SMOKE DETECTION USING TEMPORAL HOGHOF DESCRIPTORS AND ENERGY COLOUR STATISTICS FROM VIDEO

Konstantinos Avgerinakis[1,2], *Alexia Briassouli* [2], *Ioannis Kompatsiaris*[2]

(1)Centre for Vision, Speech and Signal Processing, University of Surrey, UK
(2)Information Technologies Institute, CERTH-ITI, Thessaloniki, Greece

ABSTRACT

We present a novel algorithm for detecting and localizing smoke in video. The first step of our method focuses on detecting the presence of smoke in video frames, while in the second part localization of smoke particles in the scene takes place. In our implementation, we take advantage of both appearance and motion information, so that we can extract robust and meaningful information. Machine learning is used in order to discriminate our data more thoroughly and provide accurate smoke detection. Experiments carried out with various benchmark datasets show that smoke is indeed accurately localized both in time and space via the proposed approach.

1. INTRODUCTION

In recent years, smoke detection using video surveillance cameras has attracted research attention as it promises to solve a great deal of problems that traditional techniques are not able to deal with, where smoke is detected using methods such as particle sampling, temperature sampling, relative humidity sampling, air transparency testing, in addition to the traditional ultraviolet and infrared fire detectors. The main drawback that these techniques have is the time that is required to acquire information. Furthermore, they are specifically designed for certain environments and regularly fail in real time situations. Video based smoke detection promises to boost early fire detection, as smoke comes before the fire, and to deal with open spaces, where is extremely difficult for conventional methods to extract a meaningful result.

The main challenges in video analysis for smoke detection are the variability in shape, motion and texture patterns of smoke, whose appearance is contingent on the luminance conditions, background and colours of the scene. Techniques relying on motion information are proposed in [1], [2]. Optical Flow is computed and, based on its magnitude and angle values, the presence of smoke is predicted. Although it was a promising start for smoke detection, the computational cost of the methodology led later researchers to use other methods. In [3], [4] chromatic models are introduced for smoke detection: they locate smoke blobs and, in conjunction with a background subtraction technique, a prediction is extracted. These methods lack robustness as they dont take into account other information except for colour, resulting in a high error rate when objects with similar saturation as smoke, such as clouds, appear in the scene. Colour characteristics are further studied in [4] and a smoke detection technique for forest fires is proposed, based only on colour features.

An interesting and robust approach has been defined by Toreyin et al. in [5], [6]. In that work, energy computation from wavelet coefficients is introduced in order to localize smoke, and the results are combined with background subtraction and a chromatic detector to extract a candidate blob (contour). The resulting candidate blob is then imported into an HMM model for better results. A simpler technique is proposed in [7], where a two-layer backpropagation (BP) neural network is introduced for smoke prediction. In this method no colour features are taken into account and a simplistic adaptive background subtraction method is proposed for motion segmentation. Lee [8] introduce real-time fire detection, which is specifically designed for tunnel environments, extending the related work with a module that eliminates noise caused by car lights and sudden illumination changes. More recently, Piccinini et al. in [9] extended the literature by proposing a Mixture of Gaussian (MoG) to model the wavelet energy in order to classify smoke patterns.

In this work, motion and appearance features are combined, in order to avoid the shortcomings of previous methods. The proposed approach is not limited to smoke detection under specific circumstances (e.g. a traffic tunnel) and is able to accurately localize smoke even in challenging cases, where similarly colored objects (e.g. a person with a white shirt) are moving in the scene. This paper is organized as follows: in Sec. 2 an overview of the proposed method is presented. Details on the first part of the approach, concerning background subtraction, are presented in Sec. 3. Sec. 4 is concerned with feature extraction in active blocks and their incorporation into a Bag of Words framework. Sec. 5 explains how energy and color information complements the previous steps, in order to spatially localize the smoke. Experiments with benchmark datasets and comparisons with existing methods are provided in Sec. 6. Finally, conclusions are drawn in Sec. 7, along with ideas for future research.

2. OVERVIEW

In the first stage of our algorithm, background/foreground separation takes place. For this purpose, we use a novel algorithm which is based on motion information and not on image appearance, as is usually the case. In the resulting binary images, which denote where motion occurs, we localize candidate blocks which are going to be further analyzed to determine whether smoke exists in them or not. From these blocks, histograms of oriented gradients (HOGs) and histograms of oriented optical flow (HOFs) are constructed, to take into account both appearance and motion information. Afterwards, they are averaged over time to induce temporal smoothing, leading to a spatio-temporal descriptor for each candidate block.

