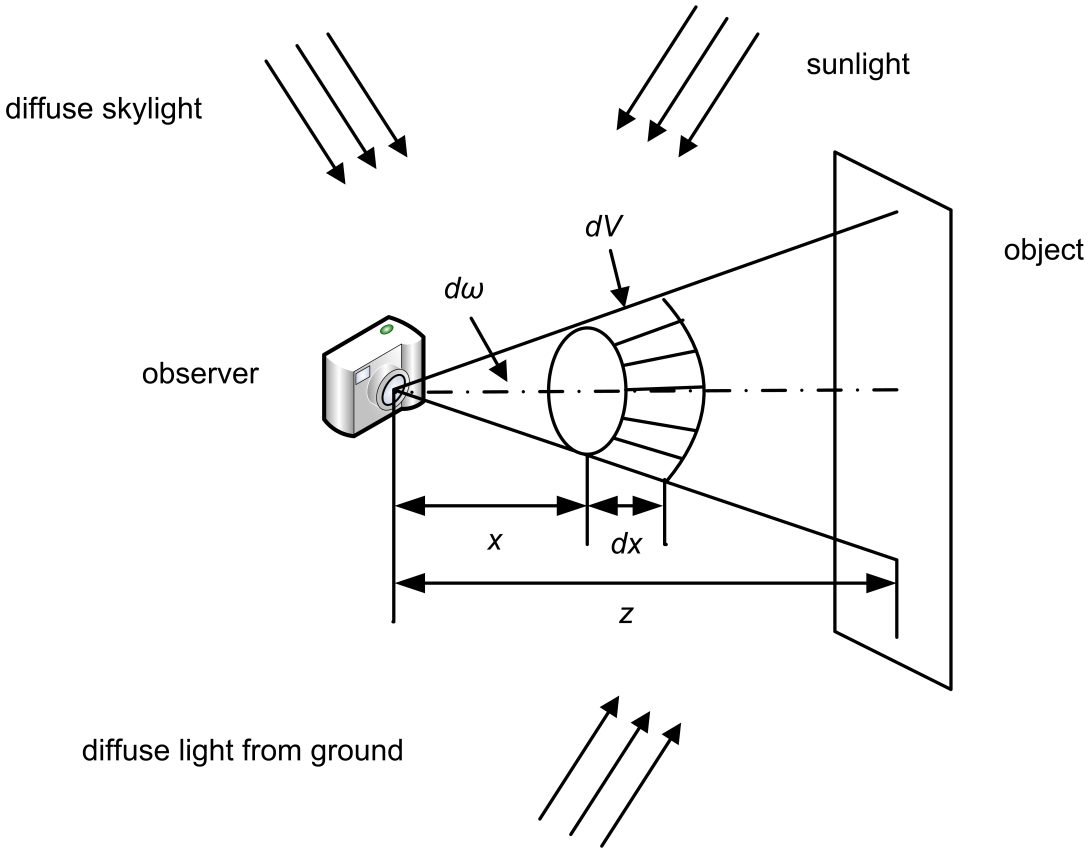


Figure 3.2: Source of the airlight between an observer and an object.

Consider a scenario of airlight as shown in Figure 3.2. Assume that the environmental illumination along the line of sight of the observer is constant but unknown in intensity, direction and spectrum. Actually the cone with the solid angle $d\omega$ subtended by the observer, and truncated by an object at distance z, can be regarded as a source of airlight. The infinitesimal volume dV at distance x from the observer may be expressed as the

product of the cross-sectional area, $d\omega x^2$, and the infinitesimal length dx:

$$dV = d\omega x^2 dx. \tag{3.7}$$

No matter what type of environmental illumination is incident onto dV, its intensity due to scattering in the direction of the observer is:

$$dI(x, \lambda) = dV k\beta(\lambda) = d\omega x^2 dx k\beta(\lambda), \tag{3.8}$$

where the proportionality factor k accounts for the illumination and the scattering properties of the medium.

If dV is regarded as a source with intensity $dI(x, \lambda)$, the irradiance it generates at the observer's end is experienced attenuation due to the medium and given by Equation (3.5):

$$dE(x, \lambda) = \frac{dI(x, \lambda)e^{-\beta(\lambda)x}}{x^2}. \tag{3.9}$$

The radiance $dA(x, \lambda)$ of dV can be calculated from its irradiance $dE(x, \lambda)$ as follows:

$$dA(x, \lambda) = \frac{dE(x, \lambda)}{d\omega} = \frac{dI(x, \lambda)e^{-\beta(\lambda)x}}{d\omega x^2} = k\beta(\lambda)e^{-\beta(\lambda)x}dx. \tag{3.10}$$

The total radiance of the length of path z from the observer to the object can be obtained by integrating Equation (3.10) between the limits $x = 0$ and $x = z$:

$$A(z, \lambda) = k(1 - e^{-\beta(\lambda)z}). \tag{3.11}$$

If the object is at an infinite distance (at the horizon), the radiance of airlight is maximum and is calculated by setting $z = \infty$ to get $L_\infty(\lambda) = k$. Moreover, in consideration of the optical settings of the camera, the radiance of airlight for any given length of path z is expressed as:

$$A(z, \lambda) = gL_\infty(\lambda)(1 - e^{-\beta(\lambda)z}). \tag{3.12}$$

Obviously, the radiance of airlight for an object which is right in front of the observer ($z = 0$) equals zero.

3.4 Imaging model in smoke medium

Similarly, 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 would behave like a source of light through scattering as well. However, unlike other media (e.g. fog and haze), smoke usually does not occupy the entire space of the scene. Assume that the distance between the scene and camera is z and smoke appears at distance z_s from a camera and its thickness along the line of sight is Δz, as shown in Figure 3.3. Also assume that 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 analogy to Equation (3.6), the attenuation model in the case of smoke is:

$$T(z, \lambda) = g\frac{e^{-\beta(\lambda)\Delta z}}{z^2}L_\infty(\lambda)\rho(\lambda). \tag{3.13}$$

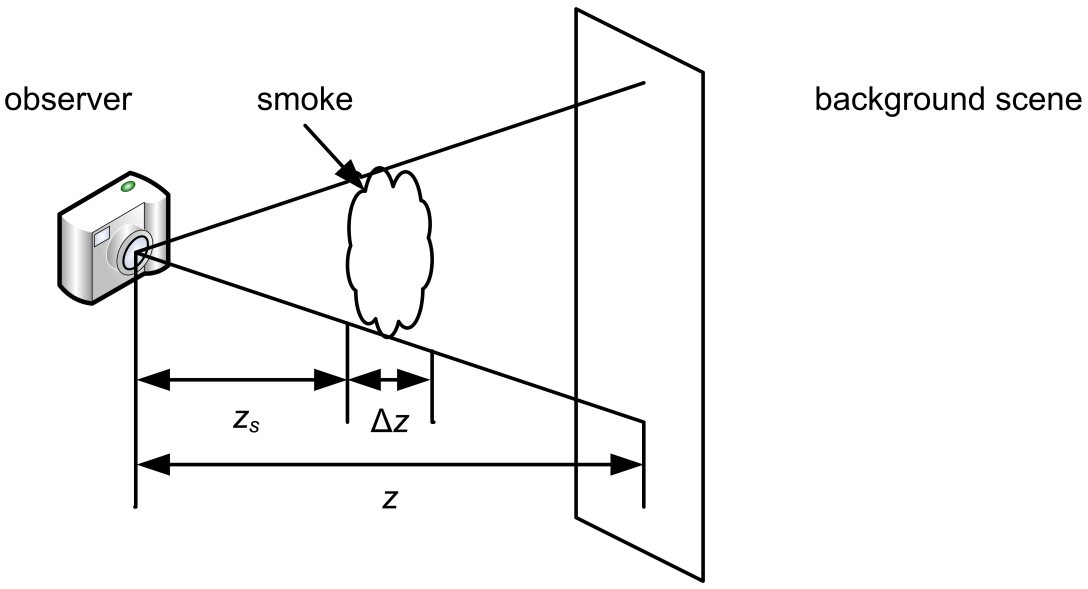


Figure 3.3: Smoke usually appears at a certain distance from the observer with limited thickness along the line of sight.

In consideration of the optical settings of the camera and the horizon radiance, the airlight model in the case of smoke can be obtained by integrating Equation (3.10) between $x = z_s$ and $x = z_s + \Delta z$:

$$A(z_s, \lambda) = g \int_{z_s}^{z_s+\Delta z} L_\infty(\lambda)\beta(\lambda)e^{-\beta(\lambda)x}dx = ge^{-\beta(\lambda)z_s}(1 - e^{-\beta(\lambda)\Delta z})L_\infty(\lambda). \quad (3.14)$$

The final spectral irradiance $F(z, \lambda)$ received by the observer can be expressed as the sum of the irradiance $T(z, \lambda)$ of directly transmitted light and the irradiance $A(z, \lambda)$ of airlight:

$$F(z, \lambda) = T(z, \lambda) + A(z, \lambda). \quad (3.15)$$

The image model that accounts for the presence of smoke with limited thickness along the line of sight can be obtained by substituting Equation (3.13) and (3.14) into Equation (3.15):

$$F(z, \lambda) = (1 - \Omega(\Delta z, \lambda))B(z, \lambda) + \Omega(\Delta z, \lambda)S(z_s, \lambda), \quad (3.16)$$

where

$$\begin{aligned} \Omega(\Delta z, \lambda) &= 1 - e^{-\beta(\lambda)\Delta z}; \\ B(z, \lambda) &= \frac{g}{z^2} L_\infty(\lambda)\rho(\lambda); \\ S(z_s, \lambda) &= ge^{-\beta(\lambda)z_s} L_\infty(\lambda). \end{aligned} \quad (3.17)$$

Equation (3.16) is the image formation model when smoke exists. $B(z, \lambda)$ accounts for the background under clear air when there is no smoke. In the scope of this thesis, it is referred to as the background component or non-smoke component interchangeably. $S(z_s, \lambda)$ 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 scattering coefficient $\beta(\lambda)$ and the thickness Δz of 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 does not vary much within a small area perpendicular to the line of sight, $\Omega(\Delta z, \lambda)$ is constant in this local area. In the scope of this thesis it is referred to as the blending parameter. It is noted that the derived model Equation (3.16) indicates an additive relationship between smoke and non-smoke components.

3.5 Summary

Light is attenuated when it travels through smoke. The related effects are described by the attenuation model. Meanwhile, smoke behaves like a source of light, which is accounted for by the airlight model. Both models contribute to the final spectral irradiance which reaches to the camera. Based on the derived image formation model and certain reasonable assumptions, any small-sized image can be expressed as a linear combination of smoke component and background component. This understanding of the smoke imaging process lays the foundation for developing reliable algorithms for vision-based smoke detection.

Chapter 4

Novel texture feature for smoke detection

Due to the dispersive distribution of smoke, different types of texture features have been explored for smoke characterization and detection (see Chapter 2). As a well-known descriptor for texture classification, local binary pattern (LBP) was applied to video-based smoke detection in [187], [185], [97] and has shown a promising performance. However, LBP has some limitations for smoke characterization, which will be identified in this chapter. Based on this understanding, a novel texture feature derived from the original LBP is adopted to better capture the textural characteristics of smoke.

4.1 Local binary pattern

Local binary pattern (LBP), which is an effective descriptor to capture local appearance information, was originally proposed for texture classification [116]. Due to its properties of high discriminative power, robustness under illumination changes, and computational simplicity, it has been applied to smoke detection in [187], [185], [97]. Intuitively, the LBP code of a specified pixel can be obtained by comparing its intensity with those of its

neighbors. Specifically, given a pixel $c = (x_c, y_c)$, its LBP code is defined as follows:

$$LBP_{P,R}(x_c, y_c) = \sum_{p=0}^{P-1} s(g_p - g_c)2^p, \tag{4.1}$$

where g_c represents the gray-scale value of the center pixel c and $g_p(p = 0, ..., P-1)$ represents the gray-scale values of P equally spaced pixels on a circle with radius $R(R > 0)$ which form a circularly symmetric neighbor set, $s(x)$ is a Sign Function which is defined as follows:

$$s(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{otherwise.} \end{cases} \tag{4.2}$$

An illustration of the LBP descriptor is shown in Figure 4.1 where the LBP code of the center pixel (in red color) is obtained by comparing its gray-scale value with the neighboring pixels. The neighboring pixels whose gray-scale values are equal or larger than that of the center pixel are marked as "1", otherwise as "0". In this case, all 8 neighboring pixels of the center pixel are considered, which means that P and R are 8 and 1 respectively.

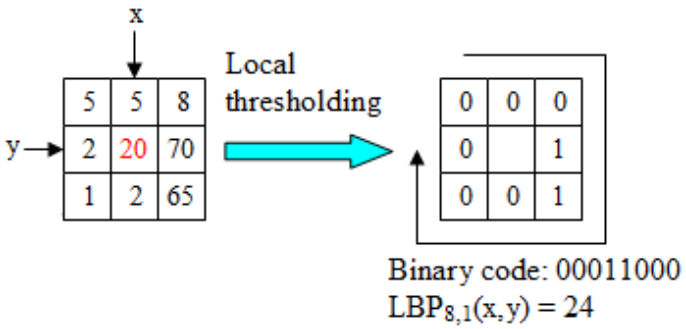


Figure 4.1: An illustration of the LBP descriptor.

As can be noticed, LBP is named to reflect the effect of the operator, viz., a local neighborhood is transformed into a binary pattern with respect to the gray-scale value of the center pixel. Scanning an image (or image block) in pixel-wise manner, the corresponding LBP codes for all the pixels are accumulated into a discrete histogram called LBP histogram. The LBP histogram distribution is considered as a feature vector to represent the texture of the image (or image block). Obviously, the dimension of the feature

vector is equal to the number 2^P of $LBP_{P,R}$ histogram bins.

One typical type of LBP is so-called uniform LBP which was defined in [116]. The term "uniform" refers to the uniform structure of the LBP, i.e., there are a limited number of transitions or discontinuities in the circular presentation of the pattern. Specifically it ensures that there are at most two bitwise transitions between 0 and 1 in a LBP code when the bit pattern is considered circularly. One significant property of uniform LBPs is that they often represent primitive structures of texture while non-uniform LBPs normally correspond to unexpected noises and hence are less discriminative. For instance, uniform LBPs account for almost 90 percent of all the texture patterns observed in typical images when P and R are 8 and 1 respectively. The reason why the histogram of uniform LBPs provides better discrimination in comparison with the histogram of all individual LBPs lies in the difference in their statistical properties. The relative proportion of non-uniform LBPs with respect to all LBPs is so small that their probabilities cannot be estimated reliably. As a result, treating all non-uniform LBPs individually in histogram computation would deteriorate the performance on the dissimilarity analysis of samples. Based on this understanding, all non-uniform LBPs are stored in one single bin in histogram computation. Another advantage of introducing uniform LBPs is to reduce the dimension of the feature vector. For example, the dimension of the feature vector is 59 when $P = 8$ and $R = 1$. In the rest of this chapter, the superscript $u2$ will be used to denote the corresponding uniform patterns. Illustratively, uniform LBP is represented by LBP^{u2}.

4.2 Non-redundant local binary pattern

Although the original LBP has been widely used in applications, two drawbacks were identified in [112]. One is that the original LBP is sensitive to the relative changes between background and foreground. As can be seen from Figure 4.2, the LBP codes of two selected pixels in these two images are quite different from each other. However, they describe the same spatial characteristics of smoke. The other disadvantage of the original LBP is the requirement about storage. The original $LBP_{8,1}$ requires 256 histogram bins for feature extraction. Even if $LBP^{u2}_{8,1}$ is adopted, 59 bins are still needed. Less storage

is preferred for real-time smoke detection. Taking these two points into consideration, we employ non-redundant local binary pattern (NRLBP) [112] to capture the textural information of smoke. As can be noticed in Figure 4.2, the LBP codes of the selected pixels from the two images are complementary to each other (i.e. the sum of them is $2^P - 1$). Motivated by this fact, NRLBP is defined as follows:

$$NRLBP_{P,R}(x_c, y_c) = \min\{LBP_{P,R}(x_c, y_c), 2^P - 1 - LBP_{P,R}(x_c, y_c)\}. \quad (4.3)$$

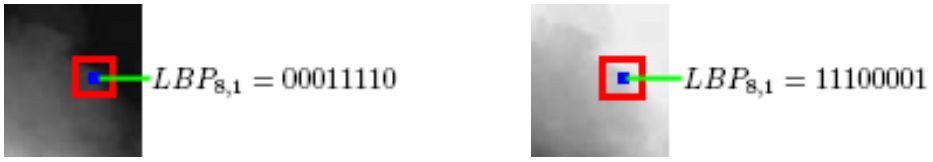


Figure 4.2: Two images represent the same smoke structure with inverted background and foreground.

In other words, the original LBP code and its complement are considered as the same in this case. Thus they correspond to the same bin in histogram distribution, which means that the dimension of feature vector is reduced by half. In addition, as NRLBP reflects the relative contrast between background and foreground, it is more robust and discriminative compared with the original LBP.

Although NRLBP has more discriminative power and lower storage requirement in comparison with the original LBP, it is still a spatial texture descriptor which only encodes local appearance information. For video-based smoke detection, motion information of smoke should be utilized as well to represent the temporal characteristics of smoke. Motivated by the work done by Viola et al. [161] in which features were computed on the difference image between consecutive frames to encode the motion patterns of pedestrians, non-redundant local motion binary pattern (NRLMBP) is adopted here. It is a NRLBP computed on the difference image between consecutive frames. In this case, the motion characteristics of smoke is encoded by NRLMBP.

4.3 Experimental results

To verify the effectiveness of NRLBP and NRLMBP for smoke characterization, video-based smoke detection experiments were conducted. The framework of the detection algorithm followed a typical pattern recognition approach with preprocessing, feature extraction and classification. At the preprocessing stage, background subtraction was used to perform foreground detection. As adaptive Gaussian Mixture Model (GMM) [143] can deal well with lighting changes, repetitive motion from clutter, and long-term scene changes, it was adopted to perform background modeling. Block-based processing was another technique employed at this stage. Each image frame was divided into non-overlapped blocks of the same size. After background subtraction using GMM, the number of foreground pixels in each block could be calculated. If the amount of foreground pixels in one block exceeds a threshold, this block will be considered as a candidate block. NRLBP and NRLMBP were extracted in candidate blocks at the stage of feature extraction. After block-based features were extracted, the feature vectors of positive and negative samples were used to train a classifier which were further employed to classify new blocks. A Kernel Support Vector Machine (SVM) was selected as the classifier.

4.3.1 Experimental setup

As there is no representative databases for evaluating the performance of smoke detection algorithms to date, some video clips [a] were chosen to test the adopted features. These video clips cover both indoor and outdoor scenarios. Some are with true smoke events, while others are with moving objects such as pedestrians and vehicles. In order to label smoke samples efficiently, a labeling tool was developed and its graphical user interface is shown in Figure 4.3. We take one video clip as an example to describe the rules and procedures employed in manually labeling frames of a video as ground truth. Each frame of the original video clip will be divided into non-overlapped blocks with the size of

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

16×16 pixels. Note the selection of block size is essential for early detection. A larger size of block is not suitable for early detection as smoke usually covers a very small area at the very beginning. Additionally, the computational cost for processing a larger-sized block will be high. Inversely, if the size of block is too small, there will be many disturbances from other foreground objects which may lead to plenty of misclassifications. Next, every block is manually labeled as smoke or other foreground objects. The smoke blocks are also labeled into different types of smoke (e.g. heavy and light). It should be noted that all the labeled smoke and non-smoke used in the experiments in the scope of this thesis are available on request.

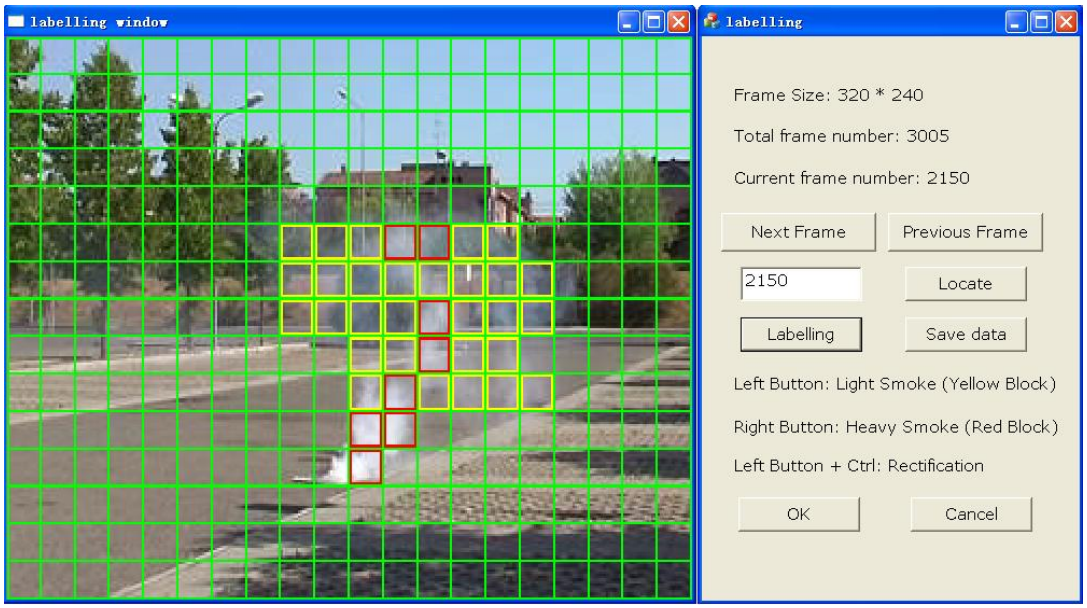


Figure 4.3: The graphical user interface of the labeling tool.

Uniform LBP-based features LBP^{u2}, $NRLBP^{u2}$, and $NRLMBP^{u2}$ were extracted for each block. In terms of texture classification, a higher accuracy was noticed when more neighboring pixels were involved in the calculation of the LBP code in [116], which indicates that a larger P may lead to a better textural representation of smoke and thus achieve a better detection accuracy. In consideration of potential real-time application, for all the LBP-based features we set $P = 8$ and $R = 1$ to improve the computation efficiency. It means that given a pixel of interest, only the nearest pixels in its circular neighborhood are used to calculate the corresponding features. The experimental results are presented in the following section.

4.3.2 Results and discussions

In our experiments, 5091 smoke and 4083 non-smoke image blocks with the size of 16×16 pixels were selected to construct the training set. The testing set is composed of 563 smoke blocks and 818 non-smoke blocks. Four basic measures true positive (TP), true negative (TN), false positive (FP), and false negative (FN) are defined to calculate performance indicators:

- *TP*: true smoke is detected as smoke
- *TN*: true non-smoke is detected as non-smoke
- *FP*: true non-smoke is detected as smoke
- *FN*: true smoke is detected as non-smoke

The performance indicators detection rate (DR), false alarm rate (FAR), and misclassification rate (MR) are defined as:

$$DR = \frac{TP}{TP + FN} \tag{4.4}$$

$$FAR = \frac{FP}{FP + TN} \tag{4.5}$$

$$MR = \frac{FP + FN}{TP + TN + FP + FN} \tag{4.6}$$

The block-based detection results are reported in Table 4.1. Although there is a trade-off between detection rate and false alarm rate, misclassification rate can be independently regarded as an overall indicator of the algorithm performance. The LBP^{u2} case is considered as a baseline. As can be seen from Table 4.1, lower misclassification rate was achieved when $NRLBP^{u2}$ was used compared with the baseline. This is probably because NRLBP is more robust and adaptive to relative changes of background and foreground. In terms of capturing appearance information of smoke, NRLBP outperformed

LBP. When the feature vectors of NRLBP and NRLMBP were concatenated to construct a spatial-temporal descriptor, the lowest misclassification rate among all the three cases was obtained. This demonstrates that NRLMBP is effective in capturing the motion information of smoke.

Table 4.1: The block-based detection results

	LBP^{u2}	$NRLBP^{u2}$	$NRLBP^{u2} + NRLMBP^{u2}$
DR (%)	77.49	80.21	79.52
FAR (%)	0.4712	0.5020	0.4227
MR (%)	0.7427	0.7396	0.6698

For real video-based smoke detection, smoke alarm should be issued based on the whole image or frame instead of block. However, a frame should not be classified as having smoke just because of the existence of one isolated "smoke" block in the frame. Based on this understanding, a frame is deemed to have smoke if at least two neighboring blocks in a circular neighborhood are detected as smoke blocks. The frame-based detection results are reported in Table 4.2. The case when LBP^{u2} was used is also considered as a baseline. Similarly, reduced misclassification rate was obtained by using $NRLBP^{u2}$ in comparison with the baseline case, which echos the robustness of NRLBP in the respect of characterizing texture. In addition, when NRLBP and NRLMBP were combined to encode both appearance and motion information, the best performance in terms of all the three indicators was achieved.

Table 4.2: The frame-based detection results

	LBP^{u2}	$NRLBP^{u2}$	$NRLBP^{u2} + NRLMBP^{u2}$
DR (%)	97.23	96.58	97.69
FAR (%)	2.696	1.165	1.165
MR (%)	2.724	2.063	1.622

Some scenarios of the video clips tested in the experiments are shown in Figure 4.4 and 4.5. Moreover, some true positive examples and false positive examples are indicated by red block in Figure 4.4 and 4.5 respectively. Although the adopted features are suitable for the case where the relative changes between background and foreground exist, they may fail to differentiate smoke from some homogeneous non-smoke objects.

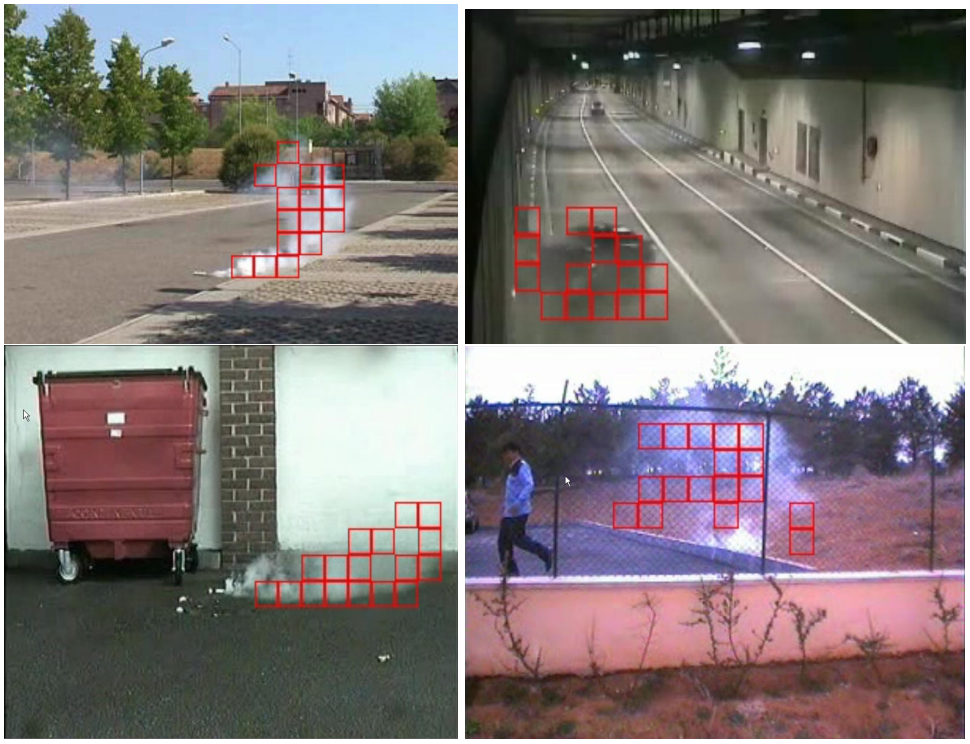


Figure 4.4: True positive examples (The block images indicated by red rectangle in each image represent the detected smoke blocks).

Figure 4.5: False positive examples (The block images indicated by red rectangle in each image represent the detected smoke blocks).

As can be noticed, NRLBP can be calculated with a few comparisons in a small neighborhood and a look-up table and, hence, it is computationally efficient. The dimensions of the related feature vectors are shown in Table 4.3. In comparison with LBP^{u2}, improved detection performance can be achieved by integrating $NRLBP^{u2}$ and $NRLMBP^{u2}$ nearly without increasing the storage cost.

Table 4.3: The dimensions of feature vectors

LBP^{u2}	$NRLBP^{u2}$	$NRLBP^{u2} + NRLMBP^{u2}$
59	30	60

4.4 Summary

To overcome the sensitivity of the original LBP to the relative changes between background and foreground, NRLBP was employed to describe the texture of smoke in a discriminative and compact way in this chapter. Additionally, NRLMBP, which is a NRLBP extended to the temporal domain, was adopted to encode the motion patterns of smoke in video. As a spatial-temporal representation of smoke, NRLBP and NRLMBP were combined and applied to smoke detection in video. Experimental results have demonstrated that they yielded superior detection performance in comparison with the original LBP.

Chapter 5

Local smoothness model for smoke detection

Based on the understanding that background covered by smoke can deteriorate the quality of the features for smoke characterization and detection in video, this chapter presents a novel method for separating the smoke component from the background such that visual features can be extracted from the smoke component for reliable smoke detection.

5.1 Motivation

Texture is a key feature of smoke. It is much more reliable than other features such as color and motion for the purpose of detection. However, it can be severely corrupted when the smoke is light and, hence, the background is visible. Opaque objects, such as human faces and vehicles, will block the visibility of 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 the covered scene (background) is not visible. 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 Figure 5.1 illustrate the possible effects that

different background images can have on the features extracted from light smoke.

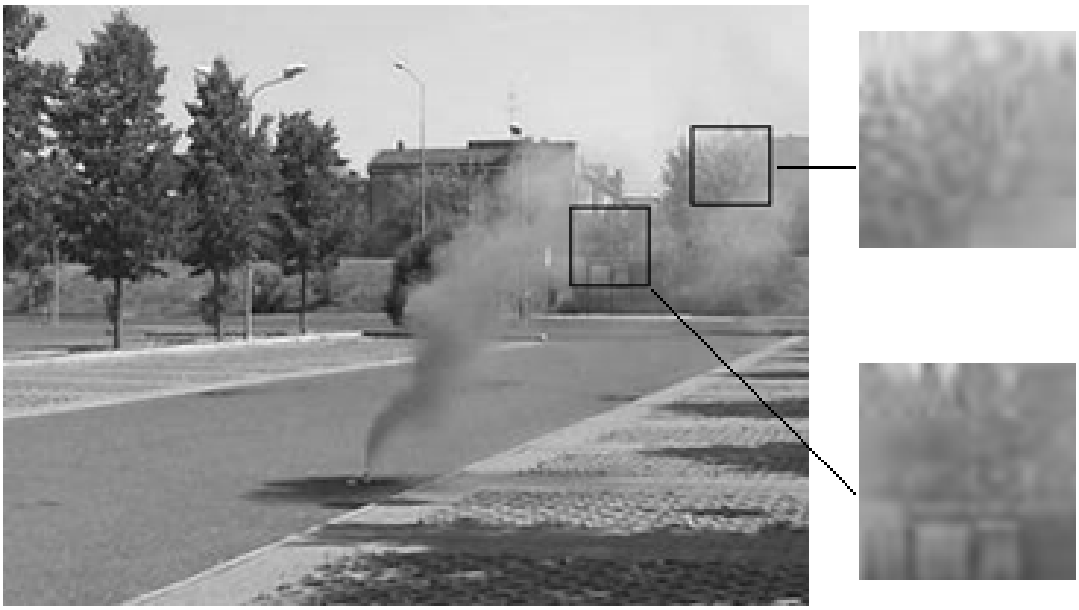


Figure 5.1: Two image blocks covered by light smoke.

