Machine Vision and Applications (2011) 22:705–719
DOI 10.1007/s00138-010-0272-1

ORIGINAL PAPER

Vision based smoke detection system using image energy and color information

Simone Calderara · Paolo Piccinini · Rita Cucchiara

Received: 11 May 2009 / Accepted: 27 April 2010 / Published online: 21 May 2010

Abstract Smoke detection is a crucial task in many video surveillance applications and could have a great impact to raise the level of safety of urban areas. Many commercial smoke detection sensors exist but most of them cannot be applied in open space or outdoor scenarios. With this aim, the paper presents a smoke detection system that uses a common CCD camera sensor to detect smoke in images and trigger alarms. First, a proper background model is proposed to reliably extract smoke regions and avoid over-segmentation and false positives in outdoor scenarios where many distractors are present, such as moving trees or light reflexes. A novel Bayesian approach is adopted to detect smoke regions in the scene analyzing image energy by means of the Wavelet Transform coefficients and Color Information. A statistical model of image energy is built, using a temporal Gaussian Mixture, to analyze the energy decay that typically occurs when smoke covers the scene then the detection is strengthen evaluating the color blending between a reference smoke color and the input frame. The proposed system is capable of detecting rapidly smoke events both in night and in day conditions with a reduced number of false alarms hence is particularly suitable for monitoring large outdoor scenarios where common sensors would fail. An extensive experimental campaign both on recorded videos and live cameras evaluates the efficacy and efficiency of the system in many real world scenarios, such as outdoor storages and forests.

Keywords Smoke detection · Image processing · MoG · DWT

Electronic supplementary material The online version of this article (doi:10.1007/s00138-010-0272-1) contains supplementary material, which is available to authorized users.

S. Calderara (✉) · P. Piccinini · R. Cucchiara
DII, University of Modena and Reggio Emilia, Modena, Italy
e-mail: calderara.simone@unimore.it

P. Piccinini
e-mail: piccinini.paolo@unimore.it

R. Cucchiara
e-mail: cucchiara.rita@unimore.it

1 Introduction

Automatic or semi-automatic environmental surveillance is a key issue of the political policies for preventing crisis situation due to natural or artificial events that can compromise safety of people and security of buildings and nature. Among these events, smoke and fire are the most important for safety and commercial motivations.

Smoke detection in video streams is still an open challenge for computer vision and pattern recognition communities. It concerns the definition of robust approaches to detect, as soon as possible, spring and fast propagation of smoke possibly due to explosions, fires or special environmental conditions. For these motivations, a smoke detection module enriches standard video surveillance systems for both indoor and outdoor monitoring.

The video analysis tasks for smoke detection are not trivial due to the variability of shape, motion and texture patterns of smoke itself, which appearance is dependent on the luminance conditions, the background manifolds and colors of the scene. Since smoke modifies the visual cues of the background, typically background suppression techniques are adopted, followed by validation/classification tasks. The smoke identification becomes more challenging in presence of other moving objects, shadows and whenever the background is variable too.

In this paper, we propose the use of image energy and color information as key features for detecting smoke. In our

approach, moving objects are extracted with a novel and robust background suppression technique, that ensures a robust segmentation even in adverse conditions, reducing the number of false positives. Subsequently blobs are classified as either real objects or artifacts due to smoke presence in the scene. The image energy is computed using the Wavelet Transform and its temporal evolution modeled using a Gaussian Mixture Model (MoG) to capture slow energy decays that typically occur when smoke covers part of the scene.

In addition, the MoG classification is improved using a Bayesian framework accounting for the scene color variations to both speed up and improve the decision process.

It is worth noticing that, differently from most of previous solutions, we do not make any assumptions both on the external conditions and on the field of view of the camera making our system flexible enough to be applied on different setups and under several illumination conditions. We evaluate the system performance on the publicly available dataset on the website http://imagelab.ing.unimore.it/visor, measuring both the detection rate and the time to detect.

The results state a satisfying detection rate in several outdoor fixed-camera setups with a reduced number of false alarms and a recall of 100% that is desiderable in any smoke detection system.

2 Related works

The problem of studying environmental, natural or artificial, effects on video scenes have been deeply investigated in literature and certainly impacts everyone life in many different situations.

Among the phenomena that may be visually analyzed we can cite natural events such as weather condition changes (due to rain, fog, etc.) and artificial ones such as smoke due to fire or different sources (Table 1).

The approaches in literature on weather conditions detection treat, basically in two ways, the different nature of the analyzed events. On one side, fog and haze events are correlated to the problem of scene recovering and color restoration while on the other side, rain and snow are usually referred to the problem of detection and removal. In Narasimhan and Nayar [1], they propose an interesting method, based on atmospheric optic, to recover the "clear-day" scene color from two or more images taken under different and unknown weather conditions. They additionally developed a method for depth-segmentating and extracting the three-dimensional scene structure using two scattering models *Attenuation* and *Airlight* (scattering caused by fog or haze), imposing the constraint that both the observer and the object observed must lie at the ground level. Schechner et al. [2] propose a method for haze removal from an image. This method relies on the property that the natural environmental light, scattered by the atmosphere, is partially polarized. This approach does not rely on previous knowledge about the scattering model or knowledge about the illumination direction, but requires only two independent images. Oakley et al. [3], instead, use a statistical model for detecting airlights imposing a linear dependency between the real and the distorted pixel values.

Table 1 Summary of reference for several natural event detection techniques

Event	Technique	Authors
Haze	Statistical model	Oakley et al. [3]
	Airtight and Attenuation scattering model	Narasimhan and Nayar [1]
	Light polarization	Schechner et al. [2]
Rain	Setting camera parameters	Garg et al. [4]
	Intensity constrain, photometric constrain, spatio-temporal correlation	Garg et al. [5]
Fire	Color information	Phillips et al. [6]
	FFT and boundary analysis	Fastcom Tech.SA [7]
	FFT and shape analysis	Liu [8]
Smoke	Non-self similarity and motion irregularities	Kopilovic et al. [9]
	Chromatic analysis, growth-rate and disorder measure	Chen et al. [10]
	Mean Crossing Rate	Xiong et al. [11]
	Wavelet transform, energy analysis and shape analysis	Toreyin et al. [12]

Regarding the visual effect due to the rain presence, Garg et al. [13] studied the appearance of a single raindrop, developing a photometric and geometric model for the raindrop refraction and reflection assuming that a raindrop behaves like a wide-angle lens and although it is a transparent entity, its brightness is independent from the background brightness. In [4], a technique for removing the rain from a video without post-processing and without altering the scene perception is depicted. In [5], authors developed an algorithm capable to detect rain in a video sequence. The detection of the rain is composed by several steps: in the first step, all the pixels that present a peak in intensity over a set of three subsequent frames are selected; in the second step, the false positives are discarded using the photometric constraint; subsequently, the spatio-temporal correlation and the direction of the rain falls are computed. Although these methods are all of some interest in many industrial scenarios, such as open air storages, the problem of fire and smoke detection is definitely crucial and represents a hard challenge to be solved using cameras.