HOGs are used as they help separate rigid objects from smoke: HOGs of rigid objects tend to have high values due to the presence of strong edges and corners, unlike the HOGs of smoke. HOF is used to help us robustly detect smoke, as its motion is usually directed upwards, while other object motions may be in various directions. Afterwards, a visual-vocabulary is built by applying hierarchical k-means clustering on these descriptors and a fast bag-of-words is implemented by using an inverted index and term frequency inverted document frequency as a metric. Finally, a Kernel matrix is created by comparing training data in a pairwise manner, and is used in an SVM classifier to detect which frame contains smoke. As soon as the frames with smoke are detected, the second stage of our algorithm takes place: in this stage we try to spatially localize smoke particles in each frame. Statistics based on the color and image energy of the candidate blocks (extracted in the previous step) are analyzed to determine whether they contain smoke or not.

3. BACKGROUND SUBTRACTION

For separating static from moving pixels, optical flow values are analysed using the Kurtosis metric. When there is no real motion, non-zero motion values are induced by noise, corresponding to hypothesis H_0, and are therefore assumed to follow a Gaussian distribution, while real motion introduce deviations from Gaussianity (hypothesis H_1):

$$\begin{aligned} H_0 &: u_k^0(\bar{r}) = z_k(\bar{r}) \\ H_0 &: u_k^1(\bar{r}) = z_k(\bar{r}) + v_k(\bar{r}), \end{aligned} \tag{1}$$

where $z_k(\bar{r}), v_k(\bar{r})$ denote noise and true motion values respectively. The kurtosis G_2 of Gaussian data is equal to zero and is used to detect whether motion is caused by noise or by changes in optical flow. The unbiased kurtosis estimator G_2 of W data values $y = \{y_i\},\ i = 1, 2, ..., W$ is given by:

$$G_2[y] = \frac{3}{W(W-1)} \sum_{i=1}^{W} (y_i)^4 - \frac{W+2}{W(W-1)} \left(\sum_{i=1}^{W} (y_i^2) \right)^2, \tag{2}$$

where W is a manually chosen temporal window from which motion values are obtained. Kurtosis values are significantly higher in regions of pixels whose motion changes. In our experiments, $W = 5$ was used, although accurate results can be obtained for various window sizes [10].

4. BAG OF WORDS AND SVM

Features are extracted in the regions of activity in the video after the latter are separated into blocks containing moving pixels of interest. Blocks are constructed every N_b pixels in regions where more than 50% of the pixels have been determined to be moving. In our experiments with a wide range of videos we found that $N_b = 16$ gave good results. HOGs and HOFs are calculated for each block and are collected for W frames to build the spatio-temporal descriptor that contains information for each block. It was determined experimentally that $W = 5$ frames provide sufficient information for localizing smoke before its location changes signfiicantly.

The descriptors derived from the training set are clustered using hierarchical k-means in order to build a vocabulary tree of visual words. Using this vocabulary, we can define what objects exist at each frame of the training set based on a term frequency-inverse document frequency metric (tf-idf). A kernel map is built by comparing the training set descriptors in a pairwise manner and is used for training a SVM classifier. Afterwards, we describe all the frames of the test set based on the constructed vocabulary tree and predict whether they contain smoke particle or not by using the trained SVM classifier. In Sec. 6 it is shown that the proposed method provides accurate results on benchmark datasets, namely videos from Bilken University and the Visor datasets.

5. ENERGY AND COLOUR ANALYSIS

After frames that contain fire are detected as in Sec. 4, the precise spatial location of the smoke is determined by calculating energy and colour statistics for each block in the scene. Image energy statistics are used because we have observed that smoke has the ability to *gradually* reduce the energy of the pixels that it covers before depleting it to zero, in contrast to rigid object occlusions that decrease image energy to zero almost instantly. Thus, each block is studied throughout time (over W frames) and afterwards it is evaluated whether it contains smoke or not. There are several techniques to estimate block energy, such as the Fast Fourier Transform(FFT), Discrete Cosine Transform(DCT), Sobel Filter, Wavelet Transform and Laplace Transform. We choose to use the Sobel edge detection filter, as it can be calculated quickly and easily. The energy is then given by:

$$E_{b_k, I_t} = \sum i, j \in b_k E_V(i,j)^2 + E_H(i,j)^2 + E_D(i,j)^2, \tag{3}$$

where E_{b_k, I_t} denotes the overall energy for video frame I_t at time t and in the k_{th} block b_k. E_V, E_H, E_D represent the energy obtained from the vertical, horizontal and diagonal edge images, respectively (Fig. (1)).