It is expected that the adverse effect of background image can be removed and texture features can be extracted 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. Considering the state-of-the-art background subtraction methods [140], 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 5.2 illustrates the presence of residual background information when a simple background subtraction method is used.

According to the smoke imaging process in Chapter 3, any small-sized image can be expressed as a linear blending of a smoke component and a background component (see Equation (3.16)). The additive relationship between the smoke and background components indicates a possibility to separate the smoke component.

(a) background

(b) given video frame

(c) difference image

Figure 5.2: Background subtraction (c) between a smoke image (b) and its background (a). The background information is clearly noticed in the difference image.

5.2 Problem formulation

At 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 smoke can also be localized. 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 chapter focuses on this problem.

Let $\mathbf{f}_t \in \mathbb{R}^N$ be an image block with N pixels at time t, $\mathbf{b}_t \in \mathbb{R}^N$ be the corresponding background under clear air (or as if no smoke exists), and $\mathbf{s}_t \in \mathbb{R}^N$ be the corresponding smoke component. 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 $\mathbf{b}_t$ according to the image formation model expressed by Equation (3.16):

$$\mathbf{f}_t = \alpha_t \mathbf{s}_t + (1 - \alpha_t)\mathbf{b}_t + \mathbf{n}_t, \tag{5.1}$$

where $\mathbf{n}_t \in \mathbb{R}^N$ represents modeling noise, $\alpha_t \in [0, 1]$ is the blending weight at time t. From Equation (3.17) as shown below again for easy reference, it is apparent that the blending parameter α_t depends on the scattering coefficient $\beta(\lambda)$ of smoke and the thickness Δz of the smoke along the line of sight.

$$\alpha_t = 1 - e^{-\beta(\lambda)\Delta z}.$$

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, α_t is a constant within the image block, and the quantity is referred to as the blending parameter on block level here. Note that it may vary from block to block. For brevity and without loss of generality, the subscript t will be dropped hereafter.

Assuming $\mathbf{f}$ is acquired by a stationary camera, background modeling techniques such as Gaussian mixture model (GMM) [144] 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:

$$\min_{\alpha,\mathbf{s}} \|\mathbf{f} - \alpha\mathbf{s} - (1-\alpha)\mathbf{b}\|_2^2 \quad s.t. \quad \alpha \in [0,1]. \tag{5.2}$$

Equation (5.2) is under-determined because there are N equations but $N+1$ free variables for gray-scale images. Additional constraints are required in order to find a solution. As the background $\mathbf{b}$ can vary significantly from one application to another and smoke has relatively consistent visual characteristics, a model, namely, the local smoothness is proposed to constrain the smoke component $\mathbf{s}$ in the next section. Based on this model, the solution to Equation (5.2) is derived.

5.3 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:

$$\min_{\alpha,\mathbf{s}} \|\mathbf{f} - \alpha\mathbf{s} - (1-\alpha)\mathbf{b}\|_2^2 + \lambda\Theta_s \quad s.t. \quad \alpha \in [0,1], \tag{5.3}$$

where λ is a weighting parameter which trades off the residual error and the smoothing constraint expressed by Θ_s. Specifically, the variance between neighboring pixels is employed to measure the smoothness here. Thus Equation (5.3) can be written as:

$$\min_{\alpha,\mathbf{s}} \|\mathbf{f} - \alpha\mathbf{s} - (1-\alpha)\mathbf{b}\|_2^2 + \lambda \sum_{i=1}^{N} \sum_{j\in\Omega_i} (s_i - s_j)^2 \quad s.t. \quad \alpha \in [0,1], \tag{5.4}$$

where s_i and s_j are the ith and jth entries of $\mathbf{s}$ respectively. Ω_i is a small neighborhood centered at the ith pixel. By defining a matrix $\mathbf{T} \in \{-1, 0, 1\}^{M \times N}$, we can rewrite the second term as:

$$\Theta_s = \sum_{i=1}^{N} \sum_{j \in \Omega_i} (s_i - s_j)^2 = \|\mathbf{Ts}\|_2^2 = \mathbf{s}^T \mathbf{T}^T \mathbf{Ts} = \mathbf{s}^T \mathbf{As}, \tag{5.5}$$

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. For example, if the neighborhood is composed of four nearest neighbors horizontally and vertically, $M = 2\sqrt{N}(\sqrt{N} - 1)$. Illustratively, for an image patch of interest with 4×4 pixels, the specific $\mathbf{T}$ and $\mathbf{A}$ have the following forms:

$$\mathbf{T} = \begin{bmatrix}
1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \\
1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1
\end{bmatrix}$$

$$\mathbf{A} = \begin{bmatrix}
2 & -1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
-1 & 3 & -1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & -1 & 3 & -1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & -1 & 2 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
-1 & 0 & 0 & 0 & 3 & -1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & -1 & 0 & 0 & -1 & 4 & -1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & -1 & 0 & 0 & -1 & 4 & -1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & -1 & 0 & 0 & -1 & 3 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 3 & -1 & 0 & 0 & -1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & -1 & 4 & -1 & 0 & 0 & -1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & -1 & 4 & -1 & 0 & 0 & -1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & -1 & 3 & 0 & 0 & 0 & -1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 2 & -1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & -1 & 3 & -1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & -1 & 3 & -1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & -1 & 2
\end{bmatrix}$$

Substituting Equation (5.5) into Equation (5.4) yields

$$\min_{\alpha,\mathbf{s}} \|\mathbf{f} - \alpha\mathbf{s} - (1-\alpha)\mathbf{b}\|_2^2 + \lambda\mathbf{s}^T\mathbf{A}\mathbf{s} \quad s.t.\ \alpha \in [0,1], \tag{5.6}$$

which can be solved through alternate optimization with regard to α and $\mathbf{s}$ respectively. Each is a convex problem and the convergence of the optimization is guaranteed [11]. First, we solve for $\mathbf{s}$ by fixing α. In this case, Equation (5.6) can be written as a quadratic function of $\mathbf{s}$:

$$\min_{\mathbf{s}} \mathbf{s}^T(\alpha^2\mathbf{I} + \lambda\mathbf{A})\mathbf{s} - 2\mathbf{s}^T(\mathbf{f} - \mathbf{b} + \alpha\mathbf{b})\alpha + (\mathbf{f} - \mathbf{b} + \alpha\mathbf{b})^T(\mathbf{f} - \mathbf{b} + \alpha\mathbf{b}), \tag{5.7}$$

where $\mathbf{I} \in \mathbb{R}^{N \times N}$ denotes the identity matrix. An analytical solution $\hat{\mathbf{s}}$ with respect to α can be derived by differentiating Equation (5.7) with respect to $\mathbf{s}$ and setting the derivatives to zero. That is,

$$\hat{\mathbf{s}} = (\alpha^2 \mathbf{I} + \lambda \mathbf{A})^{-1} \alpha (\mathbf{f} - \mathbf{b} + \alpha \mathbf{b}). \tag{5.8}$$

Next, we solve for α by fixing $\mathbf{s}$. In this case, Equation (5.6) can be written as a quadratic function of α:

$$\min_{\alpha} (\mathbf{b} - \mathbf{s})^T (\mathbf{b} - \mathbf{s}) \alpha^2 + 2(\mathbf{b} - \mathbf{s})^T (\mathbf{f} - \mathbf{b}) \alpha + (\mathbf{f} - \mathbf{b})^T (\mathbf{f} - \mathbf{b}) \quad s.t. \ \alpha \in [0, 1]. \tag{5.9}$$

The minimizer of the quadratic function is:

$$\alpha^* = \frac{(\mathbf{b} - \mathbf{s})^T (\mathbf{f} - \mathbf{b})}{(\mathbf{b} - \mathbf{s})^T (\mathbf{s} - \mathbf{b})}. \tag{5.10}$$

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

$$\hat{\alpha} = \begin{cases} 0 & \text{if } \alpha^* \leq 0 \\ \alpha^* & \text{if } 0 < \alpha^* < 1 \\ 1 & \text{if } \alpha^* \geq 1. \end{cases} \tag{5.11}$$

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

5.4 Smoke detection framework

Based on the image separation method presented in the previous section, a novel block-based smoke detection framework is proposed and depicted in Figure 5.3. 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 work, the non-overlapped division is adopted. For each frame block $\mathbf{f}$ and its associated background $\mathbf{b}$, the blending parameter α and smoke component $\mathbf{s}$ are computed using the local smoothness model presented in the previous section. To detect whether $\mathbf{f}$ 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), $\mathbf{f}$ 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 $\mathbf{s}$ is then passed on to the next step in which LBP [116] feature is extracted from $\mathbf{s}$ and the extracted feature is input to a binary support vector machine (SVM) classifier. A decision is made on whether $\mathbf{s}$ 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.

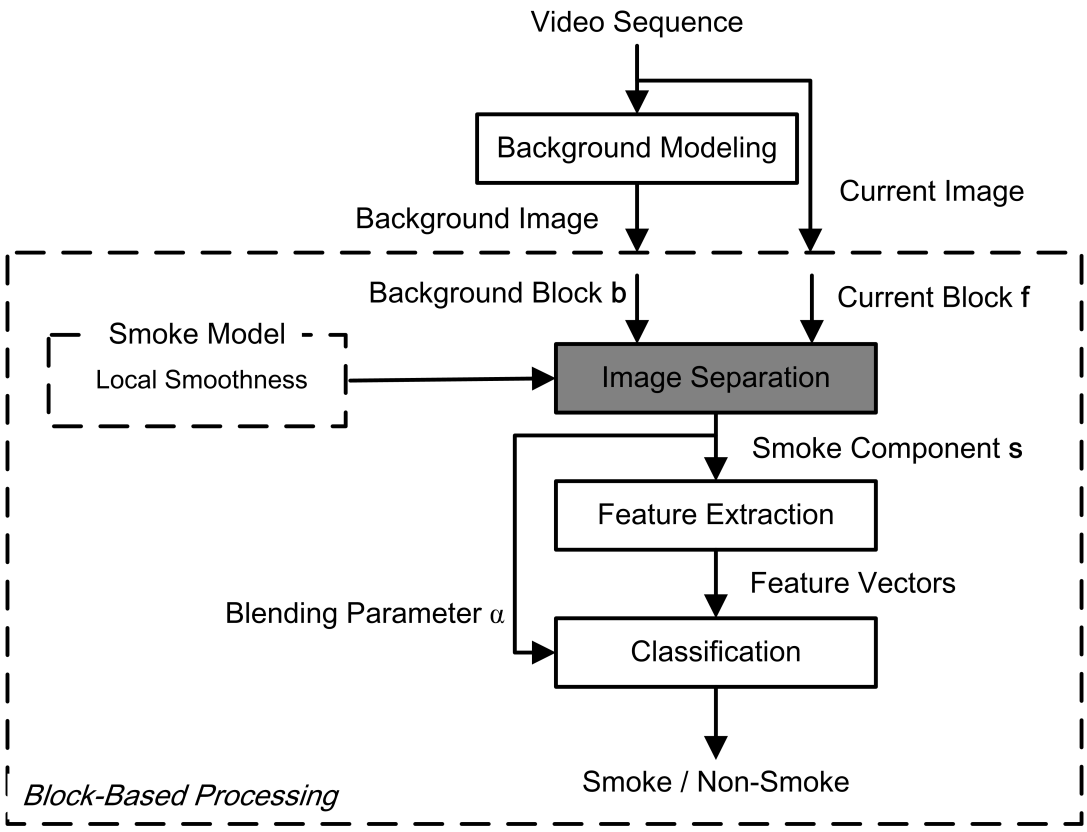


Figure 5.3: Proposed framework for video-based smoke detection using local smoothness for smoke separation.

The proposed smoke detection method consists of three major steps: background modeling, separation of smoke component $\mathbf{s}$ and classification of $\mathbf{s}$. Most computation

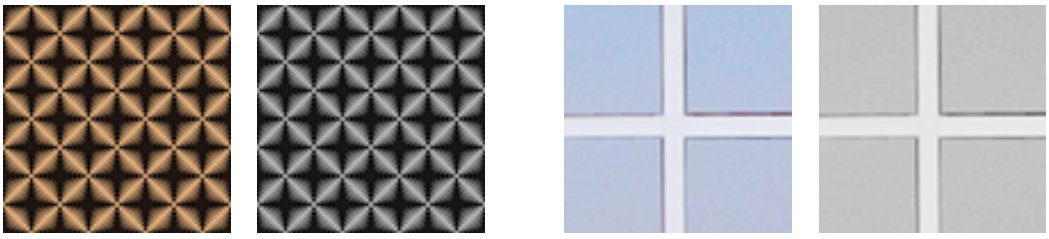
is spent on the separation of smoke component $\mathbf{s}$, where the blending parameter α and the smoke component $\mathbf{s}$ are alternately calculated using Equation (5.10) and (5.8). The complexity of this step is $O(\aleph N^3)$ where N is the number of pixels in each block image and $\aleph$ is the number of alternations.

5.5 Experimental results

In this section, the effectiveness of the proposed method for smoke separation is verified using synthesized images. Furthermore, smoke detection performance based on image separation using real video images is provided as well.

5.5.1 Image separation on synthesized images

Two image patches with 100 by 100 pixels were chosen as background image samples. As shown in Figure 5.4, the images in Figure 5.4(a) represent color and gray-scale texture background respectively while the ones in Figure 5.4(b) represent color and gray-scale piece-wise smooth background respectively. Note that the following experiments in this chapter were conducted on gray-scale images.



(a) Texture (b) Piece-wise smooth

Figure 5.4: Background image blocks.

As for pure smoke, both black and white smoke were considered in the experiments. Two types of pure smoke image patches with 100 by 100 pixels and their corresponding gray-scale images are shown in Figure 5.5.

Using these representative background and pure smoke image blocks, 40 image blocks were synthesized with different α values to construct a test data set of smoke images,

(a) Black smoke (b) White smoke

Figure 5.5: Pure smoke image blocks.

which is shown in Figure 5.6. For each combination, α value ranged from 0.1 to 1.0 with an increment of 0.1.

In the experiments, each image in the test data set was considered as an input image block $\mathbf{f}$. Given the corresponding background image block $\mathbf{b}$, the blending parameter α and smoke component $\mathbf{s}$ were estimated according to Equation (5.11) and (5.8). For the purpose of visual comparison, the corresponding separated smoke components for all the images in the test data set are shown in Figure 5.7. As can be noticed, there is high visual similarity between the separated smoke components and the ground truth pure smoke.

Next, quantitative comparison was performed to further verify the effectiveness of the proposed method for smoke separation. Specifically, the experiments were divided into four groups according to different combinations of background and pure smoke. In each group, ten pairs of α and $\mathbf{s}$ were obtained. As we have the ground truth of the blending parameters, the absolute difference between the calculated and the ground truth α values was adopted as an indicator of whether the separation is successful or not. The average absolute differences of α versus termination condition are shown in Figure 5.8. Here for computational simplicity, termination condition is defined as the absolute difference between the α values calculated in two consecutive iterations. A more sophisticated termination condition can be defined based on the difference between the smoke components estimated in two consecutive iterations. As can be seen, for four groups of experiments, if the termination condition is set as 10^{-4}, the average absolute differences of α are rather small. In addition, it is expected that the textural pattern of pure smoke can be preserved in the separated smoke component. To measure the degree of similarity between the estimated smoke component and the ground truth in terms of textural pattern, normalized

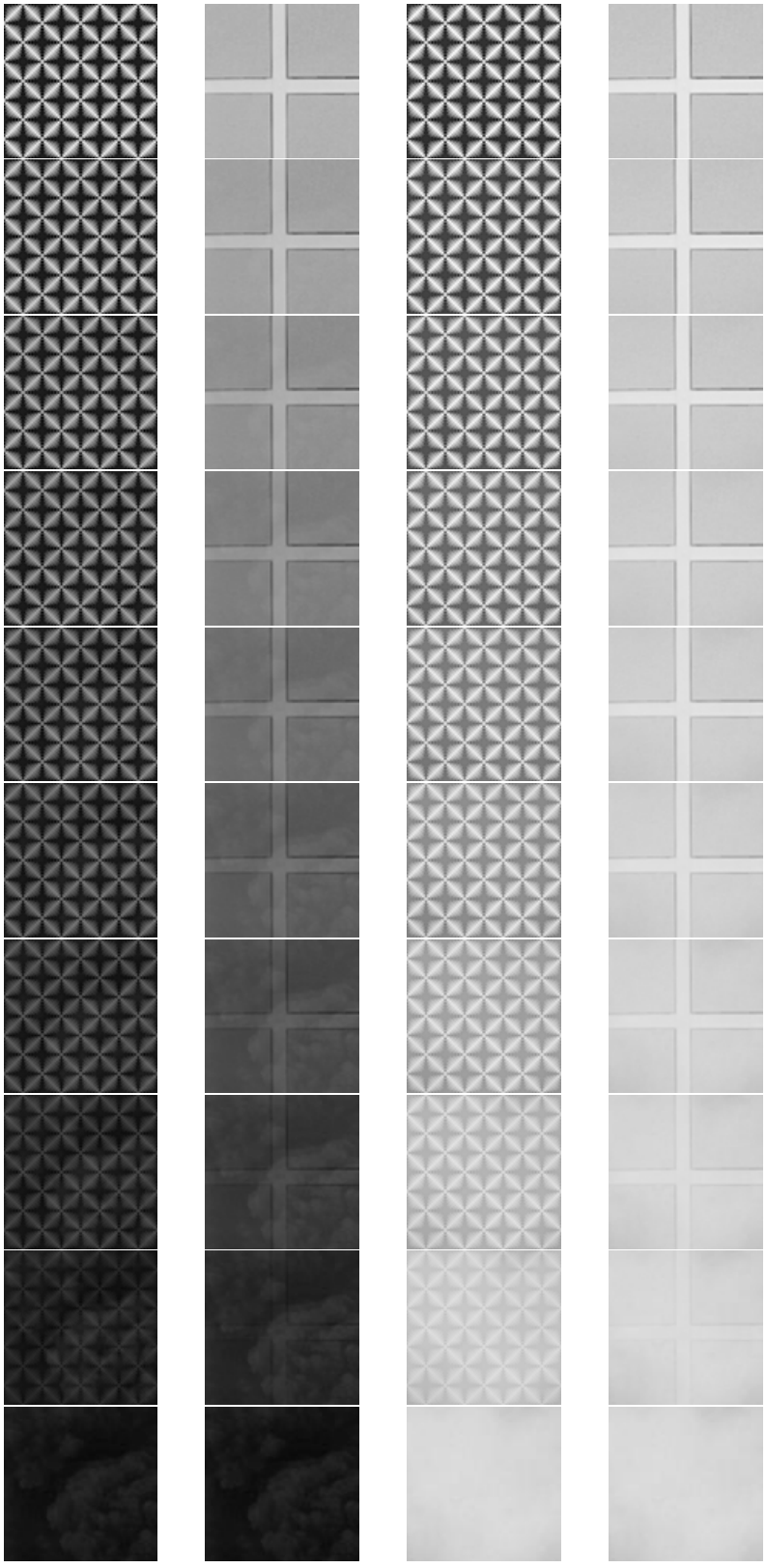


Figure 5.6: The test data set of synthesized smoke images (From left to right, the images in each column were synthesized from (1) texture background and black smoke, (2) piece-wise smooth background and black smoke, (3) texture background and white smoke, (4) piece-wise smooth background and white smoke respectively).

Figure 5.7: The separated smoke components corresponding to all the images in the test data set shown in Figure 5.6.

cross correlation (NCC) [41] was adopted, which is a widely used method for template matching. Specifically given an estimated smoke component $\mathbf{s}'$ and its corresponding ground truth $\mathbf{s}$, NCC is defined as follows:

$$NCC = \frac{1}{N}\sum_{i=1}^{N}\frac{(s_i - \bar{\mathbf{s}})(s_i' - \bar{\mathbf{s}'})}{\sigma_s \sigma_{s'}}, \tag{5.12}$$

where s_i and s_i' are the ith entries of $\mathbf{s}$ and $\mathbf{s}'$ respectively; $\bar{\mathbf{s}}$ and $\bar{\mathbf{s}'}$ are the average of $\mathbf{s}$ and $\mathbf{s}'$ respectively; σ_s and $\sigma_{s'}$ are the standard deviation of $\mathbf{s}$ and $\mathbf{s}'$ respectively. The average NCC values between the estimated smoke component and the ground truth versus termination condition are shown in Figure 5.9. For four groups of experiments, when the termination condition is set as 10^{-4}, high NCC values were obtained and this means that the separated smoke components have very similar patterns to the ground truth.

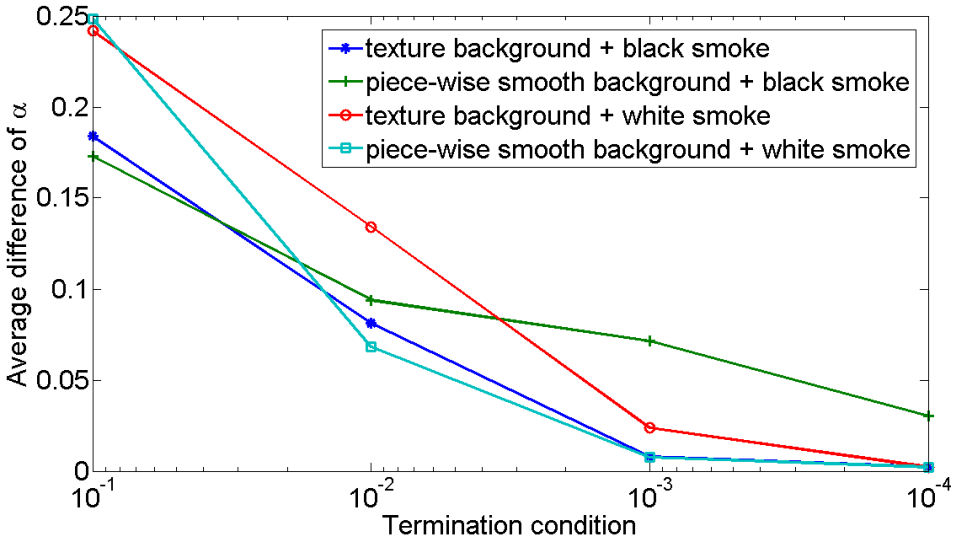


Figure 5.8: Average absolute differences of α versus termination condition.

Although smoke component can be well separated from true background using the proposed method, background images can have noise due to several reasons including imperfect background modeling in real scenarios. Thus in order to validate the proposed image separation method for practical application, some experiments involving noisy background images were conducted. Here, additive white Gaussian noise (AWGN) with zero mean was added to the ground truth background images shown in Figure 5.4 to simulate noisy background images. Specifically, texture and piece-wise smooth back-

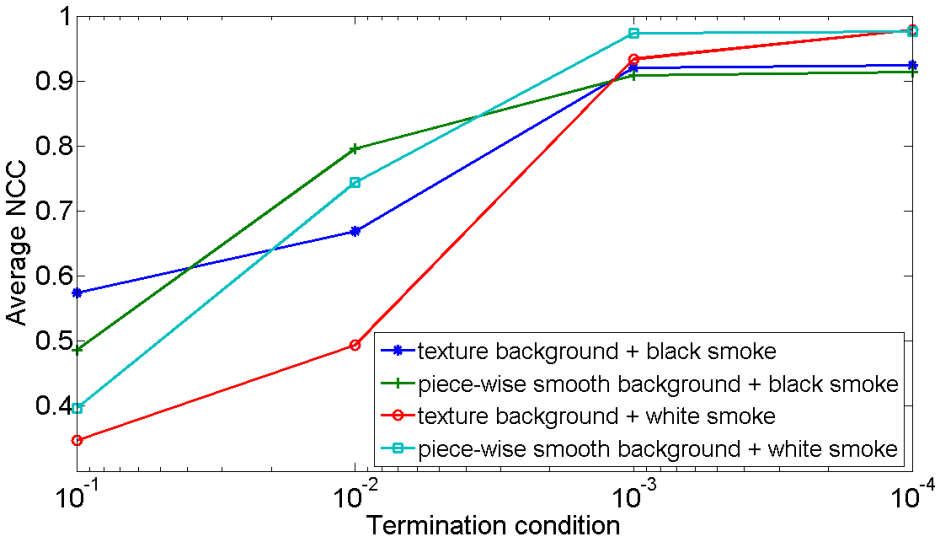
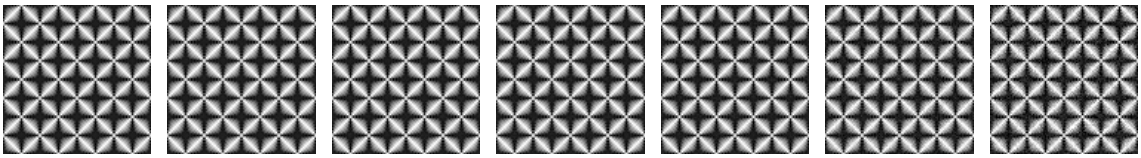
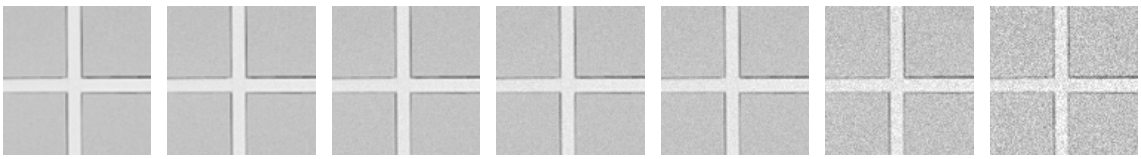


Figure 5.9: Average NCC values versus termination condition.

ground images with AWGN of different standard deviations ranging from 1 to 15 were constructed and shown in Figure 5.10. Taking these images as $\mathbf{b}$ and the ones in Figure 5.6 as $\mathbf{f}$, image separation experiments were conducted once again. It is noted that the termination condition (i.e. the absolute difference between the α values calculated in two consecutive iterations) was set as 10^{-4} in these experiments. The average absolute difference between the calculated and the ground truth α values versus the standard deviation are plotted in Figure 5.11. And the curves of average NCC between the estimated smoke component and the ground truth versus the standard deviation are shown in Figure 5.12. As can be noticed, when the level of AWGN is higher (e.g. standard deviation is 15), the increased average difference of α and the decreased average NCC values suggest that the performance of smoke separation is deteriorated. However, background modeling techniques are relatively reliable in real applications and the standard deviation is around 5 (see 6.3.4). In this case, the average difference of α and average NCC values are still acceptable, which means that the separated smoke components would have the similar texture patterns to the ground truth.



(a) Noisy texture background



(b) Noisy piece-wise smooth background

Figure 5.10: Background images with the AWGN (for each type of background, from left to right, the standard deviation of the AWGN is 1, 2, 3, 4, 5, 10, and 15 respectively).

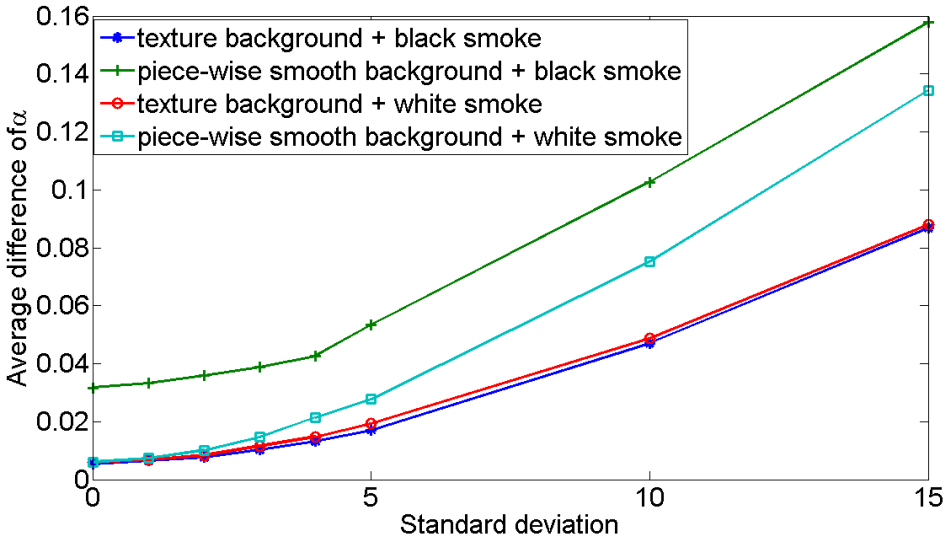


Figure 5.11: Average absolute differences of α versus standard deviation.

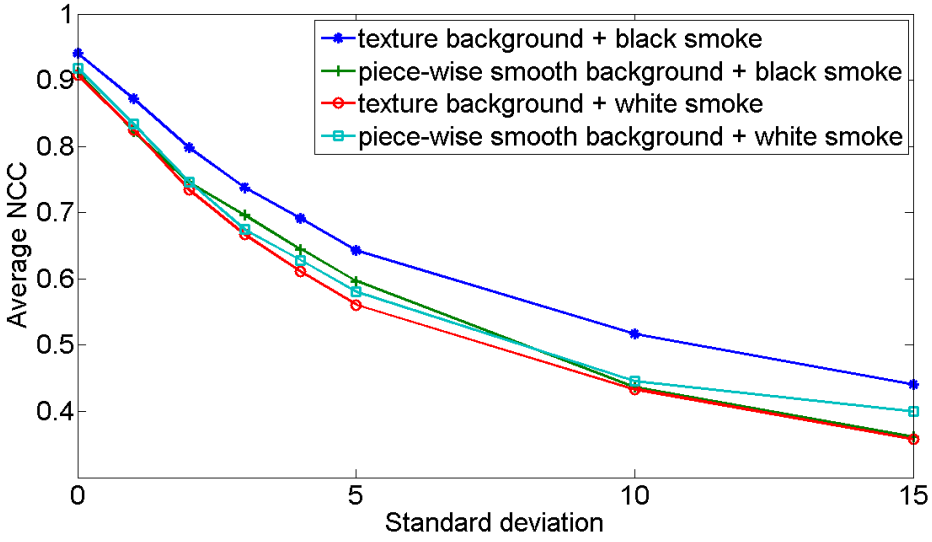


Figure 5.12: Average NCC values versus standard deviation.