For this reason, we focus our attention on the problem of smoke detection and how to fast and reliably detect dangerous smoke presence in the scene. The problem of smoke

detection has been discussed in the past in some works where local features of pixels in the images or measures on the shape temporal-variations are exploited. In an early work, Kopilovic et al. [9] took advantage of irregularities in motion due to non-rigidity of smoke. They computed the optical flow field using two adjacent images, and then used the entropy of the motion directions distribution as key feature to differentiate smoke motion from non-smoke motion. Similarly, motion was exploited in [14] where local motions from cluster analysis of points in a multidimensional temporal embedding space are extracted. The goal was to track local dynamic envelopes of pixels, and then to use the velocity distribution histograms to discriminate between smoke and various natural phenomena such as clouds and wind-tossed trees that may cause such envelopes. In this work, the presence of other moving objects, typical of video-surveillance scenes, is not taken into account.

Recently, Chen, Yin et al. [10] present a smoke detection approach working on pixel-level classification after motion segmentation based on frame difference. Pixels can be initially classified as smoke-pixels with a very simple color based static decision rule; it is based on two thresholds in the color space assuming that smoke usually displays grayish colors. A further dynamic decision rule is dependent on the spreading attributes of smoke: the ratio between the sums of segmented smoke regions circumferences and the number of smoke-pixels extracted can give a measure of disorder in the segmented objects. Similarly other works evaluate the contour of the objects that are candidate to be classified as smoke. In [11], smoke detection is based on four steps: background subtraction, flickering extraction, contour initialization, and contour classification using both heuristic and empirical knowledge about smoke.

An interesting and robust approach has been defined by Toreyin et al. [15] and further improved in [12,16]. They use the Collins background subtraction method to extract moving objects [17]. Then, a flickering analysis and a measure of turbulence is provided by evaluating the edge and texture variations using the Wavelet Transform. In every block of the resulting sub-image, the energy variation is computed. The energy is given by the sum of the high-frequency components in the wavelet domain. Finally two thresholds are given to measure an acceptable energy variation. The dynamism of the variation is modeled with a simple three state Random Markov Model (RMM), trained with smoke and non-smoke pixels. Finally, an analysis of smoke shape complexity is provided as in [11,10], based on the distance between the contour points and the center of mass of the shape. This approach is quite robust in the given examples, but a precise evaluation of the different features contributions is not provided. However, in our experiments, we observed that sometimes a strong shape variation and edge complexity of smoke regions cannot be visually revealed due to both camera field of view and wind direction. For this reason we avoid to use this feature.

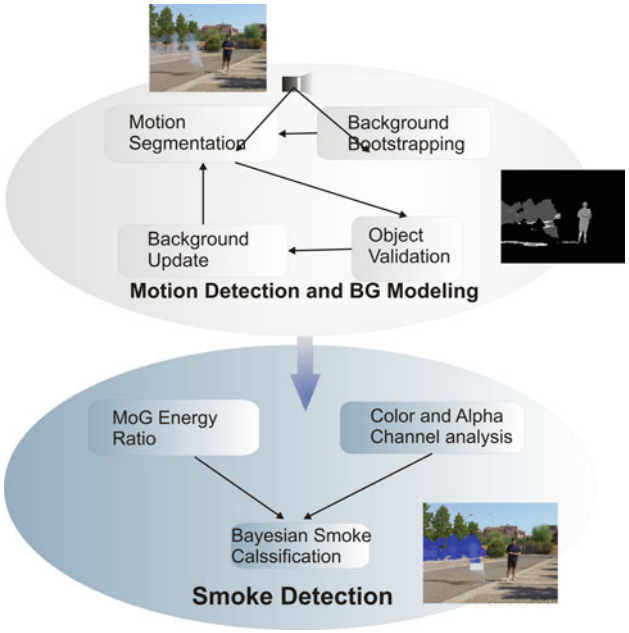

Fig. 1 Scheme overview of the smoke detection system

3 System overview

The proposed system scheme is depicted in Fig. 1. First, moving objects are extracted using a robust background model able to reliably segment motion in presence of several distractors due to noise in the scene. In particular, the proposed background is robust enough to remove all the possible sources of false alarms such as shadows or small objects that may affect negatively the precision of the smoke detection system. Image energy is then analyzed using the Wavelet Transform coefficients and the ratio between the background model energy and the current input frame. Based on this ratio, a statistical model, employing a temporal Mixture of Gaussians(MoG), is used to capture the energy decay that represents a good cue to detect possible smoke regions in the scene. In addition, color properties are used to detect the regions where a blending between a reference color and the background model occurs. This complex analysis will reveal image regions where a smoke event may occur. Finally, the smoke is detected looking at the amount of the intersection between candidate smoke regions and moving objects masks.

The paper is structured as follows.

In Sect. 4, the background model used for detecting moving objects is depicted. Section 5 describes the features used for smoke detection and the adopted Bayesian Classifier.

In conclusion, Sect. 6 details all the system parameters and discusses the experiments carried out.

4 Background model for smoke segmentation

The adopted motion detection algorithm is specifically designed to ensure a robust and reliable background estimation even in complex outdoor scenarios. It is a modification of the SAKBOT system [18], that increases the robustness in outdoor uncontrolled environment. The SAKBOT background model is a temporal median model with knowledge-based update stage. That means that regions of the image known as moving objects are not included in the background updating stage and consequently not injected in the median model. This updating strategy ensures a clean background and the ability of correctly extract slightly moving objects or static ones. Suitable modification to the model in [18] improving the background initialization, motion detection and object validation have been developed. The desired background model should exhibit some properties that are important to correctly detect smoke objects in the scene and obtain high accuracy results, in particular:

- The detection must be accurate in night and day condition
- No ghosts (i.e apparently moving objects due to an incorrect background updating or structural modification of the scene, such as a parked car that moves away), micro-objects and shadows must be present in the scene
- The background must constantly update to respond to environmental variations
- The background must not incorporate the smoke in the background model so selectivity is needed.

For these reasons, we focus on background bootstrapping, updating and ghosts removal to avoid the presence of noise in the scene with the main purpose of obtaining clear blobs for reliable smoke detection.

4.1 Background bootstrapping

The first issue in every background suppression approach is the background initialization or *bootstrapping*, which needs to be both fast and accurate. Unfortunately, it is often impossible to have a clear background for many frames in order to compute the statistics for building the model. Therefore, it is important to implement a method that can initialize the background model as quickly as possible even starting from "dirty" frames.

Our approach basically partitions the image into blocks (of 16×16 pixels) and selectively updates the background model with a block whenever a sufficiently high number of pixels within the block are detected as static. Motion is evaluated with a thresholded single difference between two consecutive frames. If more than 95% of the pixels in the block are detected as not in motion, the bootstrapping model is updated with those pixels. When this occurs for more than 10 times (even non consecutive), the whole block is considered "stable" and no longer evaluated. Once all the blocks have been set to "stable" the background model is ready; additionally, the threshold for the single difference is increased to avoid deadlocks and to speed up the bootstrapping, if no blocks change their state to "stable" for two consecutive frames.

4.2 Background updating and foreground extraction

Once the model is properly initialized, the background model is updated using a temporal median. A fixed $k+1$-sized circular buffer is used to collect the past k values for every pixel over time. In addition to the k values, the last stable background model value, for the pixel (i, j) at time $t-1$, $\mathrm{BG}_t(i, j)_{t-1}$ is added to the buffer to account for the last reliable background information available. These $n = k+1$ values are then ordered according to their gray-level intensity, and the median value is used as an estimate of the current background model $\mathrm{BG}_t(i, j)_t$ at time t for the pixel (i, j).