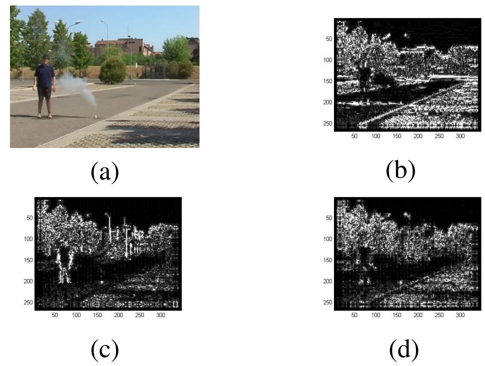

Fig. 1. (a) Original image, (b)-(d) Sobel operator applied on horizontal, vertical and diagonal axes, respectively.

Each block's energy is observed throughout time and compared against a statistically derived threshold:

$$\eta = mean(b_k) + c \cdot std(b_k), \tag{4}$$

where $mean(b_k)$, $std(b_k)$ are the mean and standard deviation for each candidate temporal block and c is a constant that indicates the angle of the energy line. It is found experimentally that $c = 5$ gives very good smoke localization results. Figure 2 shows how image block energy decreases when it is induced from rigid object motion and from smoke.

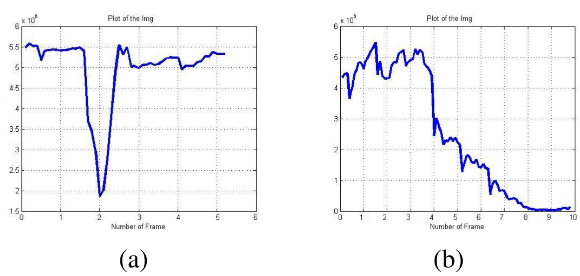

Fig. 2. Energy decrease over time induced from (a) rigid object motion and (b) smoke motion.

We also take colour statistics into account in order to localize smoke, as its color is usually whitish-blue. We convert our RGB frame into an HSV signal, where the Value channel stores the biggest value of the RGB channels: $V \leftarrow max(R; G; B)$. The Saturation channel is computed by taking into account the following condition:

$$\begin{aligned} &If\ V \neq 0 && \Rightarrow S = \frac{V - min(R, G, B)}{V} \\ &Otherwise && \Rightarrow S = 0. \end{aligned} \tag{5}$$

	Smoke videos	TP (%)	TF (%)
1	Bilkent/sBehindtheFence	96.15	3.85
2	Bilkent/sBtFence2	96.55	3.45
3	Bilkent/sEmptyR1	80	20
4	Bilkent/sEmptyR2	81.25	18.75
5	Bilkent/smoky	96.67	3.33
6	Bilkent/sParkingLot	100	0
7	Bilkent/sWasteBasket	94.23	5.77
8	Bilkent/sWindow	100	0
9	VISOR/movie08	74.86	25.14
10	VISOR/movie09	92.51	7.49
11	VISOR/movie10	58.51	41.49
12	VISOR/movie11	79.02	20.98
12	VISOR/movie12	88.52	11.48
13	VISOR/movie13	76.36	23.64
14	VISOR/movie14	89.48	10.52
15	VISOR/movie15	51.28	48.72
16	VISOR/movie16	100	0
17	VISOR/movie17	58.04	41.96
	Total Average	84.08	15.92

Table 1. Smoke Detection in Bilkent and VISOR datasets.

The Hue channel is calculated as follows:

$$\begin{aligned} &If\ V = R && \Rightarrow H = 60 \times (G - B/S) \\ &If\ V = G && \Rightarrow H = 120 + 60 \times (B - R/S) \\ &If\ V = B && \Rightarrow H = 240 + 60 \times (R - G/S). \end{aligned} \tag{6}$$

Thus, knowing that the smokes colour is whitish-blue to white when the temperature of smoke is low, we can filter the moving objects by thresholding the Saturation and the Value channel values. Smoke is then detected when:

$$S < Th_1 \quad and \quad V > Th_2, \tag{7}$$

where Th_1, Th_2 are estimated from training videos.

6. EXPERIMENTS

Experiments took place with the smoke videos from Bilkent University used in [5],[6], and the ViSOR smoke dataset that is used in [9]. Videos are separated in test and training data in a one-versus-all manner, i.e. we test each video clip assuming that we have all the rest for training. The results for temporal smoke localization using the method of Sec. 4 can be seen in Table 1. We derive an Average Accuracy of 84.08% on both datasets (without accounting for the videos with car lights). We faced difficulties when we tried to distinguish car lights from smoke or when smoke particles where very dispersed.

Spatial localization of the smoke takes place as described in Sec. 5, and some characteristic results are shown in Fig. 3, where it can be seen that the separation of the color spaces into Hue and Saturation greatly helps in detecting smoke. Localization results from the Bilkent and Visor smoke datasets can be seen in Fig. 4. We observe that the smoke is well separated from moving rigid objects, and that there is good localization even when the smoke is dispersed, under difficult weather conditions (wind blowing) and at a large distance. Videos of the localized smoke can also be found in the first part of http://mklab.iti.gr/web-demos/annotation/, where it can be seen that the smoke is indeed detected with accuracy throughout all video frames.