5.5.2 Smoke detection on real video images

In order to serve early smoke detection and localization, block images with 16 by 16 pixels were involved in the detection experiments. Totally 8000 block images were collected from 20 video clips of smoke [a]. The selected video clips cover indoor and outdoor, short and long distance surveillance scenes with different illuminations. Specifically half of the collected blocks are test images and the rest are the corresponding learned background images. Furthermore, for the 4000 test block images, there are 1000 light smoke, 1000 heavy smoke, 1000 other foreground objects and 1000 almost the original background respectively. Illustratively, some non-smoke and smoke block images used in the experiments are shown in Figure 5.13 and 5.14 respectively. A classical adaptive GMM [144] was selected to perform background modelling. As LBP [116] has several properties such as powerful discrimination, low computational complexity, and less sensitivity to variation of illumination, it was chosen to extract texture features of smoke here. And kernel SVM based on radial basis function (RBF) was adopted as a classifier. When the cost parameter and γ parameter in SVM settings were equal to 1 and 0.5 respectively, the best classification results could be obtained and are reported here.

Figure 5.13: Examples of non-smoke block images collected from real videos.

Smoke detection rates based on different images are shown in Table 5.1. Take light

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

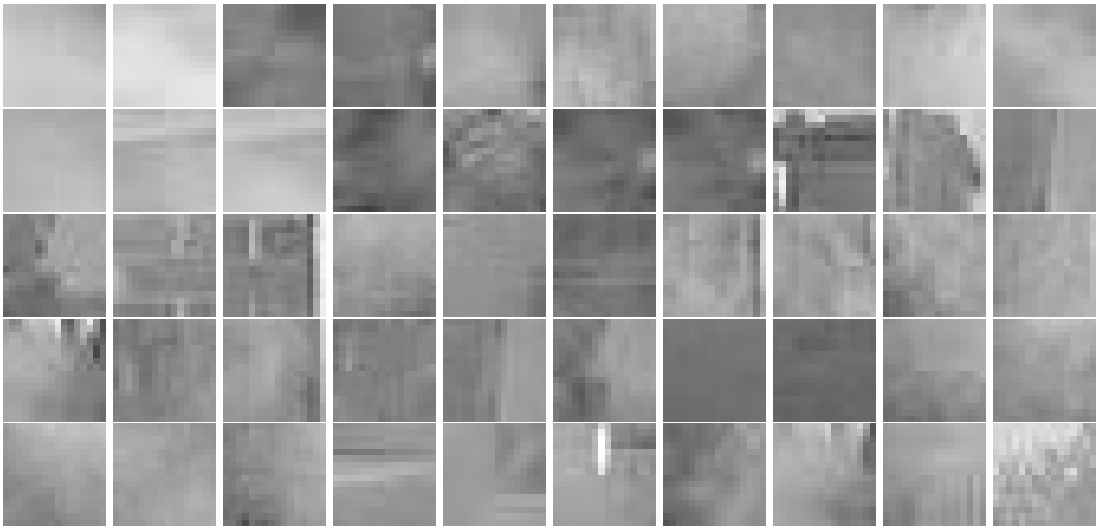


Figure 5.14: Examples of smoke block images collected from real videos.

Table 5.1: Smoke detection rates according to different images

	Light smoke	Heavy smoke	Overall
Test image $\mathbf{f}$	68.1	93.2	80.7
Smoke component $\mathbf{s}$	**94.2**	**98.6**	**96.4**

smoke for example, compared to the case when feature extraction and classification is performed on test image $\mathbf{f}$, significant improvement has been achieved if the same operation is performed on smoke component $\mathbf{s}$ obtained by using the proposed image separation method. Even for heavy smoke, some improvement was observed by performing the detection on the separated smoke component $\mathbf{s}$. Overall, more than 15% improvement on the detection rate was achieved by means of the proposed method. For the 1000 block images of other foreground objects, they include black cars, white shirts and other objects with homogeneous appearance. The false alarm rate is 4.9% when smoke component $\mathbf{s}$ was used for smoke detection, compared to 5.1% when image $\mathbf{f}$ was used. One fact should be noted that, for all the false alarms in the experiments, the solved α is equal to 1. As for the 1000 test image blocks which are almost the original background, false alarm rate perfectly reduces to zero. The reason is that all the solved α values are nearly 0 in this case and they could not pass the α checking.

5.6 Summary

In order to eliminate the adverse effects of background on smoke detection, a novel method has been proposed for smoke detection in videos by separating smoke from background in this chapter. Motivated by the image formation model for smoke, given a video frame and its background, the estimation of the blending parameter and the actual smoke component has been formulated and solved as an optimization problem. Both the blending parameter and smoke component have been used for smoke detection. The effectiveness of the proposed method has been validated by experimental results.

To capture the characteristics of the smoke component, we impose a "local smoothness" constraint on it. Actually other types of constraints can be taken into account to achieve better separation and thus improve the detection performance.

Chapter 6

Sparse representation model for smoke detection

In the previous chapter, improved performance of video-based smoke detection was achieved by separating the smoke component from background. However, the local smoothness model is a heuristic constraint for the smoke component and may not be suitable for smoke modeling in some cases. This chapter presents two models to constrain the smoke component by learning from pure smoke samples. Based on these two models adapted to real smoke images, it is expected that the smoke separation is more effective and thus the detection based on the separated smoke component is more reliable.

6.1 Principal component model

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. Here the widely used principal component analysis (PCA) [158] is employed to learn this subspace from 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 direction represented by 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 a subspace that is expected to well represent pure smoke images.

Let $\mathbf{P} \in \mathbb{R}^{N \times L} (L < N)$ be the subspace, 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

$$\mathbf{s} = \mathbf{P}\mathbf{y}, \tag{6.1}$$

where $\mathbf{y} \in \mathbb{R}^L$ is the coefficient vector of projecting $\mathbf{s}$ onto the subspace $\mathbf{P}$. Substituting Equation (6.1) into Equation (5.2) yields

$$\min_{\alpha, \mathbf{y}} \|\mathbf{f} - \alpha\mathbf{P}\mathbf{y} - (1-\alpha)\mathbf{b}\|_2^2 \quad s.t. \quad \alpha \in [0, 1]. \tag{6.2}$$

Notice that Equation (6.2) is a quadratic function of α (or $\mathbf{y}$) when $\mathbf{y}$ (or α) is fixed. Similar to the previous chapter, we can solve for α and $\mathbf{y}$ alternately, and then reconstruct $\mathbf{s}$ through Equation (6.1).

Specifically, when $\mathbf{y}$ is fixed, the current solution $\hat{\alpha}$ for α can be expressed as Equation (5.11) as well, where α^* is the minimizer of the quadratic function of α and can be obtained as

$$\alpha^* = \frac{(\mathbf{b} - \mathbf{P}\mathbf{y})^T(\mathbf{f} - \mathbf{b})}{(\mathbf{b} - \mathbf{P}\mathbf{y})^T(\mathbf{P}\mathbf{y} - \mathbf{b})}. \tag{6.3}$$

When α is fixed, Equation (6.2) can be written as a quadratic function of $\mathbf{y}$:

$$\min_{\mathbf{y}} \alpha^2\mathbf{y}^T\mathbf{P}^T\mathbf{P}\mathbf{y} - 2\alpha\mathbf{y}^T\mathbf{P}^T(\mathbf{f} - \mathbf{b} + \alpha\mathbf{b}) + (\mathbf{f} - \mathbf{b} + \alpha\mathbf{b})^T(\mathbf{f} - \mathbf{b} + \alpha\mathbf{b}). \tag{6.4}$$

Let $\hat{\mathbf{y}}$ be the current solution for $\mathbf{y}$ with respect to α, which can be obtained by differentiating Equation (6.4) with respect to $\mathbf{y}$ and setting the derivatives to zero. That is,

$$\hat{\mathbf{y}} = (\alpha\mathbf{P}^T\mathbf{P})^{-1}\mathbf{P}^T(\mathbf{f} - \mathbf{b} + \alpha\mathbf{b}). \tag{6.5}$$

Then the current solution $\hat{\mathbf{s}}$ with respect to α is

$$\hat{\mathbf{s}} = \mathbf{P}(\alpha \mathbf{P}^T \mathbf{P})^{-1} \mathbf{P}^T (\mathbf{f} - \mathbf{b} + \alpha \mathbf{b}). \tag{6.6}$$

6.2 Sparse representation model

Furthermore, an assumption can be made for pure smoke that they lie in multiple low-dimensional subspaces. In this case, a single linear subspace obtained by PCA may not be sufficient to describe all possible variations of smoke. According to the theory of sparse representation [169], 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 a collection of samples forms 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 an over-complete 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 $\mathbf{D}$;

$$\mathbf{s} = \mathbf{D}\mathbf{x}, \tag{6.7}$$

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 information about which bases and the proportion thereof contribute to the construction of $\mathbf{s}$ from $\mathbf{D}$. Based on Equation (6.7) and the sparseness conditions, Equation (5.2) can be rewritten as follows:

$$\min_{\alpha, \mathbf{x}} \|\mathbf{f} - \alpha \mathbf{D}\mathbf{x} - (1 - \alpha)\mathbf{b}\|_2^2 + \eta \|\mathbf{x}\|_0 \quad s.t. \quad \alpha \in [0, 1], \tag{6.8}$$

where η is a regularization parameter balancing the reconstruction error and the sparseness of $\mathbf{x}$. Since the ℓ_0-norm is non-convex, we follow the common method in the literature by

replacing it with the ℓ_1-norm to make the optimization problem solvable. Thus,

$$\min_{\alpha,\mathbf{x}} \|\mathbf{f} - \alpha\mathbf{D}\mathbf{x} - (1-\alpha)\mathbf{b}\|_2^2 + \eta\|\mathbf{x}\|_1 \quad s.t. \quad \alpha \in [0,1]. \tag{6.9}$$

Similar 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 Equation (5.11), with α^* having the following form:

$$\alpha^* = \frac{(\mathbf{b} - \mathbf{D}\mathbf{x})^T(\mathbf{f} - \mathbf{b})}{(\mathbf{b} - \mathbf{D}\mathbf{x})^T(\mathbf{D}\mathbf{x} - \mathbf{b})}. \tag{6.10}$$

When α is fixed, Equation (6.9) becomes an ℓ_1-regularized least squares problem:

$$\min_{\mathbf{x}} \|\mathbf{f} - \alpha\mathbf{D}\mathbf{x} - (1-\alpha)\mathbf{b}\|_2^2 + \eta\|\mathbf{x}\|_1. \tag{6.11}$$

To solve for $\mathbf{x}$ efficiently, the feature-sign search algorithm [82] is adopted in this thesis. 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, Equation (6.11) reduces to the following quadratic optimization problem:

$$\min_{\bar{\mathbf{x}}} \|\mathbf{f} - \alpha\bar{\mathbf{D}}\bar{\mathbf{x}} - (1-\alpha)\mathbf{b}\|_2^2 + \eta\bar{\mathbf{g}}^T\bar{\mathbf{x}}. \tag{6.12}$$

A solution to this problem can be analytically obtained as:

$$\hat{\bar{\mathbf{x}}} = (\alpha^2\bar{\mathbf{D}}^T\bar{\mathbf{D}})^{-1}[\alpha\bar{\mathbf{D}}^T(\mathbf{f} - \mathbf{b} + \alpha\mathbf{b}) - \eta\bar{\mathbf{g}}/2]. \tag{6.13}$$

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 Equation (6.11) is updated. Once the updated solution satisfies the following optimality conditions, the optimal $\mathbf{x}$ will be obtained.

$$\frac{\partial \|\mathbf{f} - \alpha\mathbf{D}\mathbf{x} - (1-\alpha)\mathbf{b}\|_2^2}{\partial x_i} + \eta g_i = 0 \quad \forall x_i \neq 0$$

$$\left|\frac{\partial \|\mathbf{f} - \alpha\mathbf{D}\mathbf{x} - (1-\alpha)\mathbf{b}\|_2^2}{\partial x_i}\right| \leq \eta \quad \forall x_i = 0 \tag{6.14}$$

Using the optimal $\mathbf{x}$, $\mathbf{s}$ can be calculated by Equation (6.7).

For completeness, the smoke detection framework based on the principal component and sparse representation models is shown in Figure 6.1. Compared to Figure 5.3, the only difference is the models used for smoke separation.

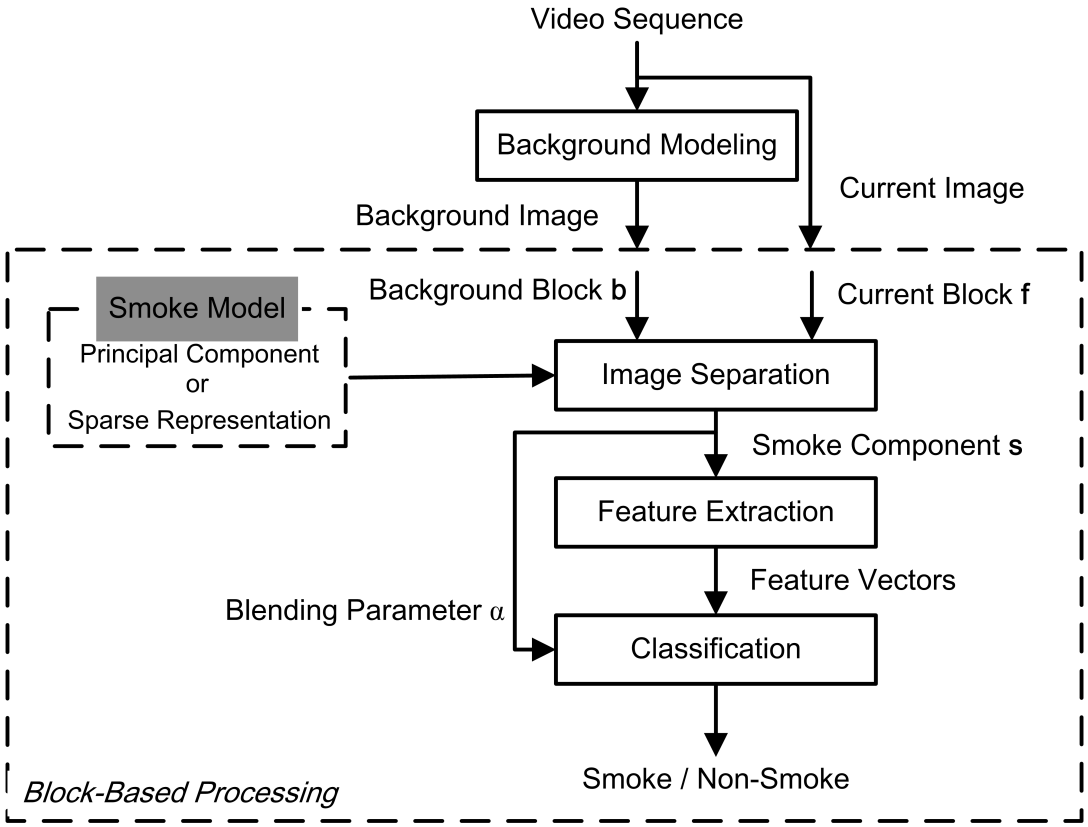


Figure 6.1: Proposed framework for video-based smoke detection using principal component and sparse representation for smoke separation.

6.3 Experimental results

Extensive experiments were conducted using both simulated and real video to evaluate the efficacy of the smoke models and the associated separation and detection algorithms. This section reports the results. Note that in addition to the two models proposed in this chapter to constrain the smoke component, the experimental results relating to the local smoothness model for smoke separation are also provided here for comparison purpose.

6.3.1 Smoke component modeling

The principal component model requires the subspace $\mathbf{P}$ of pure smoke. Similarly, an over-complete dictionary $\mathbf{D}$ is required for the sparse representation model. We present here how $\mathbf{P}$ and $\mathbf{D}$ were constructed from training samples.

Learning subspace $\mathbf{P}$

In order to learn $\mathbf{P}$, one thousand (1000) samples of pure smoke were collected. Each sample is 16×16 pixels and L (i.e. the number of columns in $\mathbf{P}$) was chosen to be much less than 256 ($L << 256$).

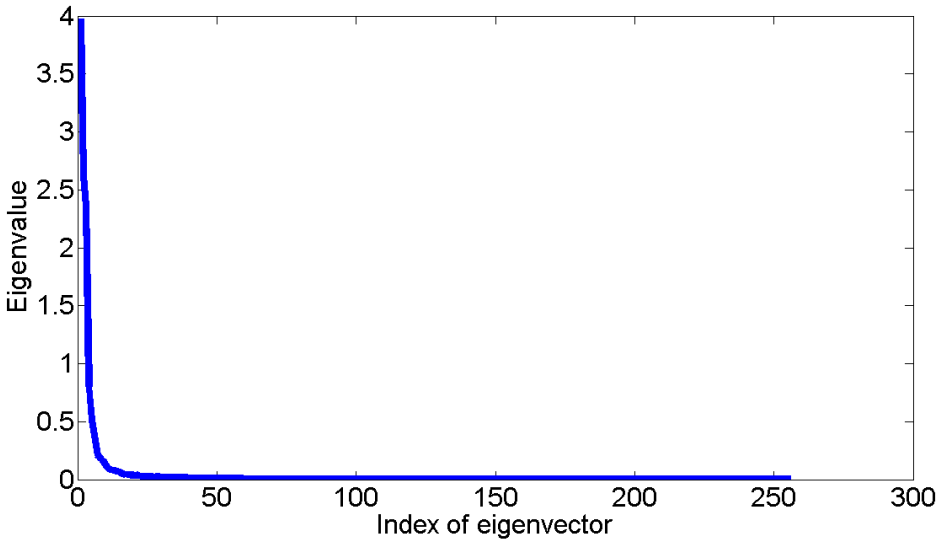


Figure 6.2: Principal component analysis on pure smoke images.

The eigenvalues obtained from the application of PCA to the training samples were ranked (maximum to minimum) and shown in Figure 6.2. The L eigenvectors corre-

sponding to the L largest eigenvalues can be selected to construct $\mathbf{P}$. 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 $\mathbf{P}$. The selected eigenvectors are shown in Figure 6.3.

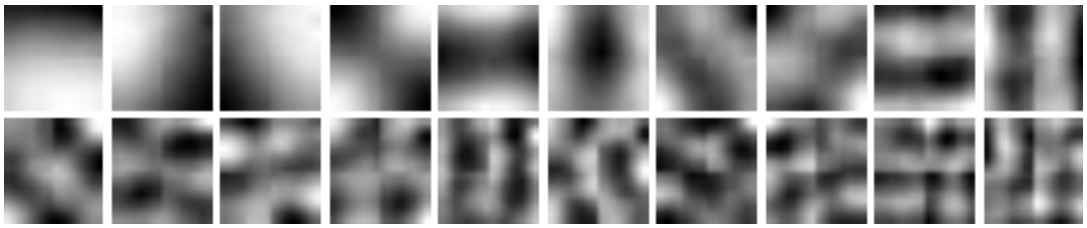


Figure 6.3: The eigenvectors corresponding to the top 20 largest eigenvalues shown as 16×16 elemental patches.

Learning dictionary $\mathbf{D}$

The same one thousand (1000) pure smoke samples were also used for training the dictionary. Additionally, another set of 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.

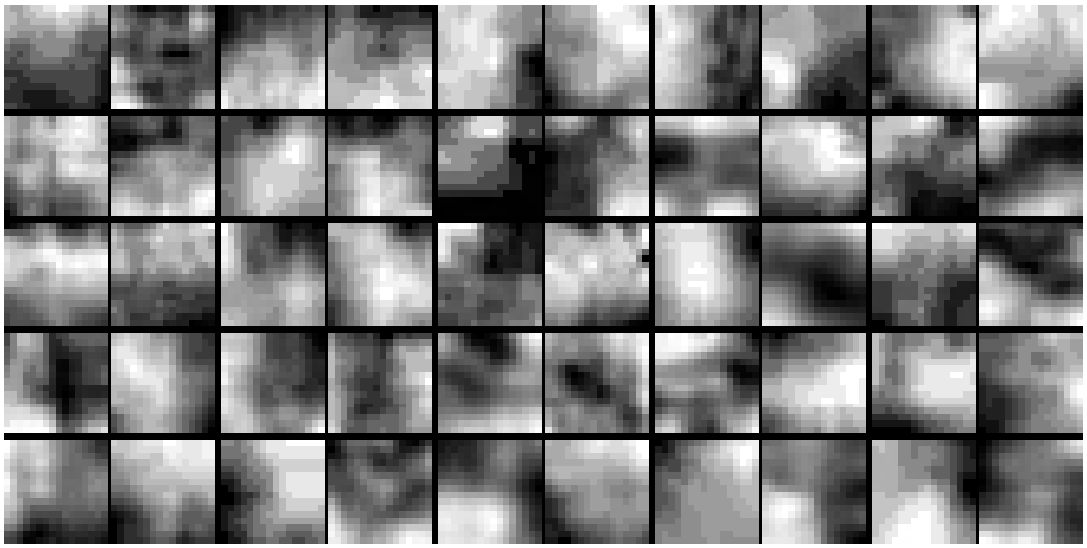


Figure 6.4: 50 basis samples, shown as 16×16 elemental patches, out of the 500 bases in an over-complete dictionary.

The widely-used K-SVD [2] 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 Figure 6.4. The NCC [41] 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:

$$\mathbf{x}_{est} = \underset{\mathbf{x}}{\operatorname{argmin}} \|\mathbf{s}_{ori} - \mathbf{D}\mathbf{x}\|^2 + \beta\|\mathbf{x}\|_1, \tag{6.15}$$

where β is a constant. The reconstructed version of $\mathbf{s}_{ori}$ was obtained as $\mathbf{s}_{rec} = \mathbf{D}\mathbf{x}_{est}$. The graphs shown in Figure 6.5 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.

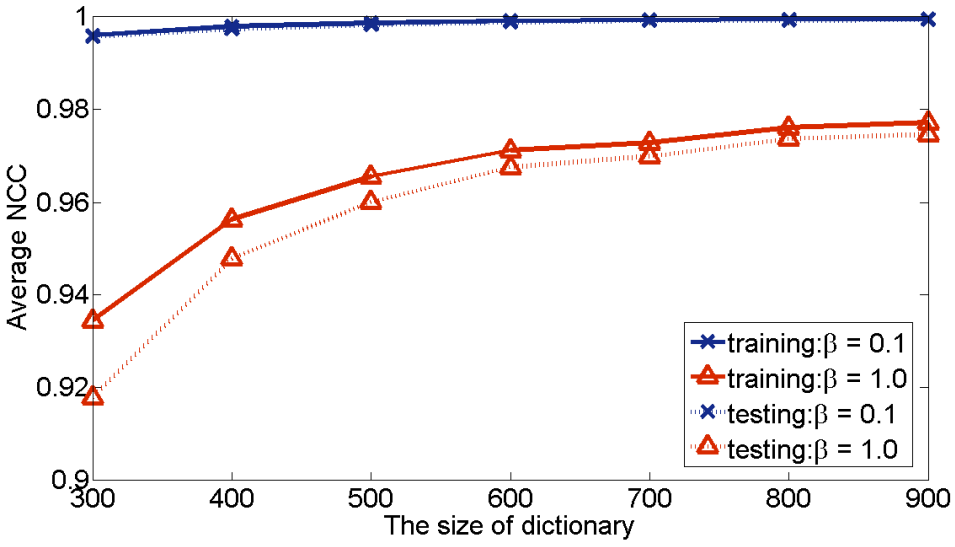


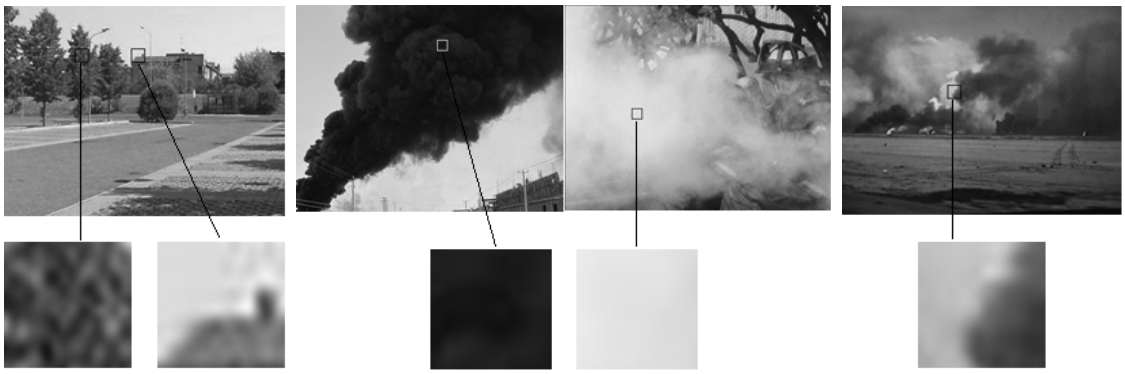
Figure 6.5: Average NCC between the reconstructed and actual pure smoke images.

6.3.2 Performance of smoke separation on synthesized images

In this section, the performance of the proposed image separation methods on synthesized images is reported. In particular, the methods were evaluated on noise-free and noisy background $\mathbf{b}$, and smoke and non-smoke input images $\mathbf{f}$.

Data sets

In order to synthesize realistic smoke image blocks (sized 16×16), we used two types of representative 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 type is gray with large variation among the pixel values. Typical background images are shown in Figure 6.6 (a), while samples of pure smoke are shown in Figure 6.6 (b) and (c) respectively.



(a) background (b) dark and bright smoke (c) gray smoke

Figure 6.6: 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).

Based on these background and pure smoke image blocks, 60 image blocks were synthesized with different α values to construct a test data set of smoke images, which is shown in Figure 6.7. For each combination, α value ranged from 0.1 to 1.0 with an increment of 0.1.

Using true background $\mathbf{b}$

In this experiment, each synthesized image in Figure 6.7 was considered as an input image $\mathbf{f}$, and the corresponding background image in Figure 6.6 (a) was considered as $\mathbf{b}$. In other words, we have true background images. Given the learned subspace $\mathbf{P}$ or over-complete dictionary $\mathbf{D}$ for pure smoke if needed, the blending parameter α and smoke component $\mathbf{s}$ were obtained by solving the optimization problems of Equation 5.6, 6.2,

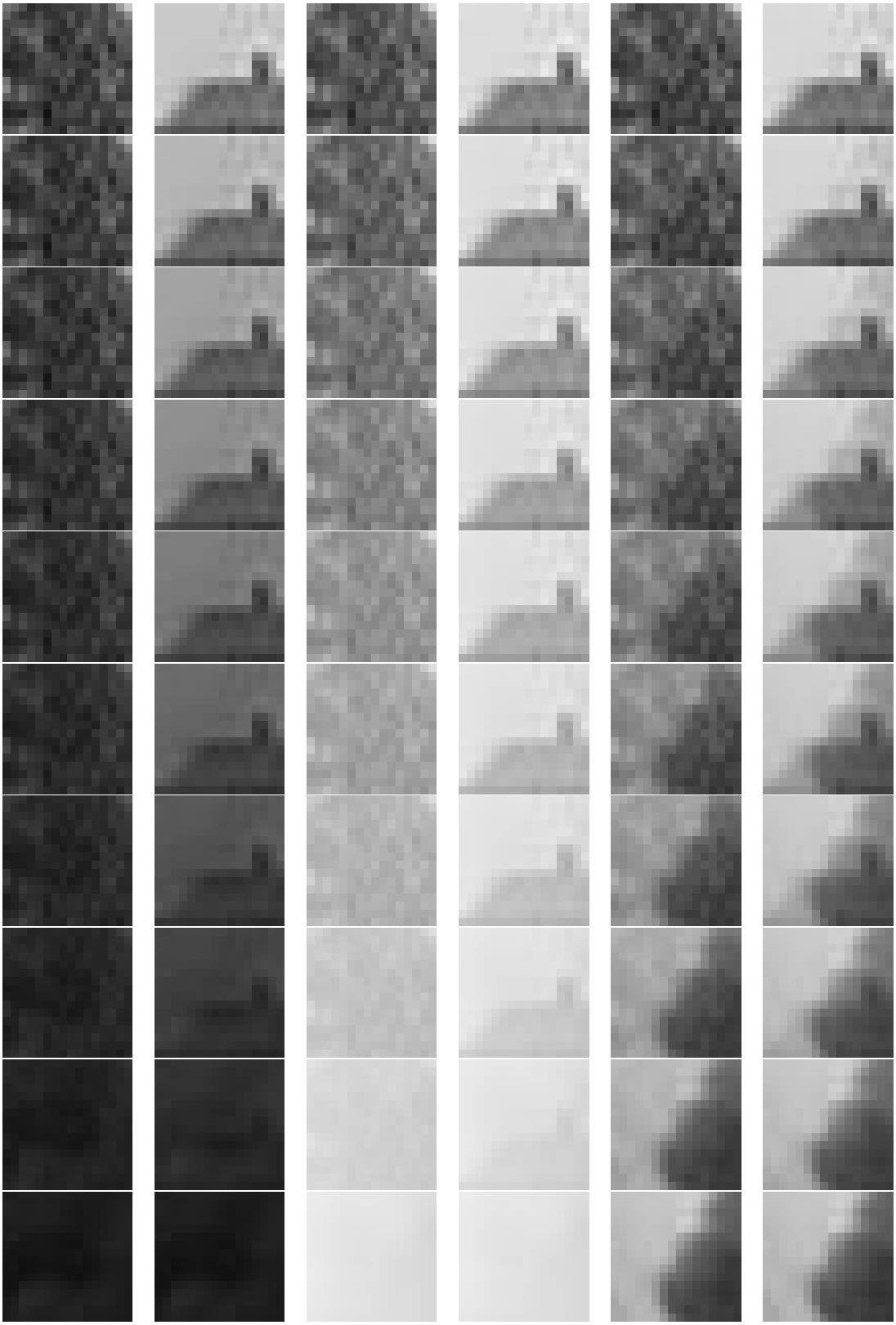


Figure 6.7: The test data set of synthesized smoke images (From left to right, the images in each column were synthesized from (1) texture background and dark smoke, (2) piece-wise smooth background and dark smoke, (3) texture background and bright smoke, (4) piece-wise smooth background and bright smoke, (5) texture background and gray smoke, and (6) piece-wise smooth background and gray smoke respectively).

and 6.9 respectively. For the purpose of qualitative evaluation, the corresponding separated smoke components of all images in the test data set using the sparse representation model are shown in Figure 6.8. Compared with the ground truth shown in Figure 6.6 (b) and (c), the separated smoke components share high visual similarity to them.