Once the background model has been created and updated, the foreground is extracted, frame by frame, using the background differencing technique. In the following, we will refer to the current background model as BG_t omitting the time index. The difference between the current image I_t and the background model BG_t is computed and the foreground mask $M_t(i, j)$ extracted:

$$M_t(i, j) = \frac{(I_t(i, j) - \mathrm{BG}_t(i, j)) \cdot \mathbf{i}^{\mathrm{T}}}{3} \tag{1}$$

where $\mathbf{i}^{\mathrm{T}}$ is the 1×3 identity vector. The mask $M_t(i, j)$ is then binarized using two different thresholds: a low threshold T_{low}, to filter out the noisy pixel extracted due to small intensity variations, and a high threshold T_{high} to identify the pixels where a large intensity variation occurs. Both these thresholds are local, i.e. they have different values for each pixel of the image. Let $b_p(i, j)$ be the value at position p inside the ordered circular buffer b of pixel location (i, j) and, consequently, $b_{\frac{k+1}{2}+1}$ the median. The thresholds are computed as the weighted difference between buffer elements around the median:

$$T_{\mathrm{low}}(i, j) = \lambda \left(b_{\frac{k+1}{2}+l} - b_{\frac{k+1}{2}-l} \right) \tag{2}$$

$$T_{\mathrm{high}}(i, j) = \lambda \left(b_{\frac{k+1}{2}+h} - b_{\frac{k+1}{2}-h} \right) \tag{3}$$

where λ is a fixed multiplier, while l and h are fixed scalar values. We experimentally set $\lambda = 7$, $l = 2$ and $h = 4$, for a buffer of $n = 9$ values.

It is straightforward to see that, being the vector b ordered, T_{high} is always higher or equal than T_{low}. Our experiments

demonstrated that these settings perform well in most common surveillance scenarios. The reason for the adoption of dynamic per-pixel thresholds is trivially explained by the fact that using fixed per-frame thresholds makes the system less reactive to local illumination changes.

The final binarized motion mask B_t is obtained as the composition of the two binarized motion masks computed, respectively, using the low and the high thresholds: a pixel is marked as foreground in B_t if it is present in the low-thresholded binarized mask AND it is spatially connected to at least one pixel in the high-thresholded binarized mask then adopting a two-pass labeling algorithm the list MVO_t of moving objects, at time t, is extracted from B_t.

Every element MVO_t^i of the list is a candidate foreground object and small objects due to oversegmentation are discarded according to their area.

4.3 Object validation

Once the moving visual objects are extracted, the object-level validation step is performed in order to remove all the moving objects generated by small recurrent motion in the scene, for example waving trees. This validation is performed accounting for joint contributions coming from color information and intensity gradient of the objects.

The gradient is computed with respect to both spatial and temporal coordinates taking the directional derivative along the spatio-temporal direction $\mathbf{v_i} = [i, t]^{\mathrm{T}}$ and $\mathbf{v_j} = [j, t]^{\mathrm{T}}$:

$$\begin{aligned} \frac{\partial I_t(i,j)}{\partial(x,t)} &= I_{t-\Delta t}(i-1,j) - I_t(i,j) \\ \frac{\partial I_t(i,j)}{\partial(y,t)} &= I_{t-\Delta t}(i,j-1) - I_t(i,j) \end{aligned} \tag{4}$$

For stationary points, we can approximate the past sample $I_{t-\Delta t}$ with the background model BG_t:

$$\begin{aligned} \frac{\partial I_t(i,j)}{\partial(x,t)} &= \mathrm{BG}_t(i-1,j) - I_t(i,j) \\ \frac{\partial I_t(i,j)}{\partial(y,t)} &= \mathrm{BG}_t(i,j-1) - I_t(i,j) \end{aligned} \tag{5}$$

The gradient module is obtained by using the following equation:

$$G_t = \left\{ g(i,j) \mid g(i,j) = \sqrt{\left\| \frac{\partial I_t(i,j)}{\partial(x,t)} \right\|^2 + \left\| \frac{\partial I_t(i,j)}{\partial(y,t)} \right\|^2} \right\} \tag{6}$$

From our tests, it emerges that this gradient module is quite robust against small motions in the background, mainly thanks to the use of spatio-temporal partial derivative. Small moving objects, subjected to micro-movement, share the property of having their gradient quite constant in the spatio-tempo direction. Moreover, the joint spatio-temporal derivative makes the object gradient computation more accurate, since it also detects gradient in the inner parts of the object.

Given the list of moving objects, at time t, MVO_t, the gradient G_t is compared with the gradient (in the spatial domain) of the background GBG_t in order to evaluate their coherence. This *gradient coherence* GC_t is computed over a $k \times k$ pixel-neighborhood as the minimum of absolute differences between G_t and GBG_t:

$$\mathrm{GC}_t(i,j) = \min_{\substack{i-k \le x \le i+k \\ j-k \le y \le j+k}} |G_t(i,j) - \mathrm{GBG}_t(x,y)| \tag{7}$$

Unfortunately, when the gradient module (either G_t or GBG_t) is close to zero, data are not reliable consequently we combine the gradient coherence with the *color coherence*, CC_t, to strengthen the validation taking differences between current input frame I_t and background model BG_t:

$$\mathrm{CC}_t(i,j) = \min_{\substack{i-k \le x \le i+k \\ j-k \le y \le j+k}} \|I_t(i,j) - \mathrm{BG}_t(x,y)\| \tag{8}$$

where $\|.\|$ represents the norm in the RGB color space.

The overall validation score VS_t^i is the normalized sum, computed for every object, of the per-pixel validation scores, obtained by multiplying both the coherence measures:

$$VS_t^i = \frac{\sum_{(i,j) \in \mathrm{MVO}_t^i} \mathrm{GC}_t(i,j) * \mathrm{CC}_t(i,j)}{N_t^i} \tag{9}$$

where N_t^i is the area of the ith moving visual objects, MVO_t^i. Validation is obtained thresholding objects' validation score and, if below the threshold, the object is discarded and its pixels are marked as belonging to background and used in the update stage.

4.4 Fast ghost suppression

When adopting a selective update scheme, ghosts may affect the extraction of moving objects. Specifically, a ghost occurs whenever a static object is included in the background model; subsequently, the object disappears, resulting in a false foreground detection and causing a deadlock in the update scheme. It is then evident that a method to detect ghosts and force them into the background model must be developed. Our approach is similar to the one we used for background bootstrapping , but at region level instead of pixel level. All the validated objects are used to build an image called A_t that accounts for the number of times when a pixel is detected as static by the single difference:

$$A_t(i,j) = \begin{cases} A_{t-1}(i,j) + 1 & \text{if } static \\ A_{t-1}(i,j)/2 & \text{otherwise} \end{cases} \tag{10}$$

A valid object MVO_t^i is then classified as a ghost if:

$$\frac{\sum_{(i,j)\in \text{MVO}_t^i} A_t(i,j)}{N_t^i} > T_{\text{ghost}} \tag{11}$$

where T_{ghost} is the threshold on the amount of time the points of the MVO_t^i are sensed static.