(a) (b) (c)

Fig. 3. (a) HSV, (b) Saturation and (c) Hue. The smoke can be clearly distinguished in the Hue and Saturation channels.

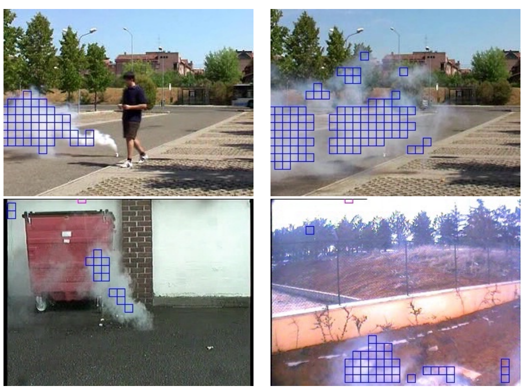

Fig. 4. Smoke localized under various challenging conditions.

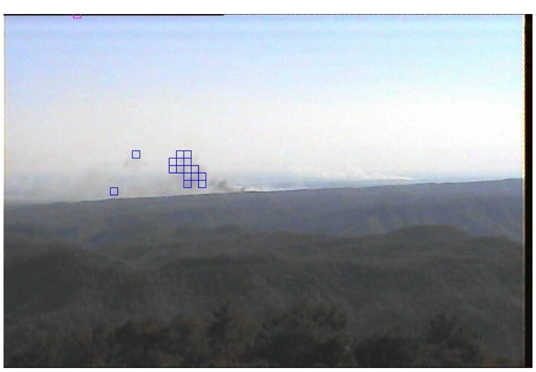

Fig. 5. Far away smoke correctly localized by our method.

Acknowledgement: This work was funded by the European Commission under the 7th Framework Program (FP7 2007-2013), grant agreement 288199 Dem@Care.

7. CONCLUSIONS

This work presents a novel method for the detection of smoke in videos. Smoke is localized in time based on appearance (HOGs) and motion characteristics (HOFs) which are unique to smoke. These are gathered from training videos and input into an SVM in order to detect frames in which smoke is present. Experiments on benchmark datasets show that accurate temporal smoke localization is achieved in this manner. Spatial localization of the smoke then takes place by taking advantage of its non-rigid nature and its characteristic color. Thus, edge energy in an image is decreased gradually when smoke passes over it. This information, used in conjunction with the color information of smoke, extracted from training videos, is shown to accurately localize smoke in various videos, even under challenging conditions, such as wind blowing, or the presence of other moving entities.

8. REFERENCES

[1] I. Kopilovic, B. Vagvolgyi, and T. Sziranyi, "Application of panoramic annular lens for motion analysis tasks: Surveillance and smoke detection," in *Pattern Recognition, 2000. Proceedings., 10th International Conference on*, 2000.

[2] F. Gomez-Rodriguez, B. Arrue, and A. Ollero, "Smoke monitoring and measurement using image processing. application for forest fires," in *Automatic Target Recognition XIII. Proceedings of the SPIE*, 2003, vol. 5094, pp. 404–411.

[3] T. Chen, P. Wu, and Y. Chiou, "An early fire-detection method based on image processing," in *International Conference on Image Processing, 2004. ICIP '04*, 2004.

[4] T. Chen, Y. Yin, S. Huang, and Y. Ye, "The smoke detection for early fire-alarming system base on video processing," in *International Conference on Intelligent Information Hiding and Multimedia Signal Processing, 2006. IIH-MSP '06*, 2006.

[5] B. Toreyin, Y. Dedeoglu, and A. Cetin, "Wavelet based real-time smoke detection in video," in *EUSIPCO 05*, 2005.

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

[7] Z. Xu and J. Xu, "Automatic fire smoke detection based on image visual features," in *Proceedings of the 2007 Int. Conf. on Computational Intelligence and Security Workshops*, 2007, pp. 316–319.

[8] B. Lee and D. Han, "Real-time fire detection using camera sequence image in tunnel environment," in *ICIC'07 Proceedings of the intelligent computing 3rd international conference on Advanced intelligent computing theories and applications*, 2007.

[9] P. Piccinini, S. Calderara, and R. Cucchiara, "Reliable smoke detection system in the domains of image energy and color," in *IEEE International Conference on Image Processing, ICIP '08*, 2008.

[10] A. Briassouli and I. Kompatsiaris, "Robust temporal activity templates using higher order statistics," *IEEE Transactions on Image Processing*, vol. 18, no. 12, pp. 2756–2768, Dec. 2009.