To evaluate the accuracy of the estimation of α quantitatively, 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 6.1 and plotted in Figure 6.9. Note that the average absolute difference is rather small. There is an exception in the case where local smoothness was used to separate gray smoke with large variation among the pixel values. This is likely because the large variation violates the local smoothness assumption too much.

Table 6.1: 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 ((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)

	Local smoothness	Principal component	Sparse representation
(1)	0.0032 ± 0.0021	0.0024 ± 0.0015	0.0022 ± 0.0019
(2)	0.0020 ± 0.0015	0.0011 ± 0.0011	0.0019 ± 0.0015
(3)	0.0021 ± 0.0016	0.0094 ± 0.0055	0.0023 ± 0.0017
(4)	0.0022 ± 0.0010	0.0089 ± 0.0042	0.0018 ± 0.0014
(5)	0.1378 ± 0.0884	0.0027 ± 0.0022	0.0337 ± 0.0266
(6)	0.0971 ± 0.0511	0.0015 ± 0.0013	0.0062 ± 0.0027

To quantitatively 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 Figure 6.10 and 6.11). In general a high NCC value, indicative of good separation performance, is achieved from a relatively low α value onwards. Nevertheless, the sparse representation model has led to the best separation performance (highest NCC value) among all three models. Compared with the principal component model, the local smoothness model is more effective when smoke is dense enough.

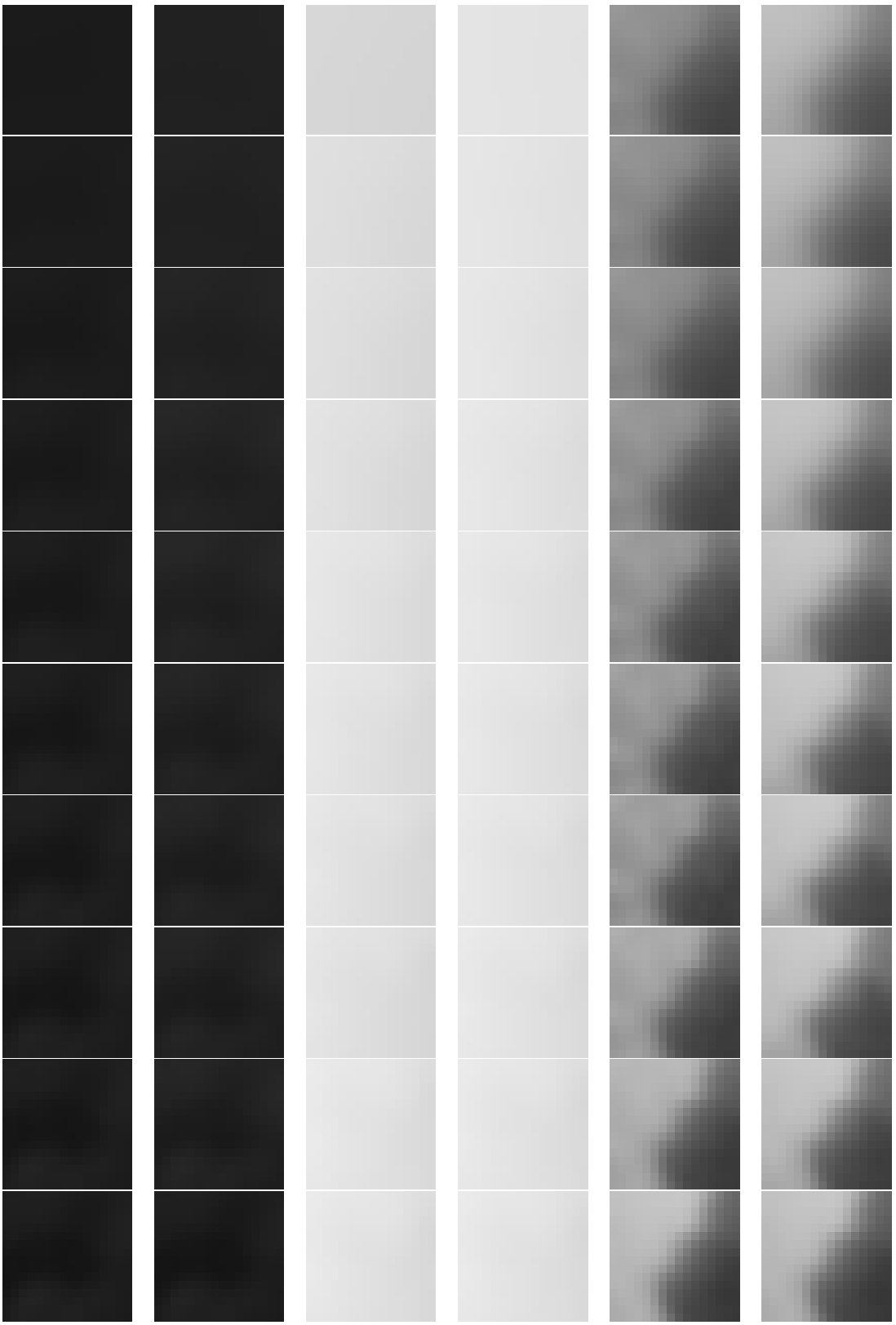


Figure 6.8: The separated smoke components corresponding to all the images in the test data set shown in Figure 6.7.

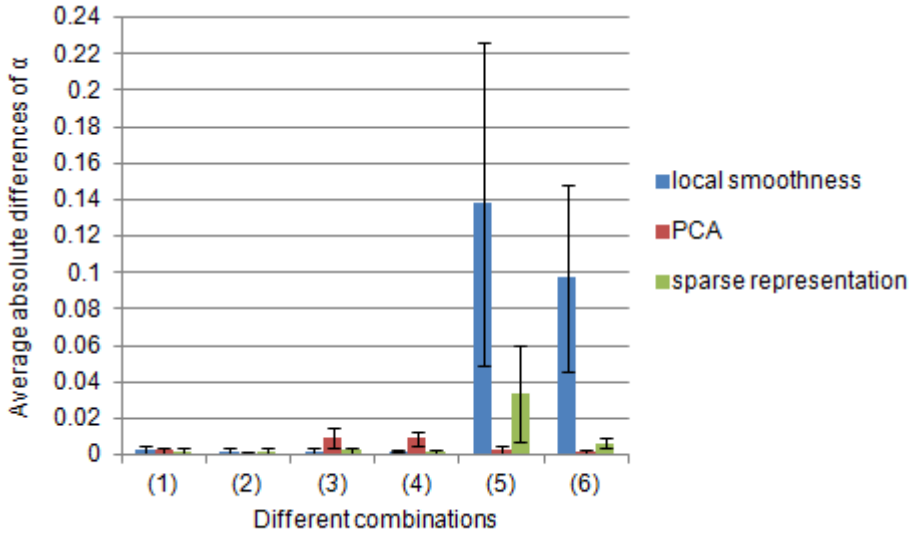
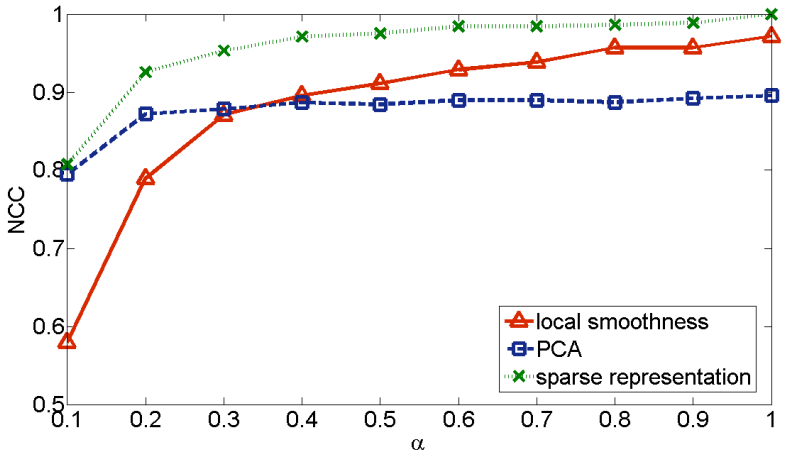


Figure 6.9: Average absolute differences between the estimated α and the actual α for different combinations of background and pure smoke using different models for the smoke component ((1) texture background and dark smoke, (2) piece-wise smooth background and dark smoke, (3) texture background and 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).

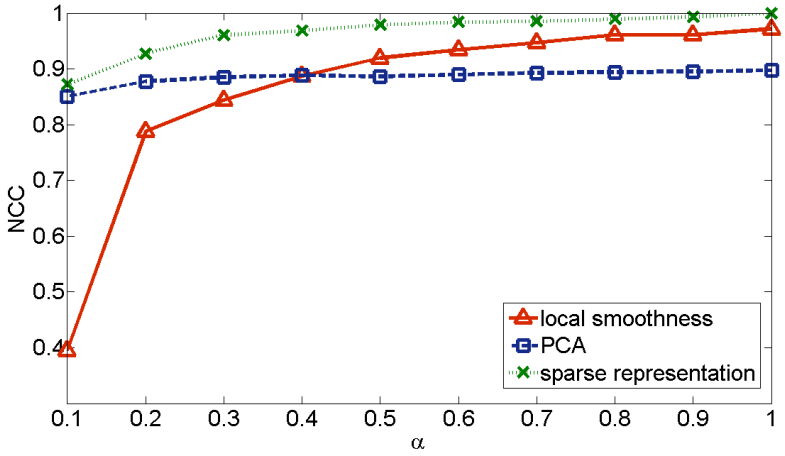
Regardless of which smoke model is used, the following phenomenon is noted in the cases that the pure smoke has small variation. When the blending parameter increases, the separated smoke component becomes more similar to the ground truth with increasing and high NCC values. Similar characteristic is still observed in the cases that the pure smoke has large variation 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 even in the cases that the pure smoke has large variation.

Using noisy background $\mathbf{b}$

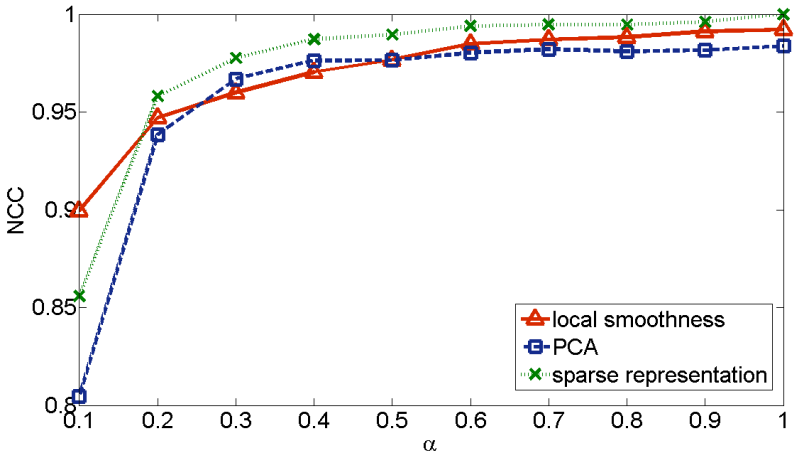
In real scenarios, background images can be noisy for several reasons including imperfect estimation of background. We evaluate the proposed methods under a simulated noisy background by adding to the true background white Gaussian noise with zero mean and standard deviation ranging from 1 to 15 (additive white Gaussian noise (AWGN) model).



(a) texture background and dark smoke

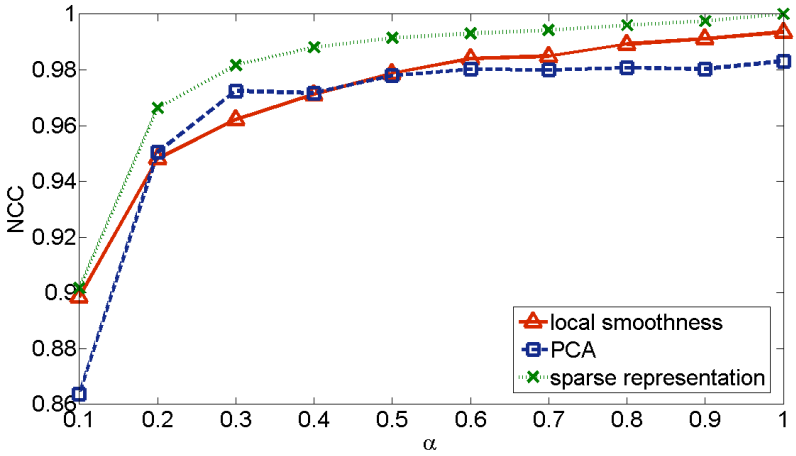


(b) piece-wise smooth background and dark smoke

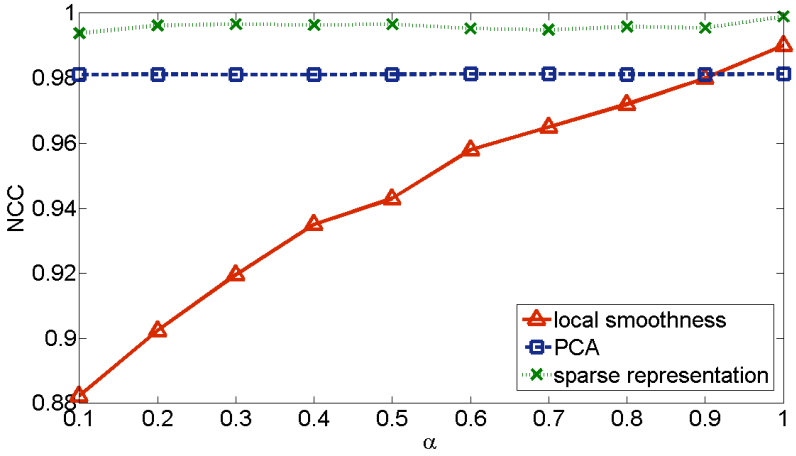


(c) texture background and bright smoke

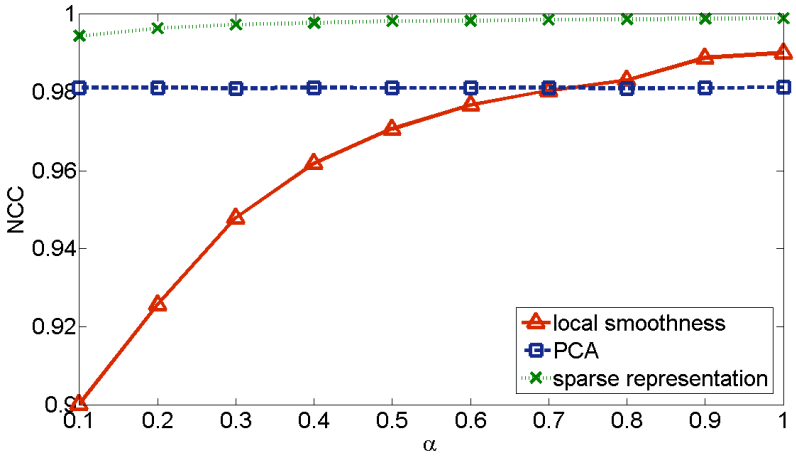
Figure 6.10: NCC between the separated smoke component and ground truth for different combinations of background and pure smoke using different models for the smoke component.



(a) piece-wise smooth background and bright smoke



(b) texture background and gray smoke



(c) piece-wise smooth background and gray smoke

Figure 6.11: NCC between the separated smoke component and ground truth for different combinations of background and pure smoke using different models for the smoke component (cont.).

The noisy background images used for image separation are shown in Figure 6.12. The average absolute differences of α were computed and shown in Figure 6.13 and 6.14, while the average NCC values were computed and shown in Figure 6.15 and 6.16.

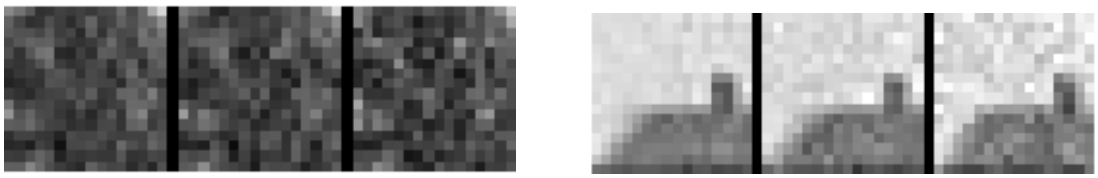


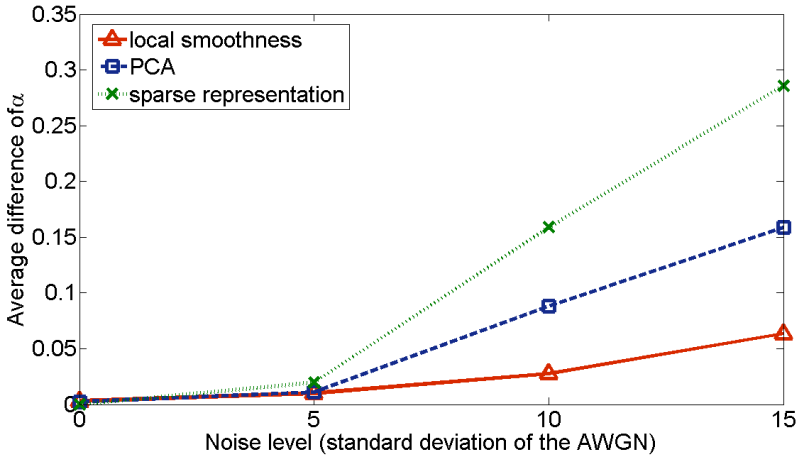
Figure 6.12: Background images with the AWGN (for each type of background, from left to right, the standard deviation of the AWGN is 5, 10, and 15 respectively).

As expected, the performance of the separation methods deteriorated as the noise level (i.e. increasing standard deviation of AWGN) increased. However, the deterioration was 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 separate smoke with large variation. The deterioration was large and remained fairly constant (Figure 6.14(b)-(c) and 6.16(b)-(c)).

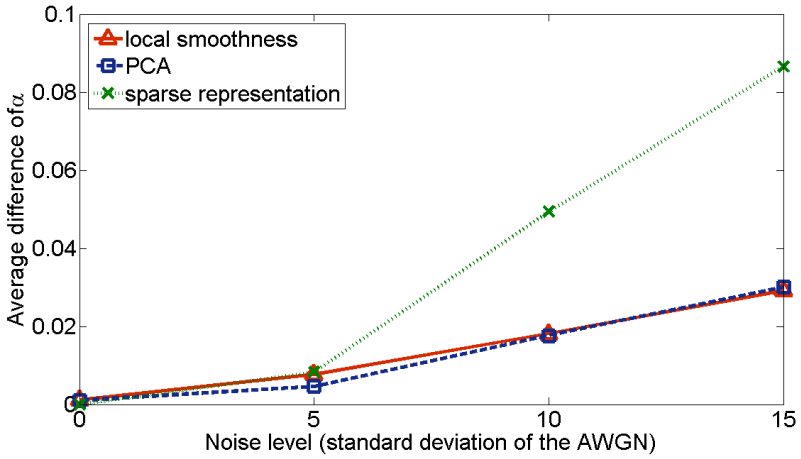
We note here, and this will be demonstrated with real videos in 6.3.4, that the noise level of an estimated background modeled by GMM is usually below a standard deviation of 5. Thus in this case the average difference of α and the average NCC will be relatively constant. It is interesting to note that there is strong visual similarity between two images when their NCC value is larger than 0.3. The trend of the graphs shown in Figure 6.13, 6.14, 6.15, and 6.16 suggests that the separation method based on the local smoothness constraint for smoke component has the highest noise tolerance among all the proposed methods.

Using non-smoke foreground $\mathbf{f}$

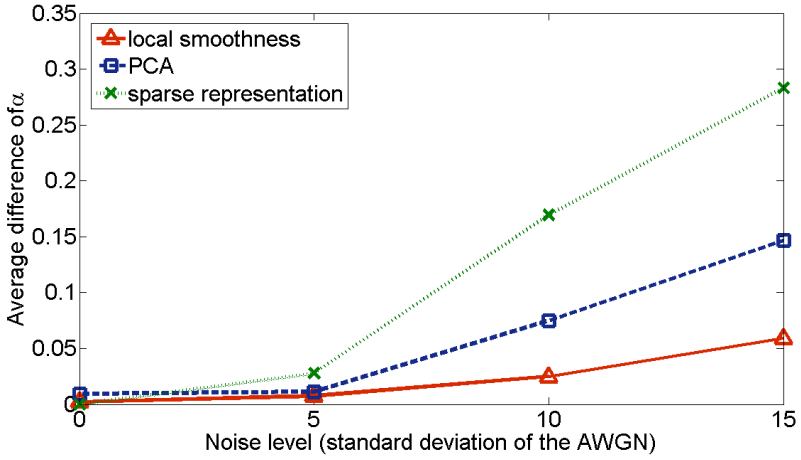
In order to obtain good smoke detection performance, pure smoke must be well separated if indeed smoke exists in the image $\mathbf{f}$. However, the question arises as to the nature of the separated "smoke component" when $\mathbf{f}$ is a non-smoke image; does the separated $\mathbf{s}$ look like smoke or not? In this section, experiments were conducted to evaluate what would be the product after image separation when input image $\mathbf{f}$ is non-smoke.



(a) texture background and dark smoke

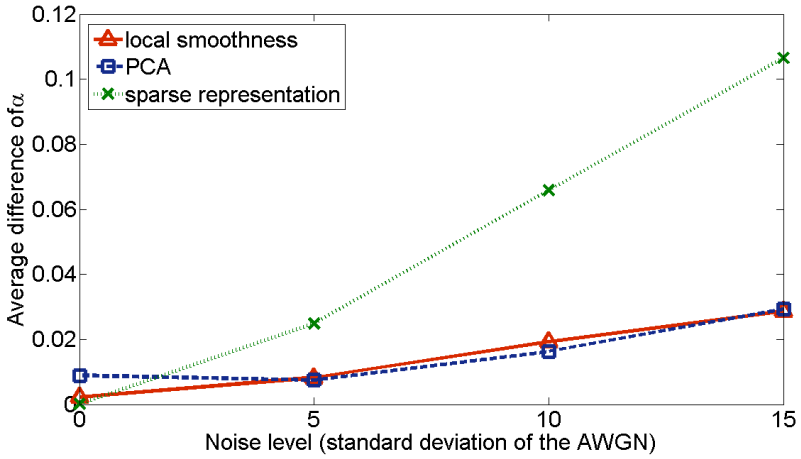


(b) piece-wise smooth background and dark smoke

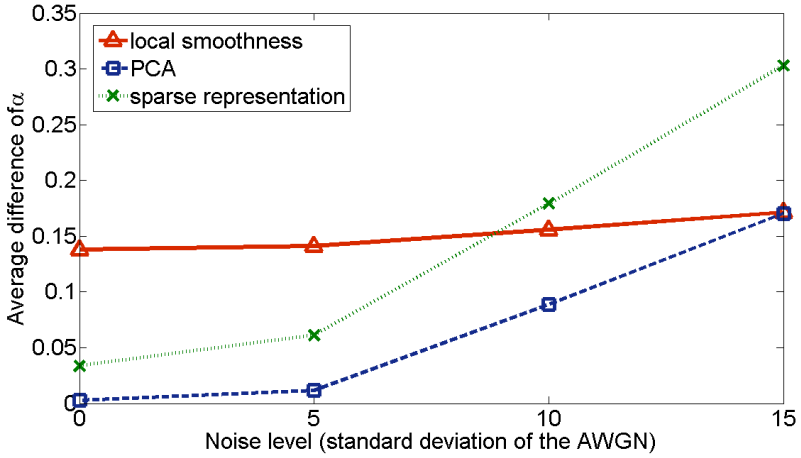


(c) texture background and bright smoke

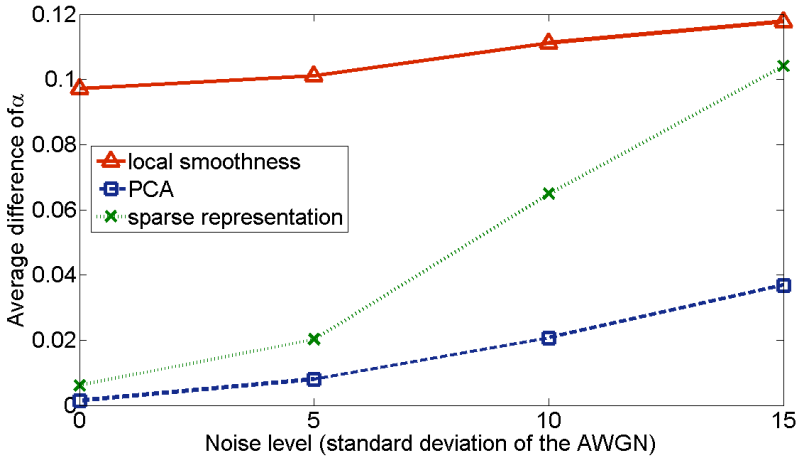
Figure 6.13: Average absolute differences of α versus noise levels for different combinations of background and pure smoke using different models for the smoke component.



(a) piece-wise smooth background and bright smoke

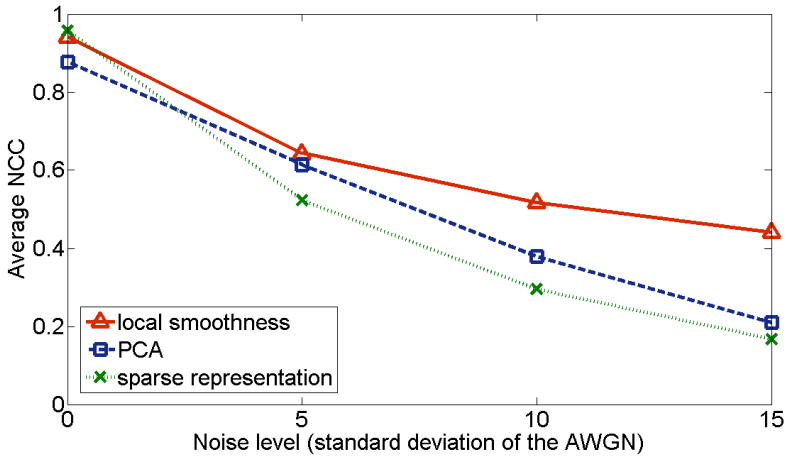


(b) texture background and gray smoke

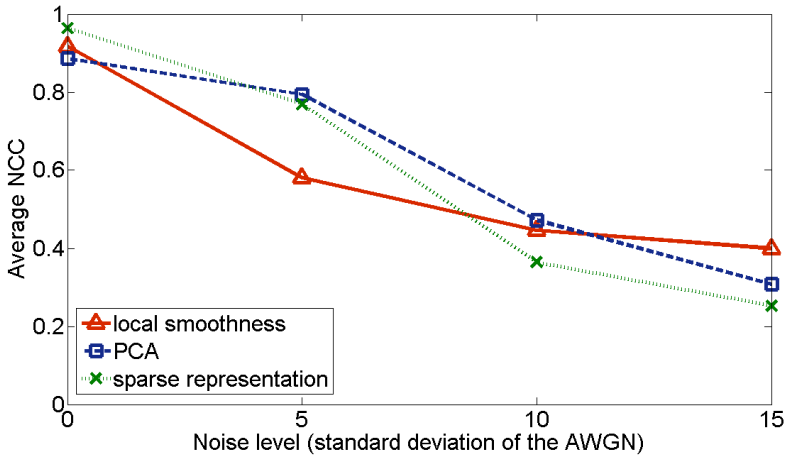


(c) piece-wise smooth background and gray smoke

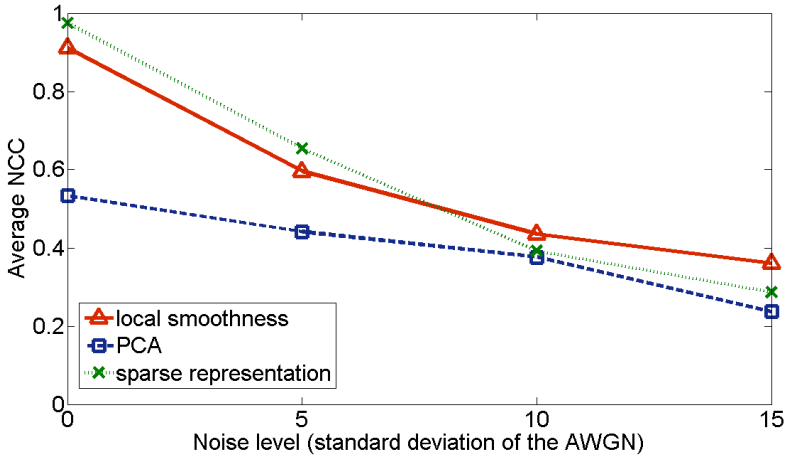
Figure 6.14: Average absolute differences of α versus noise levels for different combinations of background and pure smoke using different models for the smoke component (cont.).



(a) texture background and dark smoke

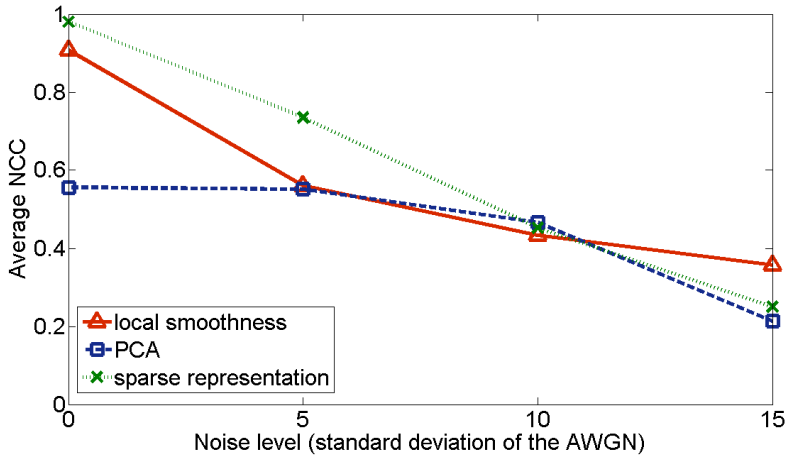


(b) piece-wise smooth background and dark smoke

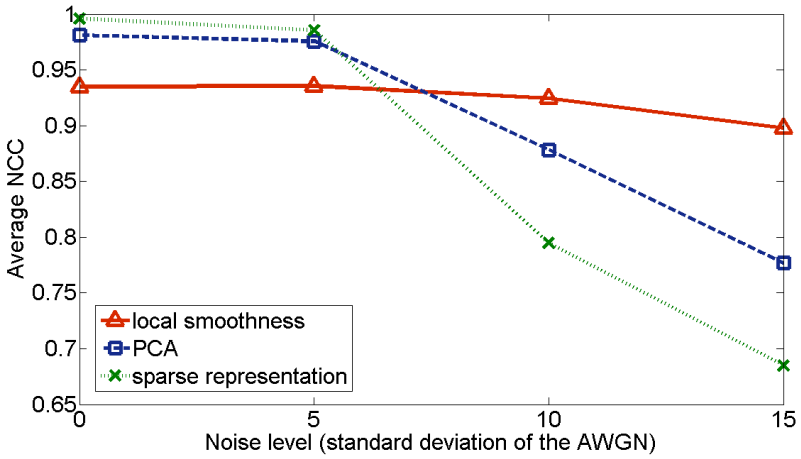


(c) texture background and bright smoke

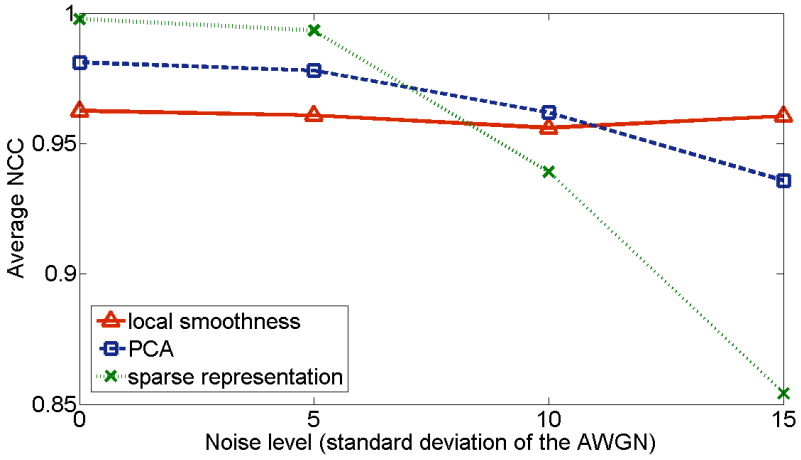
Figure 6.15: Average NCC values versus noise levels for different combinations of background and pure smoke using different models for the smoke component.