The threshold T_{ghost} depends on the frame rate and on the considered scenario. We set for a common surveillance scenario, with a frame rate of 20 *fps*, a value of 200 that means that the ghost must remain in its position for 10 s before it is removed from the background.

Practically, in the case of pixels belonging to a ghost, the single difference will be lower than the threshold and will start increasing the value in A_t. When the average sum of the accumulator values, for the moving visual objects, exceeds the threshold T_{ghost}, the object is forced into the background. With the described procedure, the system is capable of detecting many visual objects in different outdoor and indoor conditions, and can be used as a first step of a tracking system [19] or as the basis for complex object classification applications. In the smoke detection application the tracking is not necessary since the image is analyzed blockwise and the objects are used only for classifying possible smoke regions.

5 Smoke detection for foreground object classification

Once it is possible to detect automatically moving objects that differ from the background model, every object of interest must be analyzed to find smoke presence. The proposed model evaluates the joint contributions coming from the gray-level image energy and color intensity attenuation to classify a moving visual object as possible smoke under the assumption that when smoke grows and propagates in the scene its image energy is attenuated by the blurring effect of smoke diffusion.

We first detect possible candidate objects, by means of the motion segmentation algorithm previously proposed, then their energy is analyzed using the Wavelet Transform coefficients evaluating its temporal evolution. The color properties of the objects are analyzed accordingly to a smoke reference color model to detect if color changes in the scene are due to a natural variation or not. The input image is then divided in blocks of fixed sized and every block is evaluated separately. To perform the final classification, a Bayesian approach discriminates whether a foreground object is smoke or not.

5.1 Energy analysis using the discrete wavelet transform

When smoke diffuses in the scene, it partially covers background elements. This motivates the use of the image energy as an important feature to find smoke presence. An efficient way to evaluate the energy variation of an intensity image is the discrete Wavelet Transform DWT [20].

The DWT is obtained convolving the image signal with several banks of filters obtaining a multiresolution decomposition of the image. Given the input image I_t the decomposition produces four subimages, namely the compressed version of the original image C_t, the horizontal coefficients image H_t, the vertical coefficients image V_t and the diagonal coefficients image D_t. An example of decomposition, computed with the algorithm proposed in [20], is shown in Fig. 2.

The energy is evaluated blockwise dividing the image in regular blocks, b_k, of fixed size and summing up the squared contribution coming from every coefficient image:

$$E(b_k, I_t) = \sum_{i,j\in b_k} V_t^2(i,j) + H_t^2(i,j) + D_t^2(i,j) \tag{12}$$

The energy value of a specific block varies significantly over time in presence of smoke, Fig. 3.

When the smoke covers part of the scene the edges are smoothed and the energy consequently lowered. This energy drop can be further emphasized computing the ratio r_k between the image energy of the current input frame I_t and the one of the background model BG_t. The energy ratio has the advantage of normalizing the energy values and allowing the comparison between different scenes where the block

Fig. 2 Example of discrete Wavelet Transform. The leftmost image is the original image. The right image is the transformed one. The components are: *top left* compressed image C_t, *top right* horizontal coefficient image H_t, *bottom left* vertical coefficient image V_t and *bottom right* diagonal coefficient image D_t

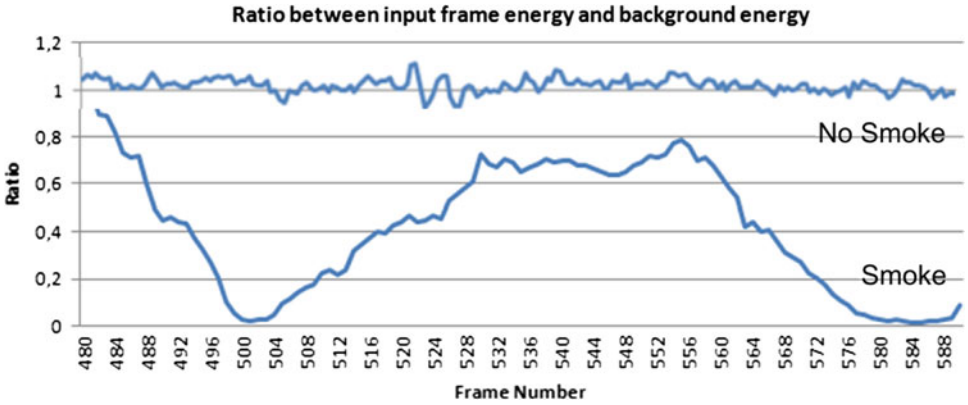

Fig. 3 Energy ratio trend of both non-smoke and smoke blocks. In presence of smoke the energy ratio is subjected to gradual drops in its value

energy itself can vary significantly. The ratio of the block b_k is given by:

$$r_k(b_k, I_t, \mathrm{BG}_t) = \frac{E(b_k, I_t)}{E(b_k, \mathrm{BG}_t)} \tag{13}$$

The analysis of the energy ratio is performed in two different contexts to account for both global and local energy decays.

The image energy ratio of Eq. (13) is computed for several clips containing at least one smoke event; the computation is performed over the entire image, summing the energy ratio contributions coming from every block, with the purpose of finding an average value of the global energy ratio in the case of smoke presence in the scene. The energy values are then collected, for all the images and all the clips, and modeled as a probability distribution to correctly capture both the average values and the variations around these. We chose to learn the probability distribution using a non-parametric approach because the distribution shape is not a priori predictable.

In particular, the *Parzen window* technique is adopted to build, using a Gaussian kernel, a non-parametric distribution from energy-ratio values computed on several clips. The Parzen window method is a kernel density estimator that computes a non-parametric distribution from a set of iid samples $X = \{x_i \mid i = 1, \ldots, N\}$ of a random variable x. The approximated pdf is computed summing the kernel for all the sampled values, using, as kernel, the standard Gaussian kernel function $K = \frac{1}{2\pi} e^{-\frac{1}{2}x^2}$:

$$\widehat{f} = \frac{1}{N\,h} \sum_{i=1}^{N} K(x - x_i) \tag{14}$$

In addition to the global energy evaluation, every block is locally evaluated frame-by-frame to capture the temporal evolution of the energy ratio. When an energy drop is observed for a significant period of time, an edge smoothing process occurs. The edge smoothing process can be affected by noise due to light variation in the scene. A Mixture of Gaussian model is adopted to improve the robustness of the analysis. The MoG has the great advantage to correctly model the variations of a multi-modal stochastic process. To compute the Gaussian Mixture's parameter the on-line Expectation Maximization algorithm proposed in [21] is used every frame. The on-line EM updating is based on the concept of *sufficient statistics*. A statistic $T(x)$ is sufficient for underlying parameter η if the conditional probability distribution of the data θ, given the statistic $T(\eta)$, is independent of the parameter η. Thanks to the Fisher–Neyman's factorization theorem [22], the likelihood function $L_\eta(x)$ of x can be factorized in two components, one independent by the parameters η and the other dependent on them only through the sufficient statistics $T(x)$: $L_\eta(x) = h(x) g_\eta(T(x))$. It has been shown [23] that in the case of distributions of the exponential family (such as the Gaussian pdf) the factorization theorem can be written as:

$$p(x|\eta) = h(x)\, g(\eta) \exp\left\{\eta^{\mathrm{T}} T(x)\right\} \tag{15}$$

Thus, the sufficient statistics for a single Gaussian distribution are:

$$T(\boldsymbol{x}) = \begin{bmatrix} x_i \\ x_i^2 \end{bmatrix} \tag{16}$$

In the case of a mixture of distributions belonging to the exponential family, the on-line updating of the mixture parameters can be obtained by simply updating the sufficient statistics (s.s.) of the mixture:

$$T_M(\boldsymbol{x}) = \sum_{k=1}^{K} \gamma_k T_k(\boldsymbol{x}) \tag{17}$$

where $T_k(\boldsymbol{x})$ are the s.s. for the kth single distribution. The updating process (having observed up to the sample $(i-1)$th) can be obtained as:

$$T_k^i(\boldsymbol{x}) = \alpha(i)\gamma_k T_k(x_i) + (1 - \alpha(i))\, T_k^{i-1}(\boldsymbol{x}) \tag{18}$$

where $T_k(x_i) = \begin{bmatrix} x_i \\ x_i^2 \end{bmatrix}$.

In detail, for every block b_k of the image I_t, at time t, the value $r_k(b_k, I_t, \mathrm{BG}_t)$ is computed and the MoG of block b_k updated using Eq. (18).

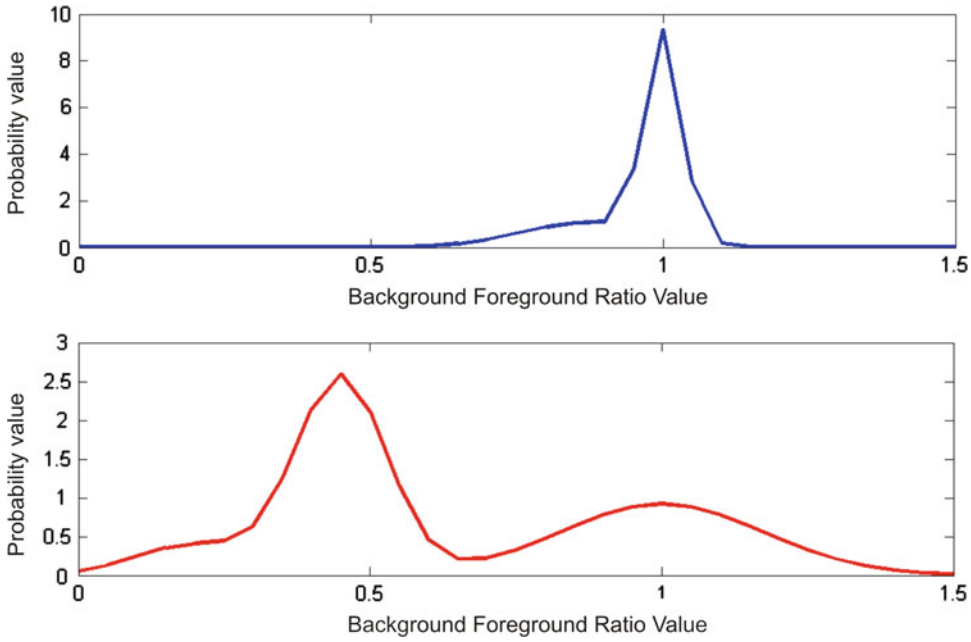

Fig. 4 Gaussian mixtures obtained observing energy ratio values at a single block. The *upper plot* shows the mixture when there is no smoke in the block. The *lower plot* shows how the mixture changes when smoke is in the scene. It is worth noting that when a block is covered by smoke the mixture components mean values move gradually toward 0. Readers please note that axis scale factor are different in the two plots

This process has the main advantage that the mixture component re-weighting process is able to model slow and gradual variations of energy ratio. Values that do not occur frequently are filtered out and assigned to the least probable Gaussian of the mixture. This property is helpful for evaluating the gradient energy lowering process of smoking regions that has the peculiarity of being slow and continuous in time, Fig. 4.

To capture the temporal variation of the energy ratio, the Gaussian Mixture Model was preferred to a Hidden Markov model (HMM). Although HMMs are widely adopted to classify and model temporal stochastic processes, the data values sequence is crucial to obtain a good classification. Instead, as previously stated, the block energy ratio is subject to strong fluctuations of energy values due to noise and natural scene lighting. This reason makes the lowering sequence unpredictably variable in different setups; thus the specific energy drop time series can produce misleading results. On the contrary, it is interesting to analyze the global trend. The image energy is indeed a good clue to detect smoke regions but it may gradually drop for many different reasons and for illumination changes, consequently additional features must be used in conjunction with energy itself to obtain a reliable detection.

5.2 Color analysis to detect blended smoke regions

When a smoke event occurs scene regions, covered by smoke, change their color properties. The smoke can either be completely opaque or partially transparent. In the former case, the covered region changes completely its color while in the latter case the color of the covered region appears to be blended with the smoke color as shown in Fig. 5.

This simple observation remains valid in all the observed cases and intuitively suggests a hint to characterize the color of a smoke region.

The proposed model simply adopts an evaluation based on a blending function inspired by computer graphics. A reference color model is chosen in the RGB color space to represent the color of the smoke in the scene. The model is selected by analyzing the different color tones produced combusting different materials. For explanatory purposes, it is possible to concentrate the analysis to the case of a light gray color model as the smoke in the leftmost image of Fig. 2. Every pixel $I_t(i, j)$, in the input frame at time t, is then checked against the smoke model and the background model BG_t to evaluate the reference color presence computing the blending parameter bl using Eq. (19). The evaluation takes into account the case where the scene color and the smoke color are mixed together.

$$bl(i, j, I_t, \mathrm{BG}_t, S) = \frac{I_t(i, j) - \mathrm{BG}_t(i, j)}{S - \mathrm{BG}_t(i, j)} \tag{19}$$

where BG_t is the current background model at time t and S is the RGB vector of the smoke reference color model.

To filter out the errors and possible measurement inaccuracies, the blending value is computed for every image block as the average of bl values in the block:

$$\beta_{b_k}(I_t, \mathrm{BG}_t, S) = \frac{1}{N^2} \sum_{i,j \in ,b_k} \frac{I_t(i, j) - \mathrm{BG}_t(i, j)}{S - \mathrm{BG}_t(i, j)} \tag{20}$$

where the block size is $N \times N$.

Fig. 5 Example of the blending effects of smoke

Table 2 Comparison of smoke detection precision and recall using MoG background and the proposed background model

	MoG model (%)		Proposed background model (%)	
	Precision	Recall	Precision	Recall
Energy	94	96	100	95
Color	92	92	100	92
Total	94	98	100	99

In conclusion β quantifies how much each block globally shares chromatic properties with the reference color model.