(a) piece-wise smooth background and bright smoke



(b) texture background and gray smoke



(c) piece-wise smooth background and gray smoke

Figure 6.16: Average NCC values versus noise levels for different combinations of background and pure smoke using different models for the smoke component (cont.).

We first proceed by letting $\mathbf{f}$ be the original background. Hence in these separation experiments, the images in Figure 6.6 (a) were regarded as both $\mathbf{b}$ and $\mathbf{f}$. 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.

Next let us assume that $\mathbf{f}$ represents other foreground images. Here, both textured and piece-wise smooth images are considered and some examples are shown in the second column in Figure 6.17. In conjunction with the images shown in the first column in Figure 6.17 as $\mathbf{b}$, the several separated "smoke components" based on different types of models for the smoke component are shown in columns 3 to 5 in Figure 6.17. As expected, regardless of the types of background and foreground images, the separated "smoke component" is similar to the original foreground image if the local smoothness or sparse representation model is used. When the local smoothness model is used, the separated "smoke component" can be considered as a smoothed version of the original foreground. However, when the principal component model is employed, the separated "smoke component" appears more like pure smoke than the original foreground image. In other words, the use of the principal component model 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 model 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.

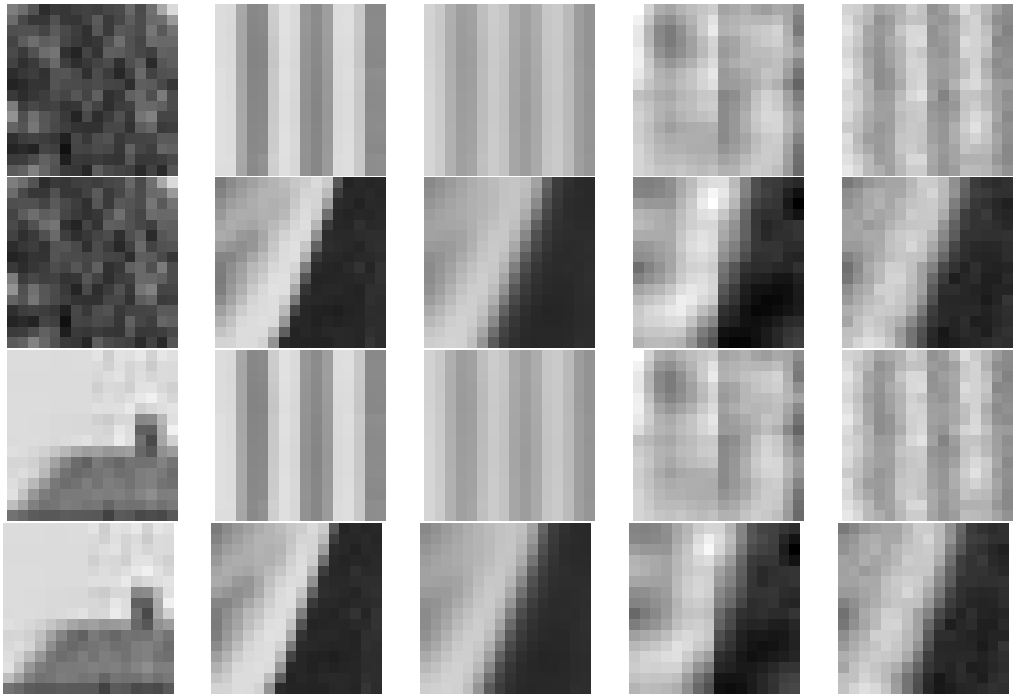


Figure 6.17: Foreground object separation (column 1: texture and piece-wise smooth background, column 2: texture and piece-wise smooth foreground, column 3-5: the separated "smoke component" using the local smoothness, principal component, and sparse representation constraint respectively).

6.3.3 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 Figure 6.18. The collage in Figure 6.19 and 6.20 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 the second scenario in Figure 6.20). Given three image blocks from the video image shown in Figure 6.21 as the current input images $\mathbf{f}$, 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 Figure 6.21), 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.

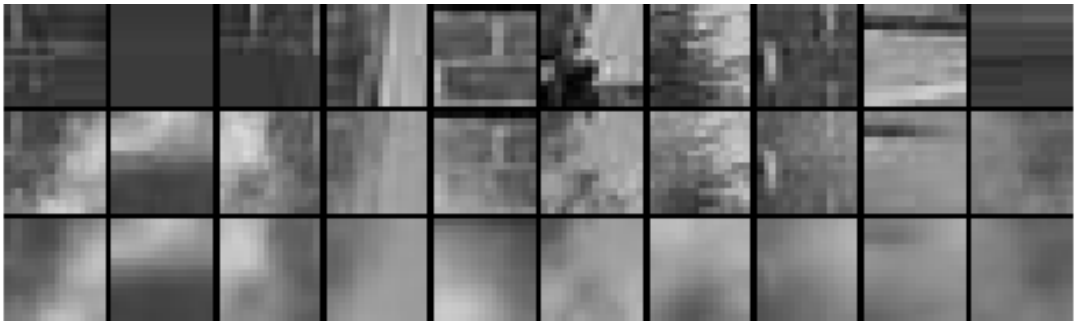


Figure 6.18: Some separated pure smoke block images using real video images (row 1: the learnt background images $\mathbf{b}$ using a GMM [144], row 2: the current input images $\mathbf{f}$, row 3: the separated smoke components $\mathbf{s}$).

6.3.4 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 $\mathbf{f}$ rather than smoke components $\mathbf{s}$.

Data sets

In total 15000 pairs of background images $\mathbf{b}$ modeled by a GMM [144] and corresponding input images $\mathbf{f}$ were created from 20 publicly available video clips of smoke [a]. These video clips cover indoor and outdoor, short and long distance surveillance scenes with different illuminations, a few of which are shown in Figure 6.19 and 6.20. 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

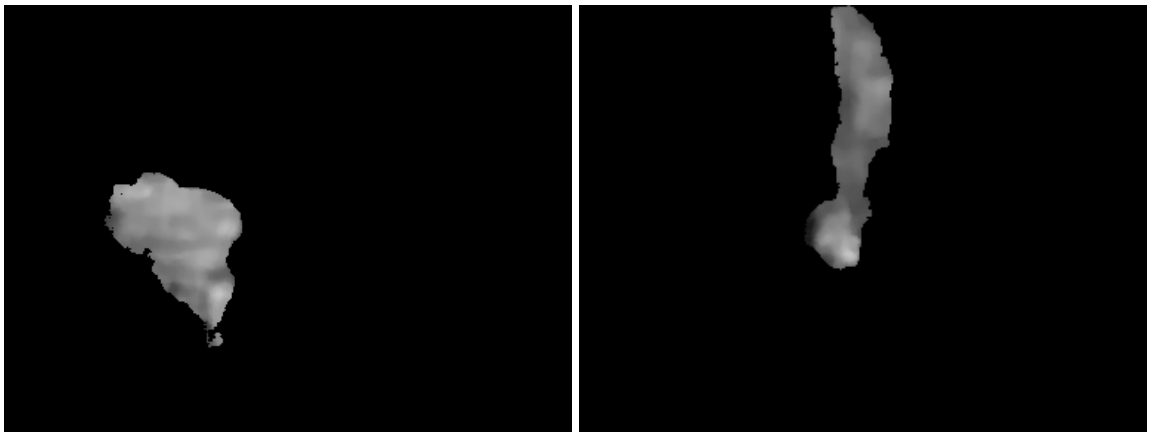
SI4: images that smoke covers less than half of their areas.

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

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

(a) the learned background images

(b) the current input images

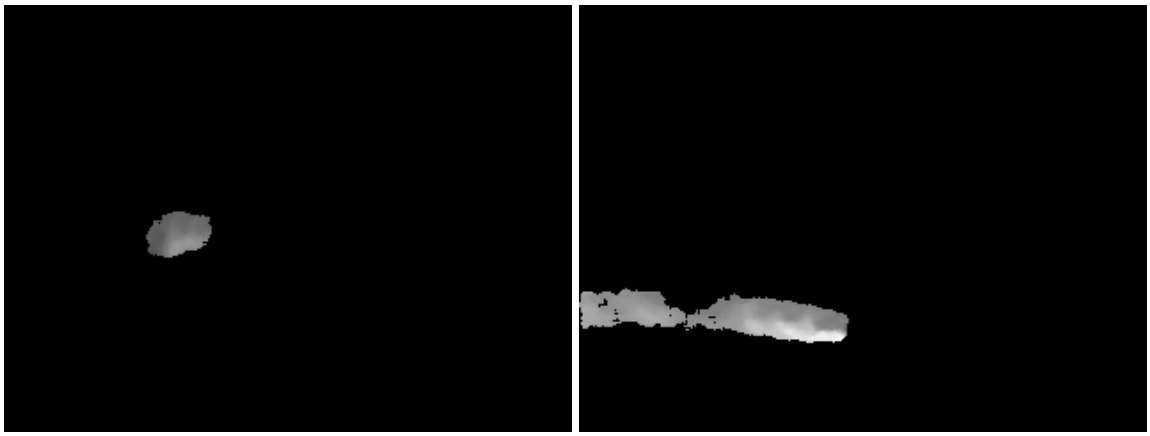


(c) the separated smoke components

Figure 6.19: Some separated pure smoke frames using real video images ((a) the learned background images $\mathbf{b}$ using a GMM [144], (b) the current input images $\mathbf{f}$, (c) the separated smoke components $\mathbf{s}$).

(a) the learned background images

(b) the current input images



(c) the separated smoke components

Figure 6.20: Some separated pure smoke frames using real video images ((a) the learned background images $\mathbf{b}$ using a GMM [144], (b) the current input images $\mathbf{f}$, (c) the separated smoke components $\mathbf{s}$) (cont.).

Figure 6.21: The α values and separated smoke components for three different image blocks in a real video frame.

NS1: images that are covered by non-smoke opaque objects, and

NS2: images that are not covered by anything.

Notice that each input image $\mathbf{f}$ in these categories has its corresponding background image $\mathbf{b}$ obtained using GMM-based background modeling.

The noise level of $\mathbf{b}$ 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 $\mathbf{b}$ 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.

Evaluation method

The local binary pattern (LBP) has been successfully used in texture classification tasks because of its ability to describe texture [116]. In this work, LBP was extracted from $\mathbf{s}$ to describe the texture of smoke and input to a Kernel SVM to decide whether $\mathbf{f}$ is covered by smoke or not. Radial basis function (RBF) kernel was tested in the experiments. By tuning the cost parameter and γ parameter in SVM settings, the best classification results are reported here.

In the experiments, each of the four categories of smoke images (SI1, SI2, SI3, SI4 described in Section 6.3.4) was combined with the non-smoke image category NS1 to form four test data sets and ten-fold evaluation was performed on each data set. By setting thresholds for the output score of SVM classifier, the Receiver Operating Characteristics (ROC) curves for the four data sets were generated as performance measurement.

Results

In our evaluation we set the methods proposed by [155] and [150] 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 $\mathbf{f}$ to perform smoke detection respectively. Further, as part of our evaluation, we compared our approach to the direct background separation in a bid to dispel the notion that it is a simple and appropriate method for smoke separation. The difference image between $\mathbf{f}$ and $\mathbf{b}$ was computed and LBP features were extracted for the experiments. Figure 6.22 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 $\mathbf{s}$ compared to the case when features were extracted directly from the image $\mathbf{f}$. Compared with the case when features were extracted from the difference images, it is noteworthy that the detection performance obtained was

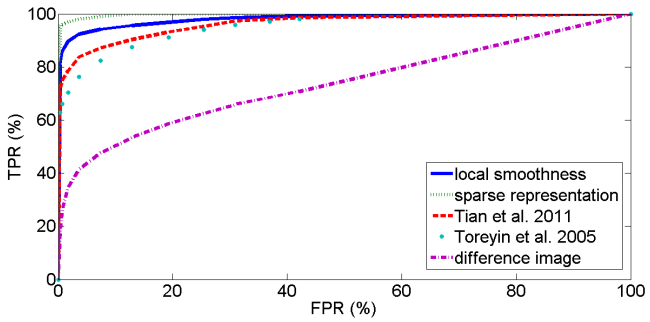
much worse than those achieved by the proposed approach; a result that justifies the unsuitability of background subtraction for smoke component separation.

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 $\mathbf{f}$ as shown in Figs. 6.22(a)-(c). In our experiments one exception was observed for the image category SI04 (Figure 6.22(d)) when only 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 $\mathbf{f}$ and $\mathbf{b}$. Thus the resulting difference image becomes a homogeneous patch.

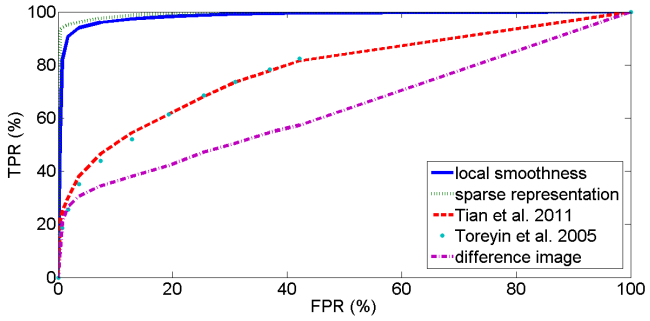
In all, it can be concluded that the comparative results presented here verify the effectiveness of the proposed approach.

6.3.5 Computational complexity

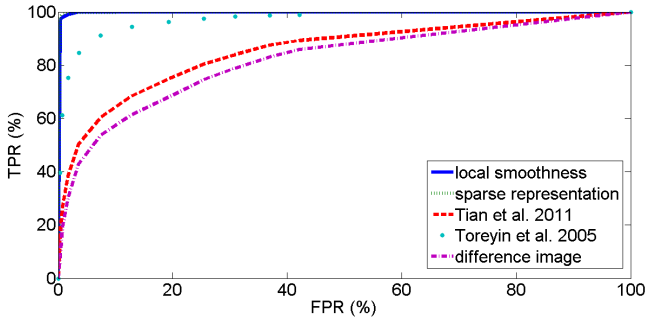
The proposed smoke detection algorithm consists of three major steps: background modeling, separation of smoke component $\mathbf{s}$ and classification of $\mathbf{s}$. Most computation is spent on obtaining the sparse codes to separate the smoke component. In this step, the blending parameter α and the sparse codes to represent the smoke component $\mathbf{s}$ are alternately solved. The parameter α is analytically calculated using Equation (6.10) and the sparse codes for the smoke component are obtained using the feature-sign search algorithm. The complexity of this step is $O(\aleph_1\aleph_2K^3)$ 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.



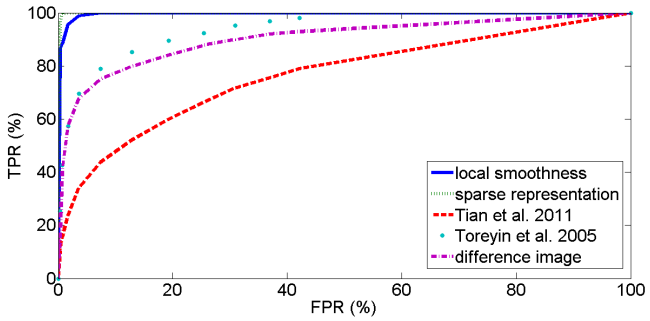
(a) heavy smoke



(b) light smoke



(c) more than half of area is covered by smoke



(d) less than half of area is covered by smoke

Figure 6.22: 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).

6.4 Summary

The fact that smoke often appears as an overlay on a background image could make the task of smoke detection difficult. In this chapter we propose a method whereby the smoke component, if any, is separated from an image before detection. The image formation model for smoke that linearly blends an amount of smoke and background image forms the basis of the proposed method. Using the sparse representation constraint, the blending parameter was estimated and the smoke component was separated. Experimental results have verified the effectiveness of the proposed method. Compared with the local smoothness model proposed in Chapter 5, the sparse representation model is more effective for smoke separation and thus can achieve better performance on smoke detection.

In general, Chapter 5 and Chapter 6 presents ideas pioneering a new direction for vision-based smoke detection. However, there are still rooms for improvement. Under this framework, at least two issues are worthy of further study. Firstly, in terms of smoke separation, local smoothness and sparse representation constraints have proved effective when the noise level of background modelling is not too high. It is worthwhile to study other types of constraints in order to achieve better separated smoke component in the presence of high level of noise in the modelled background. Secondly, 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.

Chapter 7

Smoke detection and separation in a single image

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. In this chapter, a novel feature is proposed to detect smoke from a single image. Furthermore, a method is presented to separate smoke from background in single images.

7.1 Motivation

To the best of our knowledge, 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 some scenarios. Firstly, cameras in an open space are inevitably jittering under severe and dynamic environment such as wind. Secondly, if the surveillance is based on battery-powered sensor network [3], [76], the available power supply, computing resource, or bandwidth is hardly sufficient for video real-time streaming and/or processing and smoke detection. In this case, surveillance images rather than videos are available. Thirdly, when a pan-tilt-zoom (PTZ) camera is used in surveillance, reliable background modeling is hardly achievable to serve smoke detection. In such

circumstances, detection of smoke from single images becomes highly desirable.

Compared with the case that video-based smoke detection is performed on the video frames/images directly, performance has been significantly improved when the detection is performed on the separated smoke components as shown in Chapter 5 and 6. Therefore, the same approach to separating smoke component from background before detection is adopted for single images. It should be noted that background modeling based on the information of previous video frames is a strict prerequisite in the methods presented in Chapter 5 and 6. Given only one single image, it is no longer possible to estimate the required background. Thus based on the image formation model for smoke, novel methods for separating smoke from background need to be developed. For this purpose, a method based on dual-dictionary modeling is proposed in the next section.

7.2 Dual-dictionary modeling

Similar to the video case, block-based detection scheme is adopted here as well in order to achieve early detection (smoke will usually cover a very small area at the early stage) and localization of smoke. 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 of Equation (3.16) can be written as

$$\mathbf{f} = (1-\alpha)\mathbf{b} + \alpha\mathbf{s} + \mathbf{n}, \tag{7.1}$$

where $\mathbf{n} \in \mathbb{R}^N$ represents modeling noise. Unlike the video case where the background component $\mathbf{b}$ is estimated through background modeling, in the case of single images, $\mathbf{b}$ is expected to be estimated simultaneously with the smoke component $\mathbf{s}$. Thus the problem can be formulated as follows:

$$\min_{\alpha,\mathbf{b},\mathbf{s}} \|\mathbf{f} - \alpha\mathbf{s} - (1-\alpha)\mathbf{b}\|_2^2 \quad s.t. \quad \alpha \in [0,1]. \tag{7.2}$$

Given only a single input image block $\mathbf{f}$, Equation (7.2) is under-determined because there are N equations but 2N + 1 free variables for a gray-scale image. Further constraints are required to obtain an unique and reliable solution to Equation (7.2). A good estimation of $\mathbf{b}$, $\mathbf{s}$, and α is expected if both $\mathbf{b}$ and $\mathbf{s}$ could be well modeled according to the visual property of non-smoke and pure smoke.

Due to the fact that smoke can be well modeled based on an over-complete dictionary as shown in Chapter 6, a dual-dictionary method is employed here to model $\mathbf{b}$ and $\mathbf{s}$ respectively. Similar to the case of pure smoke, if each image block is considered as a point in an N-dimensional space, non-smoke images are likely to lie in multiple low-dimensional subspaces. If samples of non-smoke images can be collected or generated to capture their distribution in the N-dimensional space, it is expected that any specific non-smoke image would have a sparse representation with respect to the samples. Such a collection of non-smoke samples represents a non-smoke dictionary. Both dictionaries, one for pure smoke and the other for non-smoke, are designed such that they lead to sparse representations over only one type of image content (either pure smoke or non-smoke). In addition, each dictionary should be trained using real samples to adapt to the specific type of image content. In this thesis, the two dictionaries are trained independently. More sophisticated schemes such as supervised dictionary learning [63], [93] can be adopted to enhance the discriminative power of the dictionaries.

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 $\mathbf{D_s}$:

$$\mathbf{s} = \mathbf{D_s}\mathbf{x_s} \quad s.t. \quad \|\mathbf{x_s}\|_0 \leq M_s, \tag{7.3}$$

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:

$$\mathbf{b} = \mathbf{D_b}\mathbf{x_b} \quad s.t. \quad \|\mathbf{x_b}\|_0 \leq M_b. \tag{7.4}$$

Here M_s and M_b are the upper bounds for the number of non-zero entries in the sparse

coefficients $\mathbf{x_s}$ and $\mathbf{x_b}$ respectively. Considering Equation (7.3) and (7.4) as the models for pure smoke and non-smoke respectively, Equation (7.2) can be rewritten as follows:

$$\min_{\alpha,\mathbf{x_b},\mathbf{x_s}} \{\|\mathbf{f} - \alpha\mathbf{D_s}\mathbf{x_s} - (1-\alpha)\mathbf{D_b}\mathbf{x_b}\|_2^2 + \eta\|\mathbf{x_b}\|_0 + \gamma\|\mathbf{x_s}\|_0\} \quad s.t. \quad \alpha \in [0,1], \tag{7.5}$$

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:

$$\min_{\alpha,\mathbf{x_b},\mathbf{x_s}} \{\|\mathbf{f} - \alpha\mathbf{D_s}\mathbf{x_s} - (1-\alpha)\mathbf{D_b}\mathbf{x_b}\|_2^2 + \eta\|\mathbf{x_b}\|_1 + \gamma\|\mathbf{x_s}\|_1\} \quad s.t. \quad \alpha \in [0,1]. \tag{7.6}$$

The optimization problem expressed by Equation (7.6) is convex with respect to one of $\mathbf{x_b}$, $\mathbf{x_s}$, and α when fixing the other two. One may propose to optimize the three terms alternately. However, α and $(1-\alpha)$ are coupled with $\mathbf{x_s}$ and $\mathbf{x_b}$ respectively by multiplication, which indicates that $\mathbf{x_b}$, $\mathbf{x_s}$, and α may not be well estimated if no other constraints are imposed. Noting that the optimal α is a scalar, we can always absorb α into $\mathbf{x_s}$ and $(1-\alpha)$ into $\mathbf{x_b}$ in Equation (7.6), and solve for $\alpha\mathbf{x_s}$ and $(1-\alpha)\mathbf{x_b}$. The only changes are to scale down γ and η by α and $(1-\alpha)$ respectively. This does not significantly change the essence of the optimization, but helps to reduce one unknown α. Based on this consideration, the following variables are defined

$$\mathbf{y_b} = (1-\alpha)\mathbf{x_b}; \quad \mathbf{y_s} = \alpha\mathbf{x_s}. \tag{7.7}$$

Then Equation (7.6) can be written as

$$\min_{\mathbf{y_b},\mathbf{y_s}} \|\mathbf{f} - \mathbf{D_s}\mathbf{y_s} - \mathbf{D_b}\mathbf{y_b}\|_2^2 + \eta'\|\mathbf{y_b}\|_1 + \gamma'\|\mathbf{y_s}\|_1. \tag{7.8}$$

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 hereafter. Given $\mathbf{f}$, $\mathbf{D_b}$, and $\mathbf{D_s}$, Equation (7.8) 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 [82]. Each is a convex

problem and the convergence of the optimization is guaranteed. Once the difference between the objective function (Equation (7.8)) values in two consecutive iterations is less than a predefined threshold, the optimal $\mathbf{y_b}$ and $\mathbf{y_s}$ can be obtained.

For any input image block $\mathbf{f}$ irrespective of whether it contains smoke, the separated quasi-smoke component tends to be more like smoke. Thus features extracted from quasi-smoke component only may not be sufficient for deciding whether there is smoke or not in $\mathbf{f}$, this is particularly true when $\mathbf{f}$ is not covered by any smoke. Similar argument can be made for the quasi-background component. However, features extracted from both components jointly characterize $\mathbf{f}$ and are expected to have a good discriminative power. One may propose to extract LBP from both components and concatenate them as a feature to characterize $\mathbf{f}$. Actually the coefficient vectors $\mathbf{y_b}$ and $\mathbf{y_s}$ indicate which bases and the proportion thereof contribute to the construction of the quasi-background and quasi-smoke components. Thus, $\mathbf{y_b}$ and $\mathbf{y_s}$ are expected to encode more discriminative information. Based on this understanding, $\mathbf{y_b}$ and $\mathbf{y_s}$ are concatenated as a novel feature to characterize $\mathbf{f}$. The extracted feature is input to an SVM classifier. A decision is made on whether there is smoke or not in $\mathbf{f}$. The proposed detection algorithm is summarized in Algorithm 1. The effectiveness of the proposed feature for smoke detection will be validated in the experiments.

7.2.1 Computational complexity

In the proposed detection 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 Equation (7.8). 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 $O(K_1K_2(K_3^3 + K_4^3))$, 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. The detection algorithm was implemented in MATLAB on a PC with 2.93GHz Intel(R)

Algorithm 1 Single Image Smoke Detection

Input:

block image $\mathbf{f}$
dictionary $\mathbf{D_b}$ for non-smoke
dictionary $\mathbf{D_s}$ for pure smoke
regularization parameters η' and γ'
threshold Th for convergence checking
initial values obj_0 and obj_1 of the objective function (Equation (7.8)) such that $|obj_0 - obj_1| > Th$

Output:

class label for $\mathbf{f}$

initialize $\mathbf{y_s}$ by solving $\min_{\mathbf{y_s}} \|\mathbf{f} - \mathbf{D_s}\mathbf{y_s}\|_2^2 + \gamma'\|\mathbf{y_s}\|_1$;
initialize $\mathbf{y_b}$ by solving $\min_{\mathbf{y_b}} \|\mathbf{f} - \mathbf{D_b}\mathbf{y_b}\|_2^2 + \eta'\|\mathbf{y_b}\|_1$;
while $|obj_0 - obj_1| > Th$ **do**
  calculate $\mathbf{y_b}$ by solving Equation (7.8) with $\mathbf{y_s}$ fixed;
  $\mathbf{y_b} \Leftarrow \mathbf{y_b}$;
  calculate $\mathbf{y_s}$ by solving Equation (7.8) with $\mathbf{y_b}$ fixed;
  $\mathbf{y_s} \Leftarrow \mathbf{y_s}$;
  $obj_0 \Leftarrow obj_1$;
  calculate the current value obj_1 of the objective function (Equation (7.8));
  $obj_1 \Leftarrow obj_1$;
end while
concatenate $\mathbf{y_s}$ and $\mathbf{y_b}$ and input them to an SVM;
return the classification label.

Core(TM) i7 CPU and 8GB memory and was able to process each block image with the size of 16×16 pixels within 0.2$\sim$0.4 seconds. Notice that the proposed algorithm is highly parallel because each block can be processed independently.

7.3 Separation of the smoke component

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

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

Such a formulation of the problem to estimate the blending parameter shares some common nature to image matting [166]. However, the separation of smoke component differs from image matting in three aspects. Firstly, a trimap that serves as the constraints for image matting is accurate or error free as it is usually manually constructed. In our case, a similar map is to be obtained through automatic detection; and this will be referred to as detection map hereafter. Obviously, it includes detection errors. Secondly, almost all the image matting algorithms are developed to deal with the cases where a substantial area of foreground objects to be separated is opaque. However, much of the area covered by smoke may be semi-transparent. Thirdly, most existing matting methods only focus on alpha extraction but not foreground estimation. In our case, smoke is also expected to be extracted from single gray-scale images in addition to the blending parameter. For these differences, it has been found through experiments that the existing image matting algorithms do not work satisfactorily in our case. Therefore, a new method is developed for separating smoke component in a single image.

For each uncertain pixel based on the detection results, its smoke component, background component, and blending parameter are all unknown. Thus some assumptions are required to solve this under-determined problem. Firstly based on the intuition that similar pixels should have similar blending parameters, the following assumption is made: the

alpha value of a pixel can be approximated by a weighted sum of those of similar pixels to the weights given by a kernel function $K(i,j)$, which can be expressed mathematically as follows:

$$\alpha_i \approx \sum_j \alpha_j K(i,j)/D_i \quad or \quad D_i\alpha_i \approx \mathbf{K}_i\boldsymbol{\alpha}, \tag{7.9}$$

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

$$D_i = \sum_j K(i,j), \tag{7.10}$$

$$\mathbf{K}_i = \Big[K(i,1), \quad \cdots, \quad K(i,n)\Big]. \tag{7.11}$$

Under this assumption, two key issues need to be addressed for alpha estimation. The first one is that how to collect the similar pixels j given a pixel i of interest. The second one is that how to construct the kernel function $K(i,j)$.

Both the reliable smoke and reliable non-smoke pixels in a detection map are modeled by GMM. The reason that GMM is adopted to cluster the known pixels is as follows. The detection map constructed from smoke detection includes some errors, which means that there are some misclassifications among reliable smoke, reliable non-smoke, and unknown pixels. Given a pixel of interest, some misclassified pixels may be selected as similar pixels as the one of interest, which will lead to an inaccurate estimation of alpha. In order to reduce or eliminate the adverse effects of misclassified pixels on similar pixel selection process, we propose to cluster the reliable smoke and reliable non-smoke pixels using GMM. In addition, the similarity measurements between pixels are deemed to include both photometric and geometric information. Specifically, a local patch centered at each known pixel is selected and the gray-scale values of all the pixels in the local patch are used to capture the photometric information of the centering pixel. The spatial coordinates of each known pixel are regarded as the geometric information of the pixel.