5.3 A Bayesian approach for classification

In the previous subsections the block-wise energy ratio r and the color blending value β have been presented as possible features to identify a smoke region in the scene. A Bayesian formulation has been chosen to identify whether a block b_k is likely to belong to a smoke region or not. For every block the posterior probability of smoke presence in block b_k, event $f = 1$, is defined:

$$P(f = 1|b_k) \propto P(b_k|f = 1)P(f = 1) \tag{21}$$

The likelihood value is obtained by combining both the contributions coming from energy ratio and color information considering both terms statistically independent to simplify the treatment.

$$\begin{aligned} P(b_k|f = 1) &= P(r_k, \beta_{b_k}|f = 1) \\ &= P_r(b_k|f = 1) \cdot P_\beta(b_k|f = 1) \end{aligned} \tag{22}$$

The likelihood contribution due to energy ratio decay is obtained by summing the weighted Gaussians of the MOG having mean value below a considered threshold computed empirically observing the average energy ratio values in smoke regions.

$$P_r(b_k|f = 1) = \sum_{i=1}^{K} w_i N(r(b_k, I_t, \mathrm{BG}_t)|\mu_i\, \sigma_i) \tag{23}$$

when the ith Gaussian mean value μ_i is below the threshold.

The color contribution to the likelihood value is directly computed as the block color blending measure β_{b_k} according to Eq. (20).

$$P_\beta(b_k|f = 1) = \beta_{b_k}(I_t, \mathrm{BG}_t, S) \tag{24}$$

The classification is biased making use of prior knowledge acquired observing several clips containing smoke. The prior probability of a smoke event, in the current frame, is directly

Table 3 System results on 15 reference clips

The detection rate was evaluated using energy and color component respectively and their joint contributions. Time to detect(TtD) and False positives rate(FP) are reported for each considered contribution

Clip	Frame No	Type	Temporal analysis		Color analysis		Global analysis	
			TtD	FP	TtD	FP	TtD	FP
Clip1	165	Outdoor	22	–	1	–	1	–
Clip2	210	Indoor	18	–	1	–	1	–
Clip3	2200	Outdoor	28	–	34	–	20	–
Clip4	3005	Indoor	212	–	273	–	285	–
Clip5	1835	Indoor	87	–	100	3	52	–
Clip6	2345	Outdoor	129	–	161	–	116	–
Clip7	2024	Indoor	57	3	99	–	35	–
Clip8	2151	Outdoor	88	2	88	–	42	–
Clip9	1880	Outdoor	59	–	56	–	45	–
Clip10	2953	Outdoor	457	–	498	–	300	–
Clip11	1485	Indoor	62	–	*x*	5	62	–
Clip12	499	Outdoor	43	–	8	–	16	–
Clip13	195	Indoor	53	–	23	–	27	–
Clip14	1226	Outdoor	77	–	370	–	69	–
Clip15	109	Outdoor	29	–	*x*	1	3	–

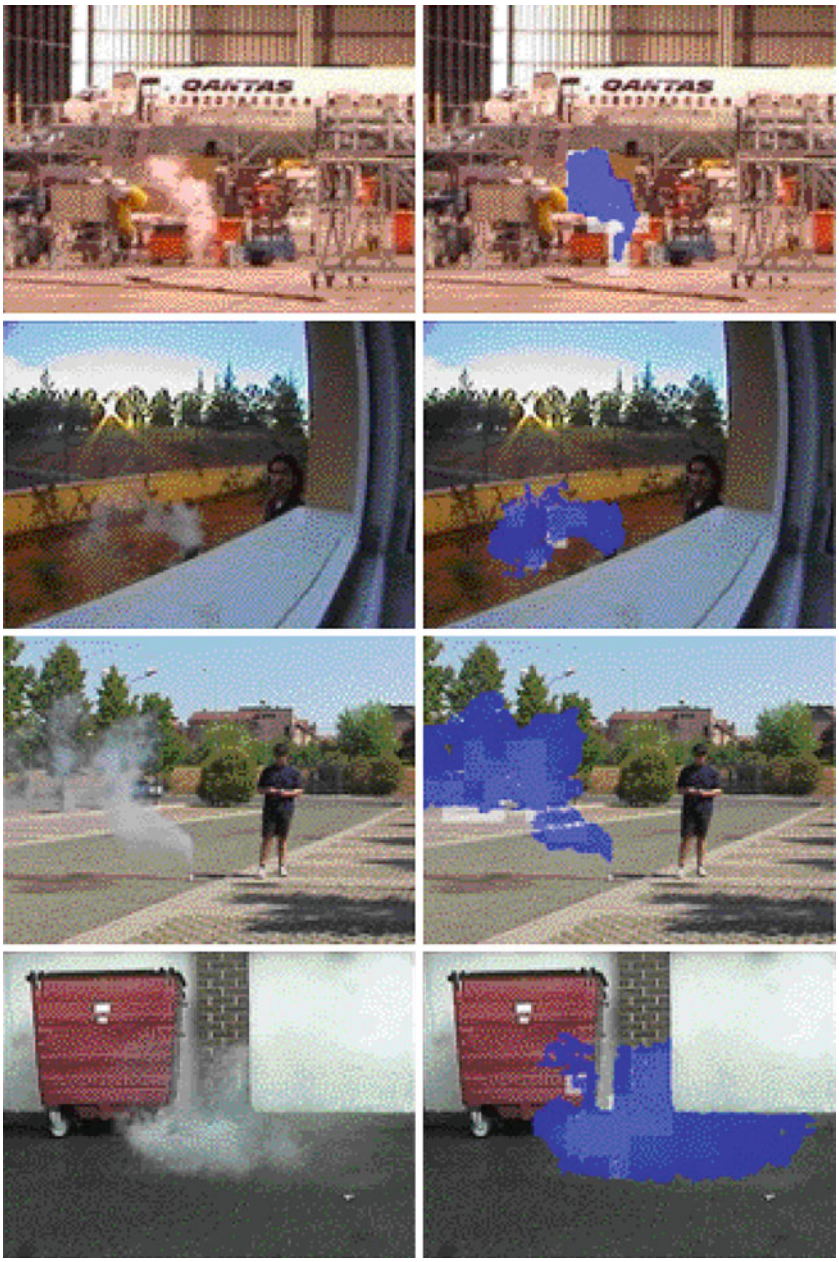

Fig. 6 Snapshots of the proposed system working on several clips in different conditions. The *dark grey area* in the rightmost images is detected as smoke

related to the mean energy ratio value of the scene and computed using the non parametric distribution obtained by Eq. (14).

$$P(f=1) = \widehat{f}\left(\frac{1}{M}\sum_{\forall b_k \in I_t} r(b_k, I_t, \mathrm{BG}_t)\right) \tag{25}$$

where I_t is composed by M blocks.

If the posterior probability is greater than 0.5 the block is then considered as a candidate smoke block. The test for smoke presence is performed after foreground objects segmentation. For every segmented object in the scene, the number of candidate blocks intersecting the object's blob is computed. Finally, a MVO is classified as smoke when the 70% of its area overlays candidate smoke blocks.

6 Experimental results and discussion

The proposed smoke detection system can be used in conjunction with a whichever video surveillance system providing moving object segmentation using a background model. The background model should be updated regularly but smoke regions should not be included in the background. This can be achieved choosing a slow background update rate and avoiding to update the background areas where a

Distance from camera	Time to detect
10m	6 sec
20m	7 sec
30m	8 sec
60m	10 sec
80m	10 sec

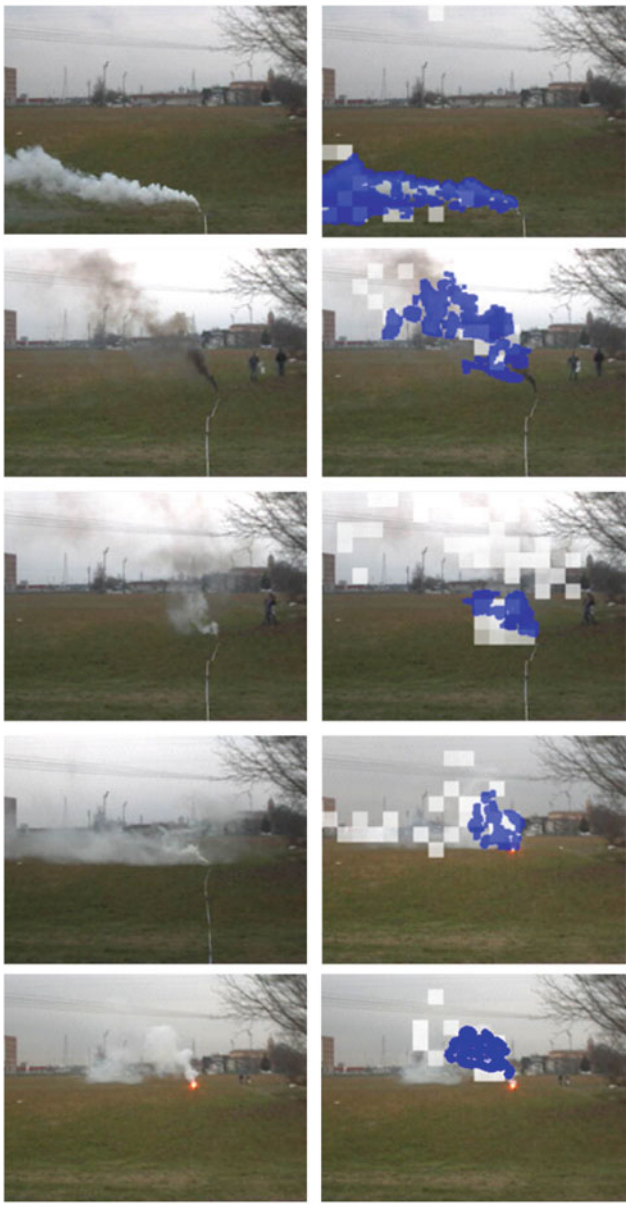

Fig. 7 System test measuring the distance from camera and the relative time to detect of the smoke event in a field scenario

smoke object is detected. The tests were performed using both the Stauffer and Grimson background model with selective update [24] and the background model proposed in Sect. 4 and results in terms of precision and recall were compared to demonstrate that the proposed background can improve the reliability of the smoke detection system as shown in Table 2.

During the tests all the system parameters were left unchanged to evaluate the degree of generalization of the whole approach in many different scenarios and under different environmental conditions. The system was tested on 50 clips of varying length in both indoor and outdoor setups where moving objects such as people or vehicles were present in the scene during the smoke event. Each clip contained a smoke event. Part of the dataset is publicly available at website http://imagelab.ing.unimore.it/visor. Each likelihood term was evaluated separately to measure the impact on the system performance.

Table 3 summarizes the results obtained on 15 reference clips.

The first column of the table reports the video type and its frame-length. The average clips framerate is 25fps. The

Distance from camera	Time to detect
6m	4 sec
15m	5 sec
25m	8 sec
45m	10 sec

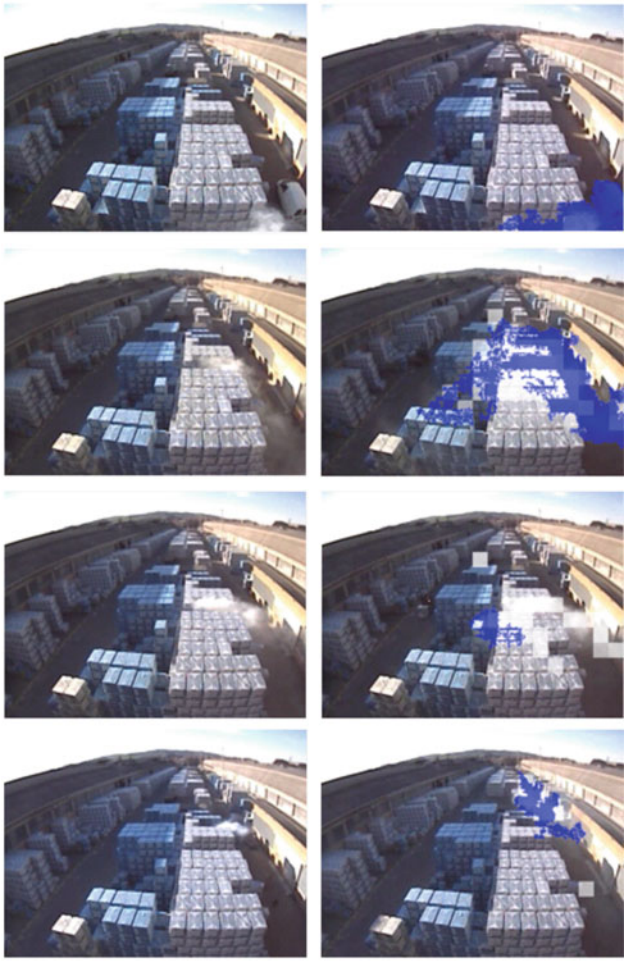

Fig. 8 System test measuring the distance from camera and the relative time to detect of the smoke event in an outdoor industrial storage scenario

remaining columns report the results obtained using every likelihood term separately and finally the results of the complete system. The detection time after the smoke event occurs is reported for all the test clips. The table clearly shows that the likelihood term due to temporal analysis (Eq.23) is effective in most of the observed cases. The main problem is the long detection time. This is caused by the time based statistics used to capture the energy ratio decay. Although the likelihood contribution due to color blending has the advantage of speeding up the detection process it tends to detect much false positives if used alone. See seventh column of Table 3. By observing the last two columns of Table 3, we can state that the complete approach is fast and reliable enough even in situations where each likelihood contribution fails. The overall system results on the 50 clips, used for testing purposes, report a detection rate of 77% 3 s later the smoke event occurs, 98.5% 6 s after and finally 100% 10 s afterwards with an average false positive rate of 4%. Figure 6 shows some snapshots of the system working on different conditions.

Additional experiments were carried out to test the system in real conditions. In particular, the smoke detection system was tested in several scenarios generating a synthetic smoke event at different distances from the cameras. First, a field scenario was used where no occlusions interfered with the smoke propagation in the scene and the time to detect of the system acquired varying the distance the smoke event generated itself. Figure 7 shows the results in the field scenario. It is remarkable to note that the system has a maximum time to detect of 10 s and is able to detect smoke in absence of occlusion 90 meters far from the camera. Several snapshots of the system triggering the alarm are presented. It is valuable to observe that the tests were performed in different days under different conditions and no false alarms were reported in a week. Another test campaign verified the system applicability in outdoor industrial storages scenarios. In these scenarios, several occlusions may interfere with the smoke detection and the detection itself becomes more challenging. To partially overcome this problem, it is possible to mount

Fig. 9 System test performed in an outdoor industrial storage in night mode. The *first row* shows how the system is capable of detecting smoke at dawn using color images. The *remaining rows* show how the system perform in a case of a night and day camera that switch to gray-level images during night

Fig. 10 System tested in a wood scenario. *Gray blocks* reveal possible smoke region and the *dark grey area* reveals smoke after the alarm has triggered

the cameras on sufficiently high poles capturing the smoke far enough from the ground. This leads to a slightly slower time to detect but does not affect the system efficacy. Results are shown in Fig. 8.