The photometric and geometric information form a feature vector and the clustering using GMM will be performed in the feature space. Illustratively for the mth Gaussian, the corresponding Gaussian probability density function (PDF) is expressed as:

$$f_m[\mathbf{p}(i)] = \frac{1}{(2\pi)^{M/2}|\mathbf{\Sigma}_m|^{1/2}} exp\{-\frac{1}{2}[\mathbf{p}(i) - \boldsymbol{\mu}_m]^T \mathbf{\Sigma}_m^{-1}[\mathbf{p}(i) - \boldsymbol{\mu}_m]\}, \tag{7.12}$$

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

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

$$K(i,j) = 1 - \frac{||\mathbf{p}(i) - \mathbf{p}(j)||_1}{C}, \tag{7.13}$$

where $\mathbf{p}(i)$ and $\mathbf{p}(j)$ are the feature vectors for the ith and jth pixel respectively, C is the normalized constant.

After addressing the aforementioned two issues, we further define:

$$\mathbf{A} = \begin{bmatrix} K(1,1) & \cdots & K(1,n) \\ \vdots & \ddots & \vdots \\ K(n,1) & \cdots & K(n,n) \end{bmatrix}, \tag{7.14}$$

and

$$\mathbf{D} = \begin{bmatrix} D_1 & 0 & \cdots & 0 \\ 0 & D_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & D_n \end{bmatrix}. \tag{7.15}$$

$\mathbf{A} \in \mathbb{R}^{n\times n}$ is often referred to as affinity matrix and $\mathbf{D} \in \mathbb{R}^{n\times n}$ is a diagonal matrix. According to Equation (7.9), we have:

$$\mathbf{D}\boldsymbol{\alpha} \approx \mathbf{A}\boldsymbol{\alpha}. \tag{7.16}$$

Thus,

$$(\mathbf{D}-\mathbf{A})\boldsymbol{\alpha} \approx \mathbf{0} \quad or \quad \boldsymbol{\alpha}^T\mathbf{L}_c\boldsymbol{\alpha} \approx 0, \tag{7.17}$$

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

$$\boldsymbol{\alpha}^* = \operatorname{argmin} \boldsymbol{\alpha}^T\mathbf{L}_c\boldsymbol{\alpha} + \lambda(\boldsymbol{\alpha}^T - \mathbf{v}^T)\mathbf{V}(\boldsymbol{\alpha}-\mathbf{v}), \tag{7.18}$$

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

$$v_i = \begin{cases} 1 & \text{if pixel } i \text{ is reliable smoke} \\ 0 & \text{otherwise.} \end{cases} \tag{7.19}$$

The solution to Equation (7.18), which is a quadratic function of $\boldsymbol{\alpha}$, can be obtained by differentiating it with respect to $\boldsymbol{\alpha}$ and setting the derivatives to zero. Thus the optimal blending parameter is:

$$\boldsymbol{\alpha}^* = (\mathbf{L}_c + \lambda\mathbf{V})^{-1}(\lambda\mathbf{v}). \tag{7.20}$$

Secondly, a similar assumption can be made for the smoke component: the smoke com-

ponent of a pixel can be approximated by a weighted sum of those of similar pixels to the weights given by a kernel function. Furthermore, a similar analysis leads to the following optimization problem:

$$\mathbf{s}^* = \operatorname{argmin} \mathbf{s}^T \mathbf{L}_c \mathbf{s} + \lambda(\mathbf{s}^T - \mathbf{u}^T)\mathbf{V}(\mathbf{s} - \mathbf{u}), \tag{7.21}$$

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

$$u_i = \begin{cases} f_i & \text{if pixel } i \text{ is reliable smoke} \\ 0 & \text{otherwise.} \end{cases} \tag{7.22}$$

The optimal smoke component can be expressed as:

$$\mathbf{s}^* = (\mathbf{L}_c + \lambda \mathbf{V})^{-1}(\lambda \mathbf{u}). \tag{7.23}$$

Once the blending parameter and smoke component are estimated through solving Equation (7.20) and (7.23), they are substituted into the image formation model and the background component can be calculated.

7.4 Experimental results

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

(1) Experiments on images captured by non-stationary cameras were performed to illustrate the limitation of video-based smoke detection and justify the need for single image smoke detection.

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

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

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

(5) Evaluation of the proposed feature on challenging images, such as fog, that have similar visual appearance as smoke.

(6) Validation of the effectiveness of the proposed feature on real data.

(7) Experiments of separating smoke from background on real images.

7.4.1 Data sets

Six data sets were constructed for the above experiments.

DS1

The data set consists of 1000 pure smoke images with the size of 16×16 pixels and was used to learn $\mathbf{D_s}$. **DS1** is the same as the one used for training the smoke dictionary in Chapter 6. Again, K-SVD was adopted to train $\mathbf{D_s}$ which has the size of 256×500. Some basis samples from $\mathbf{D_s}$ are shown in Figure 7.1.

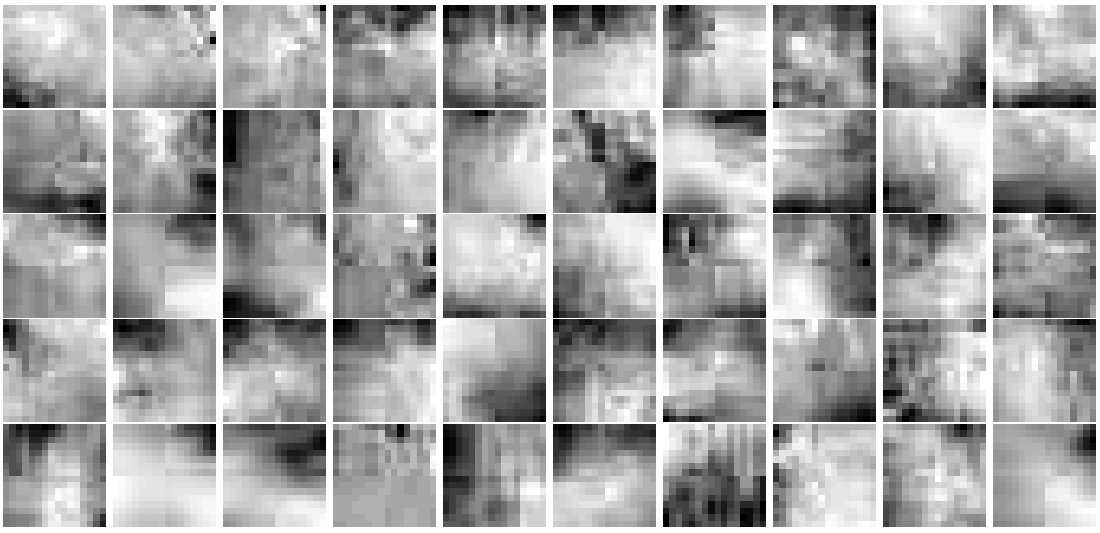


Figure 7.1: Examples of the bases from the learned smoke dictionary $\mathbf{D_s}$.

DS2

This data set consists of non-smoke images with the size of 16×16 pixels and was used to learn $\mathbf{D_b}$. To make $\mathbf{D_b}$ have good generalization ability, it was constructed by randomly cropping 60000 non-smoke images with the size of 16×16 pixels from the images in the CIFAR-100 data set [74]. Similarly, K-SVD was used to train $\mathbf{D_b}$ which has the same size as $\mathbf{D_s}$. Some basis samples from $\mathbf{D_b}$ are shown in Figure 7.2.

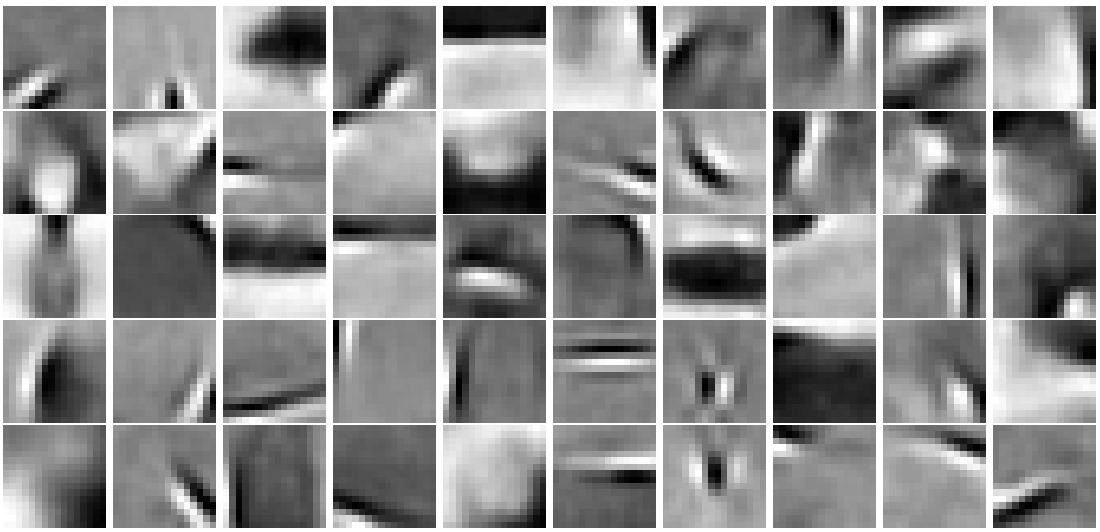


Figure 7.2: Examples of the bases from the learned non-smoke dictionary $\mathbf{D_b}$.

DS3

This data set was constructed to demonstrate how unreliable background modeling would affect adversely video-based smoke detection and thus justify the need for single image smoke detection. Specifically, **DS3** has 1000 images with the size of 16×16 pixels that were manually cropped from four videos captured by unstable cameras. These videos cover both indoor and outdoor scenes and example frames of them are shown in Figure 7.3. Moreover, 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 [144].

Figure 7.3: Example frames of the videos captured by unstable cameras.

DS4

This data set is a collection of smoke images with the size of 16×16 pixels and was used for testing the proposed detection algorithm. Specifically, it includes 5000 block images that were manually cropped based on visual observation from 25 publicly available video clips of smoke. These video clips [16, 155, 185], cover indoor and outdoor, short and long distance surveillance scenes with different illuminations, example frames of which are shown in Figure 6.19 and 6.20. Furthermore, half of the 5000 block images are heavy smoke and the other half are light smoke.

DS5

It is a collection of general non-smoke block images used for testing the proposed detection algorithm, which covers a large variety of real life image patches. Specifically, 5000 images with the size of 16×16 pixels were randomly cropped from the images in the 15-scene data set [78].

DS6

This data set is composed of challenging non-smoke block images used for testing the proposed detection algorithm. In a single image smoke detection problem there are three categories of challenges that could lead to misclassification or false detection. These challenges arise respectively from the physical formation process, transparency and homogeneity properties of smoke. As a result, each image in **DS6** must exhibit at least one of the challenges in the three categories. The first category includes some objects sharing similar physical formation process as smoke, such as haze/fog, cloud, and steam. As shadow and glass have similar transparency property as smoke, they can be considered as the representative samples of the second category. Due to the homogeneous property of some pure smoke, image patches with high homogeneity may pose a challenge to accurate detection. Thus smooth wall, sky, clothes and vehicle body are included in the third category. Water is also considered due to its transparency property and high homogeneity. In summary, to construct **DS6**, images a total of 10 challenging object classes were collected from website and 2500 non-smoke images with the size of 16×16 pixels were cropped randomly from the related challenging regions, with 250 block images in each class. Sample challenging non-smoke images and the corresponding block images cropped are shown in Figure 7.4.

In a summary, the data sets used in the related detection experiments are listed in Table 7.1.

Table 7.1: Data sets used in the related detection experiments

Experiment	Data sets involved
(1)	**DS3**
(3)	**DS4**, **DS5**
(4)	**DS4**, **DS5**
(5)	**DS4**, **DS6**
(6)	**DS4**, **DS5**, **DS6**

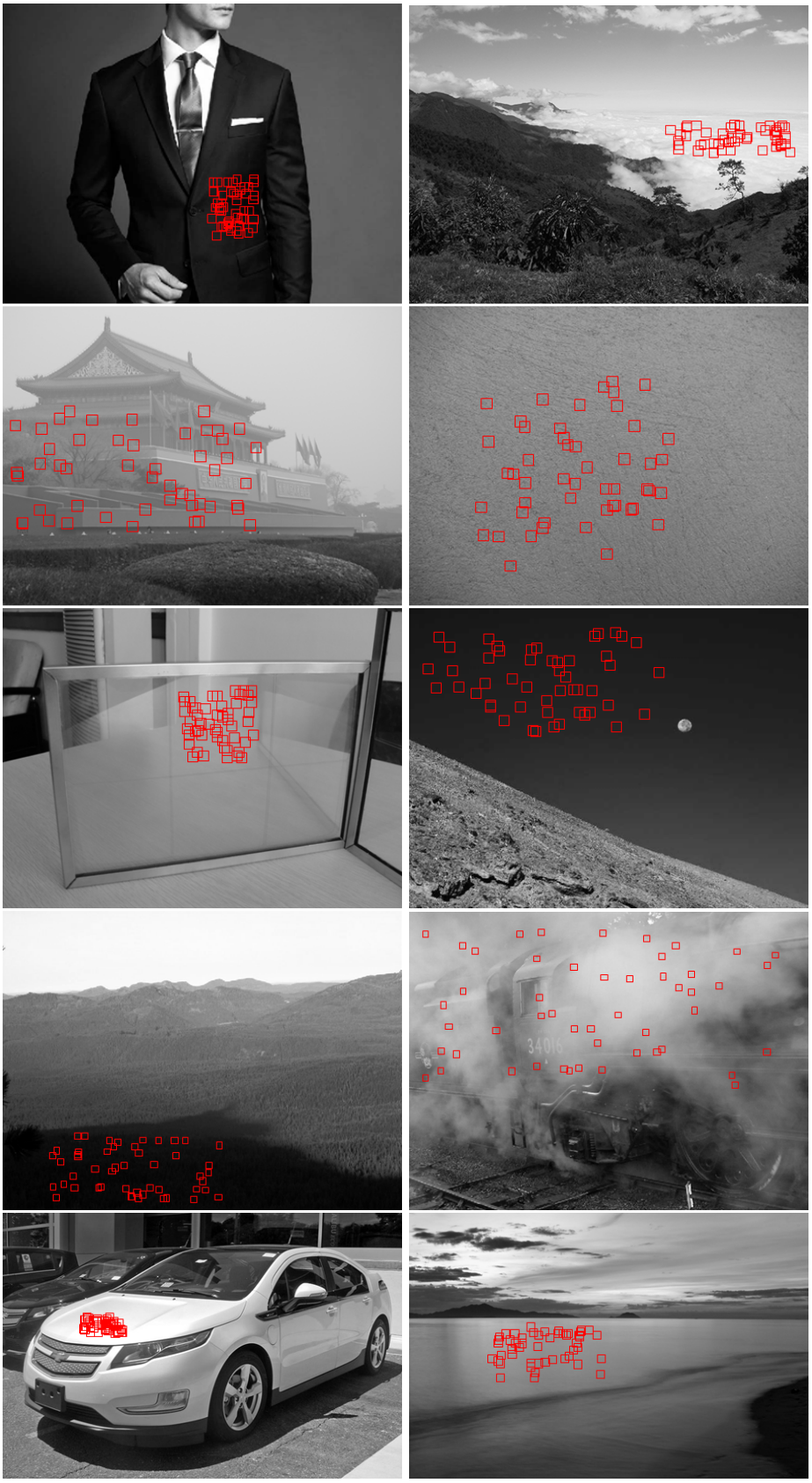


Figure 7.4: Sample challenging non-smoke images and the corresponding block images cropped (The block images indicated by red rectangle in each image represent one of the following challenging object classes respectively: clothes, cloud, haze/fog, wall, glass, sky, shadow, steam, vehicle body, and water).

7.4.2 Influence of jitter on smoke detection

Existing video-based smoke detection algorithms usually assume a stationary camera. It is interesting to test how camera jittering can affect video-based smoke detection. More importantly, whether the proposed method for single image smoke detection will outperform video-based smoke detection is verified.

The method presented in Chapter 6 was used as a baseline for video-based smoke detection in the experiments. The performance of smoke separation using this method in the case of non-stationary camera was tested on synthesized smoke images shown in Figure 6.7. To simulate the corresponding background images captured by a shaking camera, the ground truth background images shown in Figure 6.6 (a) were randomly shifted by no more than 5 pixels towards one of the eight directions. Specifically, the simulated background images under the case of non-stationary camera are shown in Figure 7.5. In the experiments, each synthesized image in Figure 6.7 was considered as an input image $\mathbf{f}$, and the corresponding background image in Figure 7.5 was considered as $\mathbf{b}$. Using the methods presented in Chapter 6, the blending parameter α and smoke component $\mathbf{s}$ were obtained. To evaluate the accuracy of the estimation of α quantitatively, 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 are reported in Table 7.2. For comparison purpose, the average absolute differences of α calculated under the situation of stationary camera are also included in the table. Note ground truth background images shown in Figure 6.6 (a) were considered as $\mathbf{b}$ in the case of stationary camera. As can be seen from Table 7.2, compared with the case of stationary camera, the increased average absolute differences of α in the case of non-stationary camera indicate a significant deterioration of smoke separation.

To quantitatively evaluate the performance of the smoke component estimation, NCC between the estimated and true smoke components was computed. For each combination of background and pure smoke, the average NCC are reported in Table 7.3. Similarly, the average NCC calculated in the case of stationary camera are also shown for comparison purpose. Compared with the situation of stationary camera, there exists decreases for the

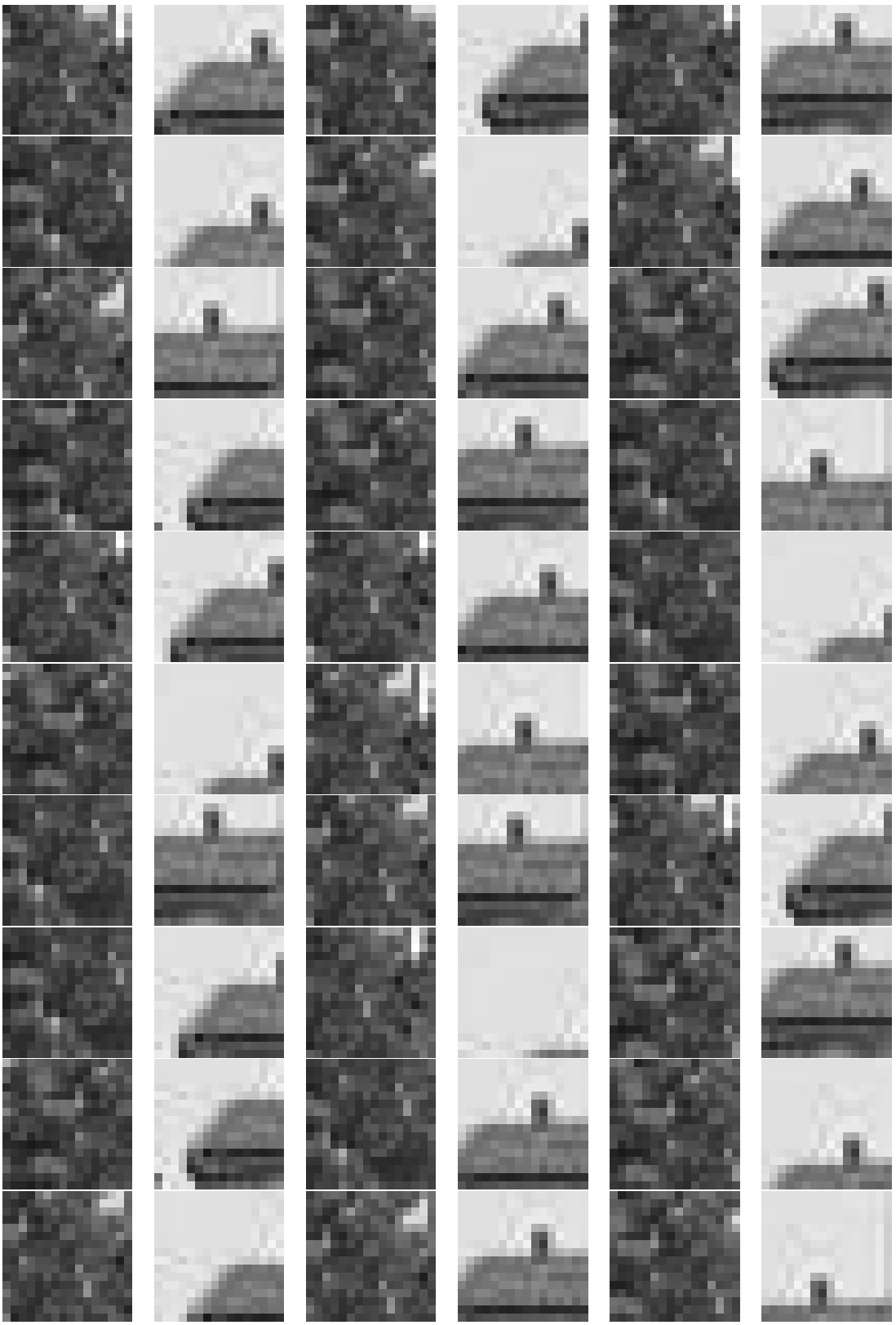


Figure 7.5: The simulated background images corresponding to all the images in the test data set shown in Figure 6.7.

Table 7.2: Average absolute differences between the estimated α and the actual α for different combinations of background and pure smoke ((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)

	non-stationary camera	stationary camera
(1)	0.4482 ± 0.2882	0.0022 ± 0.0019
(2)	0.4264 ± 0.2675	0.0019 ± 0.0015
(3)	0.4478 ± 0.2898	0.0023 ± 0.0017
(4)	0.4473 ± 0.2878	0.0018 ± 0.0014
(5)	0.4356 ± 0.2778	0.0337 ± 0.0266
(6)	0.45 ± 0.2872	0.0062 ± 0.0027

average NCC in the case of non-stationary camera, which means the estimation of smoke component is not accurate. Based on the results shown in Table 7.2 and 7.3, it can be concluded that the performance of smoke separation will be deteriorated significantly if the video is captured by a non-stationary camera.

Table 7.3: Average NCC between the estimated and true smoke components for different combinations of background and pure smoke ((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)

	non-stationary camera	stationary camera
(1)	0.5435 ± 0.0874	0.9274 ± 0.045
(2)	0.4061 ± 0.0751	0.9579 ± 0.0136
(3)	0.403 ± 0.3505	0.9137 ± 0.031
(4)	0.6406 ± 0.1606	0.962 ± 0.0334
(5)	0.8363 ± 0.2447	0.992 ± 0.0148
(6)	0.8966 ± 0.0945	0.995 ± 0.013

In the case of non-stationary camera, the inaccurate smoke component estimation would have adverse effects on smoke detection in video. 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 **DS3** (1000 block images cropped from the video clips captured by non-stationary cameras), the smoke detection algorithm presented in Chapter 6 was implemented on these block images. 5-fold cross validation was performed and the classification accuracy is

only 54.5%.

In the image-based smoke detection, for each of the 1000 block images, the corresponding sparse coefficients $\mathbf{y_b}$ and $\mathbf{y_s}$ were estimated by solving Equation (7.8). The concatenated $\mathbf{y_b}$ and $\mathbf{y_s}$ was used as the feature input to an SVM classifier. Similarly, 5-fold cross validation was performed and the classification accuracy is 95.5%. Obviously, when dealing with a video captured by a non-stationary camera, existing video-based smoke detection algorithms will perform poorly but single image smoke detection can still achieve good performance.

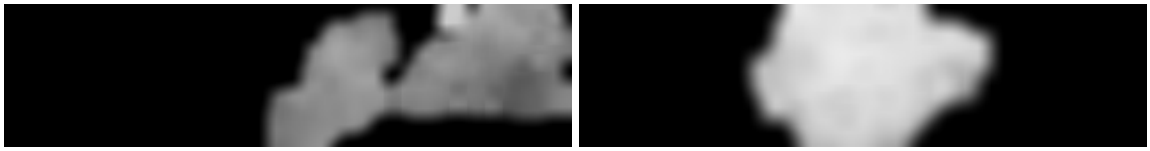
7.4.3 Separation of quasi-smoke and quasi-background

Given a test image block $\mathbf{f}$ and the trained dictionaries $\mathbf{D_b}$ and $\mathbf{D_s}$, the corresponding sparse coefficients $\mathbf{y_b}$ and $\mathbf{y_s}$ are estimated by solving Equation (7.8). Then quasi-background component $\mathbf{D_b y_b}$ and quasi-smoke component $\mathbf{D_s y_s}$ are calculated. For an image which includes many blocks, the separation can be performed on every block in a sliding window manner. To evaluate the separation performance qualitatively, visual inspection was conducted for both indoor and outdoor scenarios. Some separated quasi-smoke and quasi-background components are shown in Figure 7.6 and 7.7. As can be noticed, the pure non-smoke areas in the test images can be well represented by the corresponding quasi-background component. For the smoke-covered regions, most smoke information is included in the corresponding quasi-smoke component while the details of non-smoke are characterized by the corresponding quasi-background component.

7.4.4 Classification of smoke and non-smoke

Given 5000 smoke image blocks (**DS4**) and 5000 general non-smoke image blocks (**DS5**), the discriminative power of 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 Equation (7.8). The concatenated $\mathbf{y_b}$ and $\mathbf{y_s}$ was considered as a novel feature to characterize the test image block and as the input to an SVM classifier to determine whether it contains smoke.

(a) the test images



(b) the separated quasi-smoke components

(c) the separated quasi-background components

Figure 7.6: Quasi-smoke and quasi-background separation ((a) the test images, (b) the separated quasi-smoke components, (c) the separated quasi-background components).

(a) the test images

(b) the separated quasi-smoke components

(c) the separated quasi-background components

Figure 7.7: Quasi-smoke and quasi-background separation ((a) the test images, (b) the separated quasi-smoke components, (c) the separated quasi-background components) (cont.).

In the rest of the chapter, the proposed feature will be referred to as SC.

To compare the SC feature with conventional image features for smoke detection, the texture feature, LBP, was chosen since it offered the state-of-the-art performance in video-based smoke detection [185]. As shown in Chapter 5 and 6, 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 estimated, quasi-background component $\mathbf{D_b y_b}$ and quasi-smoke component $\mathbf{D_s y_s}$ were obtained. Similar to the trick used in Chapter 5 and 6, LBP of $\mathbf{D_s y_s}$ was considered as a feature for smoke detection, and will be referred to as LBP_S in the rest of this chapter. Additionally, the concatenated LBP extracted respectively from $\mathbf{D_b y_b}$ and $\mathbf{D_s y_s}$ may encode discriminative information and was tested as well; and this will be referred to as LBP_C hereafter. For completeness, LBP that was extracted from the original image block $\mathbf{f}$ without performing separation was also tested; and this will be referred to as LBP in the rest of this chapter.

Both linear and radial basis function (RBF) kernel SVM were tested and 5-fold cross validation was performed in our experiments in this chapter, unless otherwise specified. The classification accuracies are reported in Table 7.4. As shown in the table, among the four features tested, the proposed feature SC achieves the highest accuracy in the binary classification of smoke and non-smoke. As expected, the texture feature LBP extracted from $\mathbf{f}$ has the worst performance. With the texture information of both quasi-background and quasi-smoke components considered, LBP_C is more discriminative than LBP_S, which only represents the texture feature of quasi-smoke component. For clarity, the confusion matrix based on SC is shown in Table 7.5. Furthermore, the ROC curves are adopted as performance measurement. They are shown in Figure 7.8 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 an SVM classifier using the proposed feature. Some classification results are shown in Figure 7.9. In each scene shown in Figure 7.9, one smoke

Table 7.4: Accuracies for the classification of smoke and non-smoke (LBP: extracted from the original image block $\mathbf{f}$; 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**

Table 7.5: Confusion matrix for the classification of smoke and non-smoke based on the proposed feature SC

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

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

7.4.5 Classification of heavy smoke, light smoke and non-smoke

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

Specifically, 2500 block images were randomly selected from the data set of non-smoke **DS5**. Given these 2500 non-smoke, 2500 heavy smoke, and 2500 light smoke image blocks, quasi-smoke and quasi-background components were separated and the proposed feature SC was extracted. For comparative evaluation, LBP, LBP_S and LBP_C

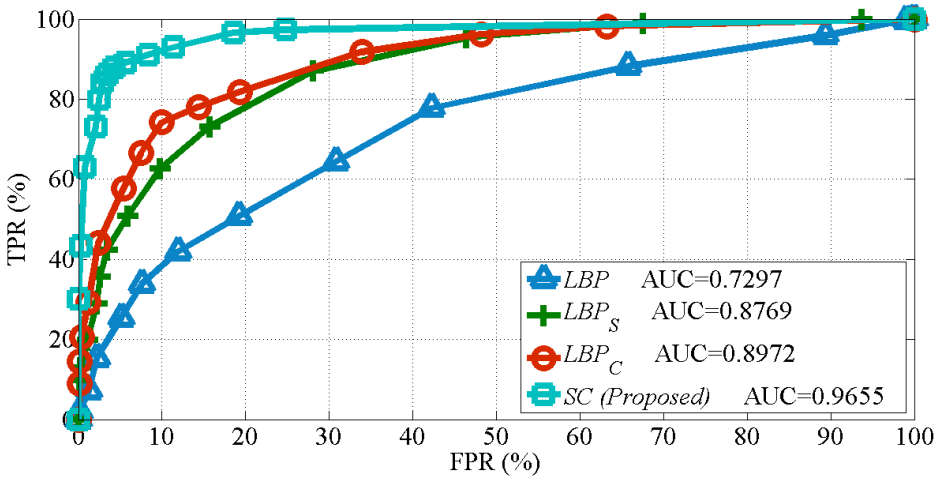


Figure 7.8: ROC curves for the classification of smoke and non-smoke (LBP: extracted from the original image block $\mathbf{f}$; LBP_S: extracted from the quasi-smoke component $\mathbf{D_s y_s}$ only; LBP_C: extracted from both the quasi-smoke component $\mathbf{D_s y_s}$ and the quasi-background component $\mathbf{D_b y_b}$ and then concatenated).

were also extracted as texture feature. The classification accuracies are reported in Table 7.6. As shown among all the four features the highest accuracy was obtained when using the proposed feature SC. It is also noted that, for the classification of heavy smoke, light smoke, and non-smoke, the features 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 7.7. 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. This result is consistent with the classification of smoke and non-smoke.