In addition, the system was tested in night condition using night and day cameras to acquire images even when a small illumination is present in the scene, Fig. 9. In this case, the system exhibits a detection rate of smoke events of 100% but the average number of false alarms, after a week working 24 h a day, increases a little. In particular, the average number of false alarms in day condition is *one per week* in an outdoor storage scenario, where operating vehicles are present and several noises affected the scene, in contrast to an ideal scenario, such as the field one, where the false alarm rate is *one per 2 weeks*. In the case of night detection, the number of false alarms in the outdoor storage scenario raises two *one per 3 days*. This is mainly due to the headlights of the vehicles that move in the scene that saturate the cameras sensors. This could be partially avoided using a proper headlights detector to improve the performance of the system.

Another test campaign was performed in a forest scenario, Fig. 10, where the woods may severely occlude the smoke. In this case the performances are the same as in the field scenarios but the average time to detect, in the worst case (80 m distance), raises to 12 s. Concluding the system was tested on both video clips of smoke events and real scenarios for several days to test the performances in terms of both accuracy and reliability of detecting the smoke events with a reduced amount of false alarms. Many challenging situation were faced during the test campaigns and the system responses were adequate in every considered case. The proposed smoke detection complete system is then able to cover a maximum area of 80 m^2 with one camera detecting smoke in less than 13 s in every operation condition, day and night and in the case of severe occlusions. The number of false alarms is acceptable and could be further improved by considering the headlight problem in the night scenario.

7 Conclusions

In conclusion, we propose a system capable of detecting if foreground objects are smoke or not using both Wavelet Transform energy coefficients and image color properties. The proposed Bayesian approach has been extensively evaluated on public data and results in terms of detection rate and time to detect have been reported. The adoption of a two-contributions likelihood measure solves most of the emerged problems of each chosen feature and boosts up significantly the detection process. The system performed well in all the tested scenarios and results are very robust.

References

1. Narasimhan Srinivasa, G., Nayar Shree, K.: Vision and the atmosphere. Int. J. Comput. Vision **48**(3), 233–254 (2002)
2. Schechner Yoav, Y., Narasimhan Srinivasa, G. Nayar, Shree, K.: Polarization-based vision through haze. Appl. Opt. **42**, 3 (2003)
3. Oakley, J. P., Bu, H.: Correction of simple contrast loss in color images. IEEE Trans. Image Process. **16**(2), 511–522
4. Garg, K., Nayar, S.K.: When does a camera see rain? IEEE Int. Conf. Comp. Vis. (ICCV) **2**, 1067–1074 (2005)
5. Garg, K., Nayar, S.K.: Detection and removal of rain from videos. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1, no. 1, pp. 528–535 (2004)
6. Phillips, W. III., Shah, M., Lobo, N.V.: FlameRecognition in Video. Pattern Recognit. Lett. **23**(1–3), 319–327 (2002)
7. Fastcom Tech.SA, Blvd. de Grancy 19A, CH-1006 Lausanne, Switzerland: Method and device for detecting fires based on image analysis. In: Patent Coop. Treaty(PCT) Appl.No: PCT/CH02/00118, PCT Pubn.No: WO02/069292
8. Liu, C.B., Ahuja, N.: Vision Based Fire Detection. In: Proceedings of International Conference on Pattern Recognition, ICPR '04, vol. 4 (2004)
9. Kopilovic, I., Vagvolgyi, B., Sziranyi, T.: Application of panoramic annular lens for motion analysis tasks: surveillance and smoke detection. In: Proceedings of 15th International Conference on Pattern Recognition, vol. 4, pp. 714–717. IEEE, 3–7 September (2000)
10. Chen, T.-H., Yin, Y.-H., Huang, S.-F., Ye, Y.-T.: The smoke detection for early fire-alarming system base on video processing. In: International Conference on Intelligent Information Hiding and Multimedia, pp. 427–430 (2006)
11. Xiong, Z., Caballero, R., Wang, H., Finn, A., Lelic, M. A., Peng, P.: Video-based smoke detection: possibilities, techniques, and challenges suppression and detection research and applications. A Technical Working Conference (SUPDET 2007). March 5–8 Orlando, Florida (2007)
12. Toreyin, B.U., Dedeoglu, Y., Cetin, A.E., Fazekas, D., Chetverikov, T. Amiaz, N. Kiryati.: Dynamic texture detection, segmentation and analysis In: Conference On Image And Video Retrieval, pp. 131–134. ACM, New York (2007)
13. Garg, K., Nayar, S.K.: Photometric Model of a Rain Drop. Technical Report, Department of Computer Science, Columbia University, September 2004
14. Vicente, J., Guillemant, P.: An image processing technique for automatically detecting forest fire. Int. J. Therm. Sci. **41**(12), 1113–1120 (2002)
15. Toreyin, B.U., Dedeoglu, Y., Cetin A.E.: Flame detection in video using hidden Markov models. In: IEEE International Conference on Image Processing IEEE (2005)
16. Toreyin, B.U., Dedeoglu, Y., Cetin A.E.: Wavelet based real-time smoke detection in video. In: EUSIPCO '05 (2005)
17. Collins, R.T., Collins, A.J., Lipton, Kanade, T.: A system for video surveillance and monitoring. In: 8th International Topical Meeting on Robotics and Remote Systems. American Nuclear Society (1999)
18. Cucchiara, R., Grana, C., Piccardi, M., Prati, A.: Detecting moving objects, ghosts and shadows in video streams. IEEE Trans. Pattern Anal. Mach. Intell. **25**(10), 1337–1342 (2003)
19. Vezzani, R., Cucchiara, R.: AD-HOC: appearance driven human tracking with occlusion handling. In: First International Workshop on Tracking Humans for the Evaluation of their Motion in Image Sequences (THEMIS'2008), Leeds, UK
20. Mallat, S.G.: A theory for multiresolution signal decomposition: the wavelet representation. IEEE Trans. Pattern Recogn. Mach. Intell. **11**(7), 674–693 (1989)

21. Sato, M.: Fast learning of on-line EM algorithm. Technical Report TR-H-281, ATR Human Information Processing Research Laboratories
22. Casella, G., Berger, R.: Sattistical Inference, 2nd edn. Duxbury Press, Belmont (2002)
23. Bishop, C.: Pattern Recognition and Machine Learning. Springer, Berlin (2006)
24. Stauffer, C., Grimson, W.E.L.: Adaptive background mixture models for real-time tracking. In: Proceedings IEEE Conference on Computer Vision and Pattern Recognition. IEEE, pp. 246–252 (1999)
25. Sato, M.: Fast learning of on-line EM algorithm, Technical Report TR-H-281. ATR Human Information Processing Research Laboratories