Table 7.6: Accuracies for the classification of heavy smoke, light smoke, and non-smoke (LBP: extracted from the original image block $\mathbf{f}$; 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**

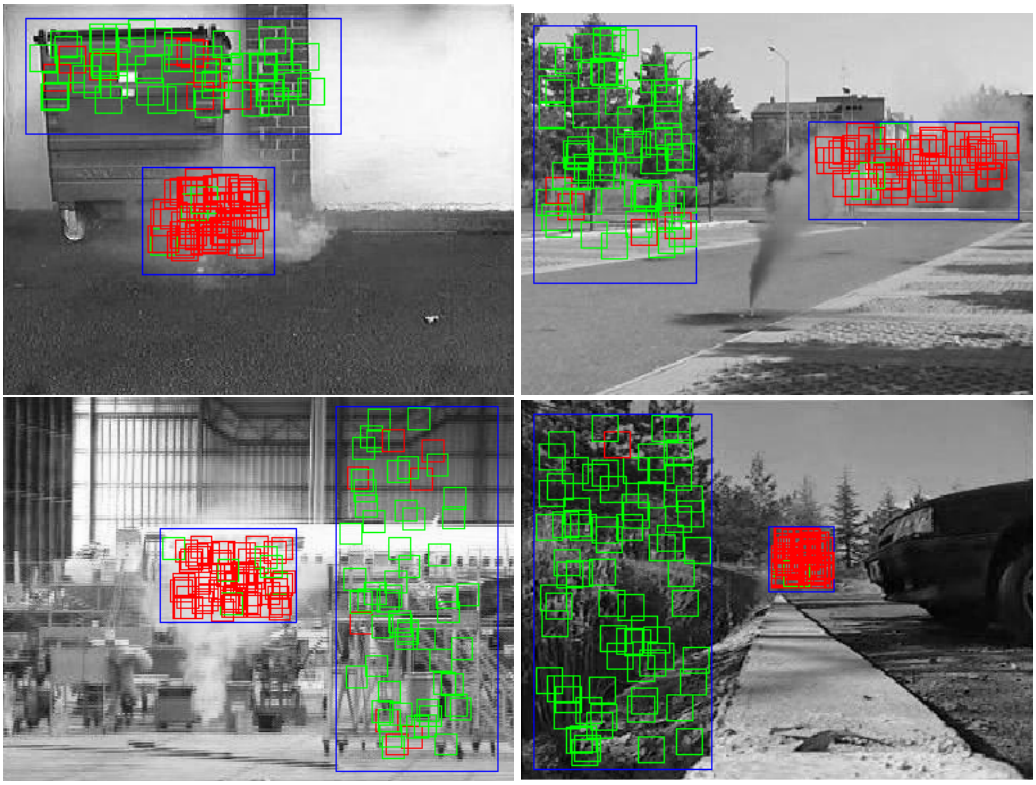


Figure 7.9: Illustrative classification results (blue rectangle: the selected region; red block: classified as smoke; green block: classified as non-smoke).

Table 7.7: Confusion matrix for the classification of heavy smoke, light smoke, and non-smoke based on the proposed feature SC

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

7.4.6 Differentiation of smoke from challenging objects

As discussed before, there are objects having similar visual properties that may challenge the smoke detector. In this section, the experiment on differentiating smoke from some challenging objects is reported. Note this is the first time such an experiment was conducted.

In this experiment, 2500 smoke (including both heavy and light) image blocks that were randomly selected from the smoke data set **DS4** and 2500 challenging non-smoke block images from **DS6** were used. 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.

The classification of these image blocks into smoke and non-smoke yielded classification accuracies of 77.16% and 79.2% when using LBP_C and SC respectively. Specifically, for each class of the challenging non-smoke objects, the accuracies (%) of being correctly classified as non-smoke are shown in Table 7.8. Notice that although LBP_C and SC lead to similar classification performance on the entire data set, their performance on each class vary significantly. Among all the challenging non-smoke classes considered, when SC is used, the test images of sky and steam have the highest and lowest probability being correctly classified as non-smoke respectively.

Table 7.8: Accuracies (%) of each class of objects in **DS6** being correctly classified as non-smoke

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

It can be expected from the above experiments that the proposed feature SC will outperform LBP-based features in a realistic case where smoke and non-smoke coexist.

7.4.7 Smoke detection: real application considerations

Based on the results so far obtained, the proposed feature SC, has been validated to effectively differentiate between the classes of smoke and common non-smoke; and the classes of smoke and challenging non-smoke. However, a classifier which can well differentiate smoke from common non-smoke may not effectively classify smoke from challenging non-smoke. Similar argument can be made for a classifier trained using smoke and chal-

lenging non-smoke only. Moreover, in a smoke detection system application it will be preferable to filter out common non-smoke at a first stage of smoke detection. Then smoke and challenging non-smoke are further differentiated at a second stage. Based on these considerations, a tree-structured classifier may have good generalization ability in classifying smoke from non-smoke and was constructed. Using the data sets **DS4**, **DS5**, and **DS6** described in Sect. 7.4.1, two partitions (training and test data) were created. In the training set, there are 1500 block images including either heavy or light smoke, 1500 common non-smoke block images selected randomly, and 1500 challenging non-smoke block images. The test set comprises 3500 smoke block images, 3500 common non-smoke block images, and 1000 challenging non-smoke block images.

Five SVMs were trained using the proposed feature SC, which are listed as follows:

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

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

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

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

Given the 3500 smoke block images and 4500 non-smoke (including 3500 common and 1000 challenging non-smoke) block images in the test set, image separation was performed and SC was extracted. The ROC curves for smoke detection based on the seven classifiers are shown in Figure 7.10. Overall tree-structured classifiers *Classifier6* and *Classifier7* outperform all the other five single SVM classifiers. Furthermore, *Classifier7* leads to the best performance among all the classifiers. Trained on the smoke and challenging non-smoke block images only, *Classifier2* gives the worst performance among all the classifiers. The ROC curve based on *Classifier7* also indicates the effectiveness of the proposed feature SC for single image smoke detection.

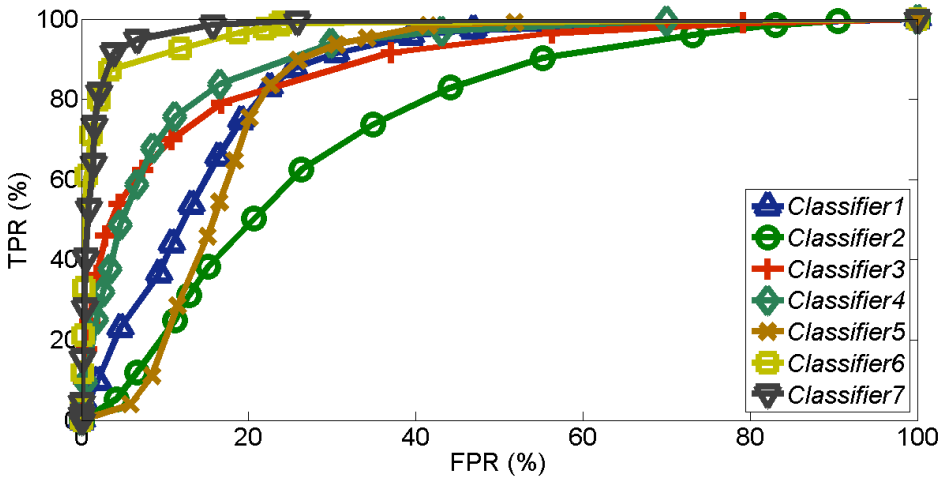


Figure 7.10: ROC curves for single image smoke detection based on different classifiers.

7.4.8 Single image smoke separation

To validate the effectiveness of constructing detection map based on the results of smoke detection, an input image and its corresponding detection maps constructed according to different thresholds for the output score of SVM classifier are shown in Figure 7.11. It is noted that reliable smoke, reliable non-smoke, and unknown pixels are indicated with white, black, and gray respectively in the detection maps. Despite of different detection maps, misclassification exists more or less.

(a) input image

(b) detection map-1

(c) detection map-2

(d) detection map-3

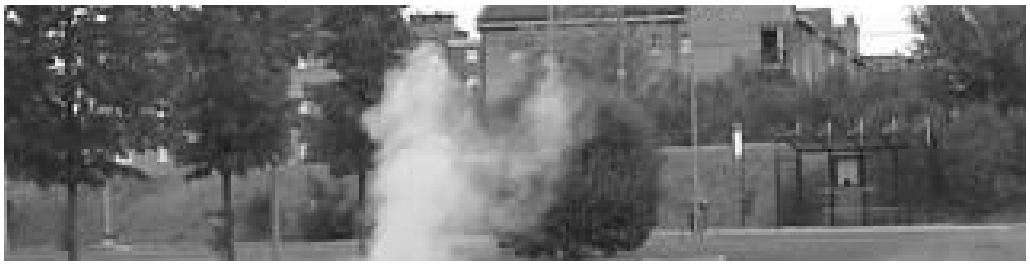
Figure 7.11: An input image and its corresponding detection maps constructed based on the results of smoke detection using different thresholds for the output score of SVM classifier.

In addition, some results for smoke separation based on the proposed algorithm are shown in Figure 7.12, 7.13, 7.14, and 7.15, which cover both outdoor and indoor scenarios with smoke at different stages. As can be noticed from Figure 7.12, 7.13, and 7.14, a small percentage of misclassification leads to a relatively reliable detection map, which further yields good estimation of the blending parameter and smoke component. When misclassification is considerable, as is shown in Figure 7.15, the estimation of the blending parameter and smoke component becomes inaccurate. In terms of background separation, although it is hard to reliably estimate the part that is covered by heavy smoke, certain

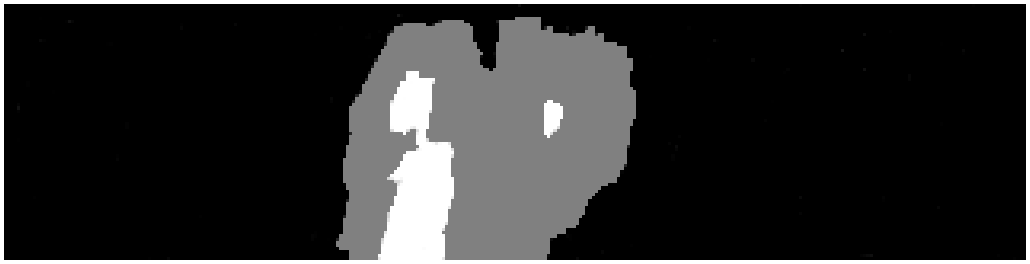
areas of the background covered by light smoke are recovered.

7.5 Summary

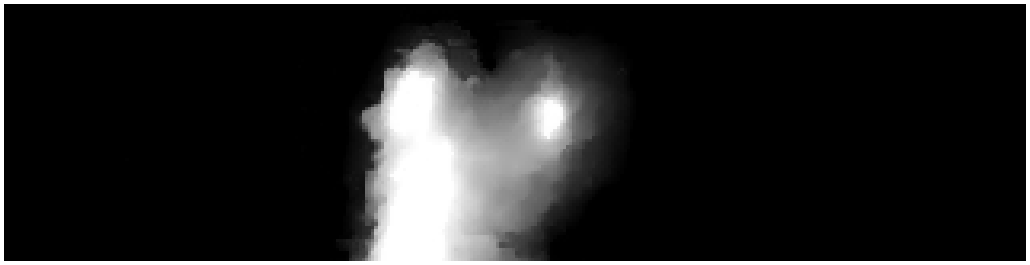
In this chapter, a novel feature, namely the sparse coefficients associated with an over-complete dictionary representation, has been proposed to detect smoke from a single image. The proposed feature arises from two parts; one representing the smoke component of the input image and the other representing the non-smoke component. Based on the imaging model, an optimization scheme allowing the separation of quasi-smoke and quasi-background components was formulated. The effectiveness of the proposed feature for single image smoke detection was validated by the experimental results. Furthermore, the design of a practical smoke detection system was presented. Preliminary results on smoke separation were obtained by means of the proposed method. Notice the blending parameter and smoke component were estimated independently and the background component was then calculated by using the image formation model. In the future, appropriate constraints may be imposed on the background component to achieve a better separation. In addition, how to minimize the adverse effects of misclassification on smoke separation needs a further study.



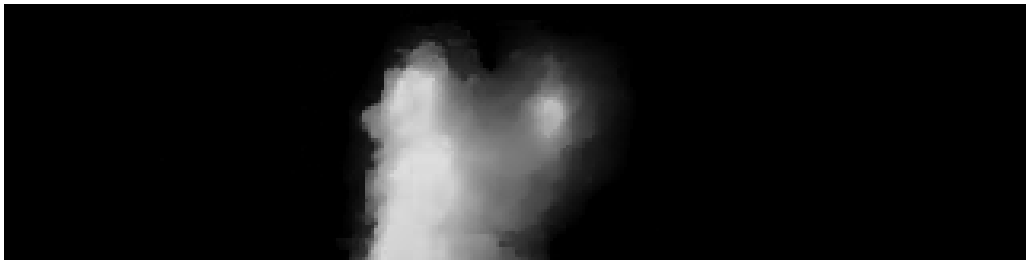
(a) input image



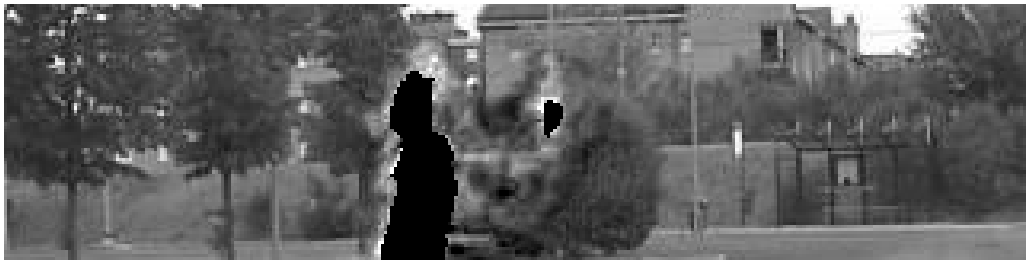
(b) detection map



(c) estimated blending parameter



(d) separated smoke component

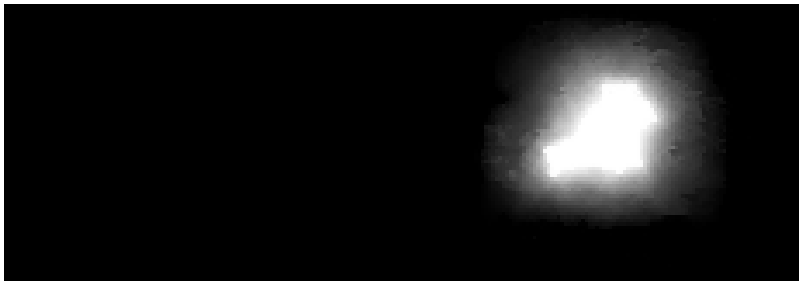


(e) separated background component

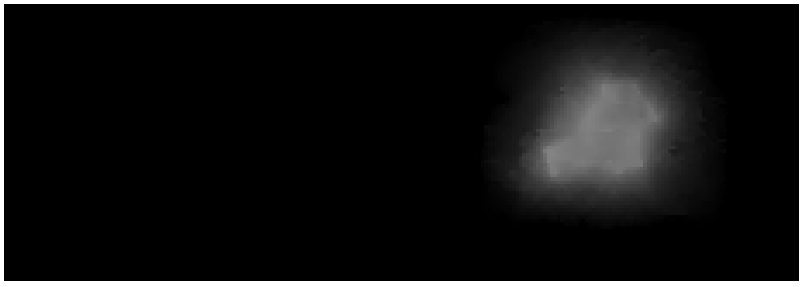
Figure 7.12: Results for single image smoke separation ((a) input image (b) detection map (c) estimated blending parameter (d) separated smoke component (e) separated background component).

(a) input image

(b) detection map



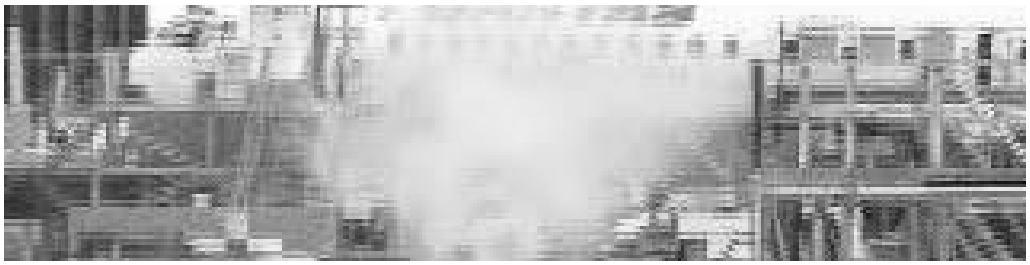
(c) estimated blending parameter



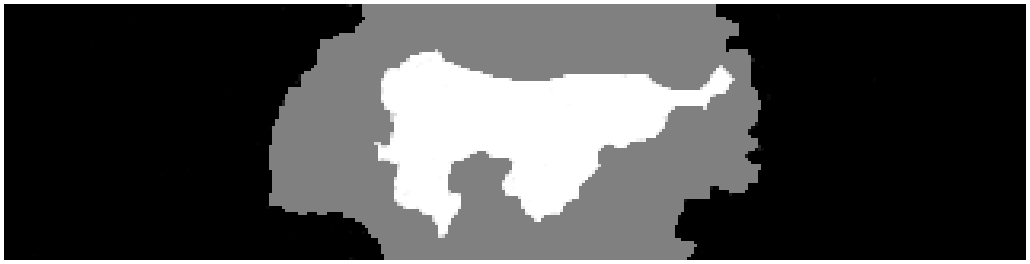
(d) separated smoke component

(e) separated background component

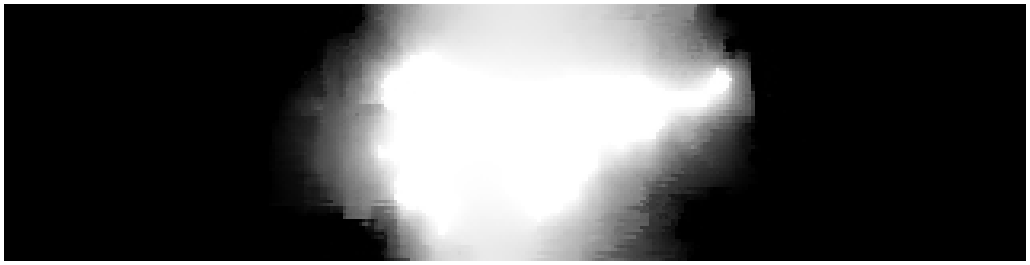
Figure 7.13: Results for single image smoke separation ((a) input image (b) detection map (c) estimated blending parameter (d) separated smoke component (e) separated background component) (cont.).



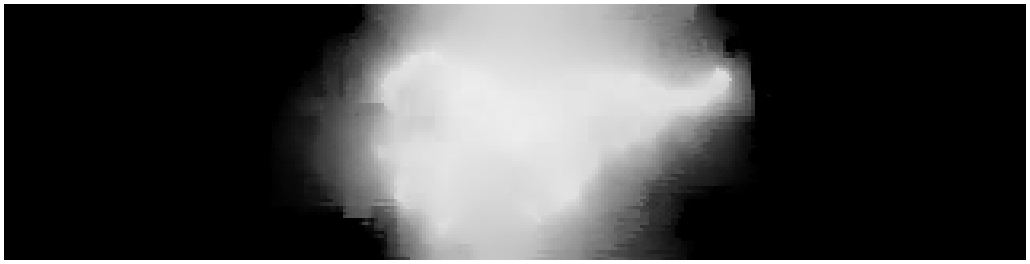
(a) input image



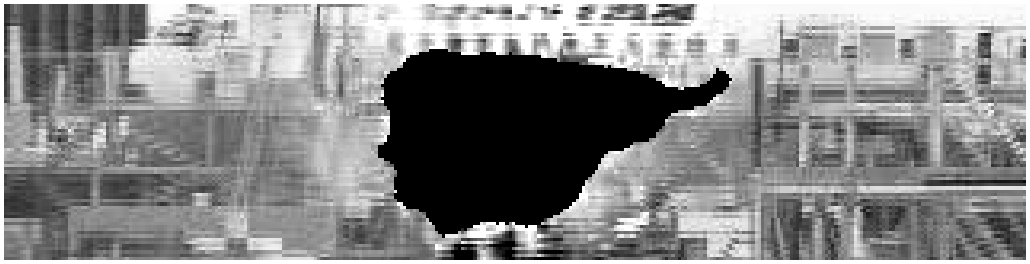
(b) detection map



(c) estimated blending parameter



(d) separated smoke component

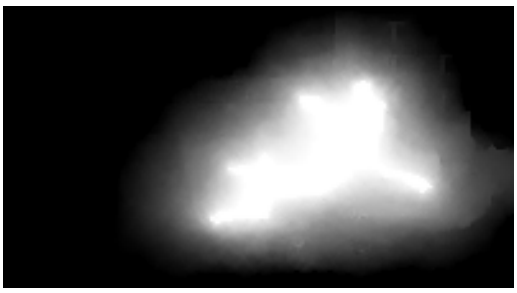


(e) separated background component

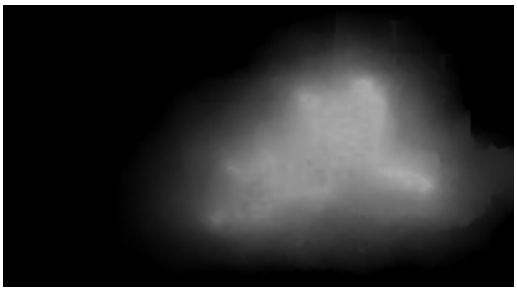
Figure 7.14: Results for single image smoke separation ((a) input image (b) detection map (c) estimated blending parameter (d) separated smoke component (e) separated background component) (cont.).

(a) input image

(b) detection map



(c) estimated blending parameter



(d) separated smoke component

(e) separated background component

Figure 7.15: Results for single image smoke separation ((a) input image (b) detection map (c) estimated blending parameter (d) separated smoke component (e) separated background component) (cont.).

Chapter 8

Conclusion

8.1 Summary

Vision-based smoke detection is of great significance for saving property and human lives from fire disasters. This thesis addresses some challenges in detecting smoke from both videos and single images. Computational models have been established to account for the imaging process of smoke and novel methods have been proposed based on the computational models for vision-based smoke detection. Extensive experiments were conducted to verify the methods proposed in this thesis. The key contributions of the thesis include:

- **A physics-based image model.** Based on the attenuation and airlight effects introduced by smoke, an image formation model has been derived to account for the imaging process of smoke. With reasonable assumptions considered, any small-sized image can be expressed as a linear combination of smoke component and background component. This understanding of the smoke imaging process forms the base for developing reliable algorithms for vision-based smoke detection.

- **A novel spatial-temporal descriptor.** To overcome the sensitivity of the original LBP to the relative changes between background and foreground, NRLBP has been employed to describe the texture of smoke in a discriminative and compact way. NRLMBP, which is a NRLBP extended to temporal domain, was adopted to encode the motion patterns of smoke in video. As a spatial-temporal representation of

smoke, NRLBP and NRLMBP were combined and applied to smoke detection in video.

- **An image separation approach.** To eliminate the adverse effects of background on smoke detection, a novel approach has been proposed for smoke detection in videos by separating smoke from background. Guided by the image formation model, given a video frame and its background, the estimation of the blending parameter and the actual smoke component has been formulated into a convex optimization framework. Three models, namely local smoothness, principal component and sparse representation, were proposed to model the smoke component for effective separation. Both the blending parameter and smoke component are then used for smoke detection.

- **A smoke detection and separation method in single images.** Based on the image formation model of smoke imaging, a method for single image smoke detection has been proposed. Given a single image, an optimization scheme allowing the separation of quasi-smoke and quasi-background components was formulated based on dual-dictionary modeling. The sparse coefficients for representing the separated components were used as features for smoke detection. Taking the advantage of the detection results, a method has been proposed for separating the smoke component.

8.2 Future work

The concept of separation-based detection proposed in this thesis has pioneered a new direction for vision-based smoke detection. In this framework, some interesting research problems still exist. Additionally, this thesis has provided some insights for dealing with vision tasks relating to semi-transparent objects.

- In the detection framework based on image separation, further constraints can be explored in the process of separating smoke to improve the detection performance. It is noted that smoke detection was mainly performed on gray-scale images in the

scope of this thesis. Actually, color cameras in video surveillance can provide more information for smoke separation and detection. How to make use of this information and incorporate it effectively into the separation-based detection framework needs a further study.

- Image sequences of smoke often exhibit certain stationarity properties in temporal domain, which can be referred to as a type of dynamic texture [29]. It is interesting to quantify the features of smoke from this perspective for detection.

- The transparency property of smoke has inspired an image separation approach for smoke detection. A similar image formation model may be derived or assumed in other vision tasks relating to semi-transparent objects. Thus the proposed methods for smoke separation and detection may be adapted to address these issues including fog/haze detection, shadow removal, and underwater image enhancement.

Bibliography

[1] V. Abolghasemi, S. Ferdowsi, and S. Sanei, "Blind separation of image sources via adaptive dictionary learning," *IEEE Transactions on Image Processing*, vol. 21, no. 6, pp. 2921–2930, Jun. 2012.

[2] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.

[3] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "A survey on sensor networks," *IEEE Communications Magazine*, vol. 40, no. 8, pp. 102–114, Aug. 2002.

[4] L. B. Almeida, "Separating a real-life nonlinear image mixture," *Journal of Machine Learning Research*, vol. 6, pp. 1199–1229, Jul. 2005.

[5] M. S. C. Almeida and L. B. Almeida, "Nonlinear separation of show-through image mixtures using a physical model trained with ICA," *Signal Processing*, vol. 92, no. 4, pp. 872–884, Apr. 2012.

[6] —, "Wavelet-based separation of nonlinear show-through and bleed-through image mixtures," *Neurocomputing*, vol. 72, no. 1-3, pp. 57–70, Dec. 2008.

[7] B. Ashe, K. J. McAneney, and A. J. Pitman, "Total cost of fire in Australia," *Journal of Risk Research*, vol. 12, no. 2, pp. 121–136, Mar. 2009.

[8] X. Bai and G. Sapiro, "Geodesic matting: a framework for fast interactive image and video segmentation and matting," *International Journal of Computer Vision*, vol. 82, no. 2, pp. 113–132, Apr. 2009.

[9] A. J. Bell and T. J. Sejnowski, "An information-maximisation approach to blind separation and blind deconvolution," *Neural Computation*, vol. 7, no. 6, pp. 1129–1159, Nov. 1995.

[10] A. B. Benyahia, N. Hamouda, F. Tlili, and S. Ouerghi, "Early smoke detection in forest areas from dct based compressed video," in *Proc. European Signal Processing Conference*, 2012, pp. 2752–2756.

[11] J. C. Bezdek and R. J. Hathaway, "Convergence of alternating optimization," *Neural, Parallel and Scientific Computations*, vol. 11, no. 4, pp. 351–368, Dec. 2003.

[12] R. Bohush and N. Brouka, "Smoke and flame detection in video sequences based on static and dynamic features," in *Proc. Signal Processing: Algorithms, Architectures, Arrangements, and Applications*, 2013, pp. 20–25.

[13] P. Bouguer, *Traite' d'optique sur la gradation de la lumiere*. 1729.

[14] S. Bronte, L. M. Bergasa, and P. F. Alcantarilla, "Fog detection system based on computer vision techniques," in *Proc. IEEE International Conference on Intelligent Transportation Systems*, 2009, pp. 1–6.

[15] S. Calderara, P. Piccinini, and R. Cucchiara, "Smoke detection in video surveillance: a mog model in the wavelet domain," in *Lecture Notes in Computer Science*, vol. 5008, 2008, pp. 119–128.

[16] —, "Vision based smoke detection system using image energy and color information," *Machine Vision and Applications*, vol. 22, no. 4, pp. 705–719, Jul. 2011.

[17] T. Celik, H. Ozkaramanli, and H. Demirel, "Fire and smoke detection without sensors: image processing based approach," in *Proc. European Signal Processing Conference*, 2007, pp. 1794–1798.

[18] J. Chen, Y. Wang, Y. Tian, and T. Huang, "Wavelet based smoke detection method with RGB contrast-image and shape constrain," in *Proc. Visual Communications and Image Processing*, 2013, pp. 1–6.

[19] Q. Chen and V. Koltun, "A simple model for intrinsic image decomposition with depth cues," in *Proc. IEEE International Conference on Computer Vision*, 2013, pp. 241–248.

[20] Q. Chen, D. Li, and C. K. Tang, "KNN matting," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 9, pp. 2175–2188, Sep. 2013.

[21] T. Chen, Y. Yin, S. Huang, and Y. Ye, "The smoke detection for early fire-alarming system base on video processing," in *Proc. International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, 2006, pp. 427–430.

[22] F. C. Cheng, C. H. Lin, and J. L. Lin, "Constant time o(1) image fog removal using lowest level channel," *Electronics Letters*, vol. 48, no. 22, pp. 1404–1405, Oct. 2012.

[23] Y. Y. Chuang, B. Curless, D. H. Salesin, and R. Szeliski, "A bayesian approach to digital matting," in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2001, pp. 264–271.

[24] T. Cleary, "Performance of dual photoelectric/ionization smoke alarms in full-scale fire tests," *Fire Technology*, vol. 50, no. 3, pp. 753–773, May 2014.

[25] M. Costagli and E. E. Kuruoglu, "Image separation using particle filters," *Digital Signal Processing*, vol. 17, no. 5, pp. 935–946, Sep. 2007.

[26] Y. Cui, H. Dong, and E. Zhou, "An early fire detection method based on smoke texture analysis and discrimination," in *Proc. Congress on Image and Signal Processing*, 2008, pp. 95–99.

[27] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2005, pp. 886–893.

[28] N. Desai, A. Chatterjee, S. Mishra, D. Chudasama, S. Choudhary, and S. Barai, "A fuzzy logic based approach to de-weather fog-degraded images," in *Proc. International Conference on Computer Graphics, Imaging and Visualization*, 2009, pp. 383–387.

[29] G. Doretto, A. Chiuso, Y. N. Wu, and S. Soatto, "Dynamic textures," *International Journal of Computer Vision*, vol. 51, no. 2, pp. 91–109, Feb. 2003.

[30] M. J. Fadili, J. L. Starck, J. Bobin, and Y. Moudden, "Image decomposition and separation using sparse representations: an overview," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 983–994, Jun. 2010.

[31] S. Fang, J. Zhan, Y. Cao, and R. Rao, "Improved single image dehazing using segmentation," in *Proc. IEEE International Conference on Image Processing*, 2010, pp. 3589–3592.

[32] H. Farid and E. H. Adelson, "Separating reflections and lighting using independent components analysis," in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, 1999, pp. 262–267.

[33] R. Fattal, "Single image dehazing," *ACM Transactions on Graphics*, vol. 27, no. 3, 72:1–72:9, Aug. 2008.

[34] S. Fazekas, T. Amiaz, D. Chetverikov, and N. Kiryati, "Dynamic texture detection based on motion analysis," *International Journal of Computer Vision*, vol. 82, no. 1, pp. 48–63, Apr. 2009.

[35] M. Funaro, E. Oja, and H. Valpola, "Independent component analysis for artefact separation in astrophysical images," *Neural Networks*, vol. 16, no. 3-4, pp. 469–478, Apr. 2003.

[36] K. Gai, Z. Shi, and C. Zhang, "Blind separation of superimposed moving images using image statistics," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 1, pp. 19–32, Jan. 2012.

[37] —, "Blindly separating mixtures of multiple layers with spatial shifts," in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2008, pp. 15–22.

[38] R. Gallen, A. Cord, N. Hautiere, and D. Aubert, "Towards night fog detection through use of in-vehicle multipurpose cameras," in *Proc. IEEE Intelligent Vehicles Symposium*, 2011, pp. 399–404.

[39] A. Genovese, R. D. Labati, V. Piuri, and F. Scotti, "Wildfire smoke detection using computational intelligence techniques," in *Proc. IEEE International Conference on Computational Intelligence for Measurement Systems and Applications*, 2011, pp. 1–6.

[40] M. Golbabaee, S. Arberet, and P. Vandergheynst, "Compressive source separation: theory and methods for hyperspectral imaging," *IEEE Transactions on Image Processing*, vol. 22, no. 12, pp. 5096–5110, Dec. 2013.

[41] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 3rd. Prentice Hall, 2007.

[42] L. Grady, "Random walks for image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 11, pp. 1768–1783, Nov. 2006.

[43] J. Gubbi, S. Marusic, and M. Palaniswami, "Smoke detection in video using wavelets and support vector machines," *Fire Safety Journal*, vol. 44, no. 8, pp. 1110–1115, Nov. 2009.

[44] R. Guidara, S. Hosseini, and Y. Deville, "Maximum likelihood blind image separation using nonsymmetrical half-plane Markov random fields," *IEEE Transactions on Image Processing*, vol. 18, no. 11, pp. 2435–2450, Nov. 2009.

[45] P. Guillemant and J. Vicente, "Real-time identification of smoke images by clustering motions on a fractal curve with a temporal embedding method," *Optical Engineering*, vol. 40, no. 4, pp. 554–563, Apr. 2001.

[46] L. Guo and M. Garland, "The use of entropy minimization for the solution of blind source separation problems in image analysis," *Pattern Recognition*, vol. 39, no. 6, pp. 1066–1073, Jun. 2006.

[47] D. Han and B. Lee, "Development of early tunnel fire detection algorithm using the image processing," in *Lecture Notes in Computer Science*, vol. 4292, 2006, pp. 39–48.

[48] —, "Flame and smoke detection method for early real-time detection of a tunnel fire," *Fire Safety Journal*, vol. 44, no. 7, pp. 951–961, Oct. 2009.

[49] K. Hara, K. Inoue, and K. Urahama, "Generalized mixture ratio based blind image separation," *IEEE Signal Processing Letters*, vol. 20, no. 8, pp. 743–746, Aug. 2013.

[50] —, "Resolving permutation ambiguity in correlation-based blind image separation," *Pattern Recognition Letters*, vol. 33, no. 5, pp. 559–567, Apr. 2012.

[51] A. C. Hardy, "How large is a point source," *Journal of the Optical Society of America*, vol. 57, no. 1, 1967.

[52] N. Hautiere, J.-P. Tarel, J. Lavenant, and D. Aubert, "Automatic fog detection and estimation of visibility distance through use of an onboard camera," *Machine Vision and Applications*, vol. 17, no. 1, pp. 8–20, Apr. 2006.

[53] N. Hautiere, J.-P. Tarel, H. Halmaoui, R. Bremond, and D. Aubert, "Enhanced fog detection and free-space segmentation for car navigation," *Machine Vision and Applications*, vol. 25, no. 3, pp. 667–679, Apr. 2014.

[54] N. Hautiere, J.-P. Tarel, and D. Aubert, "Towards fog-free in-vehicle vision systems through contrast restoration," in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2007, pp. 1–8.

[55] K. He, C. Rhemann, C. Rother, X. Tang, and J. Sun, "A global sampling method for alpha matting," in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2011, pp. 2049–2056.

[56] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 12, pp. 2341–2353, Dec. 2011.

[57] K. He, J. Sun, and X. Tang, "Fast matting using large kernel matting laplacian matrices," in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2010, pp. 2165–2172.

[58] G. M. Hidy, *Aerosols and Atmospheric Chemistry*. Academic Press, 1972.

[59] C. Ho, "Machine vision-based real-time early flame and smoke detection," *Measurement Science and Technology*, vol. 20, no. 4, pp. 1–13, Apr. 2009.

[60] C. C. Ho and T. H. Kuo, "Real-time video-based fire smoke detection system," in *Proc. IEEE/ASME International Conference on Advanced Intelligent Mechatronics*, 2009, pp. 1845–1850.

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

[62] Z. Jia, H. Wang, R. Caballero, Z. Xiong, J. Zhao, and A. Finn, "Real-time content adaptive contrast enhancement for see-through fog and rain," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2010, pp. 1378–1381.

[63] Z. Jiang, Z. Lin, and L. S. Davis, "Label consistent K-SVD: learning a discriminative dictionary for recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 11, pp. 2651–2664, Nov. 2013.

[64] D. Kim, "Smoke detection using boundary growing and moments," in *Proc. International Conference on Convergence and Hybrid Information Technology*, 2009, pp. 430–433.

[65] D. Kim and Y. Wang, "Smoke detection in video," in *Proc. World Congress on Computer Science and Information Engineering*, 2009, pp. 759–763.

[66] D. Kim, C. Jeon, B. Kang, and H. Ko, "Enhancement of image degraded by fog using cost function based on human visual model," in *Proc. IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems*, 2008, pp. 64–67.

[67] B. Ko, J. Park, and J. Y. Nam, "Spatiotemporal bag-of-features for early wildfire smoke detection," *Image and Vision Computing*, vol. 31, no. 10, pp. 786–795, Oct. 2013.

[68] B. Ko, J. Y. Kwak, and J. Y. Nam, "Wildfire smoke detection using temporospatial features and random forest classifiers," *Optical Engineering*, vol. 51, no. 1, pp. 1–10, Jan. 2012.

[69] I. Kolesov, P. Karasev, A. Tannenbaum, and E. Haber, "Fire and smoke detection in video with optimal mass transport based optical flow and neural networks," in *Proc. IEEE International Conference on Image Processing*, 2010, pp. 761–764.

[70] N. Kong, Y. W. Tai, and S. Y. Shin, "High-quality reflection separation using polarized images," *IEEE Transactions on Image Processing*, vol. 20, no. 12, pp. 3393–3405, Dec. 2011.

[71] J. Kopf, B. Neubert, B. Chen, M. Cohen, D. Cohen-Or, O. Deussen, M. Uyttendaele, and D. Lischinski, "Deep photo: model-based photograph enhancement and viewing," *ACM Transactions on Graphics*, vol. 27, no. 5, 116:1–116:10, Dec. 2008.

[72] I. Kopilovic, B. Vagvolgyi, and T. Sziranyi, "Application of panoramic annular lens for motion analysis tasks: surveillance and smoke detection," in *Proc. International Conference on Pattern Recognition*, 2000, pp. 714–717.

[73] L. Kratz and K. Nishino, "Factorizing scene albedo and depth from a single foggy image," in *Proc. IEEE International Conference on Computer Vision*, 2009, pp. 1701–1708.

[74] A. Krizhevsky, "Learning multiple layers of features from tiny images," University of Toronto, Tech. Rep., 2009.

[75] D. Krstinic, D. Stipanicev, and T. Jakovcevic, "Histogram-based smoke segmentation in forest fire detection system," *Information Technology and Control*, vol. 38, no. 3, pp. 237–244, 2009.

[76] R. V. Kulkarni, A. Forster, and G. K. Venayagamoorthy, "Computational intelligence in wireless sensor networks: a survey," *IEEE Communications Surveys and Tutorials*, vol. 13, no. 1, pp. 68–96, 2011.

[77] R. D. Labati, A. Genovese, V. Piuri, and F. Scotti, "Wildfire smoke detection using computational intelligence techniques enhanced with synthetic smoke plume generation," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 43, no. 4, pp. 1003–1012, Jul. 2013.

[78] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: spatial pyramid matching for recognizing natural scene categories," in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2006, pp. 2169–2178.

[79] B. Lee and D. Han, "Real-time fire detection using camera sequence image in tunnel environment," in *Lecture Notes in Computer Science*, vol. 4681, 2007, pp. 1209–1220.

[80] C. Y. Lee, C. T. Lin, and C. T. Hong, "Spatio-temporal analysis in smoke detection," in *Proc. IEEE International Conference on Signal and Image Processing Applications*, 2009, pp. 80–83.

[81] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, Oct. 1999.

[82] H. Lee, A. Battle, R. Raina, and A. Y. Ng, "Efficient sparse coding algorithms," in *Proc. Advances in Neural Information Processing Systems*, 2007.

[83] P. Lee and Y. Wu, "Nonlocal matting," in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2011, pp. 2193–2200.

[84] A. Levin and Y. Weiss, "User assisted separation of reflections from a single image using a sparsity prior," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 9, pp. 1647–1655, Sep. 2007.

[85] A. Levin, D. Lischinski, and Y. Weiss, "A closed-form solution to natural image matting," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 2, pp. 228–242, Feb. 2008.

[86] A. Levin, A. Zomet, and Y. Weiss, "Separating reflections from a single image using local features," in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, 2004, pp. 306–313.

[87] A. Levin, A. R. Acha, and D. Lischinski, "Spectral matting," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 10, pp. 1699–1712, Oct. 2008.

[88] C. Liu and N. Ahuja, "Vision based fire detection," in *Proc. International Conference on Pattern Recognition*, 2004, pp. 134–137.

[89] Q. Liu, J. Liu, P. Dong, and D. Liang, "SGTD: structure gradient and texture decorrelating regularization for image decomposition," in *Proc. IEEE International Conference on Computer Vision*, 2013, pp. 1081–1088.

[90] Y. Liu and X. Lu, "Fog detection and classification using gray histograms," *Advanced Materials Research*, vol. 403-408, pp. 570–576, 2012.

[91] C. Long, J. Zhao, S. Han, L. Xiong, Z. Yuan, J. Huang, and W. Gao, "Transmission: a new feature for computer vision based smoke detection," in *Lecture Notes in Artificial Intelligence*, vol. 6319, 2010, pp. 389–396.

[92] L. Ma, K. Wu, and L. Zhu, "Fire smoke detection in video images using kalman filter and gaussian mixture color model," in *Proc. International Conference on Artificial Intelligence and Computational Intelligence*, 2010, pp. 484–487.

[93] J. Mairal, F. Bach, and J. Ponce, "Task-driven dictionary learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 4, pp. 791–804, Apr. 2012.

[94] H. Maruta, A. Nakamura, T. Yamamichi, and F. Kurokawa, "Image based smoke detection with local hurst exponent," in *Proc. IEEE International Conference on Image Processing*, 2010, pp. 4653–4656.

[95] H. Maruta, Y. Kato, A. Nakamura, and F. Kurokawa, "Smoke detection in open areas using its texture features and time series properties," in *Proc. IEEE International Symposium on Industrial Electronics*, 2009, pp. 1904–1908.

[96] H. Maruta, Y. Iida, and F. Kurokawa, "Anisotropic LBP descriptors for robust smoke detection," in *Proc. Annual Conference of the IEEE Industrial Electronics Society*, 2013, pp. 2372–2377.

[97] —, "Smoke detection method using local binary patterns and adaboost," in *Proc. IEEE International Symposium on Industrial Electronics*, 2013, pp. 1–6.

[98] B. J. Mason, *Clouds, Rain, and Rainmaking*. Cambridge University Press, 1975.

[99] E. J. McCartney, *Optics of the Atmosphere: Scattering by Molecules and Particles*. John Wiley and Sons, 1976.

[100] I. Meganem, Y. Deville, S. Hosseini, P. Deliot, and X. Briottet, "Linear-quadratic blind source separation using NMF to unmix urban hyperspectral images," *IEEE Transactions on Signal Processing*, vol. 62, no. 7, pp. 1822–1833, Apr. 2014.

[101] G. Meng, Y. Wang, J. Duan, S. Xiang, and C. Pan, "Efficient image dehazing with boundary constraint and contextual regularization," in *Proc. IEEE International Conference on Computer Vision*, 2013, pp. 617–624.

[102] F. G. Meyer, A. Z. Averbuch, and R. R. Coifman, "Multilayered image representation: application to image compression," *IEEE Transactions on Image Processing*, vol. 11, no. 9, pp. 1072–1080, Sep. 2002.

[103] W. E. K. Middleton, "The effect of the angular aperture of a telephotometer on the telephotometry of collimated and non-collimated beams," *Journal of the Optical Society of America*, vol. 39, pp. 576–581, 1949.

[104] H. Q. Minh and L. Wiskott, "Slow feature analysis and decorrelation filtering for separating correlated sources," in *Proc. IEEE International Conference on Computer Vision*, 2011, pp. 866–873.

[105] M. Minnaert, *The Nature of Light and Color in the Open Air*. Dover, 1954.

[106] P. Morerio, L. Marcenaro, C. S. Regazzoni, and G. Gera, "Early fire and smoke detection based on colour features and motion analysis," in *Proc. IEEE International Conference on Image Processing*, 2012, pp. 1041–1044.

[107] G. W. Mulholland, "Smoke production and properties," in. National Fire Protection Association; Society of Fire Protection Engineers, 1995, ch. 15, pp. 2/217–2/227.

[108] A. Mustafi and S. K. Ghorai, "A novel blind source separation technique using fractional Fourier transform for denoising medical images," *Optik*, vol. 124, no. 3, pp. 265–271, 2013.

[109] J. N. Myers, "Fog," *Scientific American*, vol. 219, no. 6, pp. 75–82, 1968.

[110] S. G. Narasimhan and S. K. Nayar, "Contrast restoration of weather degraded images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 6, pp. 713–724, Jun. 2003.

[111] —, "Vision and the atmosphere," *International Journal of Computer Vision*, vol. 48, no. 3, pp. 233–254, 2002.

[112] D. T. Nguyen, Z. Zong, P. Ogunbona, and W. Li, "Object detection using non-redundant local binary patterns," in *Proc. IEEE International Conference on Image Processing*, 2010, pp. 4609–4612.

[113] T. K. T. Nguyen and J. M. Kim, "Multistage optical smoke detection approach for smoke alarm systems," *Optical Engineering*, vol. 52, no. 5, pp. 1–12, May 2013.

[114] K. Nishino, L. Kratz, and S. Lombardi, "Bayesian defogging," *International Journal of Computer Vision*, vol. 98, no. 3, pp. 263–278, Jul. 2012.

[115] N. Ohnishi, K. Kumaki, T. Yamamura, and T. Tanaka, "Separating real and virtual objects from their overlapping images," in *Lecture Notes in Computer Science*, vol. 1065, 1996, pp. 636–646.

[116] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971–987, Jul. 2002.

[117] T. Oo, H. Kawasaki, Y. Ohsawa, and K. Ikeuchi, "The separation of reflected and transparent layers from real-world image sequence," *Machine Vision and Applications*, vol. 18, no. 1, pp. 17–24, Feb. 2007.

[118] S. Osher, A. Sole, and L. Vese, "Image decomposition and restoration using total variation minimization and the H-1 norm," *Multiscale Modeling & Simulation*, vol. 1, no. 3, pp. 349–370, 2003.

[119] K. K. Pahalawatta and R. Green, "Particle detection and classification in photoelectric smoke detectors using image histogram features," in *Proc. International Conference on Digital Image Computing Techniques and Applications*, 2013, pp. 1–8.

[120] J. Park, B. Ko, J. Y. Nam, and S. Kwak, "Wildfire smoke detection using spatiotemporal bag-of-features of smoke," in *Proc. IEEE Workshop on Applications of Computer Vision*, 2013, pp. 200–205.

[121] M. Pavlic, H. Belzner, G. Rigoll, and S. Ilic, "Image based fog detection in vehicles," in *Proc. IEEE Intelligent Vehicles Symposium*, 2012, pp. 1132–1137.

[122] H. Peng and R. Rao, "Image enhancement of fog-impaired scenes with variable visibility," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2007, pp. 389–392.

[123] P. Piccinini, S. Calderara, and R. Cucchiara, "Reliable smoke detection system in the domains of image energy and color," in *Proc. IEEE International Conference on Image Processing*, 2008, pp. 1376–1379.

[124] T. Porter and T. Duff, "Compositing digital images," *Computer Graphics*, vol. 18, no. 3, pp. 253–259, Jul. 1984.

[125] A. Rafiee, R. Tavakoli, R. Dianat, and S. Abbaspour, "Fire and smoke detection using wavelet analysis and disorder characteristics," in *Proc. International Conference on Computer Research and Development*, 2011, pp. 262–265.

[126] K. S. Reisinger, "Smoke detectors reducing deaths and injuries due to fire," *Pediatrics*, vol. 65, no. 4, pp. 718–724, Apr. 1980.

[127] F. G. Rodriguez, B. C. Arrue, and A. Ollero, "Smoke monitoring and measurement using image processing. application to forest fires," in *Proc. SPIE - The International Society for Optical Engineering*, vol. 5094, 2003, pp. 404–411.

[128] B. Sarel and M. Irani, "Separating transparent layers of repetitive dynamic behaviors," in *Proc. IEEE International Conference on Computer Vision*, vol. 1, 2005, pp. 26–32.

[129] Y. Y. Schechner, S. G. Narasimhan, and S. K. Nayar, "Instant dehazing of images using polarization," in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2001, pp. 325–332.

[130] Y. Y. Schechner, J. Shamir, and N. Kiryati, "Polarization-based decorrelation of transparent layers: the inclination angle of an invisible surface," in *Proc. IEEE International Conference on Computer Vision*, vol. 2, 1999, pp. 814–819.

[131] Y. Y. Schechner, N. Kiryati, and R. Basri, "Separation of transparent layers using focus," *International Journal of Computer Vision*, vol. 39, no. 1, pp. 25–39, Aug. 2000.

[132] T. Sentenac, Y. L. Maoult, and J. J. Orteu, "Overheating, flame, smoke, and freight movement detection algorithms based on charge-coupled device camera for aircraft cargo hold surveillance," *Optical Engineering*, vol. 43, no. 12, pp. 2935–2953, Dec. 2004.

[133] E. Shahrian and D. Rajan, "Weighted color and texture sample selection for image matting," in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2012, pp. 718–725.

[134] E. Shahrian, D. Rajan, B. Price, and S. Cohen, "Improving image matting using comprehensive sampling sets," in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2013, pp. 636–643.

[135] N. Shamir, N. Kopeika, and Z. Zalevsky, "Blind source separation of images based upon fractional autocorrelation," *Journal of Electronic Imaging*, vol. 21, no. 4, pp. 1–8, 2012.

[136] Y. H. Shiau, P. Y. Chen, H. Y. Yang, C. H. Chen, and S. S. Wang, "Weighted haze removal method with halo prevention," *Journal of Visual Communication and Image Representation*, vol. 25, no. 2, pp. 445–453, Feb. 2014.

[137] S. Shwartz, E. Namer, and Y. Y. Schechner, "Blind haze separation," in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2006, pp. 1984–1991.

[138] F. N. Simon and G. D. Rork, "Ionization-type smoke detectors," *Review of Scientific Instruments*, vol. 47, no. 1, pp. 74–80, Jan. 1976.

[139] M. Singh and X. Huang, "Computing layered surface representations: an algorithm for detecting and separating transparent overlays," in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2, 2003, pp. 11–18.

[140] A. Sobral and A. Vacavant, "A comprehensive review of background subtraction algorithms evaluated with synthetic and real videos," *Computer Vision and Image Understanding*, vol. 122, pp. 4–21, May 2014.

[141] T. Srisuwan and M. Ruchanurucks, "Smoke detection using GLCM, wavelet, and motion," in *Proc. SPIE - The International Society for Optical Engineering*, vol. 9069, 2014, pp. 1–5.

[142] J. L. Starck, M. Elad, and D. L. Donoho, "Image decomposition via the combination of sparse representations and a variational approach," *IEEE Transactions on Image Processing*, vol. 14, no. 10, pp. 1570–1582, Oct. 2005.

[143] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2, 1999, pp. 246–252.

[144] —, "Learning patterns of activity using real-time tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 747–757, Aug. 2000.

[145] J. Sun, J. Jia, C. K. Tang, and H. Y. Shum, "Poisson matting," *ACM Transactions on Graphics*, vol. 23, pp. 315–321, Aug. 2004.

[146] W. Sun, "A new single-image fog removal algorithm based on physical model," *Optik*, vol. 124, no. 21, pp. 4770–4775, 2013.

[147] R. Szeliski, S. Avidan, and P. Anandan, "Layer extraction from multiple images containing reflections and transparency," in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, 2000, pp. 246–253.

[148] R. T. Tan, "Visibility in bad weather from a single image," in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2008, pp. 2347–2354.

[149] J. P. Tarel and N. Hautiere, "Fast visibility restoration from a single color or gray level image," in *Proc. IEEE International Conference on Computer Vision*, 2009, pp. 2201–2208.

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

[151] A. Tonazzini, L. Bedini, and E. Salerno, "A markov model for blind image separation by a mean-field EM algorithm," *IEEE Transactions on Image Processing*, vol. 15, no. 2, pp. 473–482, Feb. 2006.

[152] B. U. Toreyin and A. E. Cetin, "Online detection of fire in video," in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2007, pp. 3371–3375.

[153] B. U. Toreyin, Y. Dedeoglu, U. Gudukbay, and A. E. Cetin, "Computer vision based method for real-time fire and flame detection," *Pattern Recognition Letters*, vol. 27, no. 1, pp. 49–58, Jan. 2006.

[154] B. U. Toreyin, Y. Dedeoglu, and A. E. Cetin, "Contour based smoke detection in video using wavelets," in *Proc. European Signal Processing Conference*, 2006.

[155] —, "Wavelet based real-time smoke detection in video," in *Proc. European Signal Processing Conference*, 2005.

[156] A. K. Tripathi and S. Mukhopadhyay, "Single image fog removal using anisotropic diffusion," *IET Image Processing*, vol. 6, no. 7, pp. 966–975, Oct. 2012.

[157] T. X. Tung and J. M. Kim, "An effective four-stage smoke-detection algorithm using video images for early fire-alarm systems," *Fire Safety Journal*, vol. 46, no. 5, pp. 276–282, Jul. 2011.

[158] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.

[159] E. S. Varnousfaderani and D. Rajan, "Weighted color and texture sample selection for image matting," *IEEE Transactions on Image Processing*, vol. 22, no. 11, pp. 4260–4270, Nov. 2013.

[160] R. Vezzani, S. Calderara, P. Piccinini, and R. Cucchiara, "Smoke detection in video surveillance: the use of visor (video surveillance on-line repository)," in *Proc. International Conference on Content-based Image and Video Retrieval*, 2008, pp. 289–297.

[161] P. Viola, M. J. Jones, and D. Snow, "Detecting pedestrians using patterns of motion and appearance," *International Journal of Computer Vision*, vol. 63, no. 2, pp. 153–161, Jul. 2005.

[162] D. Wang, T. Chen, Y. Yin, and T. Chen, "Smoke detection for early fire-alarming system based on video processing," *Journal of Digital Information Management*, vol. 6, no. 2, pp. 196–202, Apr. 2008.

[163] J. Wang and M. F. Cohen, "An iterative optimization approach for unified image segmentation and matting," in *Proc. IEEE International Conference on Computer Vision*, 2005, pp. 936–943.

[164] —, "Optimized color sampling for robust matting," in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2007, pp. 1–8.

[165] J. Wang, B. Qi, T. Wu, and X. Liu, "Fog effects modeling and removal for real-time vision applications," in *Proc. International Symposium on Information Science and Engineering*, 2012, pp. 149–155.

[166] J. Wang and M. F. Cohen, "Image and video matting: a survey," *Foundations and Trends in Computer Graphics and Vision*, 2008.

[167] Y. Wang, T. W. Chua, R. Chang, and N. T. Pham, "Real-time smoke detection using texture and color features," in *Proc. International Conference on Pattern Recognition*, 2012, pp. 1727–1730.

[168] Z. Wei, X. Wang, W. An, and J. Che, "Target-tracking based early fire smoke detection in video," in *Proc. IEEE International Conference on Image and Graphics*, 2009, pp. 172–176.

[169] J. Wright, Y. Ma, J. Mairal, G. Sapiro, T. S. Huang, and S. Yan, "Sparse representation for computer vision and pattern recognition," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 1031–1044, Jun. 2010.

[170] D. Wu, N. Wang, and H. Yan, "Smoke detection based on multi-feature fusion," in *Proc. International Congress on Image and Signal Processing*, 2012, pp. 220–223.

[171] M. Wu, N. Han, and Q. Luo, "A smoke detection algorithm based on discrete wavelet transform and correlation analysis," in *Proc. IEEE International Conference on Multimedia Information Networking and Security*, 2012, pp. 281–284.

[172] Z. Xiong, R. Caballero, H. Wang, A. M. Finn, and M. A. Lelic, "Video-based smoke detection: possibilities, techniques, and challenges," in *Proc. Fire Suppression and Detection Research and Applications-A Technical Working Conference*, 2007.

[173] Z. Xu and J. Xu, "Automatic fire smoke detection based on image visual features," in *Proc. IEEE International Conference on Computational Intelligence and Security Workshops*, 2007, pp. 316–319.

[174] Z. Xu, X. Liu, and N. Ji, "Fog removal from color images using contrast limited adaptive histogram equalization," in *Proc. International Congress on Image and Signal Processing*, 2009, pp. 1–5.

[175] C. Yang, W. Tseng, and J. Yang, "Reducing false alarm of video-based smoke detection by support vector machine," in *Lecture Notes in Computer Science*, vol. 5075, 2008, pp. 102–113.

[176] J. Yang, F. Chen, and W. Zhang, "Visual-based smoke detection using support vector machine," in *Proc. International Conference on Natural Computation*, 2008, pp. 301–305.

[177] S. K. Yeung, T. P. Wu, and C. K. Tang, "Extracting smooth and transparent layers from a single image," in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2008, pp. 1174–1180.

[178] I. Yoon, S. Kim, D. Kim, M. H. Hayes, and J. Paik, "Adaptive defogging with color correction in the HSV color space for consumer surveillance system," *IEEE Transactions on Consumer Electronics*, vol. 58, no. 1, pp. 111–116, Feb. 2012.

[179] I. Yoon, M. H. Hayes, and J. Paik, "Wavelength-adaptive image formation model and geometric classification for defogging unmanned aerial vehicle images," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2013, pp. 2454–2458.

[180] C. Yu, Y. Zhang, J. Fang, and J. Wang, "Texture analysis of smoke for real-time fire detection," in *Proc. IEEE International Workshop on Computer Science and Engineering*, 2009, pp. 511–515.

[181] C. Yu, J. Fang, J. Wang, and Y. Zhang, "Video fire smoke detection using motion and color features," *Fire Technology*, vol. 46, no. 3, pp. 651–663, Jul. 2010.

[182] J. Yu and Q. Liao, "Fast single image fog removal using edge-preserving smoothing," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2011, pp. 1245–1248.

[183] J. Yu, C. Xiao, and D. Li, "Physics-based fast single image fog removal," in *Proc. IEEE International Conference on Signal Processing*, 2010, pp. 1048–1052.

[184] X. Yu, J. Xu, D. Hu, and H. Xing, "A new blind image source separation algorithm based on feedback sparse component analysis," *Signal Processing*, vol. 93, no. 1, pp. 288–296, Jan. 2013.

[185] F. Yuan, "A double mapping framework for extraction of shape-invariant features based on multi-scale partitions with adaboost for video smoke detection," *Pattern Recognition*, vol. 45, no. 12, pp. 4326–4336, Dec. 2012.

[186] —, "A fast accumulative motion orientation model based on integral image for video smoke detection," *Pattern Recognition Letters*, vol. 29, no. 7, pp. 925–932, May 2008.

[187] —, "Video-based smoke detection with histogram sequence of LBP and LBPV pyramids," *Fire Safety Journal*, vol. 46, no. 3, pp. 132–139, Apr. 2011.

[188] W. Yuan, C. Yu, and Y. Zhang, "Based on wavelet transformation fire smoke detection method," in *Proc. International Conference on Electronic Measurement & Instruments*, 2009, pp. 872–875.

[189] W. Yuan, J. Li, J. Fang, and Y. Zhang, "Color model and method for video fire flame and smoke detection using fisher linear discriminant," *Optical Engineering*, vol. 52, no. 2, pp. 1–12, Feb. 2013.

[190] R. Zdunek and A. Cichocki, "Blind image separation using nonnegative matrix factorization with Gibbs smoothing," in *Lecture Notes in Computer Science*, 2008, pp. 519–528.

[191] Y. Zhai and X. Liu, "An improved fog-degraded image enhancement algorithm," in *Proc. International Conference on Wavelet Analysis and Pattern Recognition*, 2007, pp. 522–526.

[192] C. Zhang and I. Sato, "Image-based separation of reflective and fluorescent components using illumination variant and invariant color," *IEEE Transactions on*

Pattern Analysis and Machine Intelligence, vol. 35, no. 12, pp. 2866–2877, Dec. 2013.

[193] —, "Separating reflective and fluorescent components of an image," in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2011, pp. 185–192.

[194] D. Zhang, A. Hu, Y. Rao, J. Zhao, and J. Zhao, "Forest fire and smoke detection based on video image segmentation," in *Proc. SPIE - The International Society for Optical Engineering*, vol. 6788, 2007, pp. 1–7.

[195] H. Zhang and Y. Zhang, "Bayesian image separation with natural image prior," in *Proc. IEEE International Conference on Image Processing*, 2012, pp. 2097–2100.

[196] J. Zhang, L. Li, G. Yang, Y. Zhang, and J. Sun, "Local albedo-insensitive single image dehazing," *Visual Computer*, vol. 26, no. 6-8, pp. 761–768, Jun. 2010.

[197] Y. Zhao, W. Lu, Y. Zheng, and J. Wang, "An early smoke detection system based on increment of optical flow residual," in *Proc. International Conference on Machine Learning and Cybernetics*, 2012, pp. 1474–1479.

[198] Y. Zheng and C. Kambhamettu, "Learning based digital matting," in *Proc. IEEE International Conference on Computer Vision*, 2009, pp. 889–896.